Jed DeVaro

STRATEGIC COMPENSATION AND TALENT MANAGEMENT

Lessons for Managers

Strategic Compensation and Talent Management

Lessons for Managers

Written for current and aspiring managers, this textbook guides readers through the core components of compensation and puts them in the manager's chair, challenging them to apply their understanding to solve business problems such as attracting, managing, and retaining company talent. The book's central theme, supported by extensive treatment of compensating differentials, is that compensation is heavily driven by market competition. The coverage also includes analytics, negotiation and bargaining, wage theft, and non-profits and small businesses, as well as a detailed treatment of stock options. Case studies are included to demonstrate the principles in practice, and "Lessons for Managers" in each chapter provide practical advice and takeaways. A rich package of online teaching and learning materials, including teaching slides, sample syllabi, additional case studies, a test bank, and instructor notes is also provided to support teachers and students.

Jed DeVaro is the Wang Family Professor of Management and Economics, and Chair of the Department of Economics, at California State University–East Bay, where he teaches an MBA course in "Strategic Compensation and Reward Systems". He is a research fellow at Cornell University's Institute for Compensation Studies and a Senior Research Fellow at the Research Institute for the Finnish Economy. His published articles on compensation and related topics span economics, finance, strategy, organizational behavior, and human resources management.

Strategic Compensation and Talent Management

Lessons for Managers

Jed DeVaro

California State University–East Bay

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108495202
DOI: 10.1017/9781108861458

First published 2020

Printed in the United Kingdom by TJ International Ltd, Padstow, Cornwall, 2020

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-49520-2 Hardback
ISBN 978-1-108-81743-1 Paperback

Additional resources for this publication at www.cambridge.org/devaro

"To those who pay, those who have been paid, and those who should have been paid."

Brief Contents

Contents

Acronyms

ARRA = American Recovery and Reinvestment Act of 2009
ADA = Americans with Disabilities Act of 1990
ADEA = Age Discrimination in Employment Act of 1967
AFSCME = American Federation of State, County and Municipal Employees (a labor union)
BART = Bay Area Rapid Transit (public transportation in San Francisco area)
BEA = Bureau of Economic Analysis
BFOQ = Bona fide occupational qualification
BLS = Bureau of Labor Statistics
BP = British Petroleum
BSM = Black–Scholes–Merton (options pricing formula)
BSO = Boston Symphony Orchestra
CalPERS = The California Public Employees' Retirement System
CBA = Collective bargaining agreement
CBE = College of Business and Economics
CBR = *Compensation & Benefits Review* (a practitioner journal)
CD = certificate of deposit
CEO = Chief Executive Officer
CFO = Chief Financial Officer
CFA = California Faculty Association (the faculty union in the CSU system)
CMR = *California Management Review* (practitioner journal)
COLA = Cost-of-living-adjustment
COO = Chief Operating Officer
CPI = Consumer Price Index
CPI-U = Consumer Price Index (for all urban consumers)
CPS = Current Population Survey
CSU = California State University (system)
CSUEB = California State University–East Bay
CSUF = California State University–Fresno
CSULA = California State University–Los Angeles
CSULB = California State University–Long Beach
CSUSF = California State University–San Francisco
DB = Defined benefit (pension system)
DC = Defined contribution (pension system)
DOL = Department of Labor

DOMA = Defense of Marriage Act
EBRI = Employee Benefits Research Institute
ECEC = Employer Costs for Employee Compensation (index)
ECI = Employment Cost Index
EEOC = Equal Employment Opportunity Commission
EPA = Equal Pay Act of 1963
ESPP = Employee stock purchase plan (a form of equity-based compensation)
FAA = Federal Aviation Administration
FLSA = Fair Labor Standards Act of 1938
FMLA = Family and Medical Leave Act of 1993
GPA = grade point average (a measure of a student's performance in school)
HBR = *Harvard Business Review* (practitioner journal)
HR = Human resources
HRS = Health and Retirement Study (longitudinal data set of older workers)
HWHFA = Healthy Workers, Healthy Families Act of 2014
ICS = Institute for Compensation Studies (Cornell University)
IRA = Individual Retirement Account (either "traditional" or "Roth")
IRS = Internal Revenue Service (the agency that collects taxes in the United States)
IT = Information Technology
KC = Kansas City, Kansas (and Missouri)
ML = Merrill Lynch (an American wealth management company)
NCS = National Compensation Survey (conducted by the BLS)
NLRA = National Labor Relations Act
NLRB = National Labor Relations Board
NLS = National Longitudinal Surveys (sponsored by the BLS)
NYPD = New York City Police Department
OES = Occupational Employment Statistics
OLS = Ordinary least squares (regression)
OSHA = Occupational Safety and Health Administration
PATCO = Professional Air Traffic Controllers Organization
PB = Pilgrim, Baxter and Associates (an American mutual-fund company)
PDV = Present discounted value
PhD = Doctor of Philosophy (the highest-level academic degree in most fields)
PHL = Philadelphia International Airport
PSID = Panel Study of Income Dynamics (a longitudinal household data set)
RA = Research assistant (a job title)
RPP = Regional price parity (a regional price index produced by the US BEA)
RSU = Restricted Stock Units (a form of equity-based compensation)
SAS = Statistical Analysis System
SEC = Securities Exchange Commission

SEIU 32BJ = Service Employees International Union, Local 32BJ (a labor union)
SEP = Simplified Employee Pension (a type of traditional IRA)
SF = San Francisco, California
SIPP = Survey of Income and Program Participation
SNAP = Supplemental Nutrition Assistance Program
STEM = Science, Technology, Engineering, and Mathematics
TDA = Tax-deferred annuity (more commonly called tax-sheltered annuity, or TSA)
TSA = Tax-sheltered annuity (also called tax-deferred annuity, or TDA)
VP = Vice President
WHD = Wage and Hours Division (of US Department of Labor, i.e., DOL)
WRDS = Wharton Research Data Services

Preface

Gaining and sustaining competitive advantage requires that managers understand how to use compensation strategically to attract, manage, and retain their organization's talent. This book will help you to develop and refine that understanding, equipping you to think in a sophisticated way about compensation, to recognize the implications of compensation systems for employee behavior, and to use compensation to solve problems and achieve business objectives in your current or future organization.

This book is written for managers around the world. It can also be used as a text for a one-semester compensation course with no pre-requisites that is accessible to readers with the analytical preparation and quantitative skills that are now standard in business. Instructor resources are available on the course website, including PowerPoint slides, sample course syllabi, 30 case discussions, and a multiple-choice test bank. I target three audiences, all of whom have, or aspire to have, careers in management:

(1) MBAs and other professional masters students
(2) advanced business undergraduates
(3) experienced managers, including:
  - those who are enrolled in executive education programs
  - those who simply want to learn more about compensation to further their professional development.

Speaking to these audiences requires a new approach to teaching compensation. This book starts from the following concept: *Anything that fails to reach and be relevant for an audience of professional managers will also fail to reach an audience of business students*. After all, most business students will soon become professional managers, if they are not already. Business students are at their most attentive and engaged when they're treated as managers rather than as students. Moreover, most managers don't want to read anything that looks like a textbook. In addition to lacking the time and inclination to wade through stuffy and impersonal prose that's heavily laden with footnotes citing the dry academic literature, managers have little use for the usual key terms, outlines, study tips, homework questions, end-of-chapter quizzes, etc., that clutter the page, disrupt the flow of the narrative, and kill the momentum. This book omits the usual trappings of a textbook and adopts an informal, narrative, conversational style that puts the reader in the manager's chair.

It also focuses on content and themes that are highly relevant for managers but that receive little or no attention in standard compensation textbooks.

For nearly 20 years I have taught different versions of this material at the undergraduate through PhD levels, at Cornell, California State University–East Bay (CSUEB), and the University of Cologne. In CSUEB's business school, I use the book to teach an MBA elective in the Management Department called "Strategic Compensation and Reward Systems" and its undergraduate counterpart. I have also presented this material to groups of business executives.

Employers who engage with compensation-related issues are of two types: the general managers who run the organization, and the professionals who administer compensation systems in the human resources (HR) and payroll offices. This book speaks to both groups but particularly to the first. For that reason, and in contrast to most compensation textbooks, I focus more on general concepts, theory, and ideas than on facts and details of compensation systems. This approach also makes the book more relevant for an international readership than most compensation books. Although most of the labor law and institutional context focuses on the United States, all of the insights and ideas are easily transportable to other countries, because my focus is less on the specifics of regulations and institutions than on how to think about them. Other unique features that distinguish this compensation text from others are:

- ***Conversational writing style.*** Compensation can be pretty dry material. Textbook discussions easily become encyclopedic, verbose, and tedious to read. I've tried to avoid that by writing in a conversational style. I rely heavily on first and second person and usually avoid third person. "I" (first person) am speaking directly to "you" (second person) informally, just as we would in an in-person chat. Moreover, "you" are assumed to be seated in the manager's chair throughout the book, because that's where you're likely to ultimately end up, if you're not already there. This approach, though unconventional for a textbook, has the advantage of disciplining my writing. By speaking to you directly as a manager, I am forced to keep the discussion sharply focused on exactly what is relevant for business managers. To further that end, every chapter ends with a "lessons for managers" section that gives you (the manager) the key takeaways.
- ***Case discussions.*** Actively and collaboratively engaging with the material is the best way to learn about compensation. To facilitate such learning, I have created 30 mini cases that are designed to stimulate group discussion. One case appears at the end of each chapter, and the entire set of 30 is available online on the course website. Many of the cases are relevant to several chapters, and they span a variety of industries, occupations, and sectors. They are designed to be completed in class, spontaneously, in about an hour. Students break off into groups for 30 or so minutes to discuss the case and prepare answers, and the

remainder of the hour is devoted to class discussion. That format works well in my MBA class. Alternatively, cases can be assigned for homework, with students bringing written answers to class for discussion. Many of the cases have more discussion questions than can reasonably be completed in an hour, to give instructors choice.

- ***Emphasis on market competition***. The central theme that pervades this book is that compensation is heavily driven by market competition. To develop that theme, I draw on economic concepts to an unusual extent, though striving to do so in a way that will resonate with those who have had no prior exposure to economics or whose prior exposure has been unpleasant. My multidisciplinary background positions me well to pull this off. In the business school at CSUEB, I hold a joint appointment in the Departments of Management and Economics. My home department is Management, and my compensation courses are offered in the Human Resources and Organizational Behavior (HR/OB) subdivision of that department, but I am also the chair of the Economics Department. Multidisciplinary perspective is crucial in the area of compensation, and my academic research on compensation spans economics, finance, human resources management, organizational behavior, and strategy.
- ***Content.*** The book contains material that is important for managers in practice but that is typically absent from compensation texts. Some examples are:
  - ***Compensation analytics***. Chapters 6 and 7 on compensation analytics reflect the increasingly central role of business analytics in modern managerial decision-making. Recent technological advances in hardware and software, along with expanded availability of "big data" human resources records in electronic form, facilitate the development of innovative, analytically deep, data-driven solutions to business problems. Many companies are already doing this, and creating and sustaining a competitive advantage now requires it. Readers of this book who will become consultants need to understand how to leverage a client's internal compensation data, and those who will become general managers or run their own organizations need to know how to leverage their own company's internal data and draw on external benchmarks. The two chapters are designed to quickly equip readers to conduct and evaluate competent, impactful analysis of compensation data, and they are unique in offering lots of practical advice (e.g., on organizing and cleaning the data prior to analysis) that is critical for data analysis.
  - ***Negotiation and bargaining*** with current and prospective employees over compensation is one of the most important challenges that managers face. I cover the subject in Chapter 14.
  - ***Compensating differentials***, which can derive from any characteristic of a job (good or bad), form the core analytical support structure for the book's

central theme, namely the role of market competition in shaping compensation. A real appreciation for how other companies' compensation practices affect those of your own requires a thorough understanding of compensating differentials. Chapter 3 develops that material in depth, and it is consistently reinforced throughout the book.

- ***Compensation contract failure and wage theft*** is a neglected subject that I cover early. Compensation is about *contracts* (or agreements involving an exchange of labor for pay) between employers and employees. A great way to learn how something works is to study it when it's broken, and wage theft is the consequence of a broken (i.e., breached) contract. Chapter 2 sets the stage for the core chapter on compensating differentials that immediately follows it, because employees view the threat of wage theft as a negative job characteristic, and any negative job characteristic can create a compensating differential.

Selected recommendations for further reading appear at the end of each chapter. I do not attempt to provide an exhaustive list, though I do strive for breadth in disciplinary perspectives. Practitioner outlets are included as well as academic journals.

The limited amount of quantitative material in this book is not daunting and involves only basic mathematics. Chapters 6 and 7, on compensation analytics, cover the basics of regression analysis, and Chapters 8 (on training) and 12 (on turnover management and talent retention) cover present value analysis, which is needed for discussing streams of compensation over time. Compensation is inherently quantitative, as a glance at your last W-2 form reveals! As a current or prospective manager, you will benefit from enhancing your quantitative skills, because quantitative problems arise increasingly often in business, particularly in compensation.

The book is best read cover to cover, like a novel. The presentation is cumulative, and as the discussion progresses I highlight the interconnectedness of the topics. A modular design of any compensation textbook, that attempts to treat the chapters as self-contained, would encourage the tunnel vision that can lead to catastrophic mistakes in organizations. Compensation systems offer powerful tools for solving business problems and fueling organizational success, but they also create fields of landmines . . . an understanding not only of the individual components of compensation systems but of how they interact in the entire system, and with other human resource management practices, is the surest way that a manager can avert the landmines.

I intentionally draw many examples from my own employment experiences, particularly from the California State University (CSU) system where I currently work. One reason for this is simply convenience, because all of the anecdotes are at my fingertips from personal experience. But a more important reason is to illustrate that most chapters of this book have relevance for *any given organization*. I have connected many of the chapters to examples from the CSU system, and, as you read, you should be able to connect many of the chapters to your own organization.

My principal intellectual debts are to the scholars whose work forms the foundation of this book and who shaped my understanding of compensation, including my former teachers and mentors at Stanford (Ed Lazear, Tom MaCurdy, John Pencavel, and Luigi Pistaferri). I am also indebted to Stanley Wang (President, Pantronix Corporation) and his family for endowing the chair I've held since joining CSUEB in 2008, and to Terri Swartz for creating the joint appointment in the Management and Economics Departments that equipped me uniquely to write this book. I'm also thankful for the support of friends, family members, and my present and former colleagues from CSUEB and Cornell. Particularly detailed and helpful feedback was provided by Jagdish Agrawal, Alan Benson, Alex Bryson, Anil Comelo, Eve DeVaro, Scott Fung, John Heywood, Antti Kauhanen, Jack Kilgour, Mee Sook Kim, Jason Perry, the anonymous reviewers at Cambridge and elsewhere, and my former MBA students who provided research assistance (Tim Gugenhan, Kilby Hammond, and Vasana Ly). Valerie Appleby has been a superb and enthusiastic editor at every stage of the publication process. Finally, I am grateful to my students, whose needs, interests, aspirations, and questions inspired this project.

The following pages are intended to be the first step in an ongoing enterprise. I welcome emails or other communications offering suggested improvements, criticisms, ideas for new case discussions, sample course syllabi, instructor notes, and any other commentary.

Jed DeVaro
The Wang Family Professor
Departments of Management and Economics
Chair, Department of Economics
College of Business and Economics
California State University–East Bay
Hayward, CA 94542, USA
Email: jed.devaro@csueastbay.edu

Case Discussions

Each chapter is followed by a related case discussion. Those 15 cases appear in boldface in the table below, and the entire set of 30 is available on the course website. The cases are designed to be prepared and discussed in class. Most require about an hour, though some require more time. The lists of questions are intentionally kept rather extensive to allow instructors to be selective. Students break off to work in groups, preparing answers for the first half of the allotted time, and the remaining time is spent on class discussion. In the following table, open circles indicate chapters that are relevant to the case, and closed circles indicate a particularly strong connection.

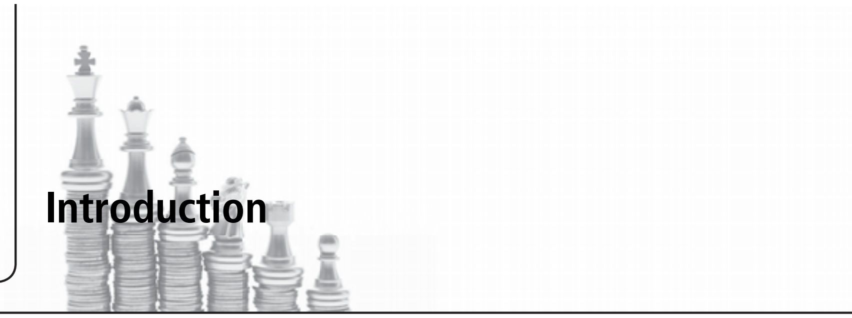
	Chapter															
Case Discussions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Page
1. **Buffer, Inc.**	●				○	○	○									20
2. Flying Pig		●		○												online
3. **The Deadliest Catch**			●						○							51
4. Uber				●												online
5. Philadelphia Airport Workers				○	●									●		online
6. **CSUEB CBE**					●				○				○		○	103
7. **Wage–Insurance Tradeoff (Part A)**			●			●	●				○					139
8. **Wage–Insurance Tradeoff (Part B)**			●			●	●				○					169
9. **Google**								●				○				188
10. **Lindy's Seafood**				○					●		○					214
11. **Tesla Motors**									●	●						241
12. BP and Royal Dutch Shell				○					●	●		○				online
13. Amazon versus Sanders				○		○	○									online
14. Amazon versus Sears				○					○		○		○			online
15. Netflix									○			○				online
16. Federal Reserve Bank of Richmond					○			○			○	○			○	online
17. **Walmart**	○			○							●	○				268

(cont.)

Case Discussions	Chapter 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Page
18. New Jersey State Pension System											●				●	online
19. Subway				○								●				online
20. **Merrill Lynch (Part A)**						○	○					●		○		297
21. Merrill Lynch (Part B)												●		○		online
22. Up-or-Out in the US Military												○	●		○	online
23. **New York City Police Department**				○	○								●		●	321
24. **Boston Symphony Orchestra**				○	○									●	○	342
25. **Salesforce.com versus Salesforce .org**			●									●			●	360
26. The Cheese Board Collective				○					○	○		○			○	online
27. Bimbo Bakeries								●				●				online
28. **The Walrus and the Carpenter**	●			●					●			○			○	87
29. Negotiation and Salary History				●								○		●		online
30. **Weaver *v.* Legend Senior Living, LLC**		●		●					●	○						32

○ = Chapter in the column heading is relevant to the case in the row heading.
● = Chapter in the column heading is strongly connected to the case in the row heading.

1 Introduction



You have almost certainly received pay for work you have done. Similarly, you have likely paid someone else for work they have done for you. For example, every other week I pay $40 to a gardener who comes to my house, every few weeks I pay $20 for a haircut at someone else's house, and every so often I hire a student to work as my research assistant at $12 per hour, with all of these payments serving as compensation for the services rendered. Such payments are what jump to most people's minds when they hear the words "pay" or "compensation" . . . they think of cash, or alternatively, hard-copy paychecks or electronic deposits that can be easily converted to cash. But compensation is a broader concept encompassing more than cash payments.

1.1 What Is Compensation?

A person's total compensation is properly understood as including *everything that the person likes about a job*. Examples include wages and salary, bonuses, health insurance, a 401(k) plan, on-site childcare, paid vacation, promotion opportunities, a collegial work environment, an understanding boss, job stability and security, flexible hours, a corner office with a big window and a nice view, low commuting costs from the employee's home, a relaxed dress code, good weather, and an appealing geographic location. That's already a lot, and it's only a partial list. Some of those items are direct costs that appear on your organization's balance sheet (e.g., salaries, bonuses, and 401(k) plans), whereas others are not (e.g., good weather and a relaxed dress code). From your employees' perspectives, compensation should usually be thought of using the broad definition just given, but sometimes it's appropriate to speak of it more narrowly in terms of the monetary and non-monetary payments you directly provide your employees.

From your perspective as a manager, when discussing compensation costs the relevant concept is usually the actual dollars you must spend as a consequence of employing an employee. Some of those compensation costs may provide your employees with no value, or even negative value. For example, in the United States, Social Security taxes are partially paid by employers. Those taxes are direct compensation costs to you as the employer, but they aren't directly valued by your employees because the government gets the money. On-site childcare is an example of a compensation component that imposes direct monetary costs on you but might have zero or even negative value for some of your employees. Whereas your employees with children may value the component positively, those who are childless may view it as a negative if the noise from screaming children on the premises creates a distracting work environment. Your childless employees might also worry that your childcare expenditures may leave you with less money to spend on salaries.

When it comes to components of compensation that do not involve direct monetary costs to you as the employer (e.g., good weather, a collegial work environment, or a relaxed dress code) the potential for your employees to disagree over the value of those items is even greater. While everyone agrees that monetary bonuses are desirable, opinions may differ wildly concerning what constitutes a collegial work environment. What one employee finds fun and highly social, another might find annoying and distracting. Some of the compensation components that involve direct monetary costs are at least partially within your control as a manager (e.g., wages, salaries, and bonuses), whereas others are not (e.g., mandatory overtime pay, and Social Security taxes). The same is true for the components that do not involve direct monetary costs. For example, as a manager you can control the dress code, but you can't control the weather and the appeal of the geographic location. Both for the components you can control and for those you can't, you should remember that your employees may have different valuations for them. What is vice to one worker is virtue to another.

The fact that employees and firms have different perspectives on compensation, as explained above, does not mean that the two parties ignore each other's perspectives. Suppose that the government imposes a new payroll tax on employers, calculated as a percentage of the salaries that you pay your workers. Your first impulse might be to say that your employees are unaffected, because this is a tax directly on, and paid by, you as the employer. But that's wrong. By lowering your employees' compensation, you can pass along to your employees at least part of any increase in costs, and you might even fire some of them or cut back their hours if they become too expensive. So your employees should care about increases in your labor costs.

Similarly, consider a change in your employee's valuation of a job characteristic that is beyond your control and not paid for by you (e.g., desirability of the geographic location). Your first impulse might be to say that factors beyond your control as an employer, and that aren't paid for by you, should be of no concern to you. But that's wrong. Suppose, for example, that your employee's affinity for the geographic location of your organization is driven by the fact that her husband's family lives there. If your employee gets divorced, which causes her affinity for the geographic location to diminish, then it becomes easier for competing firms in different locations to poach her. So from the standpoint of retention and talent management, you need to think about compensation from your employees' perspectives as well as from your own.

1.2 What Is "Strategic" Compensation?

Although "compensation" is the most important word in the book's title, some of the other words also deserve comment. "Strategic" is a buzzword that is used often in the management domain, but it is not always used consistently, and the intended meaning is not always clear. There's also some inherent redundancy in the term "strategic ______". For example, consider the term "strategic human resources management". As opposed to "non-strategic" human resources management?? Any employer who manages the company's human resources "non-strategically" should be fired immediately, and the same goes for any employer who designs and administers the company's compensation system "non-strategically"! The adjective "strategic" can usually be omitted in management contexts, because it is implied and should be understood.

My decision to start the title with "strategic" was strategic, i.e., it was "purposeful", or "with an eye toward achieving some objective(s)". Even at the risk of some inherent redundancy, there is value in reminding the reader that compensation systems must be designed and managed in a manner that furthers the organization's *objective*, of which I will say more shortly.

1.3 What Is Talent Management?

Talent management is defined in various ways by different authors. For concreteness, let's consider the definition offered by Wikipedia (accessed August 20, 2019):

> *Talent management is the science of using strategic human resource planning to improve business value and to make it possible for companies and organizations to reach their goals. Everything done to recruit, retain, develop, reward and make people perform forms a part of talent management as well as strategic workforce planning.*

That definition is as reasonable as others I've seen (despite two instances of the inherently redundant "strategic"!), so let's just roll with it. If the crucial role that compensation plays in talent management isn't already clear to you, hopefully it will be by the end of this book.

1.4 What Is Your Organization's Objective?

The preceding definition of "talent management" uses the important phrase "to make it possible for companies and organizations to reach their goals." A proper discussion of compensation must begin with a clear statement of the organization's goals or objectives. For most of this book, we will assume that the organization seeks to improve its bottom line, i.e., to maximize profit. That's a bit of an oversimplification, and it requires some qualifications, but in most organizations, at least in the private sector, it's a reasonable approximation to what's really happening and should be happening.

In the phrase "maximize profit" the word "maximize" is more important than the word "profit". The approach we will take in this book is to clearly identify an objective that the organization is trying to maximize or "manage towards", and all decisions that are made can be evaluated by the extent to which they further that objective. Organizational objectives other than profit are found in nonprofit and public-sector organizations (Chapter 15). The concept of "profit" exists in both the short-term and the long-term. Generally, when we say that the organization maximizes profit we mean long-term profit. A business decision that would lead to a loss in the short-term (perhaps because of a large fixed cost that must be paid upfront) would be wise if it could be expected to yield a future profit stream that exceeds the initial fixed investment cost. When evaluating future expected profit, it is important to "discount" properly, as I explain in Chapter 8.

My claim that organizations maximize profit might seem strange for the following reason. If they're truly successful in maximizing profit, then there's no role for you as a manager to improve matters. If the maximum profit is already achieved, profit obviously can't go any higher. You should think about this in the following way. Profit maximization is something that good managers and successful firms aspire to do, but not always with perfect success. So even if they are close to achieving maximum profit, but not achieving it perfectly, there is still room for you as a manager to improve the situation through good business decisions. And even small increases in profit can be of major consequence, particularly in highly competitive industries in which organizations are constantly struggling to survive.

Finally, regarding this section's title, note that throughout the text I use the words "organization", "firm", and "company" more or less interchangeably, whereas I use the words "establishment" or "workplace" or "production unit" to describe an individual production unit at a particular physical location. For example, McDonald's is a firm (or organization) that has many establishments (or workplaces) worldwide, including a restaurant right down the street from my house. I also use the term "employer" to describe the entity that makes decisions about compensation, and in situations in which it matters, the context will make clear whether the term refers to a firm or to an establishment. In many workplaces, decisions about different components of the compensation system are made at different levels of the organization. For example, in a multi-establishment firm operating in many locations, the senior management at each establishment might have the discretion to set salaries at their locations, whereas the benefits packages are set at the company's headquarters and apply company-wide.

Another note on terminology is that I use "employee" and "worker" interchangeably even though a distinction between these terms is sometimes implied by the Fair Labor Standards Act of 1938 (Chapter 4). In other books the term "worker" sometimes implies hourly (or "non-exempt") employees who must be paid time-and-a-half for overtime, whereas the term "employee" implies both "exempt" (i.e., salaried people who are not eligible for overtime) and non-exempt statuses.

1.5 Who Cares about Compensation?

Workers obviously care a lot about compensation. For most of us, the bulk of our income is from paid work. Employers also care a lot about compensation, for several reasons. Compensation is a substantial chunk of total costs for most firms. Compensation also affects the behavior of a firm's current workers, so changing the compensation system may cause those workers to change their behavior in ways that may help or hinder the firm's objectives. Moreover, compensation affects the type of worker the firm can attract, so changing the compensation system may encourage some types of workers to leave the firm while making other types more eager to join the firm. For all of these reasons, you cannot be a competent manager without understanding the design, operation, and implications of compensation systems.

1.6 Who Receives Compensation, and Who Doesn't?

Most people get compensation, or did in their past, but several types of people get no compensation. Some people have marketable skills for which they could be compensated in paid employment, but they choose not to work (or, equivalently, they cannot find work at the compensation level they demand). Examples would be retirees and most of the unemployed. Other people have no marketable skills for which they could be compensated in paid employment. Examples include young children or unemployed people who are physically or mentally incapacitated and who lost skills, or never acquired them.

The preceding types of people don't get paid because they don't work, either voluntarily or involuntarily. Another type of person actually works (potentially very hard) but still doesn't get paid, namely slaves. Of course, we don't see much slavery in modern economies. But a close counterpart to slavery in the modern economy can be found in prisons. Prisoners, at least in the United States, typically have the option to not work, so compensation (albeit very meager compensation) must be offered to entice them to work. Prisoners have no "outside options" for employment; all of their employment options are inside the prison. Their only "outside" option is, perhaps, to not work at all. The consequence of this lack of outside options is exceedingly low wages, commonly pennies per work hour, which is legal because prisoners' labor is exempt from federal and state minimum-wage laws. Because slaves effectively lack even the most basic outside option of not working at all, their wages are zero. The examples involving slaves and prisoners offer a useful conceptual benchmark for thinking about compensation, because these individuals lack, or are severely restricted in, *mobility*. That is, they cannot "leave and go work for someone else", nor can they become entrepreneurs and start their own businesses.

Compensation is a "three-legged stool" that requires *desire* (to work for pay), *skills*, and *mobility*. If any of the three legs are missing, the stool collapses and there is no compensation. In the case of retirees, desire is missing, at least in the United States where mandatory retirement laws have been abolished since 1986. For retirees who have desire but are involuntarily retired because of mandatory retirement laws, mobility is effectively missing, because the law severely constrains it by preventing the worker from being employed in a company. However, even in this case mobility probably isn't completely absent, because the person still has informal work options, such as giving piano lessons out of his or her house. In the case of children or the incapacitated, skills are missing. And in the case of slaves, mobility is missing. So remember:

$$\textit{Compensation} = \textit{Desire} \times \textit{Skills} \times \textit{Mobility}$$

The legs of the stool are *multiplied*, not added, which means that if any one leg is zero, compensation is also zero. Like slaves and prisoners, wage-theft victims (see Chapter 2) actually work, potentially very hard, despite not getting paid. But wage-theft victims differ in some ways from slaves and prisoners. Whereas slaves and prisoners *expect* not to be paid, or expect to be paid very little, victims of wage theft are typically caught by the surprise of a broken promise. Thus, while the employment relationships involving slaves and prisoners can be very long lasting even in the wake of sustained absence of compensation, that doesn't normally happen with wage-theft victims. Workers suffering wage theft typically have mobility, and even a single instance of wage theft might be enough to prompt them to quit and work elsewhere. This is a central idea underlying section 10.5, which discusses the threat that CEOs and other executives might quit if the bonuses that they were promised are withheld, in whole or in part.

Note that for a particular worker with given levels of skills and mobility, the third leg of the stool (desire) determines the *minimum* compensation level that they'd be willing to accept. That minimum level is often called the worker's ***reservation wage*** or ***reservation compensation***. Workers who have low desire to work have high reservation wages. You'd need to pay them a big reward to entice them to take the job!

1.7 How Does Compensation Relate to Incentives and Productivity?

Compensation is the most powerful tool at a firm's disposal for creating and shaping worker incentives. When workers behave in ways that hurt the firm, most of the time the problem can be traced to poor incentives resulting from a badly designed compensation system. Compensation is both the disease and the cure. It's often the cause of the bad worker behavior. But usually the bad behavior, regardless of its cause, can be improved by fixing the compensation system.

Worker productivity also connects closely to compensation, because firms hire workers, and pay them, to be productive. Productivity is both a cause and a consequence of compensation. Productivity causes compensation because the main components of compensation are usually payments given in exchange for past productivity. But compensation causes productivity because often the design of the pay system affects workers' productivity (see Chapters 9 and 10). For example, sales workers who are paid on commission may feel compelled to sell more items than those paid hourly wages.

1.8 Four Recurring Themes

As noted in section 1.5, compensation affects the behavior of your current and prospective workers in two ways. First, the behavior of your current workers is affected by the design of the compensation system, and typically the behavior of greatest interest is *effort*, i.e., how hard people work. I refer to this as the *incentive effect*. Second, the types of workers who are attracted to your firm are affected by the design of the compensation system. I refer to this as the *sorting effect*.

The book's first recurring theme is that incentive effects and sorting effects arise when your company changes its compensation system. The key distinction we really want to draw between the two effects concerns changes in the behavior of your existing workers versus changes (via turnover) in who actually works for your firm. When we're focusing on your existing workers, typically the behavior of greatest interest is their level of effort on their job tasks, though other behaviors are sometimes also of interest. The following mnemonic might help you to remember and distinguish between the two effects.

The sorting effect concerns who shows up for your party, whereas the incentive effect concerns how they behave once they arrive.

A second recurring theme is that pressure from market competition largely dictates the *level* (i.e., overall generosity) of the compensation packages that you must pay your workers, whereas you have more control over the *design* of the compensation system, given its level of generosity.

A third recurring theme is that, as an employer, whether you like it or not, you are effectively forced to care about what your workers want. The reason derives from the "three-legged stool" mentioned earlier, and in particular the "mobility" leg. When hiring and retaining workers, you must compete with other employers in the labor market. And your workers are mobile, meaning that they are free to quit and go work for someone else or simply to do their own thing. So if you don't pay attention to what your workers want, and your competitors do, you're likely to lose them to your competitors. You are, therefore, forced by worker mobility and market competition to care about what your workers want, even if that is not your preference.

A fourth recurring theme is that *bargaining power* also determines the level of compensation, and the strength of your workers' outside markets affects their bargaining power vis-à-vis you, the employer. Chapter 14 elaborates.

1.9 What Constitutes "Fair" Compensation?

There is a lot of discussion in the popular press, in the political realm, and from some corners of academia, concerning the philosophical question of what constitutes "fair" compensation. Such discussions tend to become emotional and incite the expression of strong opinions. A related question concerns the fairness not of my compensation, as an individual, but of mine compared to my co-workers'. For example, as we'll discuss in detail in the next section, in the CSUEB business school where I work, as in most others, it is not uncommon for professors who do virtually identical work (i.e., teach the same number of classes, along with the standard expectations for research/publishing and committee work) to have vastly different compensation levels. Is that fair? Similarly, is it fair that CEOs earn many multiples of the pay received by even the hardest-working employees at the bottom of the organizational hierarchy?

Such questions of fairness will receive little attention in this book, which is not to diminish the value of studying them. They would take us too far into the realm of social and political philosophy. This is a book written for managers, and that should be our focus. The central compensation issues for you as a manager don't usually concern your workers' perceptions of fairness, except to the extent that those perceptions have implications for worker productivity and turnover. You should be focused on improving your firm's bottom line, which usually means paying as little as possible for a given amount of worker productivity. I do not mean to suggest that you should consider your workers' perceptions of fairness irrelevant! To see why, refer to the third recurring theme of section 1.8. Understanding what drives your workers' perceptions of fairness in compensation may enable you to design your company's compensation system in a way that "looks fairer", potentially reducing compensation costs and improving the bottom line.

But while we're on the subject of fairness in pay, I can't resist saying a few words about CEOs. Because, after all, CEOs get paid a lot. I mean, *really*, a lot! Elon Musk, CEO of Tesla, received over $513 million in total compensation in 2018. On the other hand, poor Stephen Angel, CEO of Linde PLC, earned a mere fraction of that in 2018, at just over $66 million, though even that paltry sum still put him among the 10 highest-paid CEOs in 2018. So there is enormous variance across companies in what CEOs get paid, though, no matter how you slice it, they still get paid way more than the rest of us, and usually way more than all of their employees. *Are they paid too much?* If you're ever asked that question, your immediate response should be, "Relative to *what*?" We can't address the question of whether CEOs are paid "too much" unless we agree on the correct benchmark for comparison. Opinions on the proper benchmark may vary, but I'll give you the quick version of my own opinion, just so that we can lay the issue to rest and move on to more pressing matters . . .

A CEO is paid "too much" if the company could readily find a new CEO who's just as good as the current CEO and who will create just as much value for the company, but who is willing to do the job for less compensation. A question I'll leave you to ponder is, if such a person is indeed readily available, why hasn't the company already made the switch?

1.10 Secrecy versus Full Disclosure of Compensation

At the reference desk on the second floor of the CSUEB library, if you provide your university identification card as collateral, you can request the "faculty salary data folder". It names each member of the CSUEB faculty and what their cash compensation was in the previous year. The same information can be found online. After a few years of noble resistance following my arrival at CSUEB in 2008, I finally broke down and satisfied my curiosity by spending part of an afternoon with that folder, in a lounge in the library. In an email instructing me on the location of the folder and how to access it, a colleague of mine in the business school's Management Department (who, incidentally, felt underpaid) wrote, "Be prepared to be amazed."

I was, indeed, amazed! What was so amazing? Mainly the gross mismatch between the salary figures and my own perceptions of the productivities of the individual faculty members. For example, the highest-paid professor in the business school at that time wasn't doing any research, had never published in a major journal, and wasn't making any major contributions in either teaching or administrative work. In contrast, there were individuals who excelled on all of those dimensions despite being near the bottom of the salary scale.

A short time after I discovered the salary data, a senior professor independently sent an email to the entire business school faculty, including the salary information as an attachment. Her goal was to stir up trouble, and she sure succeeded. Some professors became upset and disgruntled, and one of the business school's best professors (who happened to be among the lowest-paid) actually left CSUEB for another university. The departure of the talented professor was the first time I had witnessed, firsthand, the potential costs to organizations (in terms of low morale, and turnover) of having internal compensation information publicly available and known.

Prior to joining CSUEB, I spent seven years on the faculty at Cornell, and there the compensation data were private. Only the dean knew what individual faculty were getting paid. Because the dean at the time knew of my interest in compensation, he revealed to me once in conversation how large the variance in salaries was within the school, without revealing any individual salaries. That is, if you consider the highest-paid and lowest-paid professors of a given rank, the highest-paid one receives K times the salary of the lowest-paid one. Although I will not disclose that K factor here, the number he reported was surprisingly large.

Recalling this story several years later while sitting on the second floor of CSUEB's library with the faculty salary folder, I decided to compute K for the CSUEB business school. The number was definitely not small, but it was way smaller than the corresponding number from Cornell. This is no surprise. Organizations like CSUEB that have public compensation information tend to have more "compressed" compensation distributions (i.e., there is less distance between the highest-paid and lowest-paid worker) than those like Cornell where almost everyone's pay is a secret. There are other differences between the two institutions that likely play a larger role in explaining the difference in K values (e.g., CSUEB is unionized, and in Chapter 5 we'll discuss the important role that unions play in compensation compression), but the difference in information disclosure likely contributes.

The fact that compensation disclosure goes hand-in-hand with compensation compression raises the question, does the disclosure cause the compression, or does the compression cause the disclosure? The answer is, both! When you disclose salary information to your employees, two considerations lead to compensation compression. First, your lowest-paid workers will learn that they are the lowest-paid, and they will either aggressively lobby for raises or they will quit, and either of those outcomes contributes to compression by raising the compensation of the lowest-paid workers that are currently employed by your organization. Second, your highest-paid workers will learn that they are the highest-paid workers, and they will also learn by what factor their compensation exceeds that of their lowest-paid co-worker (i.e., they will learn the value of K). This information may make your highest-paid workers less inclined to lobby hard for even larger raises, lest they be perceived as greedy. That would also contribute to compression by lowering compensation growth among your highest-paid workers. Thus, disclosure causes compression. But the compression also facilitates disclosure, because the more similar everyone's compensation becomes, the less tension arises from envy and embarrassment, and the more comfortable all of your workers feel sharing information about their compensation.

Neither disclosure nor non-disclosure is the best policy for a firm in every setting. You must weigh the costs and benefits of both approaches and decide which is best in your particular organization. In some settings, such as the public sector, disclosure is mandated by law, as is the case at CSUEB. In the private sector, some companies voluntarily disclose compensation information (e.g., see the Buffer case discussion at the end of this chapter). Suppose that your organization decides that secrecy is the best approach (e.g., there might be significant compensation dispersion in your organization, and you and the rest of the senior management want to hide this information to prevent costly turnover) but your highest-paid worker decides to ignore your wishes and brags to co-workers about his or her compensation. Can you prevent this?

Probably not. Your workers' voluntary disclosures are difficult to police, and some states even have laws protecting workers' rights to disclose their compensation. For example, section 232 of California's wage and hour law prohibits employers from requiring their workers to keep their compensation secret or from punishing workers who disclose the information. The law also prohibits employers from requiring workers to sign a waiver that explicitly relinquishes the right to disclose the employee's compensation. When it comes to disclosing the information to an external party (e.g., to a prospective employer) it is virtually impossible to monitor what the worker does anyway, regardless of the law. So the main value of the law is in protecting workers who make disclosures internally, since those disclosures are much more easily detected (and potentially punished) by the employer than disclosures to external parties.

1.11 Lessons for Managers

As a manager, you must clearly understand your organization's objective and, underneath that umbrella, the objectives of your own division, department, project, or product line. You must also understand how those smaller objectives beneath the umbrella relate to and support the broader organizational objective, which in most organizations is to maximize profit. All of your decisions concerning the design and management of your organization's compensation system should be made so as to advance those objectives to the greatest extent possible. That is the essence of strategic compensation and talent management. Successfully achieving that requires always bearing in mind that the design of compensation affects worker behavior, particularly their productivity and their desire to join and remain with your organization. You should anticipate how your workers, or potential workers, will respond to the features of your organization's compensation system, and keep those behavioral responses firmly in mind when considering changes to the compensation system.

View your employees' compensation broadly, not in terms of this worker making $75,000 and that one making $83,500 but in terms of the full set of job characteristics that your workers value. Also recognize that your workers may have very different preferences, and this is particularly the case for non-monetary components of compensation. The location of your office in downtown Manhattan might be considered a prized perk by one of your workers but a major inconvenience by another.

Think about talent management in terms of the three-legged stool of compensation. For example, pay close attention to anything that impedes or enhances your workers' mobility, because such developments have implications for what you can (and must) pay them. Your workers' accumulated skills also have important implications for their compensation, and, because you're competing with other employers who could potentially steal your workers away, you need to pay attention to how marketable your workers' skills are outside of your organization. You must make such evaluations with some regularity, because workers' skill sets evolve, as do the needs of competing employers in the market. The preceding lessons, i.e., thinking about compensation broadly, recognizing differences in preferences among your workers, and heeding the three-legged stool, are crucial for talent management and retention, because workers who are unhappy with their compensation relative to what they could earn elsewhere are likely to quit.

There's no easy answer to the question of whether compensation information should be public or secret among your employees. Both approaches come with their own plusses and minuses, and you have to determine which path would work best in your organization. Chances are you'll never have to make that decision for the entire organization, or even for part of it. Usually managers "inherit" these practices once they become managers. Your organization probably made that decision before you became a manager there, and now your job as a manager is to play the hand you've been dealt. First and foremost, be sensitive to the significant implications that disclosure has for worker morale and for the degree of compensation compression within the organization. If your organization's disclosure-versus-secrecy policy predated your tenure as manager, you can always lobby the most senior management if you believe that there's a compelling case for changing current policies. Even without changing your firm's policy, you might have authority to tip the balance in one direction or the other. For example, even when employees have salary information, there may be other components of compensation that are less visible, and emphasizing those more heavily in the compensation packages you offer will allow you to take things one step closer to an environment of secrecy. Such arguments also apply in reverse if you manage in an environment with secrecy, and you believe that greater transparency among the workers you manage would benefit your organization.

Be aware of the distinction between nominal and real compensation and of the role of inflation in eroding nominal compensation, to be discussed in the following appendix. Familiarize yourself with the various price indexes available on the BLS website (see Further Resources at the end of this book), and understand how to use those price indexes to convert back and forth between nominal and real compensation levels.

Appendix: Nominal versus Real Compensation

Nominal monetary compensation refers to the actual amount of monetary compensation paid or received, stated in dollars, pesos, euros, or some other monetary unit, without regard to how "useful" that monetary compensation is from the standpoint of what could actually be purchased with it. Most of the time, when we speak about compensation day-to-day, we are speaking about nominal compensation. For example, "I just got a job offer, and it pays $73,000." Or, "The California minimum wage was recently increased from $10.50 to $11 per hour, for firms employing at least 26 workers." Or, "My end-of-year bonus was over $15,000!" All of those are statements about nominal compensation.

When it comes to making decisions, however, the concept of *real* monetary compensation, which takes into account what can actually be purchased with the funds, is usually more important. After all, compensation is desirable only because it will eventually be used to buy stuff that workers want, so the question of exactly how much stuff can be bought with it is important. If the prices of goods and services that people want to buy were to remain unchanged, then there would be no difference between nominal and real monetary compensation. But prices do change. They change over time, and they change across locations (e.g., if we move from San Francisco to Kansas, we notice a big drop in the cost of living).

In 1938, the federal hourly minimum wage was \$0.25, whereas 81 years later, in 2019, it was \$7.25. That's a 2900% increase in the nominal federal minimum wage. That may sound impressive, until you learn that the average price of, say, a movie ticket was about \$0.25 in 1938 and about \$9 in 2019. That's a 3600% increase in the average price of movie tickets. If movie tickets were the only things that minimum-wage workers ever wanted to buy, then they'd be better off living in 1938 than in 2019. In 1938, in a minimum-wage job, one hour of work earned you one movie ticket. But in 2019, in a minimum-wage job, one hour of work would earn you only about 80.6% (i.e., \$7.25 / \$9) of one movie ticket.

Obviously, people want to buy many things other than movie tickets. But the basic argument I just made continues to hold when our attention extends beyond movies. The prices of most goods and services, not just movies, increased significantly from 1938 to 2019. So to accurately compare the minimum-wage levels between those two years, those price increases must be considered. Considering the price changes of all possible goods and services would be an overwhelming and virtually impossible task. So what's done is to rely on a "market basket" of goods and services – a representative set of things on which people spend their money. The result is the Consumer Price Index (CPI), which is computed and released monthly by the Bureau of Labor Statistics (BLS). The BLS computes several different versions of the CPI, and all of them can be found, with detailed explanations about their construction, on the following website:

www.bls.gov/cpi/

Let's consider an example of how price indexes like the CPI can be used to convert nominal compensation data to real compensation data. From the preceding website, you can obtain the April 2019 news release for the CPI. Table 1 of that document (reproduced here as Table 1.1, with the numbers of particular interest highlighted in boldface) displays the CPI "for all urban consumers", known as the CPI-U.

The CPI-U was 255.548 in April 2019. You might be wondering how to interpret the number 255.548. What exactly does it mean? What are its units? Is the number 255.548 large or small? Don't worry about these questions! The particular values of this index, or any index, aren't very interesting. Rather, *comparisons* between the index values in different years are interesting. That's why Table 1.1 also displays the CPI-U in the prior month (i.e., March 2019, when it was 254.202) and the prior year (i.e., April 2018, when it was 250.546). The CPI-U increased by about 2% between April 2018 and April 2019, which is the inflation rate for consumer prices during that period. That is,

$$[(255.548 \ / \ 250.546) - 1] \times 100 \approx 2\%,$$

where I use the symbol "≈" rather than "=" to remind you that the numbers aren't perfectly precise because of rounding. Table 1.1 (very slightly) rounds the inflation rate to 2%.

Suppose that one of your workers had a salary of \$53,259 as of April 2018, and the same salary as of April 2019. Although the nominal salary remained at \$53,259 in both years, the real salary decreased, because, in the wake of an increase of about 2% in consumer prices, compensation of \$53,259 doesn't go as far in 2019 as it did in 2018. For *real* compensation to remain unchanged between the two years, your worker would need a salary increase of 2%, i.e., the nominal salary would need to increase from \$53,259 to about \$54,324. A salary of \$53,259 in 2018 is equivalent to one of \$54,324 in 2019, in terms of what your worker can actually *buy* with that income. There's usually some inflation in consumer prices from one year to the next, which means that when nominal pay remains unchanged, real pay actually falls. This also means that federal and state minimum wages (see Chapter 4), which are legislated by the government in nominal terms and remain unchanged in nominal terms from year to year, actually diminish in real terms from year to year, until the government enacts another nominal increase.

Although it's historically rare, deflation (as opposed to inflation) in consumer prices sometimes occurs. The most dramatic example of deflation in the United States occurred during the early years of the Great Depression. From 1930 to 1933, consumer prices dropped by around 10%. During periods of deflation, real compensation increases when nominal pay remains unchanged. In fact, real compensation might increase even in the wake of nominal pay *cuts*, as long as the nominal cuts are milder than the concurrent deflation. Again, this situation is historically unusual.

The CPI can be used to figure out what a current (nominal) compensation figure would be worth at a particular date in the past. To do this, we just reverse the procedure from the preceding example. Suppose that your worker is paid a nominal salary of \$61,358 in April 2019 and you wish to know what that salary would be in real terms, in "2018 dollars". Just compute

$$\$61{,}358 \times (250.546\ /\ 255.548) \approx \$60{,}157.$$

So a nominal salary of \$61,358 in 2019 is equivalent to a salary of \$60,157 in real 2018 dollars.

Alternatively, we could (re)compute what a 2018 (nominal) salary of \$53,259 would be in real terms, in "2019" dollars. We already know that the answer is about \$54,324, because we found it above simply by applying an approximate inflation rate of 2% to the 2018 nominal salary. But here it is again, just so you can see that the computation has the same form as the preceding equation, but switches the numerator and denominator of the CPI ratio, i.e.:

$$\$53{,}259 \times (255.548\ /\ 250.546) \approx \$54{,}322.$$

The slight difference between this number and \$54,324 is due to rounding, because we found \$54,324 by applying a (rounded up) inflation rate of 2%, whereas the actual inflation rate is slightly above 1.996% but still a bit shy of 2%. The preceding two equations suggest a more general formula. Suppose that $Salary_t$ represents a

Table 1.1 Consumer price index for all urban consumers (CPI-U): U.S. city average, by expenditure category, April 2019 [1982-84=100, unless otherwise noted]

Expenditure category	Relative importance March 2019	Unadjusted indexes April 2018	Unadjusted indexes March 2019	Unadjusted indexes April 2019	Unadjusted percent change April 2018-April 2019	Unadjusted percent change March 2019-April 2019	Seasonally adjusted percent change Jan. 2019-Feb. 2019	Seasonally adjusted percent change Feb. 2019-March 2019	Seasonally adjusted percent change March 2019-April 2019
ALL ITEMS	100	**250.546**	**254.202**	**255.548**	2	0.5	0.2	0.4	0.3
Food	13.315	253.209	257.724	257.708	1.8	0	0.4	0.3	−0.1
Food at home	7.278	240.129	242.555	241.878	0.7	−0.3	0.4	0.4	−0.5
Cereals and bakery products	0.961	271.766	277.275	276.64	1.8	−0.2	0.7	0.3	−0.1
Meats, poultry, fish, and eggs	1.59	251.922	249.062	249.203	−1.1	0.1	0.2	−0.2	−0.2
Dairy and related products	0.723	216.906	217.531	217.51	0.3	0	0.3	0.6	0.1
Fruits and vegetables	1.301	296.863	305.399	303.503	2.2	−0.6	0.9	1.6	−0.9
Nonalcoholic beverages and beverage materials	0.877	167.041	172.06	170.727	2.2	−0.8	0.7	0	−0.9
Other food at home	1.826	210.632	211.55	211.078	0.2	−0.2	0.3	0.1	−0.6
Food away from home[1]	6.037	274.393	281.887	282.798	3.1	0.3	0.4	0.2	0.3
Energy	7.466	218.83	211.724	222.499	1.7	5.1	0.4	3.5	2.9
Energy commodities	4.128	246.387	231.228	253.547	2.9	9.7	1.5	6.2	5.4
Fuel oil	0.113	293.991	287.496	291.424	−0.9	1.4	2.6	2.1	1.3
Motor fuel	3.94	242.056	226.616	249.55	3.1	10.1	1.5	6.4	5.6
Gasoline (all types)	3.853	240.962	225.282	248.499	3.1	10.3	1.5	6.5	5.7
Energy services	3.339	201.245	202.461	201.341	0	−0.6	−0.8	0.3	−0.1
Electricity	2.58	209.215	210.937	210.525	0.6	−0.2	−0.3	0.4	0
Utility (piped) gas service	0.759	174.318	174.089	171.004	−1.9	−1.8	−2.4	−0.1	−0.8
All items less food and energy	79.218	257.025	261.836	262.332	2.1	0.2	0.1	0.1	0.1
Commodities less food and energy commodities	19.566	145.131	144.994	144.851	−0.2	−0.1	−0.2	−0.2	−0.3

Apparel	3.052	129.365	125.785	125.542	−3	−0.2	0.3	−1.9	−0.8
New vehicles	3.722	146.069	147.788	147.84	1.2	0	−0.2	0.4	0.1
Used cars and trucks	2.405	139.368	140.513	140.458	0.8	0	−0.7	−0.4	−1.3
Medical care commodities	1.693	380.479	379.568	381.256	0.2	0.4	−1	0.4	0.9
Alcoholic beverages	0.965	248.818	252.087	251.758	1.2	−0.1	0.2	−0.1	−0.2
Tobacco and smoking products	0.666	1,056.87	1,103.24	1,105.20	4.6	0.2	0.5	1.6	−0.1
Services less energy services	59.652	326.252	334.518	335.468	2.8	0.3	0.2	0.3	0.3
Shelter	33.238	305.716	315.135	316.263	3.4	0.4	0.3	0.4	0.4
Rent of primary residence	7.916	316.763	327.513	328.678	3.8	0.4	0.3	0.4	0.4
Owners' equivalent rent of residences[2]	23.954	312.935	322.482	323.426	3.4	0.3	0.3	0.3	0.3
Medical care services	6.973	517.228	528.575	529.371	2.3	0.2	0	0.3	0.2
Physicians' services	1.717	381.076	381.045	382.15	0.3	0.3	0.1	−0.4	0.2
Hospital services[3]	2.324	331.289	336.648	335.198	1.2	−0.4	−0.7	0.3	−0.5
Transportation services	5.936	321.437	323.982	325.114	1.1	0.3	−0.1	0	0.1
Motor vehicle maintenance and repair[1]	1.132	284.367	294.226	295.266	3.8	0.4	0.4	0.8	0.4
Motor vehicle insurance	2.388	562.749	573.087	570.492	1.4	−0.5	0.1	−0.1	−0.2
Airline fares	0.682	273.817	259.698	268.767	−1.8	3.5	0.5	−0.6	−0.1

Footnotes:

(1) Not seasonally adjusted.

(2) Indexes on a December 1982=100 base.

(3) Indexes on a December 1996=100 base.

NOTE: Index applies to a month as a whole, not to any specific date.

Last modified: May 10, 2019

Source: Bureau of Labor Statistics (www.bls.gov/cpi/)

nominal salary in year t, and CPI_t represents (some version of) the CPI in year t. To convert nominal salaries in year t to real salaries in year $t+k$, use the following general formula:

$$RealSalary_{t+k} = Salary_t \times (CPI_{t+k} / CPI_t).$$

Note that k can be either positive or negative. If k is positive, then $t+k > t$, so we are converting nominal salaries in one year to real salaries in a future year. If k is negative, then $t+k < t$, so we are converting nominal salaries in one year to real salaries in an earlier year. Suppose that k is zero. In that case, $t+k = t$, so we are "converting" nominal salaries in one year to real salaries in that same year. Obviously, there is nothing to convert in that case, because the salaries are already measured in the dollars of year $t+k$. So, in the formula, CPI_{t+k} equals CPI_t, and the ratio of those two CPI values equals 1. Then the general formula reduces to:

$$RealSalary_{t+k} = Salary_t \times 1.$$

Recalling that $k = 0$, this equation reduces to:

$$RealSalary_t = Salary_t,$$

so that the real and nominal salaries are identical, just as we expected.

Let's do two more examples. For both of them, recall that the federal (nominal) hourly minimum wage in 1938 was \$0.25, whereas in 2019 it is \$7.25. Also note that the CPI in 1938 was 14.1, whereas in 2019 it is 255.548. For the first example, let's figure out what the 1938 minimum wage would be in real, 2019 dollars. Applying the general formula, we get:

$$\begin{aligned} RealMinimumWage_{2019} &= MinimumWage_{1938} \times (CPI_{2019}/CPI_{1938}), \text{or} \\ RealMinimumWage_{2019} &= \$0.25 \times (255.548/14.1), \end{aligned}$$

which is about \$4.53.

This means that \$0.25 in 1938 had about the same purchasing power as \$4.53 in 2019. This also means that, between 1938 and 2019, the federal minimum wage has increased in both nominal *and* real terms. In nominal terms it has increased very dramatically from \$0.25 to \$7.25, and in real terms (in 2019 dollars) it has increased much more modestly, from \$4.53 to \$7.25.

For the second example, let's figure out what the current (nominal) 2019 federal minimum wage would be in real 1938 dollars. Again, apply the formula:

$$\begin{aligned} RealMinimumWage_{1938} &= MinimumWage_{2019} \times (CPI_{1938}/CPI_{2019}), \text{or} \\ RealMinimumWage_{1938} &= \$7.25 \times (14.1/255.548), \end{aligned}$$

which is about \$0.40.

This means that \$7.25 in 2019 had about the same purchasing power as \$0.40 in 1938. This also means that, between 1938 and 2019, the federal minimum wage has increased in real terms (in 1938 dollars) from $0.25 to $0.40. The two examples convey the same information but stated in a different way. In the first example, 1938 nominal compensation data are *inflated* up to real 2019 dollars, whereas in the second example, 2019 nominal compensation data are *deflated* down to real 1938 dollars.

The preceding discussion is based on the CPI, but all price indexes work in basically the same way. The idea is to create "apples-to-apples" (rather than apples-to-oranges) comparisons between compensation numbers. In the preceding two examples, comparing the nominal minimum wage between 1938 and 2019 would be an apples-to-oranges comparison, because prices for consumer goods and services were so much lower in 1938 than in 2019. Comparing real minimum wages (in either 1938 or 2019 dollars) between the two years, however, is the desired apples-to-apples comparison.

Two other price indexes that are relevant for compensation come from the National Compensation Survey (NCS), which is conducted by the BLS. The first is the *Employment Cost Index* (ECI), which measures the change in labor costs over time. The second is the *Employer Costs for Employee Compensation* (ECEC) index, which measures the level of average costs per hour worked. Both price indexes are measured quarterly, i.e., four times per year. Both indexes, like the CPI, are available from the BLS website (as shown above and, again, in Further Resources at the end of this book).

In all of the preceding discussion, an apples-to-oranges problem arises when compensation levels are compared between two points in time that have different consumer prices. But an apples-to-oranges problem can also arise when compensation levels are compared between two geographic regions, even if that comparison is made at the same point in time. This is because consumer prices differ across locations as well as over time. If two prices are compared that differ both in location and in time period, that creates even more complications and greater distance from an ideal apples-to-apples comparison.

The basic logic for creating apples-to-apples comparisons in such cases mirrors what we have already shown for dealing with compensation at different points in time. Basically the compensation levels must be adjusted using some type of price index. The US Bureau of Economic Analysis (BEA) produces *regional price parities* (*RPPs*) that are useful for this purpose. RPPs measure the difference in price levels across states and metropolitan areas in a particular year. They are expressed as a percentage of the overall national price level, so RPPs above 100 indicate areas with above-average prices, and those with RPPs below 100 indicate areas with below-average prices. For example, across large metropolitan areas, the highest RPP (based on a June 2017 BEA news release) was 186.0, for the San Francisco-Oakland-Hayward metro area. The lowest was 78.7 for the Cleveland-Elyria area in Ohio.

Using the RPPs from the BEA, the BLS produces a price-adjusted wage index that allows for comparisons of wages across metropolitan statistical areas. The starting point that the BLS uses is the Occupational Employment Statistics (OES) program that produces employment and wage estimates for more than 800 occupations and allows for comparisons of nominal wages across geographic regions. Those values, alone, don't account for differences in prices across the geographic regions, so the RPPs from the BEA are introduced to make those adjustments. What results is a set of price-adjusted wages, for a large set of occupations, that more accurately represents the real value of earnings for comparisons across geographic regions.

Case Discussion 1: Buffer, Inc.

Buffer, Inc., is a company that was started in San Francisco in October 2010 to create and develop the Buffer software application. The software, available as both web and mobile applications, allows users to manage and schedule posts to their social media accounts (e.g., Facebook, Twitter, and LinkedIn). According to the company's website, "Buffer is the best way to drive traffic, increase fan engagement and save time on social media." As of July 2019, Buffer was comprised of 85 employees, including co-founder Joel Gascoigne (CEO). Employees work remotely from 50 cities around the world. Buffer offers an online spreadsheet showing each worker's first name, job title, location, and, ever since late 2013 . . . annual salary!

The highly unusual decision for a private-sector company to post such information online is part of a concept Buffer calls "Open Salaries". In addition to the individual salaries, the website makes public the simple formula that is used to compute each salary. This high level of transparency aligns with the second of Buffer's 10 Values, namely "Default to Transparency". The simple formula used to compute salaries is referred to on the company website as a "living document", and indeed there have been three versions of the formula in the four years since it was introduced. Version 3.0, as Buffer calls it, aimed to simplify its predecessor, and was designed with four priorities in mind: (1) It should be simple enough for anyone to use; (2) take-home pay shouldn't drop for any worker as a result of the new formula; (3) it needs to be flexible enough to adapt and evolve; (4) compensation packages must remain competitive. The resulting formula as of September 2018 was:

$$\begin{aligned} Salary = {} & San\ Francisco\ 50\% Benchmark \times Cost\ of\ Living\ Multiplier \\ & \times Role\ Multiplier \times Experience\ Factor. \end{aligned}$$

Buffer's website elaborates on the four variables in the formula, but the basic idea is that salaries start with the San Francisco labor market as a benchmark and are then adjusted for a worker's geographic location, role in the company, and amount of experience. The role multiplier is the primary channel via which discretion in pay can be incorporated. The latest formula eliminated an automatic "loyalty" raise of 3% per year for all employees, because it was financially unsustainable and made the compensation system less flexible, and, therefore, less able to achieve other objectives. It also phased out a feature of the original formula that allowed employees to choose either equity as a component in their compensation or an additional $10,000 in annual salary.

Questions

1. What are the advantages and disadvantages of making everyone's salary at Buffer public on the company website?
2. Suppose that, instead, the company had disclosed the salary information to all of its workers internally, but not to the outside world. For example, since 1986, Whole Foods has allowed all of its employees to learn the salary of any worker in the company, simply by visiting the HR office. In what ways would such an approach be better for Buffer than its current policy of full publicity, and in what ways would it be worse?
3. What are the advantages and disadvantages of using a formula like Buffer's to compute compensation for every worker in the firm, regardless of whether or not the formula is divulged to employees or the general public?
4. If a formula is used, what are the advantages and disadvantages of divulging the formula to employees? To the general public?
5. On the company website Buffer's finance team describes the salary formula as a "work in progress", and it has gone through three versions in four years.
   (a) What are the advantages and disadvantages of regularly updating and fine-tuning the salary formula?
   (b) When the salary formula was created, and each time it was revised, it was done in such a way that each employee's salary was guaranteed either to increase or stay the same. Was that a wise decision? If so, why?
   (c) If you were on Buffer's finance team, what would be your primary concerns about the salary formula? Can you recommend any improvements to the formula to address those concerns?
6. In which industries, or particular firms, would Buffer's approach be particularly problematic, and why?
7. According to Buffer's finance team (the group that created and maintains the salary formula), when it sought external input and advice, people tended to ask, "Why would you pay your team members with more generosity than you need to?" Their response, as stated on Buffer's website, was, "because it's the right thing to do". Is this an appropriate response in light of the lessons from Chapter 1? Explain.

8. If you have read Chapters 6 and 7, explain how you might use the methods in those chapters to develop a better formula. What type of data would you need, how would you obtain the data, and how would you conduct the analysis?

Further Reading

Bamberger, P. and Belogolovsky E. (2010). “The Impact of Pay Secrecy on Individual Task Performance.” *Personnel Psychology*, 63, 965–96.

Belogolovsky, E., Bamberger, P., Alternman, V., and Wagner D. T. (2016). “Looking for Assistance in the Dark: Pay Secrecy, Expertise Perceptions, and Efficacious Help Seeking Among Members of Newly Formed Virtual Work Groups.” *Journal of Business and Psychology*, 31, 459–77.

Cappelli, P. (2008). “Talent Management for the Twenty-First Century.” *Harvard Business Review*, 1–9.

Colella, A., Paetzold, R. L., Zardkoohi, A., and Wesson M. J. (2007). “Exposing Pay Secrecy.” *Academy of Management Review*, 32, 55–71.

Collings, D. G. and Mellahi K. (2009). “Strategic Talent Management: A Review and Research Agenda.” *Human Resource Management Review*, 19, 304–13.

Lewis, R. E. and Heckman R. J. (2006). “Talent Management: A Critical Review.” *Human Resource Management Review*, 16, 139–54.

Tekleab, A. G., Bartol, K. M., and Liu W. (2005). “Is it Pay Levels or Pay Raises that Matter to Fairness and Turnover?” *Journal of Organizational Behavior*, 26, 899–921.

2 Compensation Contract Failure and Wage Theft

Recall our working assumption that employers make compensation decisions to maximize profit. This means paying workers as little as possible to achieve a given level of productivity. That might sound stingy and cruel, and it raises questions such as those in section 1.9 that I'm generally sidestepping in this book, but that's the reality. For example, if Oracle can hire a new computer programmer for an annual salary of $120,000, or an alternative programmer who is equally productive but who demands only $100,000, then Oracle makes more profit by hiring the latter programmer than by hiring the former. For that matter, Oracle would be even happier if it could hire an equally qualified programmer who demands no salary at all and will work for free. Of course, it will be virtually impossible to find such a programmer, because there are enough competing employers willing to pay five-or-six-figure salaries that competent programmers (all of whom have mobility, as discussed in section 1.6) do not have to work for free.

2.1 Compensation Contract Failure and Wage Theft

So even though Oracle would love to pay its programmers salaries of $0, this is infeasible due to mobility and an active labor market for programmers, even ignoring the minimum-wage regulations we'll discuss in Chapter 4. But Oracle, like all other firms, is still always looking for ways to pay less for more. One way to pay less is simply to steal the money, and there are some unscrupulous employers who do just that. Although such events are relatively rare in a well-functioning labor market, they obviously impose significant costs on workers when they happen. I will use the term "wage theft" to describe such instances, though keep in mind that the theft can be of any component of compensation, not only cash components like wages and salaries.

Wage theft is a relatively rare occurrence that typically goes unmentioned in compensation books, and it is even more unusual to see it highlighted as early in the discussion as I do here. But I think it's instructive. Sometimes the best way to get a feel for how something works is to study it when it's broken. In the case of wage theft, what is "broken" is the compensation contract between a worker and an employer. Studying the nature of such ruptures, and why they are so rare in practice, will yield some important insights that recur throughout the book.

2.2 Timing of Compensation

Think about your first day of work in the most significant job you have ever held. Did you find a paycheck waiting on your desk that day? Probably not. If your job was like most jobs, you had to wait awhile – days, weeks, or a month – before getting your first check. Although there are exceptions like signing bonuses, the usual convention is that employees work for a period of time and then receive payment for their past labor. That's why, when people quit or are fired, arrangements must be made for them to collect their "last check". The reason for this timing convention is that it protects employers from the risk that workers will "take the money and run". But it raises the possibility that unscrupulous employers might "take the labor and run" by purposely denying workers payment for services they have already rendered. For an example of this for low-wage workers in the restaurant industry, see the Flying Pig case discussion.

So whichever party (the worker or the employer) "puts stuff on the table" first runs the risk of having that stuff stolen by the other party. The problem arises because the compensation and the labor services are not simultaneously exchanged . . . one or the other is put on the table first. Simultaneous exchange, although it would eliminate the risk of theft, is rare and is usually impractical. The reason is that most work involves a sequence of tasks spread out over time.

For example, the last time I taught the MBA compensation course at CSUEB, it was as a "teaching overload" for extra compensation. The intensive course met for five consecutive Saturdays, with an eight-hour meeting each day. I was paid in two installments, the first of which occurred around the midpoint of the course, and the second of which occurred at the end, once final grades were submitted. CSUEB's first opportunity for wage theft came at the midpoint of the course, when it could have chosen to withhold my first check. Doing so would have been unwise. I might have become upset and stopped teaching the course, which would have imposed heavy costs on the students and reputational costs on CSUEB. Failing to pay me my second check would have been more tempting for CSUEB, and safer, because at that point I would have submitted course grades. I'm happy to report that I received both checks on schedule, and I will soon explain the reasons why wage theft was never a serious concern.

This example highlights the impracticality of simultaneous exchange. For true simultaneous exchange, CSUEB would not have given me two checks, but rather five (one for each Saturday), to ensure that every class meeting was compensated exactly when it occured. But even that wouldn't suffice. Because when should I get the check for a given meeting? If I get it at the beginning of the meeting, I might take the money and run, collecting compensation for an eight-hour lecture I never delivered. If I get it at the end of the meeting, I run the risk of suffering wage theft for that day, because I might be denied pay for the lecture I just gave. If I get it after the fourth hour of eight, I might take the money and run for the rest of the day, collecting compensation for eight hours after having worked only four. True simultaneous exchange would require someone from CSUEB's payroll office to continuously drop pennies into a piggybank, in real time, during the eight-hour class meeting. Needless to say, I would have found that quite distracting, though greatly amusing. You see how ridiculous this gets, and why simultaneous exchange is rarely a practical solution to the wage-theft problem.

Sometimes we can get a step closer to simultaneous exchange by shortening the window during which the exchange occurs. For example, CSUEB could have chosen to pay me the entire amount of money only after I submitted course grades. That would have exposed me to the risk of teaching an entire course without pay. By splitting the pay into two parts and paying the first part at the midpoint of the course, I was only exposed to the risk of teaching half a course without pay, because if CSUEB failed to pay me at the midpoint of the course, I could have cut my losses and walked away.

Collateral could also be used to mimic a situation with simultaneous exchange. In that case, the party who suffers the greatest risk of theft asks the other (lower-risk) party to post some collateral, i.e., give something of value to the higher-risk party that will be returned to the lower-risk party only when the compensation is paid. Because the lower-risk party wishes to reclaim the collateral, the compensation gets paid as agreed. This solution to wage theft might create new problems. For example, the collateral itself might get stolen. Ideally the collateral would be something that the lower-risk party (who issues the collateral) values more than the compensation to be paid, and that the higher-risk party (who receives the collateral) values less than the compensation to be paid.

2.3 Solutions to the Wage-Theft Problem

The wage-theft problem is one of *contract failure*. That is, the employer reneges on a contract with the worker that involves an exchange of labor services and compensation at some agreed-upon price. The contract is sometimes explicit, or formal, meaning it is written down on paper, probably signed by both parties, and could be presented to a court as proof that the agreement existed and was broken. Other times the contract is informal, implicit, or verbal, such as the contract I have with my gardener, who makes a couple of trips to the house each month in exchange for $80. In either case, if one party reneges, the contract is broken. Incidentally, in a swath of the academic literature, informal contracts are called relational contracts, though you're unlikely to encounter that terminology in business practice. I will return to the distinction between formal and informal contracts in section 10.5 when we discuss the "pay for luck" phenomenon in executive compensation.

Workers and employers are both better off when they can make agreements (contracts) without fear of someone reneging. Can they do this without help from a third party? Generally, workers and employers are pretty creative at figuring out ways around problems on their own, to facilitate mutually profitable exchanges. Nonetheless, institutions (e.g., federal, state, and local laws prohibiting wage theft) have evolved to address the problem of contract failure. These laws are usually designed to protect workers rather than employers, because usually the compensation occurs after the labor services have been provided. These laws may sound helpful and nice, but their effectiveness is another matter. Workers incur various costs and risks by invoking the protections of these laws. First and foremost, they may face retaliation by their employers. Certain worker groups are more vulnerable to retaliation than others (e.g., undocumented immigrants, particularly in the current political climate of the United States in 2019).

Other employer deterrents to wage theft are high turnover costs. Even a single occurrence of wage theft might induce a worker to quit, which hits employers with additional costs of recruitment, screening, training, and retention. The recruitment costs may be exacerbated if the wage-theft occurrence is known to prospective job applicants, both through a lower volume of high-quality applicants and through higher compensation demands of workers who ultimately join the firm and expect to be compensated, in the form of higher pay, for the risk of suffering wage theft. This extra compensation that workers demand is an example of a "compensating differential", an important concept I will introduce in Chapter 3. But avoiding turnover costs is only a deterrent to wage theft if the employer actually wants to avoid turnover. In some cases the employer is indifferent or actually wants the worker to leave. Or, whether turnover is desired or not, the employer may realize that the worker is about to leave anyway (e.g., to retire, or to relocate to another part of the country to follow a spouse who received a lucrative offer). The employer's temptation to steal wages will be strongest when the end of the employment relationship is in sight or when the employer doesn't care if the worker quits.

By the way, you might have noticed that in the preceding paragraph and throughout much of his chapter, I'm using a lot of third person rather than addressing you, "the manager", directly in second person as I typically do. The reason is that wage theft is unethical and illegal behavior, you shouldn't be doing it, and I want to keep us both out of trouble! So let's proceed, letting the third-person "employer" take the blame, or hiding safely behind passive, impersonal constructions in which "wages are stolen" (by some evil person).

A big deterrent to wage theft is fear of worker retaliation, which has become easier to do in the age of social media. Workers can retaliate in many damaging ways, some of which are hard to observe. The public retaliation that occurs using social media damages the employer's reputation and may hamper future recruitment and retention, as well as raising the compensation demands of workers who join the firm. A specific example is offered by online "gig" workers who enjoy no formal protection against wage theft but who can inflict punishment on bad employers by assigning negative online ratings to those employers. Employers with negative ratings need to pay more to attract work, which is a concept developed more fully in Chapter 3. Other real-life examples of online forums that affect employers' reputations by allowing workers to publicly punish bad behavior include Glassdoor and "best employer" rankings.

2.4 Do Laws Prohibiting Wage Theft Increase Workers' Pay?

Anti-wage-theft laws are designed to protect workers who are particularly vulnerable to exploitation by employers. Let's consider an example. Several years ago, California introduced legislation that prohibits wage theft against undocumented immigrants. What effect should those laws have on the average compensation level of undocumented immigrants? If the laws discourage instances of wage theft that would have happened in the absence of the laws, it would seem that this should raise the average wage of undocumented immigrants.

To illustrate the logic via a simple example, consider an undocumented immigrant who has a one-day job paying \$100. Suppose that in the absence of a law, the worker knows there is a 10% chance of wage theft, meaning the worker does the work but doesn't get paid. Like most workers, and particularly those in lower-skilled jobs, this one dislikes uncertainty with respect to his compensation. In other words, he would be willing to pay something to avoid the income risk that I just described and to instead receive a somewhat lower payment that is 100% guaranteed. But in his current job, which carries a 10% risk of wage theft, the worker's *expected income* is:

$$(\$100\ \hat{A}\ 90\%) + (\$0\ \hat{A}\ 10\%) = \$90.$$

In that calculation, we just added up all the possible wage outcomes (namely \$100 and \$0), weighting (i.e., multiplying) each by their respective probabilities of happening. Thus, the possibility of wage theft means that the job only offers the worker \$90 in *expected* income, rather than the \$100 that was promised. If the law successfully eliminates the threat of wage theft (i.e., reducing the probability of theft from 10% to 0%) then the worker's expected income is:

($100 Â 100%) + ($0 Â 0%) = $100.

In that case, the law guarantees that the promised wage gets paid, so expected income is the full $100 that was originally promised.

The preceding discussion appears to show that the law increases the average daily compensation of undocumented immigrants by $10, because without the law they could only expect $90 in daily income, whereas with the protection of the law they can expect $100. But this argument is flawed! To see why, we must start by asking, *why does the job pay $100?*

The job pays the *market wage* of $100 because that's presumably the amount that employers have to pay to get the job done. Workers are not slaves and aren't interested in doing the job for less than $100, or at least it would be too much of a hassle to search for a suitable worker willing to do the job for less than $100. Why aren't workers interested in doing it for less than $100? There are probably many reasons, but surely the threat of wage theft is one of them. That threat means that a job that pays $100 is less valuable to the worker than a job that pays $100 for sure. Maybe the job that pays $100 but carries the 10% chance of wage theft is viewed by the worker as roughly equivalent to a job that pays, say, $85 for certain. The market wage of $100 includes a *risk premium* embedded within it that compensates workers for a 10% risk of wage theft. If a law were to eliminate the wage-theft threat, then the market wage of $100 would suddenly be "too high", because it would still incorporate a "risk premium" to compensate workers for a risk that no longer exists. After all, if the worker views a guaranteed $85 job as equivalent to the $100 job that carries a threat of wage theft, then once the wage theft threat is eliminated it's only necessary to pay the worker $85.

In summary, when there's a threat of wage theft, the market wage incorporates a risk premium to compensate workers for it. Once the threat is eliminated (perhaps by legislation) workers no longer need to be compensated for it, so the market wage decreases when the risk premium disappears. This conclusion seems counterintuitive to people who haven't studied compensation theory. A law is enacted to protect a certain group of workers and to *raise* their compensation. In actuality, it can have the effect of *lowering* their compensation, by eliminating a risk premium that had previously been "built into" the market wage. Please understand that this is not an argument that the law is "bad" or that it makes workers worse off! It is simply an argument that its effect on compensation levels isn't as simple and straightforward as you might have guessed.

The principle developed in the preceding example is quite general and has a number of applications, including investing. For example, the long-run rate of return in the stock market is higher than what you'd get if you were to put your money into a certificate of deposit (CD) or a savings account. The reason is that stocks are a lot riskier than CDs and savings accounts. If the expected return on stocks were the same as on CDs and savings accounts, everyone would put their money into CDs or savings accounts and no one would buy stocks. So stocks must incorporate a risk premium to compensate investors for the additional risks they incur by investing in stocks rather than placing their money in CDs and savings accounts.

2.5 Cuts in Nominal and Real Monetary Compensation

There are two ways to cut your employees' compensation. Recalling the appendix to Chapter 1, you can cut their pay in nominal terms or in real terms. Nominal cuts are rare. Real cuts are far more common, so let's discuss them first.

Suppose you hired a worker at an annual salary of, say, \$58,398. If there's inflation of consumer prices of about 2% per year, and if you don't give your worker a 2% raise after the first year of employment, then you've effectively given your worker a decrease in real compensation, whether you intended to or not. This may feel like a very passive way to cut pay . . . after all, as a manager you're not actively changing the compensation. You're just sitting back and allowing inflation to erode its real value. However, as a good manager, you must always remain conscious that such passive pay cuts are happening, because they have important implications for talent management and retention. What your workers ultimately care about is the real value of their compensation, not the nominal value, because the real value determines how much they can actually consume in goods and services. You risk losing your stars if you cut their real pay.

Inflation can be helpful, however, in allowing you to cut real compensation, when doing so would advance your company's aim of making a higher profit. Some of your workers might be overpaid for various reasons. Perhaps one of your workers drove a really hard bargain during salary negotiations (see Chapter 14) and got the better of you, which you didn't fully realize until observing the worker's productivity during their first few months on the job. Or perhaps the worker, for whatever reasons, doesn't exert as much effort on the job as he once did, so his current salary is no longer justified by his productivity. In such cases, you can cut the worker's real compensation passively, simply by giving no raises, or by giving raises that are below the inflation rate. You can achieve deeper real pay cuts, and you can do it faster, when the inflation rate is high. Recently in the United States, inflation rates have been pretty low, at between 2% and 3% per year, so it would take a few years to achieve, say, a 10% cut in your worker's real compensation. If there were no inflation, then nominal and real wages would coincide (see the chapter 1 appendix) so the only way to cut your workers' real pay would be to cut their nominal pay.

Cuts in nominal pay are pretty rare. One reason is that workers hate them, and recall from the third recurring theme in section 1.8 that it's costly to do things that your workers hate. Remember that a nominal pay cut is particularly punishing in an inflationary environment. When there's inflation, even holding the pay level constant in nominal terms amounts to a real pay cut. So a nominal cut means an even deeper real cut. But nominal pay cuts sometimes happen. One common way to execute nominal pay cuts is via job reassignment, i.e., demotions (see Chapter 13). Another common way is to cut non-salary components of compensation, such as bonuses. Nominal pay cuts from demotions are "permanent" in that they typically apply to base salaries, whereas nominal cuts to non-salary components might be temporary. For example, if sales were particularly weak in a given year, your company might decide to cut back on, or even eliminate, annual bonuses, but those bonuses might be restored next year if sales recover. Nominal cuts in pay typically happen either to address performance issues with individual workers (i.e., a worker is paid too generously relative to their productivity) or because of performance issues at the company level (e.g., business may be bad because of a recession).

In unionized settings (see Chapter 5) pay tends to be high relative to workers' productivity, because unions strive to negotiate above-market compensation contracts, and they often succeed. This situation becomes particularly problematic during recessions because, when demand for the organization's products or services is slack, it's no longer possible to generously compensate union workers. Real, and even nominal, pay cuts can happen in those situations, and sometimes the union advocates for nominal pay cuts if doing so can protect jobs and avoid layoffs. See section 5.5 for an example of that in the CSU system.

Compensation cuts, both nominal and real, bear some resemblance to wage theft, which is why I'm covering them in this chapter. After all, your worker's employment contract (whether written or unwritten) states that the worker will get paid, say, a salary of $50,000 in exchange for a certain set of job tasks. If the tasks and expectations remain the same, but inflation erodes the real value of $50,000, then the terms of the contract have been changed in a way that disfavors your worker. If you actually cut their nominal salary to $42,000, that is an even more blatant instance of reneging on the original contract.

I said that such cuts "bear some resemblance" to wage theft in the sense that they deviate from the original employment contract in a way that hurts the worker and that might be unexpected. However, precisely speaking, they are not the same as wage theft. True wage theft happens when an employer "takes the labor and runs". That is, the labor has already been expended, and then the employer fails to pay the price to which both parties agreed. In contrast, in the case of nominal and real pay cuts, workers usually retain the option of walking away before the "new" contract terms kick in and before additional labor is expended. If the worker you hired at $50,000 was expecting you to maintain their real compensation level, and if you fail to grant a 2% raise at the end of their first year (to match inflation), then they can simply submit a resignation letter before starting their second year of work. A nominal pay cut can be thought of as a less extreme version of firing a worker, because firing a worker is equivalent to cutting their nominal compensation all the way to $0. We don't think of firings as wage theft, so the same is true for the less extreme forms of nominal cuts that allow workers to remain employed.

Inflation is also a threat to other (non-salary) components of compensation. Shortly after I joined CSUEB, I was chatting about retirement savings with a colleague who was nearing retirement age. Our conversation focused on CalPERS, i.e., the public, defined-benefit pension system in which the university participates (see Chapter 11). The budget situation in California was particularly dire at the time, and I asked my colleague whether she thought that California would ultimately renege on its pension obligations. She immediately exclaimed, "They already have!" She then went on to explain that, because the union had been unsuccessful in negotiating any pay raises for the last several years (meaning that nominal salaries had remained unchanged for several years, while real salaries dropped markedly) the pension payout – which is based on a formula that is tied to nominal salaries – would be considerably less, in real terms, than what workers anticipated.

The discussion in this section and in Chapter 1's appendix makes clear that real compensation hinges crucially on inflation. The tricky part is that no one – neither employers nor workers – can predict exactly what inflation will be in the future. So both sides must form some expectations about future inflation – basically educated guesses – and those expectations will have a significant bearing on the compensation contract that both parties negotiate. For example, in a unionized setting in which three-year contracts are often negotiated, the annual compensation increases that are stipulated in the contract for the next three years will take into account what is expected to happen with inflation. And if expectations turn out to be substantially wrong, that will affect the next round of negotiations in the subsequent contract. A negotiated annual increment of 2% or so that's built into some union contracts merely keeps *real* compensation relatively constant, assuming a 2% annual inflation rate (see Chapter 1's appendix). So it should not be thought of as increasing the compensation of union members over time.

This is all I have to say about wage theft, for now, though we will refer back to the topic periodically throughout the book (e.g., section 10.5 explores the threat that CEOs or other executives may be deprived of part or all of the bonuses promised them). One reason I chose to discuss wage theft so early in the book is that it nicely illustrates an important general concept from compensation theory that applies to many aspects of compensation other than wage theft. That is the concept of *compensating differentials* (see Chapter 3).

2.6 Lessons for Managers

Employment can be a dangerous business that comes with many risks, including injury, exhaustion, stress, boredom, wage theft, demotion, firing, sexual harassment, and emotional abuse, all of which are undesirable job characteristics that can be expected to require a wage premium to attract workers to fill them. As a manager, you should intimately know the laws, both federal and state, surrounding wage theft. You should also understand that wage theft carries costs beyond legal repercussions, including turnover costs, the threat of worker retaliation, and damage to your reputation that can have a chilling effect on your future recruiting and talent retention. If you are managing in an organization that has a history of wage theft, be aware that you will face serious challenges in recruiting workers. Prospective employees will likely demand a wage premium to compensate them for the risk, and to the extent that you can allay their fears and assure them that wage theft was a historical phenomenon that will not be repeated under the current leadership, you can shrink that risk premium and save your organization money.

Think of your employees' compensation in real terms, because that's what's relevant for talent management. Your workers care the most about their real compensation, because that's what determines how much stuff they can buy, so you have to care about it too. Recognize the role of inflation in helping you to achieve pay cuts, when desirable. And when negotiating over future compensation levels, bear in mind the importance of accounting for future expected inflation. Be aware that you can increase the flexibility of compensation, which makes nominal cuts easier to achieve when desirable, by relying more heavily on non-salary components (like bonuses, that can be easily varied from year to year) than on base salary, which is more "permanent".

Case Discussion 30: *Weaver* v. *Legend Senior Living, LLC*[1]

Spoiler Alert: The court decision underlying this case is referenced below, but it should not be read until after the case discussion is completed!

Legend Senior Living, LLC, (hereafter "Legend") which is headquartered in Wichita, KS, is a privately held senior housing and services company that owns and operates 40 locations in CO, FL, KS, OK, PA, and TX. The nursing home provides services for the elderly, including independent living, assisted living, memory care, and personal care. A group of employees at one of Legend's nursing homes in Oklahoma filed a collective legal action under the Fair Labor Standards Act (FLSA)

---

[1] Source: *Weaver* v. *Legend Senior Living LLC*, W.D. Okla., No. Civ-16-1230-R (July 20, 2017).

against the company and its owner. The various complaints concerned unpaid bonuses. The employees' offer letters entitled them to nondiscretionary bonuses that were based on a predetermined formula.

In "Count 1", the employees alleged that Legend effectively made improper deductions from their pay, because they were entitled to the bonus payments and never received them. This bonus theft, the employees alleged, caused their status under FLSA to change from "exempt" to "nonexempt" (recall from sections 1.4 and 4.4 that nonexempt employees must be paid overtime pay at a rate of "time-and-a-half" for hours worked in excess of 40 per week, whereas exempt workers are not entitled to overtime pay). The employees, therefore, claimed that they were entitled to overtime compensation from October 2013 for any weeks in which they worked in excess of 40 hours. Legend asked that the employees' request be dismissed by the Court on the grounds that, even though the promised bonuses went unpaid, the employees were incorrect in their claim that an unpaid bonus triggers a change in a worker's status from "exempt" to "nonexempt". Legend argued that there were no deductions from workers' *salaries* and that the unpaid bonuses did not impact the payments that Legend made to the employees' *salaries*.

The employer bears the burden of establishing that an employee is "exempt" under the US Department of Labor's criteria, which concern: (1) *how much* an employee is paid, (2) *how* they are paid (which is referred to as the "salary basis", and (3) *what type of work* they perform. The employees' claim in Count 1 amounts to a challenge to their "salary basis". Section 541.602(a) of the FLSA, which concerns the definition of "salary basis", notes that for an employee to be paid on a salary basis (which is required for "exempt" status) an employee must receive the full salary for any week in which the employee performs work. It also notes that an employer who makes an improper deduction from salary "shall lose the exemption if the facts demonstrate that the employer did not intend to pay employees on a salary basis". Such deductions will result in loss of the exemption for the period of time in which the improper deductions from salary were made, for employees in the same job classification and working for the same managers who imposed the improper deductions. Other key quotes from the FLSA are:

(a) *"An employee will be considered to be paid on a 'salary basis' . . . if the employee regularly receives each pay period on a weekly, or less frequent basis, a predetermined amount constituting all or part of the employee's compensation, which amount is not subject to reduction because of variations in the quality or quantity of the work performed."*

(b) *"[a]n employee is not paid on a salary basis if deductions from the employee's predetermined compensation are made for absences occasioned by the employer or by the operating requirements of the business. If the employee is ready, willing and able to work, deductions may not be made for time when work is not available."*

(c) *An employer can "provide an exempt employee with additional compensation without losing the exemption for violating the salary basis requirement, if the employment arrangement also includes a guarantee of at least the minimum weekly-required amount paid on a salary basis."*

In "Count 7", the employees alleged that Legend was guilty of breach of contract for changing the method for calculating nondiscretionary bonuses. Specifically, the employees alleged that after March 2015, higher performance thresholds were implemented by Legend for certain bonuses (see the discussion of the ratchet effect in section 9.9). Legend argued that Count 7 should be dismissed because the employees' offer letters stated: "In accepting our offer of employment, you certify your understanding that your employment will be on an at-will basis and that neither you nor any Company representatives have entered into a contract regarding the terms or the duration of your employment" and "Legend Senior Living reserves the right to review and make alterations to the bonus programs prior to the next fiscal year."

Questions

1. Which side has the stronger legal case on Count 1? Explain.
2. Should the Court dismiss Count 7, as requested by Legend? Explain.
3. If you are the employees' manager, you have a challenging situation on your hands that will persist even after the court case ends. The employees are disgruntled, and they remain your employees. How will you handle this situation?
   (a) Is your answer to that question different if you win in court (meaning Counts 1 and 7 are dismissed) versus if you lose? If so, explain how.
4. In general, does the fact that you were sued by your employees (regardless of the outcome) automatically imply that you should change your compensation and HR policies?
   (a) In the present case, are there any changes to compensation and HR policies that you'd recommend? If so, explain what they are, how they'd help, and how they depend on the Court's decision, if at all?
5. Although this case involved wage (i.e., bonus) theft, the legal claim was not about recovering the unpaid bonuses. Instead, it was about an indirect consequence of that wage theft that (allegedly) entitled the employees to overtime pay. Focusing on the theft itself, is there anything in particular about this industry and about the jobs that are typically held in a nursing home that makes such theft more likely to occur? Explain.
6. Focus on a particular job that is typically held in a nursing home, that meets the FLSA's salary basis, and that may receive bonus pay. Describe the tasks (as best as you understand or can imagine them) and explain how you would structure

the bonus system. Explain what performance measure you would use and how it would translate into bonus payments. Highlight the strengths and weaknesses of the measure.

7. When designing a bonus system, you want to reserve the right to make adjustments to fine-tune and improve it over time, but such flexibility makes your employees nervous that you might renege via wage theft or the ratchet effect. Given that tradeoff, what advice would you give Legend's management if they sought your advice as a compensation consultant?

Further Reading

Benson, A., Sojourner, A., and Umyarov, A. (2019). "Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market." *Management Science*.

Bobo, K. (2011). *Wage Theft in America: Why Millions of Americans Are Not Getting Paid – And What We Can Do About It*, New York: The New Press.

DeVaro, J. (2014). "Stealing Wages from Immigrants." *Compensation and Benefits Review*, 46, 96–102.

Doussard, M. and Gamal, A. (2015). "The Rise of Wage Theft Laws." *Urban Affairs Review*, 52, 780–807.

Fussell, E. (2016). "The Deportation Threat Dynamic and Victimization of Latino Migrants: Wage Theft and Robbery." *The Sociological Quarterly*, 52, 593–615.

Meixell, B. and Eisenbrey, R. (2014). "An Epidemic of Wage Theft is Costing Workers Hundreds of Millions of Dollars a Year." Issue Brief #385, Economic Policy Institute.

Milkman, R., González, A. L., and Narro, V. (2010). "Wage Theft and Workplace Violations in Los Angeles: The Failure of Employment and Labor Law for Low-Wage Workers." *Report, Institute for Research on Labor and Employment Relations*, University of California, Los Angeles.

Minkler, M., Salvatore, A. L., Chang, C., Gaydos, M., Liu, S. S., Lee, P. T., Tom, A., Bhatia, R., and Krause, N. (2014), "Wage Theft as a Neglected Public Health Problem: An Overview and Case Study From San Francisco's Chinatown District." *American Journal of Public Health*, 104, 1010–20.

Rogers, B. (2010). "Toward Third-Party Liability for Wage Theft." *Berkeley Journal of Employment and Labor Law*, 31, 1–64.

Tippett, E. C. (2018). "How Employers Profit from Digital Wage Theft Under the FLSA." *American Business Law Journal*, 55, 315–401.

3 Compensating Differentials

In my last year of college, I was offered jobs as a research assistant at several regional banks in the Federal Reserve System. At the time, most of the Feds employed research assistants, and those jobs were basically the same from one Fed to the next, in terms of required qualifications, tasks, work hours, career prospects, *nominal* compensation, etc. But the *real* compensation differed substantially between banks in different locations. The nominal salary offers that I can recall (in 1994 dollars) were $30,000 from Richmond, $31,000 from Kansas City, and $32,000 from New York City. I didn't actually interview with the San Francisco Fed (for reasons I'll explain shortly), but the nominal salary there was probably comparable to that of New York City. In this chapter, I'll develop an example based on the difference between the Kansas City and San Francisco jobs. Although the nominal salaries were roughly similar in those two jobs, after accounting for the big differences in cost of living between those locations the *real* compensation was considerably higher in Kansas City (KC) than in San Francisco (SF).

Recall from Chapter 1's appendix that nominal compensation refers to the actual dollar amount of compensation, without regard to how useful that compensation is in terms of its purchasing power. In contrast, real compensation takes into account the prices of goods and services, which may vary over time and across locations. The Bureau of Labor Statistics (BLS) produces data that allow for meaningful comparisons of compensation across different geographic locations, as I discussed in Chapter 1's appendix. But for the sake of convenience in this illustrative example, let's just use the online calculator provided by *CNN Money*, which reveals that, in 2019 dollars, a salary of $70,000 in KC is equivalent to a salary of about $157,470 in SF. The gap between those cities back in 1994 would have been similar.

In short, there was a substantial difference in *real* compensation between two jobs – SF and KC – that are essentially identical except for one characteristic . . . location. KC pays a lot more than SF, in *real* compensation. Why?

3.1 Compensating Differentials: a Definition

The answer, at least in part, is that SF is a tourist destination that is considered an exciting, desirable place to live, particularly among fresh college graduates, whereas KC is generally thought to be less appealing. If the real compensation were identical between the two cities, relatively few people would choose to work in KC. So to compete in the national labor market successfully, KC must offer higher (real) compensation than SF, to compensate workers for the relative disappointment of having to live in KC.

Such a difference in compensation that arises because of a particular job characteristic – positive or negative – is called a *compensating differential*. An equivalent way to say the same thing is that SF is able to get away with paying less, in real terms, because part of the compensation that workers receive by living there is non-monetary . . . it reflects the sheer "joy" of living in SF. Recall that the definition of compensation in section 1.1 is "everything a worker likes about a job", which includes geographic location.

I cannot overstate the following point. A crucial requirement of this example of a compensating differential is that, apart from their locations, the two jobs are *exactly the same*. If instead, for example, the job in KC was for a senior researcher whereas the job in SF was for an entry-level research assistant, then a large difference in real compensation would be expected simply because the KC job is a higher-level position. When discussing compensating differentials, it is important to compare apples-to-apples rather than apples-to-oranges, so that jobs are allowed to differ only in the one respect of interest, which in this case is desirability of geographic location.

Incidentally, employers in desirable geographic locations can get away with skimping on recruiting costs as well as on compensation costs. I was offered an on-site job interview at the SF Fed, but I turned it down because that bank's policy was not to reimburse travel expenses for job candidates, and I was a broke college student on the other side of the country. In contrast, banks like KC and Richmond reimbursed all of my travel expenses. SF didn't need to reimburse travel expenses, because of the heavy supply of interested applicants, plus the Fed wanted to avoid having people accept an on-site interview just to get a free vacation to SF, a popular vacation destination. In contrast, relatively few graduating college students on the East coast would accept a job interview in KC just for the opportunity to get a free trip to Kansas.

3.2 Mobility and Information

Recall the three-legged stool of compensation (i.e., desire, skills, mobility). One leg of that stool, i.e., mobility, is the engine that fuels compensating differentials.

Workers' ability to move from KC to SF effectively forces KC to pay premium compensation to compete with SF. If workers were, hypothetically and, obviously, illegally, *forced* to remain in KC, they would not need to be paid premium compensation, because they would be unable to escape, just like the prisoners and slaves in section 1.6. Whereas mobility is the engine that fuels compensating differentials, *information* is what enables mobility. To have an interest in moving, workers must first be able to observe the relevant job characteristics (positive or negative).

In section 2.4, we discussed an example in which two jobs are identical (i.e., an "apples-to-apples" comparison) except that one carried a higher risk of wage theft because a law had not yet been enacted to prohibit such theft. If job applicants have the relevant information, i.e., they know which of the two jobs carries a risk of wage theft, then if the jobs offer the same compensation, everyone will want to work at the job without the extra risk. So, because workers are mobile and can work at either job, the one that carries the risk of wage theft must pay a *risk premium* to attract any applicants. This is another example of a compensating differential. The risk of wage theft is a negative job attribute that, just like living in a less desirable location, requires premium compensation. Therefore, this chapter on compensating differentials, and Chapter 2 on wage theft, are connected in that the threat of wage theft is a negative job attribute (from the standpoint of workers) that can give rise to a compensating differential.

3.3 Work Environment and Compensating Differentials

Lots of job and workplace characteristics, positive and negative, shape the organization's work environment. Some organizations are well known to have great working environments and cultures. For example, *Fortune* publishes an annual list of "Best Companies to Work for in the US". That information spurs mobility (i.e., workers clamor to join those companies, while avoiding others) that gives rise to compensating differentials, allowing those firms to pay less than firms that are *otherwise identical* but that are not on the annual list.

SAS Institute is an American multinational developer of statistical software, headquartered in Cary, North Carolina, that was founded in the mid-1970s. SAS (which stands for Statistical Analysis System) has regularly featured in *Fortune*'s list for many years, and was #1 in 2010 and 2011. The company offers a generous array of perks, ranging from free healthcare to free M&Ms, unlimited sick days, short work weeks, etc. All of these workplace amenities, given that they are non-cash forms of compensation, fall under the heading of "benefits", which we will discuss in Chapter 11. The important point for now is that such job amenities give rise to compensating differentials, i.e., companies like SAS can get away with offering lower wages and salaries than other companies that are *identical* except for not offering such benefits.

SAS is also known for paying its employees well. But do its high salaries contradict my earlier statement that, due to its appealing work environment, SAS can afford to depress wages and salaries? No! And this tends to be a point of confusion for students when first introduced to compensating differentials, so you should make a mental note of it. Compensating differentials do *not* mean that companies that offer positive job characteristics offer *low* wages. Rather, they mean that companies that offer positive job characteristics can offer *lower* wages than *identical* companies that do not offer those positive job characteristics. There are plenty of companies, like SAS, Google, and others, that offer both high compensation and appealing work environments.

3.4 The Marginal Worker

My remarks at the start of section 3.1 might create the misleading impression that everyone's tastes are the same. But even though SF is *generally* more popular than KC, people obviously have different preferences. Among those who favor SF, some have such strong affinities for that city that they would be willing to sacrifice a lot of income to live there rather than in Kansas. Others prefer SF more mildly and would be willing to sacrifice only a modest amount of income to live there. Still others are indifferent and would be equally happy living in either city. And there are even some people, perhaps those who grew up in the Midwest, or who have family there, (or who have a particular taste for world-class barbeque!), who would actually *prefer* to live in KC. Of all these workers, the one who is closest to being indifferent between the two cities has a special name ... the *marginal worker*, not to be confused with the "marginal performer" in the context of performance evaluations, which is an entirely different concept.

3.4.1 Comparison to Swing Voters in an Election

The concept of the marginal worker is similar to the concept of the swing voter in an election. Let's start with that familiar example to set the stage. Consider the 2016 US presidential election, and to keep things simple let's ignore third-party candidates and assume everyone voted for either Trump or Hillary. Some of the Trump voters were very passionate about their choice and vigorously supported Trump, whereas other Trump voters had a much weaker preference for him. The same is true of the Hillary voters. Imagine that we line up all the voters in that election, from left to right, according to the strength of their preferences. The color of the voters' shirts reveals the strength of their preferences for their chosen candidates. Blue shirts represent Hillary supporters, red shirts represent Trump supporters, and the darker the color, the more passionate the support. The voter at the far left end of the line wears a shirt that's the darkest possible blue, and the one at the far right end of the line wears a shirt that's the darkest possible red.

As we walk down the line, from left to right, the shirts become lighter and lighter shades of blue, until finally we reach the last (and least enthusiastic) Hillary voter . . . the person wearing a shirt that's the lightest shade of blue. In the spot just to the right of that person is the first (and least enthusiastic) Trump supporter, who wears a shirt that's the lightest shade of red. So those two voters, who stand side-by-side somewhere in the "middle" of the line, cast different votes, but both of them had only very mild preferences for the choices they made, and it probably wouldn't have taken much to sway them to the other side. In fact, between the two of them, one of them – the one whose shirt is lightest in color – is even closer to being "on the fence" than the other. That person is the swing voter, or marginal voter. Swing voters and those around them are the most important voters in an election, and they determine its outcome. That's why you heard so much about "swing states" (also known as "battleground states") in that election. Just like swing voters (i.e., those who are closest to indifferent between Trump and Hillary) determine the election outcome, the marginal workers (i.e., those who are closest to indifferent between KC and SF) determine the size of the market wage differential, as I'll now show.

3.4.2 Identifying the Marginal Worker

Suppose we take a bunch of workers, say 1000 of them, and line them up in a row from left to right. Worker 1, who has the strongest desire to live in SF, is at the far left end of the line, and worker 1000, who has the weakest such desire (and, indeed, who may even prefer to live in KC), is at the far right end of the line. Everyone in between is appropriately ordered so that if we start at the left end of this line and walk to the right, the workers we successively meet are less and less interested in SF and more and more favorably disposed towards KC.

In which city do each of these workers want to work? The answer depends on the compensation levels and on the strengths of their preferences. In the extreme and unrealistic case in which the annual compensation is $0 in SF and $3.2 million in KC, all 1000 workers would eagerly rush to work in KC. At the other extreme, if the compensation offers were identical between the two cities, then clearly the person at the far left would choose SF . . . that person strongly prefers SF to KC and would be willing to incur a large pay cut to live in SF, so if there's an opportunity to live in SF with no pay cut at all, that worker will obviously jump at the chance. The same will be true for many other workers standing in line. If they don't have to sacrifice any compensation to live in SF, many and perhaps even all workers in the line will prefer SF.

Between those two extremes, let's think about the more interesting case in which the real compensation offer from KC exceeds that of SF by a substantial (but not outrageously high) amount. Then we can expect that the workers towards the left end of the line will prefer to work in SF, and those towards the right end of the line will prefer to work in KC. But that must mean that as we walk from left to right, from worker 1 towards worker 1000, we eventually reach a point in the line where worker preferences (at the offered compensation levels) switch from an SF preference to a KC preference, just like in the previous subsection we eventually reached a point in the line where the voters switched from Hillary to Trump.

For concreteness, suppose that worker 750 prefers SF, whereas worker 751 (and, to an even greater extent, all workers to the right of worker 751) prefers KC. One of these two workers is the *marginal worker*, meaning the *worker who is closest to being indifferent* between living in SF versus living in KC. In other words, the marginal worker is the one who is closest to changing his or her mind and switching to the other location if the compensation tilts just a little bit more in favor of the other location.

So which of the two workers (750 or 751) is the marginal worker? Based only on the information I've provided so far, we can't say. We just know that it must be one of the two. To identify which one, we need to know how attached both of those workers are to their chosen location, just like, for the two voters in the "middle" of the line in the previous section, to identify the swing voter we needed to know which one had the lighter-colored shirt.

In the present example, for workers 750 and 751, we need to know the minimum amount that the geographic wage difference would have to change to entice each worker to switch locations. Inducing worker 750 to move from SF to KC would require an *increase* in the geographic salary gap by (at least) a "certain amount" (i.e., the salaries in KC would have to increase relative to those in SF), whereas inducing worker 751 to move from KC to SF would require a *decrease* in the geographic salary gap by (at least) a "certain amount" (i.e., the salaries in KC would have to decrease relative to those in SF). Whomever reports the smaller "certain amount", which is analogous to the "lightest-colored shirt" in the voting example from the preceding section, is the marginal worker.

3.4.3 Further Describing the Marginal Worker

Suppose, for example, that if the geographic wage gap *increases* by \$2581 – meaning that the SF real salary decreases by exactly that amount, or the KC real salary increases by exactly that amount, or salaries simultaneously go down in SF and up in KC by just enough that the geographic salary difference increases by exactly \$2581 – worker 750 is willing to move to KC, but if the geographic wage difference increases by any amount smaller than \$2581 then worker 750 would prefer to stay put in SF.

Suppose further that worker 751 would be willing to move from KC to SF if the geographic salary gap were to *decrease* by \$2687, but any smaller decrease would be insufficient to induce that worker to leave KC. Again, the decrease of \$2687 in the salary gap could happen because salaries in KC decrease by exactly \$2687, or because salaries in SF increase by exactly \$2687, or because the salaries in KC decrease and those in SF increase simultaneously such that the geographic salary gap shrinks by exactly \$2687.

In this example, worker 750 is the marginal worker, because \$2581 is smaller than \$2687, meaning that it's easier to coax worker 750 to leave SF than it is to coax worker 751 to leave KC. This echoes the voting example from section 3.4.1; there, the voter with the "lightest-colored shirt" was the marginal one who was closest to being indifferent between Hillary and Trump, whereas here worker 750 is the marginal one who is closest to indifferent between SF and KC.

3.4.4 Changing the Compensation Levels Might Change the Marginal Worker

If we change the compensation levels in the two cities, then the identity of the marginal worker(s) might also change. Suppose, for example, that we change the compensation levels in the preceding paragraphs by raising the SF compensation level by 8%, while leaving the salary in KC unchanged, and further suppose that even after this change the KC salary remains the higher of the two in real terms. This change amounts to a shrinking of the geographic salary gap, so that the SF job becomes even more appealing relative to the KC job. Let's also suppose that the 8% increase in SF's compensation is large enough that the geographic salary gap shrinks by much more than \$2687, which, again, is the minimum amount that would tempt worker 751 to move from KC to SF.

In this situation, worker 751 now definitely prefers to move from KC to SF, and probably the same is true for workers 752, 753, and others. In fact, let's suppose that in this situation even worker 811 now prefers to move to SF, though worker 812 remains wedded to KC along with (to an even greater extent) everyone to the right of that worker. It would take an even larger increase in the SF compensation level (or decrease in the KC compensation level) to convince the workers to the right of 811 to move to SF. Thus, the 8% increase in the SF compensation changed the identity of the marginal worker from 750 to either 811 or 812.

Again, to identify whether the marginal worker is 811 or 812, we need to know the minimum amount that the geographic salary gap would have to expand to tempt worker 811 to move to KC, and the minimum amount that it would have to shrink to tempt worker 812 to move to SF, and whichever amount is *smaller* (meaning the worker is closest to indifferent, in the sense of having the lightest-colored shirt in the voting example) determines the marginal worker. If those two amounts happen to be exactly the same, or if we don't know exactly what those amounts are, then we can just refer to *both* workers as marginal.

The previous examples illustrate that the marginal workers are those who are closest to being indifferent among the various options. They are the easiest to sway to an alternative choice, if the compensation changes, just like swing voters in an election are the most likely ones to change their votes in light of new information. In contrast, workers who are closer to the extremes of the geographic "line" have much stronger preferences and require much larger differences in compensation to change their locations.

3.5 Marginal Worker(s) and the Size of the Market Wage Differential

The concept of the marginal worker is important, because it helps us to understand how large the market wage differential will be. I'm going to explain this in two different ways. Both explanations will use an example like the preceding, with 1000 workers standing in a geographic line. For variety, let's change all the numbers, so forget everything that was said about the preferences of workers 750 and 751. With the new numbers, let's say that at the going market wages, workers from 1 to 697 prefer SF, and workers from 698 to 1000 prefer KC.

3.5.1 Explanation 1

For my first explanation, let's suppose that worker 697, who lives in SF, is willing to sacrifice up to $10,000 in salary (but no more!) in exchange for the opportunity to work in SF, whereas worker 698, who lives in KC, is willing to sacrifice up to $9500 in salary (but no more!) in exchange for the same opportunity. The amount that worker 698 is willing to sacrifice is less than the amount that worker 697 is willing to sacrifice, because worker 698 stands to the right of worker 697 in the line of 1000 workers, and, therefore, has a weaker desire to live in SF. Similarly, worker 699 would be willing to sacrifice even less than $9500, and so on. Ignoring ties, the amount that each successive worker is willing to sacrifice shrinks as we move further to the right in the line. As we get far enough to the right, it is even possible (though not guaranteed) that the amount shrinks to zero or even becomes negative. It would become negative in the case of workers who actually prefer KC to SF, even if both locations paid the same compensation. Section 3.6 covers such an example.

The preceding information tells us that the *market wage differential*, i.e., the market wage in KC minus the market wage in SF, must be at least $9500, and it cannot exceed $10,000. How do we know? Well, if it were greater than $10,000, say $10,001, then worker 697 would prefer to move to KC, so my earlier statement that this worker chooses to live in SF would be false. And if it were less than $9500, say $9499, then worker 698 would prefer to move to SF, so my earlier statement that this worker chooses to live in KC would be false. Only market wage differentials within the interval from $9500 to $10,000 (inclusive of those two endpoints) will prevent those two workers from trying to move, which justifies my original statements about where both workers choose to live. We know that one of those two workers is the marginal worker. In this sense, the geographic preferences of the marginal worker dictate the size of the actual wage differential we observe in the labor market.

Based on the information I've provided so far, the market wage differential could be anywhere between $9500 and $10,000. For example, it might be $9735. To summarize, "Explanation 1" started with the marginal workers and then showed how their preferences give rise to a market wage differential somewhere between $9500 and $10,000, such as, for example, $9735.

3.5.2 Explanation 2

For my second explanation, let's suppose that we have no idea who the marginal workers are . . . forget all about workers 697 and 698. Instead, all we know is that the market wage differential happens to be $9735. We can use this information, along with the maximum amounts that each of the 1000 workers would be willing to sacrifice for the opportunity to live in SF, to identify the marginal worker(s). How? Let's start at the beginning of the line, at worker 1, and we will keep walking to the right until we identify the marginal worker. We can say for sure that worker 1 will prefer SF when the market wage differential is only $9735. Why? Because we know that even worker 697 would be willing to sacrifice up to $10,000 for the opportunity to live in SF, so worker 1 (who stands far to the left of worker 697 and who, therefore, has a much more intense desire to live in SF) would be willing to sacrifice even more than $10,000.

The same argument applies (albeit with slightly lesser force) to worker 2, and so on, all the way up to worker 697. All 697 of these leftmost workers would be willing to forgo at least $10,000 for the opportunity to live in SF, whereas the market compensation levels only require them to sacrifice $9735, so for sure they will prefer SF! However, once we reach worker 698, the tide turns. That worker would only be willing to sacrifice $9500 (at the absolute most!) to live in SF, but the market compensation levels are such that the worker would need to sacrifice $9735 to live in SF. At a price of $9735 that worker says, "Forget it. I'm staying in Kansas City." That argument would be even stronger for worker 699, and stronger still for worker 700, and so on.

Is worker 697 the marginal worker, or is it 698? We know that worker 697 would be willing to sacrifice up to $10,000 to live in SF, but the market wage differential only requires that person to sacrifice $9735. If the market wage differential were to increase by just $265, to $10,000, worker 697 would be indifferent between the locations, and if it were to increase by even a penny more that worker would hop on a flight to Kansas. In contrast, worker 698 would be willing to sacrifice up to $9500 to live in SF, but the current market differential of $9735 would require a greater sacrifice (which is why the worker chose KC). The market differential would have to shrink by $235 (i.e., $9735 minus $9500) before worker 698 would be tempted to move to SF. Since $235 is a smaller amount than $265, we conclude that worker 698 is the marginal worker.

3.5.3 Comparing Explanations 1 and 2

In both explanations we know workers' reservation wages for moving, i.e., we know how much they'd be willing to sacrifice in pay to live in SF. In Explanation 1, we also know the identity of the marginal worker and, therefore, the number of workers actually living in each location, which allows us to identify the market wage differential. In Explanation 2, we know the market wage differential and can use that information to identify the marginal worker and, therefore, the number of workers living in each location. Neither of these explanations is better than the other. They are opposite sides of the same coin. Explanation 1 says that the preferences of the marginal workers identify the size of the market wage differential, and Explanation 2 says that the market wage differential identifies the marginal workers. The market wage differential and the marginal workers are *jointly determined* . . . neither one logically precedes the other.

Recall that worker 697 (who chose SF) was willing to sacrifice up to $10,000 to live in SF, whereas worker 698 (who chose KC) was willing to sacrifice up to $9500. We know that the market wage differential must lie between $9500 and $10,000, and in the preceding example it was $9735. Since that market wage differential is closer to worker 698's "maximum sacrifice" (i.e., $9500) than to worker 697's (i.e., $10,000), the marginal worker is 698. If instead the market differential were $9840, then worker 697 would be the marginal worker, because $160 (i.e., $10,000 minus $9840) is smaller than $340 (i.e., $9840 minus $9500). As long as the market wage differential lies somewhere between $9500 and $10,000 either worker 697 or 698 will be marginal.

However, if the market wage differential lies outside of that interval, someone else would be the marginal worker. For example, a market wage differential of $7532 would have led to a marginal worker somewhere to the right of worker 698, and a market wage differential of $13,799 would have led to a marginal worker somewhere to the left of worker 697. If the market wage differential is extremely small, many workers (maybe even all of them) would prefer to live in SF. And if it is extremely large, many workers (maybe even all of them) would prefer to live in KC. As I mentioned in section 3.4.2, an outrageously large market differential of $3.2 million would move the marginal worker so far to the left that all workers would eagerly choose KC.

3.5.4 Relative Demand for Labor in Both Cities

The preceding discussion illustrates the connection between the market wage differential and the marginal workers' preferences. But what exactly determines the size of the market wage differential? The interaction of the supply and demand for labor. In the preceding examples, labor supply (i.e., the amount of available labor) is represented by a fixed stock of 1000 workers standing in a line. So once we know the demand for labor in the two cities, we know what the market wage differential will be between them. Suppose, for example, that market demand is such that 850 workers are demanded in SF and only 150 in KC. In that case, the market wage differential discussed in the preceding example (which was somewhere between $9500 and $10,000) is too large, i.e., it induces too many workers to prefer KC. The market wage differential must shrink enough so that workers 1 through 850 prefer SF, and the remaining 150 workers prefer KC. Any changes in relative labor demand between the locations will cause the market wage differential to grow or shrink accordingly, so that the correct number of workers end up in the right locations, and the identities of the marginal worker(s) change accordingly.

A mistake students often make is to assume that as long as the *majority* of workers prefers SF to KC, the market wage differential must involve higher real wages in KC than in SF. But that doesn't necessarily need to happen. It totally depends on how much labor is demanded in the two locations. To see why, suppose that worker 923 is indifferent between the two locations, meaning that if the compensation were the same in both places, the worker would be willing to flip a coin to decide where to work. Further suppose that all of the workers who stand to the right of worker 923 in line actually prefer KC to SF (maybe because they're huge fans of world-class barbeque and hate sourdough bread!). To entice the workers to the right of worker 923 to relocate to SF, it would actually be necessary to pay *premium* wages.

Now suppose that there's an explosion in demand for workers in SF, relative to KC, so that 988 workers are demanded in SF and only 12 in KC. The market wage differential in this case would involve a higher compensation level in SF than in KC. The reason is that to attract 988 workers to SF, the wage levels must be such that worker 988 prefers SF to KC, and that will only happen if SF pays sufficiently more than KC in real terms. Thus, even though most of the 1000 workers prefer SF to KC, there is no compensating differential to compensate people for working in KC. The reason is that demand for labor in KC is so low relative to the demand in SF that it's possible to staff all of the KC jobs with people who actually prefer to work in KC . . . not only do they *not* require a wage premium to work in KC; they'd actually be willing to sacrifice some compensation in exchange for the opportunity to work in KC.

In the preceding example, if relative demand for labor in the geographic regions were such that 923 workers were needed in SF and 77 in KC, the market wage differential would be zero . . . the locations would pay exactly the same salary, even though most workers have a preference for SF. The reason is that worker 923 would be the marginal worker, and that person is indifferent between the locations, so it is possible to attract worker 923 to SF (and all workers standing to the left of this person) even when the locations pay identical salaries. If more than 923 workers are required in SF, then the market wage differential involves higher pay in SF than in KC, and if fewer than 923 workers are required in SF, then the market wage differential involves lower pay in SF than in KC, which is the typical situation we have in mind when we discuss compensating differentials, i.e., if workers generally prefer one location to another, then the preferred location typically pays less, if the comparison is apples-to-apples with all other factors identical between the locations.

3.6 Another Example: CSUEB versus CSUSF

Let's have a look at a local labor market involving some management professors who must decide where to live and work within the San Francisco Bay Area. Their choices are California State University–East Bay (CSUEB), or California State University–San Francisco (CSUSF). Both universities are in the CSU system, and suppose that they are *identical in all respects* except for location within the Bay Area; CSUEB is located in Hayward, whereas CSUSF is located in SF. Hayward is the "Kansas City" of this example, in the sense that Hayward is generally thought to be among the least exciting and desirable places to live in the Bay Area, (though as a Hayward resident I personally like it!). Assume that professors live where they work, so the advantage of working at CSUSF is the luxury of being able to live in desirable SF.

In the Department of Management at CSUEB, as of 2019, there are ten professors, including me. We're listed in the left column of Table 3.1. In the right column are ten fictitious management professors at CSUSF. The dollar amount next to each professor's name represents the amount (out of a $100,000 annual salary) that the professor is willing to sacrifice to live and work in SF.

Although the 20 professors are listed in two columns of a table, this example is no different from the ones in sections 3.4 and 3.5. To see why, note that we could easily rearrange the 20 professors in a single line, from left to right. At the far left end would be the professor with the strongest desire to live in SF (i.e., Professor A, who is willing to sacrifice $45,123 in income to live in SF), followed by Professors B, C, D,

Table 3.1 Willingness to pay to live in San Francisco

CSUEB (Hayward)		CSUSF (San Francisco)	
Gregory Theyel	$30,000	Professor A	$45,123
Asha Rao	$25,000	Professor B	$42,985
Zinovy Radovilsky	$18,000	Professor C	$41,999
Glen Taylor	$12,000	Professor D	$41,800
Ken Pefkaros	$5000	Professor E	$40,000
Vish Hegde	$2000	Professor F	$35,200
Xinjian Lu	$1000	Professor G	$34,100
Jed DeVaro	$500	Professor H	$33,500
Bijan Mashaw	$0	Professor I	$32,000
Hongwei Du	-$5000	**Professor J**	**$31,000**

and so on, up to Professor J. Immediately to the right of Professor J would be Gregory Theyel, who would be willing to sacrifice $30,000 of income to live in SF. To the right of Gregory Theyel would be Asha Rao, then Zinovy Radovilsky, and so on, with Hongwei Du at the far right end of the line.

To simplify the example, suppose that all professors in a given department must receive the same salary, but that salary might differ between CSUEB and CSUSF. Incidentally, readers who are familiar with the CSU system know that it is unionized (as I mentioned in section 1.10 and will discuss further in Chapter 5), which means that compensation is constrained by the rules of the collective bargaining agreement and is not entirely market driven. For now, we're going to ignore those constraints, because our focus in this chapter is on understanding how market wage differentials arise in an unconstrained setting of competing employers. In the next two chapters we will deal with constraints, and in particular, Chapter 5 deals with internal union constraints such as those that apply in the CSU system.

Back to the question at hand, which department – CSUEB or CSUSF – should pay a higher average salary? To answer this we need to know the professors' geographic preferences and the two departments' relative demand for professors. All of this information is contained in the table. The professors' preferences are represented by the dollar figures next to their names. And the departments' relative demands for professors are implicitly revealed by the fact that ten professors are employed in each department.

3.6.1 Relative Demand for Professors Is the Same in Both Locations

In the next section we will see what happens if the relative demand for professors shifts so that one of the universities demands more labor than the other. But for now, because each university demands exactly ten professors, we know that CSUEB will pay $100,000, and CSUSF will pay $69,000. How do we know that? A salary of $69,000 (i.e., $100,000 minus $31,000) is just generous enough to prevent Professor J (the "marginal worker", who is indicated in boldface) from quitting and moving to CSUEB. Professor J is willing to sacrifice up to $31,000 of a $100,000 annual salary for the privilege of living and working in SF . . . but no more. If CSUSF were to cut salaries below $69,000, Professor J would move to CSUEB, because that professor is indifferent between working at CSUSF for $69,000 and working at CSUEB for $100,000. Thus, the preferences of the marginal worker (J) determine the size of the salary differential between the two locations.

The professors who are willing to sacrifice the most to live in SF are the ones who join CSUSF, and those who are willing to sacrifice the least join CSUEB. Only Professor J is indifferent. All the others at CSUSF are sacrificing less to live there than they'd be willing to sacrifice, so they're thrilled. Professor A, who is the biggest fan of SF, is the most thrilled of them all. That professor would be willing to sacrifice up to $45,123 (i.e., to accept an annual salary of $54,877) but is instead getting paid $69,000. That's like getting an extra $14,123 per year as a free bonus!

3.6.2 Relative Demand for Professors Differs between the Two Locations

Suppose next that the relative demand for labor changes so that CSUSF requires 13 professors instead of 10, which will leave CSUEB with 7. What would CSUSF have to pay to attract 13 professors? Well, certainly more than $69,000, because at a salary of $69,000 we saw, in section 3.6.1, that only ten professors are willing to work at CSUSF. To hire an eleventh professor (which would be Gregory Theyel, who is the "next closest" to the marginal worker J, meaning the CSUEB professor who places the greatest value on living in SF) $70,000 is needed, because Gregory is only willing to sacrifice (up to) $30,000 of a $100,000 annual salary to live in SF. But to hire 13 professors, it will be necessary to hire Asha Rao and Zinovy Radovilsky, in addition to Gregory Theyel. Zinovy is the hardest of these to attract, because he is only willing to sacrifice $18,000 to live in SF. Therefore, to hire 13 professors, CSUSF must pay $82,000 (i.e., $100,000 minus $18,000). Notice that the pay for all professors at CSUSF is determined entirely by the preferences of the marginal worker, which in this case is Zinovy.

If CSUSF only needed five professors, it would only need to offer a salary sufficient to attract Professor E, so it would offer only $60,000. Then Professor E would be the marginal worker and would be indifferent between working at CSUSF for $60,000 or CSUEB for $100,000. Even though most professors (18 out of 20, to be exact) have a preference for living in SF, there might not be any salary difference between the departments. Suppose that CSUSF demands 19 professors. It would need to offer just enough to attract Bijan Mashaw, who doesn't value SF at all! So both departments would pay $100,000. Alternatively, if demand at CSUSF is so strong that it has to hire people who actually dislike SF (like Hongwei Du) then the salary actually exceeds $100,000. CSUSF would need to pay $105,000 to hire Hongwei.

3.7 Lessons for Managers

What are some important takeaways from the preceding examples? Workers sort across firms according to their preferences. Pay has to be generous enough to attract the marginal worker that the firm demands. Therefore, the marginal worker's preferences determine the size of the market wage differential. Even if almost all workers like (or dislike) a particular job characteristic, there will be no market wage differential if the marginal worker doesn't care one way or the other about that characteristic.

As a manager, you should understand that every job characteristic, both positive and negative, can give risk to a compensating wage differential. Actually measuring those differentials accurately is extremely difficult, even for experienced academic researchers. The biggest challenge is ensuring a true "apples-to-apples" comparison, which is essential to identify a compensation differential. Statistical methods (in particular regression analysis) are used for this purpose, and we will develop those tools in Chapters 6 and 7. Actually directly measuring a compensating differential is probably not something you'll be called upon to do as a manager. Similarly, you won't need to identify the so-called "marginal worker", and doing so is extremely difficult. But the concept is still a useful conceptual benchmark that helps you to understand how competition between you and your competitors, combined with differences in workers' preferences for various job characteristics, gives rise to the compensation levels we actually see in the labor market.

Recalling the definition of compensation from section 1.1, you should think about each of your employee's compensation in the context of the entire set of job characteristics that those workers like (or dislike). Getting precise information on your workers' preferences is difficult, and workers often have incentives to misrepresent their true preferences. But you can sometimes obtain pieces of information that help to sharpen the focus of the picture, as in the Walmart case discussion at the end of Chapter 11.

The strength of your workers' preferences for your company's geographic location is a good example of a factor that has important implications for how likely the worker is to be successfully recruited by another employer, which in turn has important implications for the worker's current compensation in your organization. Learning that your star employee's youngest child has graduated from high school and moved away to college might tip you off that your star's attachment to the geographic location may have weakened, making you more vulnerable to having your star poached by a competitor. As a manager, thinking in terms of compensating differentials can be helpful even if you're not called upon to directly measure them. Such thinking is also helpful when negotiating over compensation (see Chapter 14) on either side of the bargaining table.

Understand that not all job amenities that workers value are expensive to offer. Simply being honest, nice, and treating your workers with respect is a valued non-monetary benefit that can make your workers tolerant even if you're a little cheap with the monetary compensation. Being nice and treating people with respect is free, and it can allow savings on monetary compensation that improve your organization's bottom line. Be ever alert to other cheap and easy ways to make your workers happy. Note that this means that as a manager you not only determine your employees' compensation, but *you yourself* are actually a non-monetary component of their compensation! Remember that your behavior as a manager has profound effects on your workers' behaviors (both incentive and sorting effects, as discussed in section 1.8). You should regularly perform self-reflection on whether you are a reason your workers value their jobs.

Case Discussion 3: The Deadliest Catch[1]

Dutch Harbor, Alaska, has been the largest seafood processing port in the United States for decades. Commercial fishing is a dangerous business, and the deckhands (most of them males between the ages of 22 and 55) risk severe injury and even death while fishing the waters of the Bering Sea. The risks are substantial, but so is the pay, and deckhands travel to Dutch Harbor from distant locations to undertake the work. Only about 20% of them actually live in Alaska, and two-thirds come from CA, OR, or WA. When fatalities occur, they are most often due to drowning or hypothermia. Bad weather, and in particular cold weather, dramatically elevates the risk. The fatality rate is about five times higher in the winter than in the summer, due to a substantial drop in water temperature and more than a doubling of the height of waves.

Regulations dictate when the fishing must occur, with the seasons differing for different species of fish. These calendars are drawn with the preservation of the species in mind, but one implication is that fishing occurs year-round, including in the dangerous winter months. All deckhands are compensated based on revenue sharing (i.e., the deckhands equally share the revenue generated from a fishing expedition) rather than fixed salaries. Seafood producers sign contracts with the

---

[1] Note: This case is based on "The Estimation of Compensating Wage Differentials: Lessons from the *Deadliest Catch*", by Kurt Lavetti, in the *Journal of Business and Economic Statistics* (2019).

fishing boats to deliver a specified amount of fish by a certain date, for a fixed price. That makes the total revenue of a boat easy to predict. What's harder to predict is how many person-hours of labor will be required to catch the specified amount of fish. So a deckhand's hourly earnings vary a lot throughout the year, even for deckhands who remain with the same boat throughout the year. Similarly, the revenue sharing rate (i.e., the percentage of total revenue that the deckhands get) varies substantially throughout the year.

Risk (of injury or death) is a negative job characteristic that can be expected to give rise to compensating differentials. The greater the risks, the greater the pay that workers will demand to compensate them for that risk.

Questions

1. Explain how and why the compensation of deckhands in a given season depends on the supply of available deckhands willing to work during that season.
2. Explain how and why the compensation of deckhands changes over the course of the calendar year (i.e., from winter to summer).
3. Henceforth, suppose there are two types of deckhands, both equally productive. "Type As" are willing to take significant risks, whereas "Type Bs" are very concerned about their personal safety. When making hiring decisions, fishing companies cannot observe who is of Type A and who is of Type B, and because revenue sharing is used, all deckhands on a ship must get paid the same.
   (a) Explain how the composition of employed deckhands (i.e., the percentage of As versus Bs) varies over the year's seasons. Also, explain how the resulting pay level varies across the year's seasons.
   (b) How do your explanations in #2 depend on the relative supply of "A" and "B" workers and the difference between As and Bs in their tolerance for risk?
   (c) If the difference in preferences between As and Bs is extreme, are there any changes to the compensation system (and to recruiting and screening processes) you'd recommend? If so, explain them. What if the difference in preferences between As and Bs is very mild?

Further Reading

Brown, C. (1980). "Equalizing Differences in the Labor Market." *Quarterly Journal of Economics*, 94, 113–34.

DeVaro, J. and Maxwell, N. (2014). "The Elusive Wage-Benefit Trade-off: The Case of Employer-Provided Health Insurance." *International Journal of Industrial Organization*, 37, 23–37.

Gertler, P., Shah, M., and Bertozzi, S. M. (2005). "Risky Business: The Market for Unprotected Commercial Sex." *Journal of Political Economy*, 113, 518–50.

Gibbons, R. and Katz, L. (1992). "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?" *The Review of Economic Studies*, 59, 515–35.

Glomb, T., Kammeyer-Mueller, J., and Rotundo, M. (2004). "Emotional Labor Demands and Compensating Wage Differentials." *Journal of Applied Psychology*, 89, 700–14.

Kilbourne, B. S., England, P., Farkas, G., Beron, K., and Weir, D. (1994). "Returns to Skill, Compensating Differentials, and Gender Bias: Effects of Occupational Characteristics on the Wages of White Women and Men." *American Journal of Sociology*, 100, 689–719.

Konrad, A. M. and Langton, N. (1991). "Sex Differences in Job Preferences, Workplace Segregation, and Compensating Earning Differentials: The Case of Stanford MBAs." *Academy of Management Proceedings*, 1, 368–72.

Lavetti, K. (2019). "The Estimation of Compensating Differentials: Lessons from The Deadliest Catch." *Journal of Business and Economic Statistics*, forthcoming.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Wallace, J. E. and Kay, F. M. (2009). "Are Small Firms More Beautiful or is Bigger Better? A Study of Compensating Differentials and Law Firm Internal Labor Markets." *The Sociological Quarterly*, 50, 474–96.

4 External Constraints on Pay

In the summer of 1988, I held a job at a Taco Bell restaurant in Somerdale, New Jersey, earning an hourly wage of $3.35, which was the federal minimum wage (and also the NJ state minimum wage). Working there was a memorably regrettable experience. I got off on the wrong foot, so to speak . . . or at least on the wrong shoes. The store manager objected to my white sneakers, which were the only shoes I owned at the time. He told me that I was expected to purchase brown shoes and wear them to work daily with my uniform (which the restaurant provided for free, with the exception of shoes). I didn't immediately comply. He reminded me several times. Finally, when the situation became dire and I feared getting fired, I headed to the mall and bought the shoes.

4.1 Wage Theft: a Reprise

His demand was probably illegal and amounted to wage theft that deprived me of nearly an entire day's pay. I don't remember what those brown shoes cost me back in 1988. But nowadays, in 2019, a reasonable cost for a new pair of brown shoes might be around $55. Adjusting for an annual inflation rate of about 2.64% during that 31-year period (see Chapter 1's appendix), those shoes would have cost a little over $25 in 1988 dollars. Dividing $25 by the hourly wage of $3.35 is around 7.5 work hours, or basically a full-time shift. Suppose that instead of requiring me to buy the shoes myself, the manager had purchased them for me and then deducted the cost from my next paycheck. That would have been an enforceable violation of the Fair Labor Standards Act (FLSA) of 1938, because my pay stub would have reflected a reduction in my hourly wage below the federal minimum. What my manager did was tantamount to that, but it was more discreet because no visible payroll deduction was ever applied. Such instances of wage theft are difficult to detect and punish. Most workers don't even know that such behavior is illegal.

But suppose that instead of a mundane pair of brown shoes, I had needed a special proprietary uniform that only my employer could have provided. An example would be a Mickey Mouse or Pluto costume, like those worn by the employees at Disney World. Because such costumes can't just be bought at a shopping mall like my pair of brown shoes, I couldn't have been asked to purchase the costume on my own. The company, if it had wanted me to pay for the uniform, would have had to provide it and then charge me for it via a payroll deduction. That's exactly what Disney did to its workers! And in March of 2017, that move cost the company nearly $4 million, which it had to pay back to 16,339 employees. The agreement was struck with the US Department of Labor, and it involved several violations of the FLSA. In addition to the costume fees that caused some workers' hourly pay to dip below the federal minimum of $7.25, Disney was faulted for failure to maintain required time and payroll records, and failure to compensate workers for tasks performed during "pre-shift" and "post-shift" periods that were "off the clock".

The Disney example illustrates that running afoul of labor laws can be very costly to an organization. Part of your job as a manager is to prevent that from happening. You won't be able to achieve that just by reading this chapter (or any chapter of any book, for that matter). Labor law is regional. It differs from one country to the next. Even within a country it varies. In the United States, each of the 50 states has its own tapestry of labor laws that overlays the federal laws. And within a given state there may be further layers of legislation at the county or city levels. As a manager, you need to worry about all the laws that apply to your organization in all of the locations in which it operates. That set of laws won't be the same for all readers of this book, and I can't cover all of the laws that will be of interest to every reader. Moreover, the legal landscape isn't static. Existing laws get amended or overturned, new laws get created, and the courts' interpretations of existing laws evolve over time as new precedents emerge. So even if I were to achieve the nearly impossible goal of covering every law that's of interest to each reader, some of that material would be obsolete by the time the book went to press.

So far that sounds pretty bleak, but it's actually not so bad, because my goal in this chapter isn't to get you up to speed on all of the labor law you need to know to be a good manager in your organization. Rather, it's to train you to think correctly about labor law in general. Then, when you're confronted, as you will be, with new legislation or with evolving interpretations of existing legislation, you'll be well-equipped to make good management decisions in uncharted legal terrain. If that's our objective, it doesn't matter so much which specific laws we draw on to frame the discussion, because the principles that emerge will have wider applicability. So I'll focus mostly on US federal labor law, with some occasional references to state (usually California) law. And if you're a manager in Iowa, or even China or France, the chapter is still worth reading, because many of the issues we discuss will pertain to the laws facing your firm.

To stay current on federal labor law in the United States, consult the following website of the US Department of Labor: www.dol.gov/elaws. For labor law at the state and local levels, visit the relevant government websites. State-level information can also be found at www.dol.gov/whd/contacts/state_of.htm.

4.2 What's the Purpose of Labor Law?

Labor law is the set of government-imposed laws and regulations that affect employment relationships. Most labor laws are designed to protect workers from poor treatment by employers, and many concern compensation. Labor law exists because of a presumption that simply letting employers and workers loose in the labor market to do their own thing might lead to bad outcomes, either for society as a whole or for certain groups of people. So the laws are designed to impede and regulate *free exchange* (of compensation for labor services) between employers and workers, so as to protect the parties, usually workers, from such bad outcomes.

Free exchange, in its purest form, means that the parties are completely unrestricted and unregulated in their abilities to form contracts with each other (i.e., to design the terms of their relationship), without the interference of rules and laws. Basically anything goes, as long as both parties are on board with it, and that includes the terms of the employment relationship and whether the parties remain in that relationship at all. Although free exchange offers a useful conceptual benchmark for considering the implications of labor law, it's a pretty extreme situation. After all, most labor markets are governed by at least some rules and laws.

One dimension of free exchange between the parties to an employment relationship concerns their right to end that relationship whenever they want, without repercussions. "At-will employment" (or "employment-at-will") is the principle that allows both parties to walk away freely at any time. Employment-at-will is a narrower concept than free exchange, because it only pertains to the parties' rights to terminate their employment relationship whenever they like, whereas free exchange pertains to those rights *and* to the parties' rights to design the terms of their relationship however they like, should they wish to remain in it. I mentioned that labor laws interfere with, and limit, free exchange. Similarly, those labor laws that concern the terms under which you can dismiss your workers should be understood as incursions into the employment-at-will doctrine. Even though employment-at-will has been understood since the late nineteenth century as the "default" doctrine that governs employment relationships in most states of the United States, it is limited by labor laws, so the situation that exists in reality is more constrained than pure employment-at-will.

This chapter covers many laws that undermine free exchange and (in the case of laws that limit your right to dismiss workers) the employment-at-will doctrine. Often labor laws, particularly the anti-discrimination laws I'll cover in section 4.5, are prohibitions of the form "*You are not allowed to [X] for the reason [Y].*" For example, *X* might be "*fire your workers*" and *Y* might be "*that they are older than 65*". The relevant labor law that makes this prohibition is the Age Discrimination in Employment Act (see section 4.5). ADEA restricts free exchange and, because it limits the conditions under which you can fire your workers, it also constrains the employment-at-will doctrine. In contrast, overtime laws are incursions into free exchange but not into employment-at-will, because those laws constrain the parties' rights to design the terms of their employment relationship rather than their rights to end it.

What a law says on paper isn't always what it means in practice. One central issue is how effectively the law is enforced, and another is how it's interpreted by courts or other arbiters when challenges arise. For example, the FLSA allows employers to pay a minimum hourly wage of $4.25 (which is below the regular federal minimum of $7.25) to workers under age 20 during their first 90 consecutive calendar days of work with an employer. The legislation prohibits you from taking any actions to displace your regular workers so as to hire these young workers (for their first 90 days) at the cheaper rate. It also prohibits you from engaging in "partial displacements" like reducing your workers' hours, wages, or employment benefits, again for the purpose of hiring young workers at the cheaper rate. But such prohibitions are difficult to enforce in practice. Creative and careful employers who want to implement such displacements can probably get away with it pretty easily.

Similarly, Section 304 of the Federal Wage Garnishment Law prohibits you from firing a worker just because some of the worker's pay is being garnished to pay back a single debt to some creditor. The concern that motivates this law is the following. If you observe that one of your workers is having their wages garnished to pay off an unpaid debt, you might think negatively of the worker. After all, the reason they're having their wages garnished is because they failed to pay their debt, which means they reneged on a contract. If your disapproval of that behavior is strong enough, you might even feel disinclined to continue employing that worker. But Section 304(b) states that:

> *Whoever willfully violates [that prohibition from firing a worker whose wages are garnished to pay back a single debt] shall be fined not more than $1000, or imprisoned not more than one year, or both.*

In most cases, a creative employer should be able to get around this law, making the firing look like it was for a reason other than wage garnishment. Except for employers who are ignorant of this law, or sloppy, such laws don't really have teeth.

In Chapter 2, I explained that anti-wage-theft legislation designed to protect the wages of undocumented immigrant workers might actually *lower* the average wage for that group. Such an outcome would probably not have been anticipated by the legislators who crafted the law. This is an example of how labor laws sometimes have unintended consequences. Laws that are well intentioned don't always achieve their desired purposes. Or they may solve one perceived problem while creating another that might be even worse. For example, minimum-wage laws that are designed to increase the welfare of low-income workers might reduce employment rates for those workers by making them more expensive in the eyes of employers, or they might increase the wages of low-income workers who are financially well off (e.g., teenagers who live in upscale households with rich parents).

Evidence of such concerns about negative consequences of the law can even be found directly in the law itself. For example, the FLSA contains "subminimum wage provisions" that allow you to pay certain groups of workers less than the federal minimum wage. These groups include various types of students, and those with mental or physical disabilities that impair their productivity. Even the literature of the US Department of Labor's Wage and Hour Division (the entity responsible for administering and enforcing federal labor laws concerning wages and hours) explains that such exceptions to the minimum wage are "authorized to prevent curtailment of opportunities for employment". In other words, the law explicitly recognizes the concern that a minimum wage can reduce employment.

Labor law can originate at all levels of government (e.g., in the United States, at the federal, state, and local levels). Sometimes different levels of government create laws with the same intent but with varying degrees of stringency. When that happens, the more stringent law typically prevails. For example, as of 2019 the US federal hourly minimum wage remains $7.25, where it has been for a decade, whereas on January 1, 2018 the California State hourly minimum wage increased from $10.50 to $11 for firms with at least 26 workers (and it increased from $10 to $10.50 for smaller firms). California employers must currently pay the state minimum wage, because it exceeds the federal minimum. A more dramatic recent example concerns the roughly 40,000 workers at the three main airports serving New York City (i.e., La Guardia, Kennedy International, and Newark Liberty International), who will get a new hourly minimum wage of $19. The new minimum wage will be phased in over five years and will be the highest minimum wage by any public agency in the United States. The increase was approved unanimously by the commissioners of the Port Authority of New York and New Jersey in late September of 2018. The minimum wage applying to those airport workers far exceeds the US federal minimum wage and also the state minimum wages of both New York and New Jersey.

As I explain in Chapter 5, much of what unions do is in the same spirit as what labor laws do. Unions try to regulate pay and hours in a way that will benefit workers, and also protect workers from various types of discrimination and maltreatment. Once a collective bargaining agreement is signed, from your perspective as a manager you can think of it in virtually the same way that you think about labor law. The collective bargaining agreement is the body of "internal rules" that you must follow, and the government-imposed labor laws are the "external rules" that you must follow, but in the end they're all just rules that you must follow (or risk consequences by violating them).

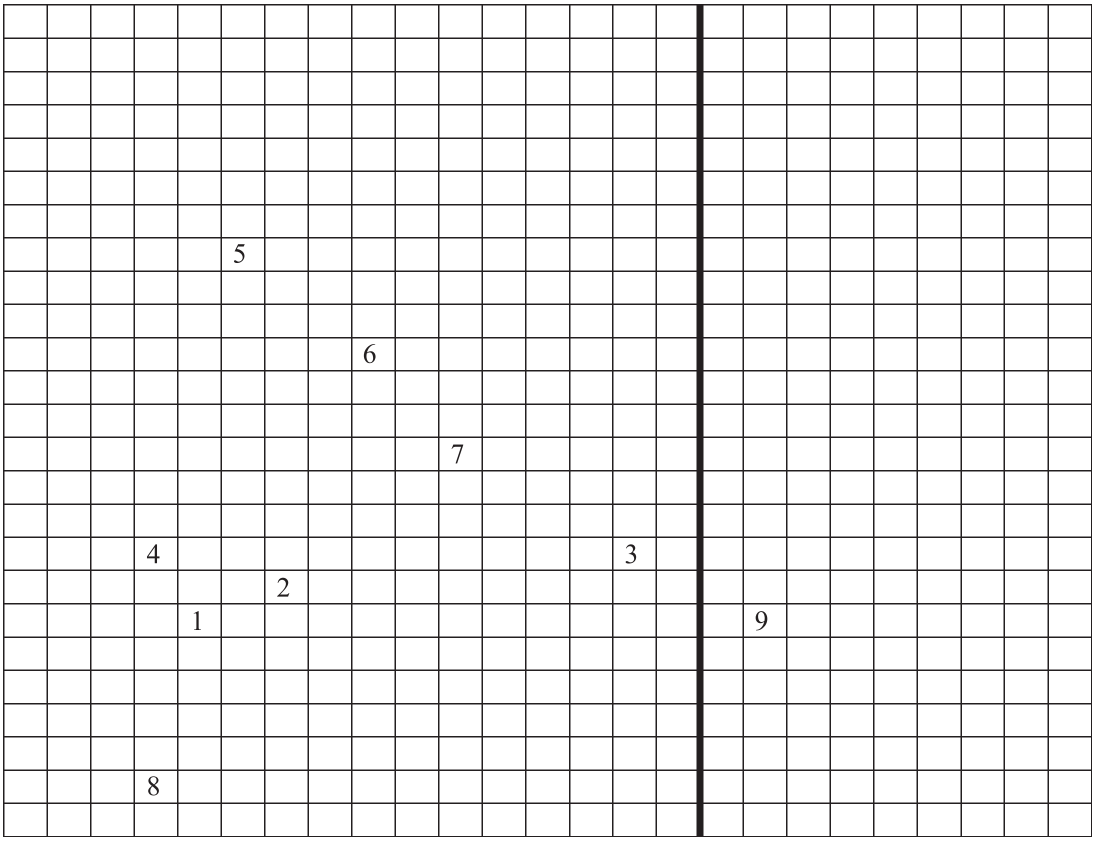


Figure 4.1 Managerial decisions with free exchange and no labor laws.

Let’s a take a look at a visual representation of how labor laws reduce profit by impeding free exchange and potentially employment-at-will. Figure 4.1 represents a business environment of pure free exchange, in which there are no labor laws or regulations. There are 625 small white squares, each of which represents an entire package of decisions you might make as a manager, with the goal of elevating the company’s profit. That package of decisions includes compensation policy, human resource management practices, job assignments, etc. I have drawn a heavy black line to separate those managerial decisions involving layoffs and firings (i.e., any white squares to the right of that line) from those that do not. The reason for this separation of the “decision squares” into two groups is so that we can talk about employment-at-will.

Each of the squares contains an integer from 1 to 625, and I’ve labeled the first 9 of them. To understand the meaning of these numbers, imagine a (very jagged!) mountain of crystal clear ice that rises up from the surface of Figure 4.1, covering it entirely, so that no matter where on the mountain you stand you can look straight down through the ice and see the number that appears on the white square directly below you. The numbers represent elevations, with 1 being the highest and 625 the lowest. The ice mountain represents profit, and your goal as a manager is to get to the highest elevation possible (see section 1.4). This is easy in Figure 4.1, because you are completely unconstrained and can see all the numbers. You'll choose the peak of the mountain. From there, if you look straight down the through the ice, you will see the square labeled "1".

Figure 4.2 is not, in fact, a screenshot from a game of Tetris! Rather, it introduces labor laws, represented by grey patches that prohibit you from choosing certain decision squares that were legal in Figure 4.1. I've depicted 5 labor laws. Regulation *A* is in the northeastern region, Regulation *B* lies to the left and contains squares 5 and 6, Regulation *C* contains squares 1 and 2, Regulation *D* straddles the heavy line, and Regulation *E* "greys out" the bottom two rows of white squares. For example, Regulation *E* might be the federal minimum wage law (currently $7.25 per hour in 2019), and the bottom two rows of squares might be

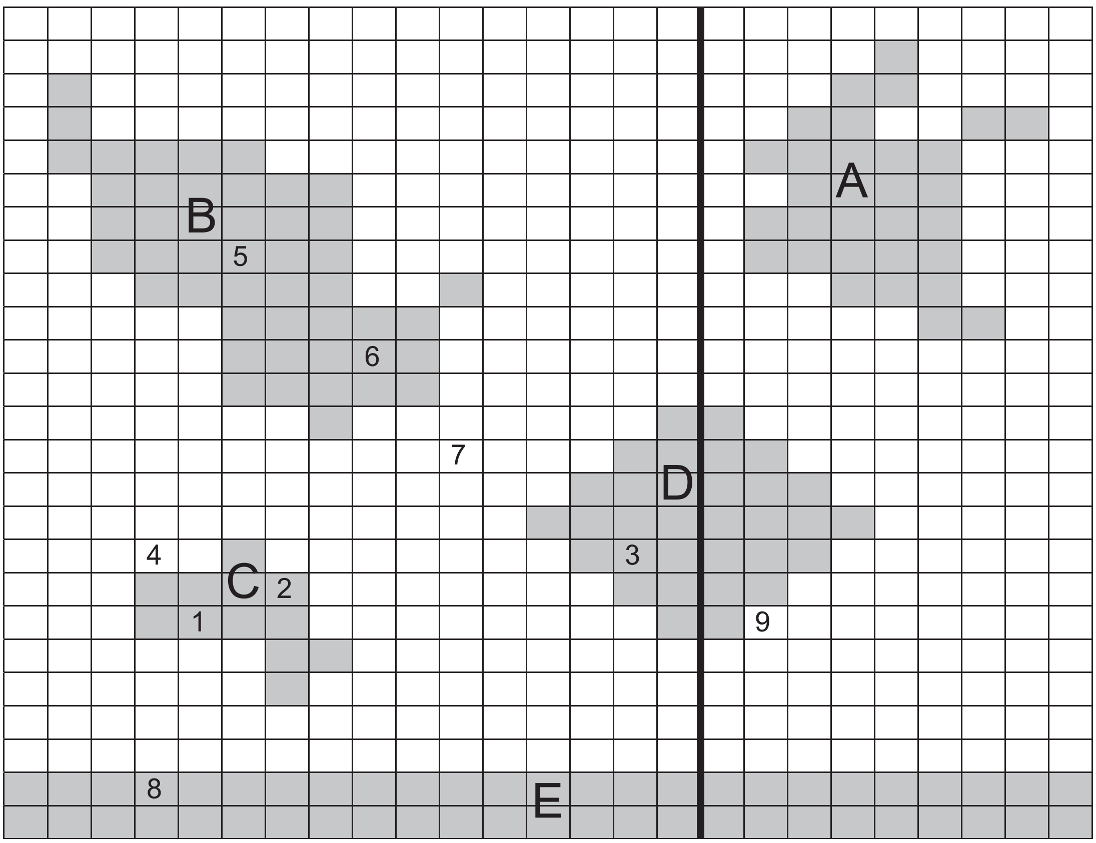


Figure 4.2 Managerial decisions with labor laws that impede free exchange.

managerial decisions that would involve illegally paying your employees below $7.25 per hour.

Regulation *C* prohibits you from choosing the package of decisions in square 1 that would give your company the highest profit, as you did in the unconstrained world of free exchange in Figure 4.1. In fact, you can't even choose the second-best package of decisions, because "2" is also greyed out by Regulation *C*. Your third-best option is allowed by Regulation *C* but blocked by Regulation *D*, so the best you can do for the company is the fourth-best option. If there is a legislation change that tightens Regulation *C* (expanding its reach) so that square 4 suddenly becomes illegal, you would have to drop all the way down to square 7, because 5 and 6 are both blocked by Regulation *B*.

Sometimes labor laws overlap. For example, you could imagine Regulations *B* and *D* both being expanded to cover square 7. Then even if one of those laws (say, *B*) is later relaxed so that square 7 is no longer covered, that square still remains grey because of Regulation *D*. For example, Idaho's state minimum wage in 2019 is $7.25 per hour. But even if Idaho eliminates its state minimum wage, the federal minimum-wage law would prohibit you and other Idaho managers from paying your employees below $7.25 per hour.

The labor laws in Figure 4.2 impede free exchange by greying out decisions that were (as shown in Figure 4.1) legal in the absence of those laws. Those regulations (i.e., *A*, *D*, and *E*) that "grey out" squares to the right of the heavy line also impede the employment-at-will doctrine by limiting your ability to (legally) make managerial decisions that involve dismissals.

4.3 Compensation Constraints and "the 3 Cs"

Rules, whether imposed by unions (as discussed in Chapter 5) or the government, are generally undesirable from your standpoint as a manager, despite the fact that advocates of those rules often argue vociferously to the contrary. Rules constrain whatever actions you're taking to improve your organization's bottom line, and if you were doing a great job of that before the rules got imposed (meaning that you were managing in such a way that profit was as high as possible) then restricting your actions with rules can only hurt profit.

The advocates of rules often try to claim that the rules create "win-wins", i.e., they benefit employers and workers alike, so that everyone's a winner. For example, in the debate over whether the federal minimum wage should be increased to $15/hour, advocates sometimes argue that employers will benefit because the higher wage will mean happier workers and, therefore, lower turnover costs. The problem with that claim is that if you, as a manager, were really so concerned about high turnover costs and truly believed that hourly wages of $15 would be such a winning, profitable move, you could have just done that on your own, without being forced. It's tough to argue that smart managers who know what they're doing will be made better off by forcing them to do what they don't want to do.

To be clear, I'm not claiming that rules (whether imposed externally by the government or internally by a union contract) are "bad" for society, or bad in general. But this book is written for business managers such as yourself, and from your perspectives, rules are generally unwanted constraints that hinder your efforts to increase profit. There are exceptions. For example, suppose you're the market leader in your industry and that you face one major competitor who entered the market after you and who has captured a fifth of your market share. An increase in the minimum wage would harm both you and your competitor by raising labor costs, but as the more established and profitable market leader you're likely in a better position to withstand the financial hit. If your competitor was already operating near the brink of closure, the increased labor costs might be too much to withstand and might require cutting back on production or exiting the market entirely, ceding market share to you. The short-run losses you incur from higher labor costs might be outweighed by the long-run benefits of capturing greater market share.

Such exceptions notwithstanding, the main point is that from a managerial perspective, rules and regulations are generally unhelpful and unwelcome constraints. In the face of those constraints, you have two responsibilities as a good manager. First, and foremost, is to know the rules and be sure that your organization maintains compliance. Second, you should try to find creative ways (*that are both legal and ethical*) to loosen or circumvent those constraints, to the benefit of your organization. Much of what it means to be a good manager is finding creative ways to relax or work around the rules that constrain your organization, without doing anything illegal or that would compromise your integrity.

In short, your managerial mantra concerning compensation constraints – both the external ones covered in this chapter and the internal ones covered in Chapter 5 – should be "the 3 Cs":

The 3 Cs of Compensation Constraints: **C***omprehend,* **C***ircumvent,* **C***omply.*

Your first job is to thoroughly *comprehend* all relevant labor laws, provisions of union contracts, and other constraints. Once you comprehend them, you should devise creative ways – that are both legal and ethical – to *circumvent* the rules to improve the bottom line. To the extent that you cannot circumvent the constraints, you must *comply* with them, to save your organization from hefty litigation costs and moral turpitude. Speaking of moral turpitude . . .

You might be troubled by an apparent inconsistency concerning the "second C". After all, I said that you should circumvent the rules without doing anything unethical, but isn't skirting the rules itself inherently unethical? Well, the fact of the matter is that we confront ethical dilemmas all the time, in all areas of life, and often we resolve those dilemmas by making reasoned decisions that skirt or break rules. For example, suppose that your child experiences a serious allergic reaction and starts having trouble breathing while you are driving her to school. If you park your car illegally in a handicapped parking space at the nearest doctor's office, so as to get your child medical treatment as fast as possible, few would question your decision on ethical grounds, even though it involved breaking a law. By the way, the distinction between circumventing the rules and actually breaking them – i.e., breakage violates a law's letter, whereas circumvention violates its spirit – involves a fine line, and similar ethical issues arise in either case.

You can read entire books, and take entire courses, on the subject of business ethics. That important material is worthy of your close study, though it is beyond the scope of this book. Ethical dilemmas are tough problems (which is why they're called dilemmas!), and resolving them in a reasonable way requires getting deeply into the particular details of each dilemma. For our current purposes, I'll just give you a couple of guiding principles. First, identify the potential winners and losers of each decision that you are contemplating, and the costs that they would suffer if the decision doesn't go their way. In the example from the preceding paragraph, the main winner from breaking the law would be the child who gets prompt medical treatment, and the cost to the child of respecting the law by avoiding the handicapped parking spot might be a serious deterioration in health, and potentially death, from a delay in medical treatment. The main losers would be any handicapped patients who might arrive and be unable to find a parking spot near the main entrance of the doctor's office, and the cost to them would be the inconvenience and possible pain of having to park further way from the entrance. Second, identify the key stakeholders, and ask yourself how comfortable you would feel justifying to them your resolution to the ethical dilemma.

I have found both guiding principles to be useful in my own managerial role as a department chair in a state university. Sometimes my decisions have implications for financial resources or for people's careers, either directly or indirectly. In such cases I consider all of the potential winners and losers, with particular attention to students and to California taxpayers. I then imagine a "town hall meeting" filled with students, prospective students, California taxpayers, and other stakeholders, and ask myself whether I'd feel comfortable defending my decision to that audience or to a judge and jury. There is no easy answer to an ethical dilemma. Your obligations as a manager are to become versed in the fundamental issues in business ethics, to develop some core guiding principles, and to apply those principles consistently to resolve, to the best of your ability, the ethical challenges arising from the "second C".

You can think about the 3 Cs in the context of Figures 4.1 and 4.2. The first C, i.e. "comprehend", involves figuring out exactly which of the 625 squares are greyed out and which remain white. That task is sometimes straightforward, as in the case of minimum-wage regulations. Other times it's harder, particularly in "grey areas" (pun fully intended!) that are murky and that require interpretation, ultimately by courts or regulatory authorities. The second *C*, i.e., "circumvent", involves finding ways to choose a "greyed out" square without technically violating the regulation, or, another way to say this is finding ways to make a grey square white. For example, imagine that Regulation *A* in Figure 4.2 covers square 10, which is a package of decisions that (because it lies to the right of the heavy line) involves worker dismissals, and suppose that squares 1 through 9 are all prohibited by other regulations. Perhaps square 10 involves dismissals of some of your older workers, and Regulation *A* is the Age Discrimination in Employment Act (see section 4.5) that limits such dismissals. If you carefully document performance-based reasons for dismissal that are unrelated to age, then square 10 (which offers higher profit than square 11, wherever that lies in Figure 4.2) may become feasible with minimal legal risk. Finally, when circumvention is impossible, the third *C*, i.e., "comply", involves choosing a square that is unambiguously white. When considering choosing a square that is, or might be, grey, you must consider all costs of non-compliance, including expected legal costs and public relations costs (both of which depend on the likelihood that your non-compliance is detected) and ethical costs (which are incurred whether or not you get caught).

4.4 What Are the Main Types of Labor Law?

Most labor laws can be categorized into one of two types. The first type aims to protect certain groups of workers from discrimination. For example, the Equal Pay Act of 1963 aims to protect women from discrimination in pay, and Title VII of the Civil Rights Act of 1964 aims to protect workers from discrimination based on race, gender, color, religion, or national origin. Similarly, the Age Discrimination in Employment Act (ADEA) aims to protect workers from age discrimination, and the Americans with Disabilities Act (ADA) aims to protect them from discrimination based on disability. I cover these laws in section 4.5.

The second type of labor law concerns restrictions on wages or work hours. I cover these laws in sections 4.6 and 4.7. I use the term "wages" to also mean "salaries". Sometimes those two terms are distinguished because of the FLSA, which requires that overtime pay of time-and-a-half be given to all "non-exempt" workers (who typically earn an hourly *wage*). Employers do not have to provide overtime pay to "exempt" workers, and most people earning an annual *salary* are exempt. If overtime pay is the subject of discussion, the distinction between wages and salaries becomes important, but otherwise it's usually harmless to use the terms interchangeably.

Restrictions on wages and salaries often involve the imposition of either *floors* or *ceilings*. As these names imply, floors represent minimum permissible compensation levels, and ceilings represent maximum permissible compensation levels. Examples of floors include laws establishing federal and state minimum wages, living wages, and prevailing wages. An example of a ceiling is the compensation cap imposed in 2008 on the CEOs of US firms that received public bailout assistance; the relevant piece of legislation was the American Recovery and Reinvestment Act of 2009 (ARRA). Caps can also be applied to groups of workers rather than to individual ones. This happens whenever a manager of a particular group is given a salary (or bonus) budget that can be allocated across the workers in that group but that cannot be exceeded. In the National Football League, for example, each team is limited to paying a fixed amount for the 53 players on its roster. Some teams spend 20% on a quarterback and very little on running backs, whereas other teams place a quarterback on a cheap "rookie contract" and spend 60% of the "cap space" on defense rather than offense.

Examples of restrictions on wages that effectively restrict work hours are the federal and state overtime regulations that require employers to give premium pay once workers' hours exceed a certain threshold. A more extreme and direct example of hours restrictions would be child labor laws that restrict the work hours of people below a certain age to equal zero. Another example would be mandatory retirement laws that restrict the hours of people above a certain age to equal zero. Mandatory retirement was abolished in the United States in 1986 but still exists for certain jobs, such as airline pilots.

4.5 Protections against Employment Discrimination

In the United States, labor laws that are designed to protect various groups of workers against employment discrimination are administered and enforced by the Equal Employment Opportunity Commission (EEOC). The first major piece of legislation is the Equal Pay Act of 1963 (EPA), which is an amendment to the FLSA. The EPA prohibits sex-based wage discrimination between men and women in the same establishment who perform jobs that require substantially equal skill, effort, and responsibility, under similar working conditions. In other words, it requires "equal pay (between the sexes) for equal work". The "equal pay" part of that equation is a lot easier to measure and verify than the "equal work" part. Unequal pay is often easily noticed, and indeed that's what leads workers (usually, but not always, women) to pursue litigation under the EPA. But what has to be battled out in court is whether or not the work is indeed "equal". A significant body of legal precedent has accumulated during the last few decades that helps to clarify how the courts interpret "equal work".

What motivated the EPA was a sizeable gap in average pay between men and women, with the men getting paid more. The gap has narrowed over time but still remains, and it also exists in most countries outside the United States. The EPA is somewhat narrow in scope, because the “equal work” requirement is a fairly stringent test. What happens when the work isn’t exactly equal? There may be men and women working in the same establishment whose pay differs but whose work differs as well. Then the EPA doesn’t apply, but Title VII of the Civil Rights Act of 1964 may apply, and its scope extends far beyond gender.

Title VII prohibits employers from discriminating against workers on the basis of sex, race, color, national origin, and religion, and it applies to most employers employing at least 15 workers for 20 or more weeks in the current or preceding calendar year. The protected groups have been expanded by subsequent legislation protecting older workers (Age Discrimination in Employment Act of 1967), pregnant women (Pregnancy Discrimination Act of 1978), and the disabled (Americans with Disabilities Act of 1990). One noteworthy exception is that Title VII offers no protection against discrimination based on sexual orientation, as emphasized in July 2017 in an amicus brief filed by the Department of Justice. The brief argued that “The sole question here is whether, as a matter of law, Title VII reaches sexual orientation discrimination. It does not, as has been settled for decades. Any efforts to amend Title VII’s scope should be directed to Congress rather than the courts”, and concluded that “Title VII does not prohibit discrimination because of sexual orientation.”

There are some exceptions that allow discrimination against a protected group if the defining characteristic of that group is a “bona fide occupational qualification (BFOQ)” reasonably necessary to the normal operation of that job. It’s a narrow exception, and to qualify for it requires you to prove three things: There’s a direct relationship between your worker’s personal characteristic in question and their ability to perform the job, the BFOQ relates to the essence or central mission of the business, and there’s no less restrictive or reasonable alternative. Certain exceptions to Title VII are made for the federal government, federally recognized Native American tribes, religious groups performing work connected to the group’s activities, and bona fide nonprofit private membership organizations. Enforcement is via lawsuits which can be filed either by the EEOC on behalf of the aggrieved parties, or directly by the parties themselves.

Protections similar to those of Title VII were extended in 1990 to workers with mental or physical disabilities, with the passage of the Americans with Disabilities Act (ADA). But whereas Title VII simply aims to guarantee that workers in a protected class are treated the same as other workers, ADA goes a step further by mandating that disabled workers receive some additional resources that the employer might not otherwise provide. In particular, employers covered under ADA must provide “reasonable accommodations”, unless doing so would impose "undue hardship" on the employer. Reasonable accommodations refer to changes to a position or workplace that will enable a disabled worker to perform his or her job. The crux of the law boils down to the definitions of "reasonable accommodation" and "undue hardship", and delving deeply into how those terms have been interpreted by the EEOC and the courts would take us too far afield. But as a manager, you should familiarize yourself with the guidelines on the EEOC website and stay up to date with the relevant case law.

As a manager, you should be well versed in ADA if you employ workers with disabilities or if you encounter them as job applicants. Be aware that employing these workers comes with additional responsibilities and risks. The responsibilities are providing any reasonable accommodations that are deemed necessary at the time of employment or that may emerge subsequently, and the risks involve the threat of an ADA lawsuit if you do not fully comply or if you take some other action that might be interpreted under ADA as disadvantageous to your disabled employees.

Employers' awareness of the additional responsibilities and risks associated with employing disabled workers creates the possibility that the law may make employers more hesitant to hire disabled workers in the first place. The issue is not unique to disabled workers and extends to any legislation that affords protections to a particular class of workers. Even when, as in the case of Title VII, the law only requires that workers from the protected class be treated like all other workers, the mere threat of a potential lawsuit (even an unjustified and frivolous one that will never succeed but that is, nonetheless, costly to the employer) can make employers hesitant to engage with the workers who enjoy those protections.

4.5.1 Age Discrimination in Employment Act (ADEA)

The Age Discrimination in Employment Act (ADEA) of 1967, which applies to firms employing at least 20 workers within the current or prior calendar year (including those overseas), prohibits you from discriminating against workers who are at least 40 years old. In ADEA's original conception, discrimination was prohibited in hiring, promotions, wages, firings, and layoffs. That scope was narrowed on the "hiring" dimension on January 23, 2019 by the Seventh Circuit court, which ruled in *Kleber* v. *CareFusion Corporation* that ADEA's prohibition of *disparate impact discrimination* (as opposed to *disparate treatment discrimination*) only pertains to current employees and doesn't extend to job applicants. ADEA also prohibits employers from stating age preferences or limitations, such as in job advertisements, which would be disparate treatment discrimination. Benefits to older workers can legally be reduced only if the cost to the employer of the reduced plan (e.g., an employer-provided health insurance plan) isn't lower than the cost of providing full benefits to younger workers.

The Act was amended in 1986 and 1991. The 1986 amendments abolished mandatory retirement for most workers, with some exceptions such as airline pilots.

If you lose an ADEA litigation case against your current or former employees, you can expect to be liable for compensatory and punitive damages, and in some cases you might have to reinstate your former workers. Don't get the impression that this law, or similar ones protecting other worker groups, means that older workers are untouchable. You can still always discharge or discipline a worker for "good cause", based on "reasonable factors other than age". But you must take care that the reasonable factors are amply documented so that your organization has a robust defense in the event of litigation.

The Act is designed to prevent employers from favoring younger workers over older workers, where 40 is the age threshold. Favoring older workers over younger ones, however, is completely fine, even if the younger ones are, themselves, over 40. There are some exceptions that allow for discrimination of older workers. You can legally specify an age limit if age has been shown to be a bona fide occupational qualification (BFOQ) reasonably necessary to the normal operation of the particular business. For example, directors of musical theater productions would be on safe legal grounds discriminating against older actors auditioning for the lead roles of *Annie* or *Oliver!*, because those roles are intended to be children.

A 2008 ruling of the US Supreme Court (*Meacham* v. *Knolls Atomic Power Lab*, 554 U.S. 84) held that if there is a dispute over whether a layoff or other action that disproportionately hurt older workers was based on age, the burden of proof rests on the employer to show that the action was based not on age but on some "reasonable factor other than age". As stated, the ruling sounds friendlier to workers than to employers, and indeed it is. But how much bite does it have? Probably less than you might think. This ruling is a good example of an important, general principle that applies to a lot of labor laws, namely that employers who intimately *comprehend* the law (the first of "the 3 Cs"), can largely *circumvent* it (the second of the 3 Cs) at little cost, whereas those who don't know the law can get burned badly by it.

For example, suppose that you're an employer who wants to eliminate a department because it happens to be comprised of older workers who are driving up your company's health insurance costs and whose health problems are adversely affecting productivity. If you thoroughly comprehend the ADEA, you would proceed with great care, avoid any references to age, and meticulously "build a case" before eliminating the department, so that after the fact there would be ample documentation that "reasonable factors", and not age, were the basis for your decision. In contrast, if you're an ill-informed manager who doesn't know the law and who is prone to making unforced errors, like sending emails (that may be subject to subpoena in age-discrimination litigation) that explicitly state your intention to discriminate based on age, then you're likely to lead your company into legal trouble.

Be aware that if you want to shed workers who happen to be older, often explicit layoffs aren't even necessary. Rather, you can take actions that will induce sorting effects (see section 1.8) that achieve the desired outcome. Simply changing the job tasks, or the compensation, in a manner less appealing to older workers, may induce those workers to quit or retire earlier than they otherwise would, or not apply for the job in the first place. In the Walmart case discussion at the end of Chapter 11, for example, the company proposed redesigning jobs in a way that included more physical labor (i.e., collecting shopping carts from the parking lot) and that older, highly paid, and less-productive workers would find unappealing.

Inducing workers to quit, via the sorting effect, rather than directly firing them, offers you some protection, but you must still be careful. If you're found to have created a hostile work environment that induced your workers to quit, your former employees may have legal claims on the basis of *constructive dismissal*, also known as *constructive discharge*, which is the term used when a worker's resignation, though technically voluntary, was induced by hostile working conditions. The following is a three-part test provided by the EEOC to determine whether constructive discharge has occurred:

(1) a reasonable person in the complainant's position would have found the working conditions intolerable
(2) the intolerable working conditions were created by employer conduct that constituted discrimination against the complainant
(3) the intolerable working conditions caused the complainant's involuntary resignation.

As a manager you need to be careful not to find yourself on the losing side of a constructive dismissal case. Nonetheless, you'll generally have an easier time in court if your worker separated for voluntary reasons than because of a layoff or firing. As I mentioned, ADEA places the burden of proof on you to show that your action was based on reasonable factors other than age, whereas once your employee voluntarily quits, the burden of proof tends to shift from you to them. For example, consider the following definition of constructive discharge by the California Supreme Court in the 1994 case of *Turner* v. *Anheuser-Busch, Inc*:

> In order to establish a constructive discharge, an ***employee must plead and prove*** [emphasis added], by the usual preponderance of the evidence standard, that the employer either intentionally created or knowingly permitted working conditions that were so intolerable or aggravated at the time of the employee's resignation that a reasonable employer would realize that a reasonable person in the employee's position would be compelled to resign.

Most of the issues we've just discussed – concerning how you, as a manager, engage with your workers and with labor law – are not unique to age and also pertain to other personal characteristics that give workers protected legal status.

4.6 Wage (or Salary) and Hours Regulations

In the United States, at the federal level, the Wage and Hours Division (WHD) of the Department of Labor (DOL) is responsible for administering and enforcing the labor laws concerning wages and hours.

4.6.1 Fair Labor Standards Act of 1938 (FLSA)

The most important piece of federal legislation concerning wages and hours is the Fair Labor Standards Act of 1938. FLSA establishes the federal minimum wage (which in 2019 is still $7.25 per hour), overtime law that requires "non-exempt" workers to receive time-and-a-half pay for their weekly work hours in excess of 40, restrictions on child labor, and various record-keeping requirements that employers must follow. There are various exemptions to the minimum wage and overtime provisions, and you should regularly consult the law to see which of those, if any, apply in your organization.

The Disney wage-theft example at the start of this chapter illustrates the potentially high costs of non-compliance with the FLSA. What risks do you run if you're not in compliance? The law is enforced both via *investigation* and via *legal remedies*. Investigation is conducted by WHD employees, stationed throughout the United States, who gather data on wages, hours, and employment conditions to determine compliance. When violations are found, the investigators recommend to employers the steps they must take to achieve compliance. Legal remedies might involve administrative procedures, in which the DOL works directly with employers to secure compliance, without involving the courts. Or they might involve litigation, either initiated by DOL (on behalf of the affected workers) or by the workers themselves in private lawsuits. In any or all of these cases, when unpaid minimum wages or overtime is at issue, the employer is liable for "liquidated damages", which *doubles* the cost to the employer. Yes, you read that correctly! So if you are found to owe your workers back wages of $10,000, you will actually have to pay $20,000. If that information, plus the Disney example, isn't enough to scare you into compliance, be aware that criminal prosecution is also a possibility. If you are found to have willfully violated the FLSA, you may be subject to criminal penalties and fines, including imprisonment.

The child labor provisions of FLSA restrict work hours, and the nature of work, for young people. The restrictions get progressively more stringent the younger the worker's age. For people 18 and older, there are no restrictions on the type of work they do or their number of hours. Youths who are at least 16 but less than 18 have no restriction on hours and few restrictions on the type of work they do, but they are not allowed to do hazardous jobs. For example, I became legally eligible for my job at Taco Bell once I turned 16, but in that job there was one task that I wasn't allowed to do . . . operating the deep fryer, which was the vat of boiling oil where all of the nachos and taco shells were made. Kids a few years older than I were allowed to work the fryer, even kids who, in my estimation, weren't particularly competent or responsible. That was an interesting example of the FLSA affecting the allocation of labor within the firm . . . on some shifts, we were all under the magical age of 18 except for one guy, and in those instances he was forced to work the fryer no matter what.

Youths who are at least 14 but less than 16 face various restrictions on how many hours per day and per week they are allowed to work, with distinctions made between school and non-school days and weeks. Youths below age 14 can legally hold only four types of jobs: delivering newspapers, performing in various theatrical contexts, working for parents in the family business, or, interestingly enough, gathering evergreens and making evergreen wreaths. All of the preceding rules pertain to non-agricultural jobs. There are separate rules for farm jobs that are a bit more permissive. Incidentally, states can have their own child labor laws, which are sometimes more restrictive than the provisions of FLSA. Although FLSA imposes no restrictions on workers once they reach 18, some states impose restrictions on workers between the ages of 18 and 21. For example, the legal age at which someone can be a bartender varies from 18 to 21, depending on the state, and it sometimes also varies across cities within a given state.

The FLSA also requires that employers keep records documenting each worker's hours and pay. Records for non-exempt workers (i.e., those eligible for overtime pay) must include:

(1) personal information, including employee's name, home address, occupation, gender, and birth date if under 19 years of age
(2) hour and day when workweek begins
(3) total hours worked each workday and each workweek
(4) total daily or weekly straight-time earnings
(5) regular hourly pay rate for any week when overtime is worked
(6) total overtime pay for the workweek
(7) deductions from or additions to wages
(8) total wages paid each pay period
(9) date of payment and pay period covered.

4.6.2 Prevailing Wages, Living Wages, and Related Legislation

Prevailing wage laws establish minimum wages, but they differ from the minimum-wage provisions of the FLSA in three ways. First, they pertain to much narrower groups of workers. Whereas the federal minimum wage applies to most workers, prevailing wages only apply to work done on government projects. Second, whereas the federal minimum wage applies only to monetary compensation, prevailing wages cover benefits as well. Third, whereas the federal minimum wage is one number (periodically increased by Congress) that applies to everyone, prevailing wages vary from one locality to the next. The third difference means that the law must specify rules to guide what the prevailing wage will be in each location, and that's where things get tricky.

The law requires that surveys be conducted by the DOL to determine what the prevailing wage is in a given area. Taxpayers bear the cost of financing prevailing wages, because those wages only apply to work done for the government. For example, consider large government construction projects that require lots of workers, like building bridges. In the absence of prevailing wage laws, those workers would be paid the market wage, the federal wage, or the state minimum wage (whichever among the three is the highest). But typically the prevailing wage exceeds all three, which raises the bill for taxpayers.

The first major piece of prevailing-wage legislation is the Davis–Bacon Act of 1931, which applies to contractors and subcontractors performing on federally funded contracts exceeding $2000 for the construction, alteration, or repair of public buildings or public works. In 1934, Congress supplemented the Davis–Bacon Act with the Copeland "Anti-kickback" Act, which prohibits wage theft from public works employees in the form of kickbacks. The key provision is given in the following lengthy sentence:

> *Whoever, by force, intimidation, or threat of procuring dismissal from employment, or by any other manner whatsoever induces any person employed in the construction, prosecution, completion or repair of any public building, public work, or building or work financed in whole or in part by loans or grants from the United States, to give up any part of the compensation to which he is entitled under his contract of employment, shall be fined under this title or imprisoned not more than five years, or both.*

The second major piece of prevailing-wage legislation is the Walsh–Healey Public Contracts Act of 1936, which applies to US government contracts exceeding $10,000 for the manufacture or furnishing of goods. The Act establishes overtime pay for contractors' weekly hours in excess of 40 and requires a minimum wage equal to the prevailing wage that is set by the Secretary of Labor. The Act also contains some child labor provisions reminiscent of those in the FLSA, along with some health and safety standards. The McNamara–O'Hara Service Contract Act of 1965 requires contractors and subcontractors performing services on contracts exceeding $2500 to pay workers prevailing wages and fringe benefits, as determined by the DOL, or to honor the terms of a predecessor contract's collective bargaining agreement. For contracts below $2500, the federal minimum wage applies. For contracts exceeding $100,000, overtime pay of (at least) time-and-a-half is required for weekly hours in excess of 40.

Living wage laws establish local minimum wages, at the level of cities, counties, or states. They're similar to prevailing wages in that they vary from one locality to the next, unlike the federal minimum wages. But prevailing wages only pertain to workers on government contracts, whereas living wages sometimes cover broader classes of workers within the locality. Living wages typically exceed federal and state minimum wages, because otherwise they'd be redundant. Living wages usually only cover businesses that receive state assistance or have government contracts, though they sometimes have wider coverage.

4.6.3 Family and Medical Leave Act of 1993 (FMLA)

The FMLA entitles eligible workers of covered employers to take unpaid leave for various family-related reasons. During their leaves, they can maintain their health insurance coverage as if they were not on leave, and at the end of their leave they can return to their job. Up to 12 workweeks can be taken in a year, for any of the following reasons:

- birth of a child and to care for the newborn child within one year of birth
- acquiring a child for adoption or foster care, so as to care for the newly placed child within one year of placement
- to care for the worker's spouse, child, or parent who has a serious health condition
- a serious health condition that makes the worker unable to perform essential job duties
- any qualifying exigency arising out of the fact that the worker's spouse, son, daughter, or parent is a covered military member on "covered active duty".

Alternatively, up to 26 workweeks of leave can be taken in a year to care for a covered service member with a serious injury or illness, if the worker is that person's spouse, son, daughter, parent, or next of kin.

In 2015, the US Department of Labor amended the definition of "spouse" under FMLA to include people in same-sex marriages, so that these individuals can claim FMLA leave time to care for their spouse or family member, regardless of where they live. This change was prompted by the US Supreme Court's decision in *United States* v. *Windsor*, which found section 3 of the Defense of Marriage Act (DOMA) to be unconstitutional. Prior to the Supreme Court's decision, DOMA barred same-sex married couples from being recognized under federal law as spouses, thereby barring them from receiving federal marriage benefits. This is an example of how labor laws are continuously evolving and why it's important, as a manager, to stay current on those that affect your organization.

Some parts of FMLA are much easier to enforce than others. Enforcement is relatively straightforward for the "You [the employer] must do this" parts of the law that say that leave must be granted, who gets it, for how long, and that health insurance must continue unabated during the leave. If an employer violates any of those provisions, a worker would have a clear case in court. Things are murkier for the "You [the employer] must not do that" parts of the law. FMLA prohibits employers from discriminating or retaliating against a worker for exercising or attempting to exercise an FMLA right. Examples include using an FMLA leave as a negative factor in employment actions like hiring, promotions, disciplinary actions, or firings.

A complicating issue here is that (up to) three months is a long absence from work. When a worker is gone for that long, life goes on, the workplace adjusts to the disruption, people get used to not seeing the worker who is on FMLA leave, and that person's tasks get reassigned (temporarily), suspended, or eliminated. If the organization learns how to live without the person for a few months, the feeling may emerge that the worker isn't really needed anymore and that their compensation is a waste of money. This happened (twice!) to my sister. When she had her first child, she took eight weeks of maternity leave. Just a few months after she returned to work, she was informed that her position was being eliminated due to "downsizing". She contends that her extended absence gave the organization occasion to figure out that it didn't really need her job anymore, and that they waited a "respectable amount of time" after her return before firing her, to avoid any whiff of non-compliance with FMLA.

A few years after the first incident, my sister took eight weeks of maternity leave after the birth of her second child. Her (different) employer informed her, while she was still on leave, that they were relocating her position to a different site that would require a 1.5-hour commute each way. She contends that her employer used the occasion of the maternity leave to effectively force her to resign (because firing her would have created grounds for litigation over non-compliance with FMLA). In her words, "What mother of a newborn wants to add 70 minutes each way to her daily commute, when she works 4 miles from home?" My sister might be right about what happened in both instances, but she would have had a difficult time proving that in court. Unless employers are sloppy or ill-informed about the law, they can usually find ways to legally justify a dismissal, at least in the United States where the employment-at-will doctrine is the default.

Prohibitions against firings for specific purposes are hard to enforce. Other examples include the Federal Wage Garnishment Law discussed earlier in the chapter, or the various anti-discrimination laws from section 4.5. My sister's employer was able to induce the desired sorting effect by making the job less appealing for new mothers via a lengthened commute, much like Walmart proposed making the jobs of its older cashiers less appealing by redesigning those jobs to include more physical labor (again, see Chapter 11 and the corresponding case discussion). But in either case the workers would probably have a tough time meeting the legal standards for a constructive discharge case (see the end of section 4.5).

Although the preceding examples concern employer misbehavior surrounding FMLA, worker misbehavior also occurs. In fact, some employers cynically refer to FMLA as the Friday and Monday Leave Act due to some workers' tendencies to use the law as an excuse to take long weekends when they are covered by FMLA's "intermittent leave" provision. Worker abuses of FMLA involve taking more time off than needed, suspicious timing of the time off (e.g., Mondays, Fridays, or dates coinciding with nice weather or major sporting events), or performing poorly during or following an FMLA leave, so as to exploit the employer's fear of an FMLA lawsuit if the worker is disciplined or dismissed. A number of websites and blogs by legal professionals offer employers tips for mitigating and responding to employee abuses of FMLA. See, for example, *FMLA Insights: Guidance and Solutions for Employers* (www.fmlainsights.com).

4.7 Compensation Floors and Ceilings

As I mentioned at the start of this chapter, in 1988 I worked at Taco Bell for $3.35 per hour. I would have preferred much more, but I would have worked for less if necessary. And Taco Bell would have been delighted to pay me less. Legally, however, I couldn't have been paid less, because $3.35 per hour was the New Jersey State minimum wage at the time, which also coincided with the federal minimum wage. So I was lucky to get $3.35, because in the absence of the law, Taco Bell surely would have offered less; as it was, they offered me the absolute minimum required by law, and, to add insult to injury, they made me buy new shoes!

4.7.1 Wage Floors

The wage I was paid at Taco Bell, like all legislated minimum wages, whether state or federal, is an example of a *wage floor*. However, not all wage floors arise from government mandates. As I explain in Chapter 5, unions produce outcomes that resemble minimum-wage laws, because collective bargaining agreements can stipulate wage floors. One difference between union-induced and legislation-induced wage floors is that employers have a direct influence over the floor in the case of unions, because the ultimate agreement arises from bargaining between workers and the employer. In contrast, in the case of legislated floors, employers have no direct input (other than very indirectly via lobbying activities) and must simply accept the minimum that is dictated by the government.

Stories like mine at Taco Bell make it tempting to argue, and indeed many politicians do argue, that minimum wages help workers, and that higher minimum wages help workers even more. During the 2016 US presidential election, in fact, a hot political question was whether the candidates supported an increase in the federal minimum wage from $7.25 to $15 per hour. Some workers clearly benefit from minimum-wage laws. But many others do not. The overall picture for workers, considering all the winners and losers, is less auspicious than you'd guess based on the popular press and what many politicians claim. Before I explain why, let me say a bit more about minimum wages.

When introduced in 1938 by the FLSA, the federal minimum wage was $0.25 per hour. As of 2019, it was still $7.25 per hour, where it had been since July 2009. How did it get from $0.25 in 1938 to $7.25 in 2009? Every so many years, Congress increases it, and then it remains fixed at the new rate until the next increase.

States can also enact their own minimum-wage laws, and the main motivation for doing so is that some states feel that the federal minimum is too low for their workers. So state minimums, when they differ from the federal minimum, are typically higher than the federal minimum. When the two minimum wages differ, the higher one prevails. For example, as of January 2019, the New Jersey State minimum wage is $8.85 per hour, which exceeds the federal minimum of $7.25 per hour. So workers in Taco Bell restaurants in New Jersey cannot be paid less than $8.85 per hour.

Why would a state ever have a state minimum wage that's less than the federal minimum? That would seem to make no sense, because the state minimum would be irrelevant given that workers would need to be paid at the (higher) federal rate. One reason concerns the timing of changes in the law. Suppose that the federal minimum wage increases to $15 per hour. That would be a huge increase that would take the federal minimum above many of the state minimums that exceed the current federal level. For example, the New Jersey minimum of $8.85 lies above the federal minimum of $7.25 but would lie well below a new federal minimum of $15. Even if New Jersey decides that $15 is too low and that a state minimum of $16 is required (which is probably unlikely), the laws in New Jersey don't change overnight. It would take time for the state legislature to pass the $16 and for the new law to go into effect, and in the meantime the state minimum would lie below the federal. And if New Jersey believes that the correct minimum wage for that state should be $15 (or less), there is no reason for the state legislature to rush to change the $8.85, because the legal rate that prevails is already $15, the federal rate.

Suppose that a particular state has a minimum that differs from the federal minimum. It may be higher or lower than the federal minimum, but let's focus on the more usual case in which it's higher. For example, California's minimum as of January 1, 2018 (for employers with at least 26 workers) is $11, which exceeds the federal minimum of $7.25. Suppose that the federal minimum increases by $1 per hour, to $8.25. Does the change in the federal law have any effect on California? It's tempting to say "no", because California workers are already getting $11, so the change in the federal law is irrelevant because the new federal minimum still lies below $11.

But that's not quite correct, and one reason is that some workers may be exempt from the state minimum but covered by the federal minimum. Often minimum-wage laws specify certain exemptions (i.e., occupations that do not have to pay minimum wages, such as waiters and waitresses). If a state exempts certain jobs from the state minimum that are not exempt from the federal minimum, an increase in the federal minimum affects those workers. The bottom line is that when the federal and state minimums differ, the lower of the two is still sometimes relevant, because it might apply to a set of workers who are exempt from the higher of the two.

In the preceding example, even setting aside the issue of exempt workers, there is another reason why California might be affected by an increase in the federal minimum, even if the new federal minimum still lies below the state minimum of $11. Because the change in the federal law affects other states, the composition of workers and employers in California might be affected. In other words, there may be a "sorting effect" in a similar spirit to section 1.8, though in this case workers sort across states rather than across firms, and the change in compensation policy that induces the sorting is a change in government policy rather than a change in an individual firm's compensation policy.

To elaborate on this *cross-state sorting effect*, consider a new company that must decide whether to locate in California or in Utah. Suppose that the company, taking into account all the plusses and minuses of both states, is indifferent between them, and is about to flip a coin to decide where to locate. Incidentally, as of 2019 Utah's state minimum was still $7.25, coinciding with the federal minimum. If the company were indifferent between the two states at the current minimum wages, then news that the federal minimum is being increased to $8.25 will create a clear preference to locate in California. The reason is that the possibility of employing minimum-wage workers in Utah at the low hourly wage of $7.25 has disappeared, since now they will need to be paid $8.25. Utah becomes a less attractive location for employers to open shop, and that favors California, where employers will continue to pay the state minimum of $11 even after the change in the federal minimum.

4.7.2 Do Wage Floors Help Workers?

Some workers clearly benefit from minimum-wage laws. For example, consider workers in a California Taco Bell restaurant in 2019 who receive the state minimum of $11 per hour. Some of those workers were hired prior to January 1, 2018, which is the date on which California's current minimum wage of $11 went into effect. Such workers enjoyed a raise of $0.50 in their hourly wage, as of January 1, 2018, so they clearly reaped benefits from the increase in the minimum wage. But not all workers benefit. The increase in the state minimum means that employers' compensation costs increase, which may cause them to cut back on workers and scheduled hours. This makes it harder than before for workers to find a job or to maintain weekly work hours. So an increase in the minimum wage is great if you can find a minimum-wage job, or hang onto the one you already have, without suffering a reduction in hours. But those jobs become harder to find. Let's consider an example . . .

My job at Taco Bell wasn't my first choice. I had applied to at least a dozen other preferred jobs, unsuccessfully, before finally getting an offer from Taco Bell. During my job search, the closest I came to success was at Woolworths, a company you've probably never even heard of. I would have been much happier to work there than at Taco Bell, even at the same minimum wage of $3.35 per hour. Woolworths was one of the few companies that even granted me a job interview. After all, I was very young and had no work experience. What cost me the Woolworths job, I believe, was a timed math test. I was handed the exam and a pencil and told to return the exam to the interviewer by the set deadline, which was perhaps 15 minutes. I knelt down in an aisle and completed the exam on the floor in perhaps 3 to 5 minutes, including the time I spent checking my work. The questions were all extremely easy and involved only the most basic arithmetic. The interviewer's disapproving glare, and her line of questioning when I submitted the completed exam so quickly, made clear that she thought I had cheated. I offered to retake the exam while being watched, but my request was denied, along with my employment application.

For a number of regretful subsequent years, I felt that if I had only taken a respectable amount of time on the math test, and thrown in a few purposely wrong answers for the sake of propriety, I could have landed the coveted Woolworths job and avoided the unpleasantness of Taco Bell. Perhaps that's true. But I wouldn't bet on it. As a young and inexperienced worker, the primary obstacle I faced at Woolworths (and at the many other employers to whom I unsuccessfully applied) wasn't the math test but rather the minimum-wage law. I strongly preferred Woolworths to Taco Bell, because of the characteristics of both jobs. Recalling our discussion of compensating differentials in Chapter 3, this meant that I would have been willing to work at Woolworths for considerably less pay than the legally mandated minimum of $3.35 per hour. Even at $2 per hour, I'd have happily accepted a job there. And from Woolworths' standpoint, at such bargain-basement prices they would likely have been willing to overlook my youth and inexperience, and probably even my "cheating"! But if Woolworths is forced by law to pay $3.35 per hour, then why hire a very young person with no experience when for the exact same price you can hire someone older with experience? It took me a long time to find a job that summer, and I faced many rejections from employers who would probably have been willing to hire me at the wage level I was willing to accept. Free exchange was impeded by minimum-wage regulations that blocked the formation of mutually advantageous employment relationships.

The moral here is that the workers who tend to suffer the most from wage floors are those who have the weakest labor market prospects, due to inexperience or other factors. The laws prevent society's most vulnerable workers from making mutually profitable exchanges with employers.

As I mentioned, though, some workers clearly benefit from the minimum wage, in particular those who are lucky enough to find a job. Once the lengthy ordeal of my job search was completed and I found a job at Taco Bell, I instantly changed from a worker who was hurt by the minimum wage to one who was helped by it. The $3.35 per hour that I received was surely higher than what Taco Bell would have paid me in the absence of the law. Is that a good thing? That depends on your perspective and who you think these minimum-wage laws should protect. Should they protect people like me (a teenager living in an upscale, middle-class neighborhood, in a household with two parents who held good, white-collar jobs) or should they protect the least well off members of society? There were, I'm sure, many people who needed that job at Taco Bell more than I, including some low-income single moms. Some of those moms were probably rejected for jobs not only at Taco Bell but also at Woolworths and other companies, all because the law forced employers to offer those workers higher pay than they would have been willing to accept.

The moral here is that the minimum wage does nothing to prevent the benefits of the law from going to people in middle- or high-income households instead of to those in the workforce who are most in need of earning money.

Oh, and here's one last thing to add insult to injury for the hypothetical, low-income, single mom who might have lost out to me for a job at Taco Bell. She might say that even if she can't work at Taco Bell, she might as well eat there. After all, tacos are cheap. But not as cheap as they'd be in the absence of minimum-wage laws! The laws increase employers' compensation costs, and employers pass part of that increase in costs along to consumers in the form of higher product prices. So everyone (including the poorest and most disadvantaged segment of the workforce and of society) ends up paying more for tacos when the workers making and serving them are required by law to get paid more. Washington State already had the highest state minimum wage in the United States when Seattle enacted a series of increases starting in 2015 that would ultimately culminate in a minimum wage of $15. What do you think that means for the prices you can expect to pay when you eat in Seattle restaurants, many of which pay minimum wages? Well, have a look at "The Walrus and the Carpenter" case discussion at the end of this chapter!

4.7.3 Nominal versus Real Minimum Wages

The discussion so far has been in terms of *nominal* minimum wages. But *real* minimum wages account for inflation and are, therefore, more relevant (see Chapter 1's appendix). The federal minimum remains $7.25 as of 2019, which is the same nominal value it has held since its initiation in July 2009. But a decade of inflation has eroded the real value of $7.25, so that the effective federal minimum wage is much lower in 2019 than it was in July 2009. During the time that a nominal minimum wage stays fixed, its real value declines with inflation. After the real value deteriorates enough, the government typically authorizes an increase in the nominal minimum wage. Again, a campaign issue during the 2016 US presidential election was whether the federal minimum wage should be raised to $15 per hour.

4.7.4 Wage Ceilings

A *wage ceiling* stipulates the maximum wage that employers can pay. As is true for wage floors, wage ceilings can arise from either legislation or union contracts. An example of such legislation is the American Recovery and Reinvestment Act of 2009 (ARRA), otherwise known as the “economic stimulus package”, which was signed by President Obama in February of 2009. One of the provisions of that law was a cap on CEO compensation. The motivation for the law was a public outcry over the fact that while some firms were receiving taxpayer-financed bailouts, their CEOs were walking away with huge bonuses. The law imposed a pay cap of $500,000 for CEOs in firms that were in receipt of such bailouts.

The law had a loophole, however. Compensation paid in the form of restricted stock grants was exempt from the cap. So wages and bonuses had to total less than $500,000, but CEOs could receive unlimited amounts of compensation in restricted stock grants. If the purpose of the law is to prevent CEOs from getting rich while the firms they lead into failure are receiving bailouts, then such loopholes are problematic. Caps that contain such loopholes can be expected to change the composition of compensation packages (i.e., the packages will be heavier on restricted stock grants, and lighter on salaries and bonuses, than before) rather than their overall levels.

By analogy, suppose that you want to keep thieves out of your house, which has both a front and back door. You can invest in triple deadbolt locks on the front door, but it won’t make much difference if the back door is a screen door that is perpetually unlocked. The new locks will just encourage thieves to enter through the back door. Similarly, the ARRA encourages firms to pay their CEOs through the back door of restricted stock grants. By the way, restricted stock grants are a form of performance-based pay in which a firm-level measure of performance is used. I’ll cover such components of compensation in Chapters 9 and 10.

The loophole involving restricted stock grants was not an oversight. The language of the law explicitly referenced the exception for restricted stock grants. Why did policymakers invest in a triple deadbolt lock for the front door, while leaving the back door wide open? A potential answer is that policymakers didn’t necessarily want a successful cap; rather, they wanted the appearance of a successful cap, to appease an angry public. The problem with a totally successful compensation cap that prevents CEOs from earning more than $500,000 annually is that such a cap would make CEO jobs far less attractive, so attracting and retaining top talent would be harder. When it comes to talent management, compensation caps are the enemy and can cause you to lose your best workers, or never succeed in hiring them in the first place.

4.7.5 Floors and Ceilings in Non-Monetary Components of Compensation

Floors and ceilings can also be imposed on non-wage components of compensation packages, such as the fringe benefits that I'll discuss in Chapter 11. An example of a ceiling on non-wage compensation would be a cap on the number of paid vacation days that a worker can take annually. An example of a floor is provided by the Healthy Workers, Healthy Families Act of 2014 (HWHFA), which was enacted in California. HWHFA imposes a minimum on the number of annual days of paid sick leave that California employers must pay to most workers. Let's take a closer look ...

Every employed person in the world, when waking up on a workday, must decide whether or not to show up for work that day. The consequences of not showing up vary from job to job. In my job, if there are no important meetings and it's not a teaching day, nothing would happen. For an airline pilot, or an elementary school teacher, or a nurse, it's a different story. Those workers can use sick days, if they have them. As of 2015, California's HWHFA mandates accrual of paid sick days for most workers employed at least 30 days per year, which creates a floor for sick days, much like the minimum wage creates a floor for wages. More precisely, paid sick leave accrues for employees who worked for a California employer on or after January 1, 2015, for at least 30 days within a year, and who satisfy a 90-day employment period, like probation. Accrual of sick days begins on the hiring date or July 1, 2015, whichever is later, and sick leave accrues at a rate of one hour for every 30 hours worked. Employers can cap the use of paid sick leave to 24 hours per year (or three days per year). Unused accrued sick leave can be carried over to the next year, but the employer can limit the amount of accrued leave to six days (or 48 hours).

Although the preceding features of the law are the main ones, HWHFA also has provisions for mandatory notice and record keeping. Knowledge of, and compliance with, such provisions are very important in organizations, as we learned from Disney's wage-theft debacle that opened the chapter. HWHFA prohibits California employers from lumping together paid sick leave, vacation, personal time, etc., into one fungible set, which was a common practice before the law was enacted. The paid sick leave must be recorded separately to ensure that it's in compliance with the law. Employee requests to use sick days can be made orally or in writing, and sufficient advance notice should be given to the employer whenever feasible. If the sickness event is unforeseen, notice can be provided as soon as possible. Reasons that qualify for using sick days include preventative healthcare, or sickness, either of oneself or certain dependents (e.g., children, parents, etc.) There are also some other qualifying reasons, like victimization from domestic abuse.

People miss work for lots of reasons, only one of which is illness. Usually your boss can't really know if you're actually sick, or how sick you are, and even requiring you to furnish a doctor's note is no guarantee that you're sick. That's why many employers, prior to the enactment of the law, didn't even try to figure out if you were sick in the event of an absence. They simply lumped together all paid time off (sick, vacation, personal, etc.) into one category, before the law prohibited such lumping. But worker absences for any reason, particularly unexpected absences, are costly and disruptive to the production process. So why, before the law was enacted, did California employers offer any paid time off at all?

The answer is, because workers like it! But, recalling the third recurring theme from section 1.8, it is not the fundamental kindness of employers that causes them to pay attention to worker preferences. A competitive market *forces* them to care about workers. Recalling the discussion of compensating differentials in Chapter 3, if McDonald's offered no sick days, and Burger King offered workers in comparable jobs one week of sick days per year, McDonald's would have to pay more to compensate for its lack of sick days. So employers aren't giving away "something for nothing". Some give generous paid sick leave and lower wages, whereas others give less (or no) paid sick leave but higher wages. People who really want sick days (and are willing to pay for them) choose the first type of employer, and those who want more cash (even if it means fewer or no paid sick days) choose the second type of employer. If workers and firms "match and marry" like this according to their preferences, do we really need a government regulation?

4.7.6 Example: Floors in Paid Time Off

Let's consider an example that illustrates how the HWHFA affects the behaviors of employers and workers. The relevant employer behavior of interest concerns the design of the compensation system, in particular how the employer makes adjustments to certain components of the compensation system when other components (in particular, the amount of paid sick leave) are restricted by a government regulation like a floor. The relevant worker behavior of interest concerns the sorting effect (see the first recurring theme of section 1.8).

Suppose that there are four firms, each of which offers a job with salary, bonus, paid sick leave, and other benefits. Assume that the paid sick leave and "other benefits" are expressed in monetary values, computed in some way. These valuations represent how much the marginal worker (see sections 3.4 and 3.5) values the benefits, not how much the employer pays to provide them! The marginal worker values the compensation packages at all four firms the same and is indifferent among them, recalling the definition of a marginal worker from Chapter 3. In Table 4.1, firms 2 and 3 (indicated in boldface) are not in compliance with HWHFA.

A key point is that the non-compliant firms won't leave the other components of the compensation package unchanged when, after the enactment of the law, they achieve compliance by increasing paid sick leave to three days. Why not? Well, if they did, then we would have the situation in Table 4.2.

Table 4.1 Wage floors on paid time off (2 non-compliant firms)

	Salary	Bonus	Annual paid sick leave	Other benefits	Valuation of total compensation by the "marginal worker"	Valuation of paid sick leave by "marginal worker" ($240/ day)
Firm 1	$60,000	$10,000	5 days	$12,000	$83,200	$1,200
Firm 2	**$80,000**	**$0**	**0 days**	**$3,200**	**$83,200**	**$0**
Firm 3	**$65,000**	**$5,000**	**2 days**	**$12,720**	**$83,200**	**$480**
Firm 4	$72,000	$2,000	3 days	$8,480	$83,200	$720

Table 4.2 Full compliance (2 firms overpay)

	Salary	Bonus	Annual paid sick leave	Other benefits	Valuation of total compensation by the "marginal worker"	Valuation of paid sick leave by "marginal worker" ($240/ day)
Firm 1	$60,000	$10,000	5 days	$12,000	$83,200	$1,200
Firm 2	**$80,000**	**$0**	**3 days**	**$3,200**	**$83,920**	**$720**
Firm 3	**$65,000**	**$5,000**	**3 days**	**$12,720**	**$83,440**	**$720**
Firm 4	$72,000	$2,000	3 days	$8,480	$83,200	$720

Since workers have *mobility* (recall the three-legged stool of compensation from section 1.6) and are free to switch employers, the marginal worker now wants to move to firm 2. Firm 2, and to a lesser extent firm 3, is paying more than necessary to attract the required workforce. These firms need only match what firms 1 and 4 are paying, i.e., $83,200 in value to the marginal worker. If firms 2 and 3 reduce salaries accordingly, we then get the situation in Table 4.3.

Now there is less variety offered in the market. There is less choice to accommodate diverse worker preferences. For example, no one is able to make $80,000 in cash compensation like they could before the regulation. That's unwelcome news for all of those workers in the labor market who want a lot of cash, even at the expense of getting little or no paid time off. Note also that the average salary across the four firms decreases from $69,250 to $69,010. The average of (*Salary* plus *Bonus*) drops from $73,500 to $73,260. The average of "*Other benefits*" stays unchanged at $9100, and the average total compensation stays unchanged at $83,200.

Table 4.3 Full compliance (no firms overpay)

	Salary	Bonus	Annual paid sick leave	Other benefits	Valuation of total compensation by the “marginal worker”	Valuation of paid sick leave by “marginal worker” ($240/day)
Firm 1	$60,000	$10,000	5 days	$12,000	$83,200	$1,200
Firm 2	**$79,280**	**$0**	**3 days**	**$3,200**	**$83,200**	**$720**
Firm 3	**$64,760**	**$5,000**	**3 days**	**$12,720**	**$83,200**	**$720**
Firm 4	$72,000	$2,000	3 days	$8,480	$83,200	$720

Finally, note that although the example assumes that firms 2 and 3 make their cuts entirely on salary in response to the mandated increase in paid time off, that need not be the case. Substitution could instead occur on bonuses, or on “other benefits”, or on any blend of those three components. But however the cuts are allocated, the total of those three components must drop so that the marginal worker’s valuation of total compensation remains unchanged after the increase in paid sick leave.

4.8 “Hard” versus “Soft” Constraints

Some constraints are tighter than others. A federal minimum wage of $15 per hour, if that were to be enacted, might be thought of as a “hard” constraint, because violating it would be illegal. Other constraints are “soft” in the sense that you would bear a cost by violating them, but you would not be breaking any laws. Political pressure can be a source of soft constraints. For example, if public sentiment concerning an aspect of compensation (typically that it is either too high or too low) is particularly strong, then it might actually help to improve your bottom line by voluntarily “self-imposing” a compensation constraint in your organization.

Your calculation in that case would be that the costs of the self-imposed constraint would be outweighed by the benefits in terms of enhanced goodwill with the public, which would hopefully translate into greater customer loyalty and, therefore, higher revenue. In fact, even if those benefits fail to outweigh the costs of imposing the constraint, you still might choose to voluntary impose it if you are confident that it is ultimately going to be externally imposed. If the constraint is probably going to be imposed anyway, then you might as well garner some goodwill with the public by voluntarily imposing it before being forced. Political pressure from the public debate about a $15 hourly minimum wage in the United States might have been a factor in Amazon’s redesign of its compensation system for low-wage workers in October 2018, increasing their wages (which are highly visible to the public) while reducing their equity-based compensation (which is less visible). See the dual Amazon case discussions on the course website.

Sometimes the political pressure that gives rise to a soft constraint eventually leads to the imposition of a hard constraint. For example, the caps on executive compensation that were imposed in 2009 by the ARRA (see section 4.7.4) were a response to the public outcry over generous executive compensation packages that were being awarded in firms that were receiving public bailouts during the financial crisis.

The "hard versus soft" dichotomy is convenient but a bit artificial. A better way to think about constraints is simply in terms of the costs that you'd incur by violating them, and you can think of a continuum of those costs, from low to high. Even so-called "hard" constraints can be, and sometimes are, violated. For example, employers sometimes engage in wage theft by paying below the legal minimum wage (see Chapter 2), despite the risks of government sanctions. The costs of doing that would depend on the probability of getting caught, the fines and other penalties that the employer would incur if caught, the reputational consequences of creating disgruntled employees, and so on. In most cases, those costs are likely to exceed those from violating a so-called "soft" constraint, but there's no guarantee of that. Violating "perceived" or "soft" constraints by using compensation policies that are politically unpopular can damage your bottom line significantly.

4.9 Lessons for Managers

Labor laws are incursions into free exchange and (when they limit your ability to dismiss workers) into the employment-at-will doctrine that prevails as the default principle governing employment relationships in most US states. You should generally think of labor laws – as well as internal rules such as the collective bargaining agreements covered in Chapter 5 – as *constraints* that limit your capacity to increase your company's profit. To the extent that the rules can be relaxed or circumvented, the company's bottom line improves. What a rule says on paper, and what it means in practice, are often two different things, and the latter is something that you can influence through good managerial decisions. Thus, your success as a manager hinges on how adept you are at navigating the inevitable rules and regulations affecting your organization. Remember the "3 Cs of compensation constraints": *comprehend*, *circumvent*, *comply*!

Firing workers from protected classes is particularly tricky, even under the employment-at-will doctrine, because of the risks of costly lawsuits. Be certain to document clearly that a decision to discharge a worker – or to do other things that might displease a worker – is based on legitimate reasons that have nothing to do with the worker's personal characteristic that is covered by legal protection. When possible, rather than firing undesired workers, it is preferable to create conditions under which they find it attractive to voluntarily quit. Inducing such sorting effects requires care, to avoid ending up on the wrong side of a constructive discharge legal claim. In general, creativity, subtlety, patience, and deep knowledge of the rules go a long way towards allowing you to relax the regulatory and legal constraints that bind your organization. In particular, in most cases you should be able to engineer an unwanted worker's departure, when desirable, without running afoul of the law.

Other regulations, like compensation ceilings and floors, are virtually impossible to avoid directly. There's no creative way around paying the minimum wage, for example. On the other hand, when confronted with floors on certain components of compensation you might decide to adjust other components to relax the constraint, as in the example concerning floors on paid sick leave. Floors and ceilings create different types of problems. Floors increase your compensation costs and hinder your ability to design the compensation system in exactly the way that will most benefit the organization, whereas ceilings impede your ability to attract and retain top talent. Both types of problems hurt the bottom line. And they might make production so unprofitable that you must shut down operations if your profit margin is low enough to start.

Hunt for creative ways around floors and ceilings. You're looking for a hidden trapdoor in the floor, which would allow you to drop compensation lower. Or you're looking for one in the ceiling, which would allow you to raise compensation higher. Sometimes the trapdoor is obvious; for example, in the case of the ARRA, the only trapdoor in the CEO compensation ceiling was restricted stock grants . . . that was the only avenue (apart from fringe benefits) for increasing CEO pay. In other cases, there may be multiple trapdoors, and the challenge is finding the most efficient one. If there's no trapdoor, try to create one. In the case of ceilings, this may involve introducing additional components of pay that aren't subject to the cap. In the case of floors, this may involve subtracting some components of pay that are subject to the floor, or reducing their relative importance within the overall compensation package. Of course, if it's a minimum-wage job and the only component of pay is the wage, there are no components that can be subtracted feasibly (e.g., most minimum-wage jobs don't offer many fringe benefits).

You can never truly *escape* from a floor or ceiling. To understand why, consider ceilings. Using a "trapdoor" means that you're using more of a component of compensation that's not capped while using less of another that's capped. This changes and distorts the *design* of the compensation plan you offer; the plan now has relatively more of the uncapped component and relatively less of the capped component than you would have offered in the absence of the constraint. Maybe that new pay design will improve the bottom line. *But you'd better hope it* *doesn't, because if it does, it probably means that you're a bad manager who should be fired!* If the new design was such a great deal for the organization, then a good manager would have already chosen it voluntarily, rather than doing it involuntarily in response to an externally imposed ceiling. Similar logic applies to floors.

If you contract with other firms that violate labor laws, you may be liable. For example, in September 2018 more than 200 of Amazon's delivery drivers filed a class action lawsuit against Amazon and one of its third-party courier companies (TL Transportation), alleging wage theft (Chapter 2) in the form of unpaid overtime. The impetus for the lawsuit was an August 2018 ruling by a federal judge that TL Transportation failed to properly pay drivers for overtime hours. Regulatory compliance in your own firm is challenging enough, but the problem amplifies when you need to keep tabs on the compliance of the firms with which you contract.

Here's a free tip to close our discussion. Be careful not to ask your non-FLSA exempt employees (i.e., those who are eligible for overtime pay) to perform tasks before or after their shift. The risks of making such an error have increased in the age of email and other online work tools, because employees have become easier to access at any time and location. For the same reasons, the costs of such errors have increased (e.g., because an after-hours email request to your employee creates electronic proof of your non-compliance with FLSA).

Case Discussion 28: The Walrus and the Carpenter[1]

The Walrus and the Carpenter (W&C) is an oyster bar in Seattle, Washington. Before April 2015, the restaurant faced a challenge. The wait staff was paid the Washington State minimum wage (which exceeded the federal minimum wage and, at $9.32, was the highest state minimum wage in the United States at the time). This created both high labor costs for the restaurant and also an excessively large gap in compensation (i.e., wage plus bonus) between wait staff and kitchen staff, because the wait staff gets tips and the kitchen staff does not. The massive pay gap created morale problems among the kitchen workers, and it existed because state law required that wait staff be paid the state minimum wage no matter how high their compensation was after including tips. If it were not for that law, W&C could have lowered the hourly wage of wait staff (potentially all the way down to zero, because tip income is so high that even at an hourly wage of zero the wait staff would make more

[1] Note: This case is based on an article published in *The New York Times* by Patricia Cohen: "As Minimum Wages Rise, Restaurants Say No to Tips, Yes to Higher Prices" (August 23, 2015).

compensation than the state minimum wage, as required by law) and used the savings in compensation costs to pay more to the kitchen staff. But state law prohibited such rebalancing, which created two problems for W&C: (1) excessive labor costs, because wait staff was overpaid in the sense that (in the absence of the law) W&C could have easily hired wait staff willing to work at lower pay levels than what W&C was forced to pay currently, and (2) the artificially high pay for the wait staff created morale problems for the kitchen staff.

In April 2015, W&C's problem got even worse when Seattle's new minimum wage ordinance went into effect, raising the minimum wage even higher than the Washington minimum and setting it on a path to reach $15 within several years. April 2015 marked the first phase of the increase. "Schedule 2" employers (i.e., those with 500 or fewer workers) such as W&C had to pay a minimum wage of $10, rising each year until hitting $15 in 2021. They also had to pay a minimum *compensation* (not to be confused with wage) of $11, rising each year until hitting $15 in 2019. This means that at W&C, in 2015, kitchen staff needed to be paid a wage of at least $11 per hour (because their compensation consists only of wages) whereas wait staff had to be paid a wage of at least $10 per hour (because tips would bring them up to, and indeed far beyond, the required $11 in *compensation*).

W&C responded to the law by immediately raising the price of all menu items by 18% (i.e., adding an automatic 18% gratuity) and eliminating tips (i.e., customers were told that the tips were built into the price). The extra revenue from the price hike was used to finance higher wages for all workers (both kitchen staff and wait staff) and to shrink the gap in pay between kitchen staff and wait staff. But that wasn't enough to cover the added labor costs from the minimum-wage ordinance, so W&C increased the automatic charge on menu items from 18% to 20% and also shrank the owner's share of profit. W&C also made a series of adjustments to rebalance the pay (giving wait staff less and the kitchen staff more) to achieve greater equality and getting the restaurant staff on board through "open-book management", i.e., by showing everyone the payroll spreadsheets so that they could see how the money was being allocated.

Questions

1. Imagine a hypothetical situation in which W&C faces no constraints on pay, i.e., no minimum wages (federal, state, or local).
   (a) How would menu prices be set?
   (b) What would be the level and the design of the pay for wait staff and kitchen staff?
   (c) Explain the reasons for any gap in pay between wait staff and kitchen staff.
2. After the 2015 Seattle minimum-wage ordinance is introduced and the pricing and pay policies at W&C are changed, describe any sorting and incentive effects (see section 1.8) on the part of *customers* and the resulting implications for

W&C's total revenue. Sorting effects refer to which customers visit W&C and how often they visit, and incentive effects refer to what they order from the menu.

3. After the 2015 Seattle minimum-wage ordinance is introduced and the pricing and pay policies at W&C are changed, describe any sorting and incentive effects on the part of *workers* (both wait staff and kitchen staff) and the resulting implications for W&C's total revenue and total costs.
4. Your answers to #2 and #3 hinge on what other Seattle restaurants are doing. Explain how your answers to both preceding questions differ between the following two situations:
   (a) W&C's policy changes are pretty rare, i.e., most Seattle restaurants do not eliminate tips by building them directly into menu prices.
   (b) Most Seattle restaurants adopt pricing and pay policies similar to W&C's, by eliminating tips.
5. What are the advantages and disadvantages of W&C's policy changes in response to the Seattle minimum-wage ordinance?
6. Do you think W&C's policy changes are the best approach for dealing with the Seattle minimum-wage ordinance? What other strategies might W&C have considered, and what are their advantages and disadvantages compared to the path chosen?
   (a) If you were a direct local competitor of W&C, what strategy would you have taken?
7. The first paragraph of the case says that, were it not for the law, W&C could lower the wages of the wait staff, "potentially all the way down to zero". But why stop at zero? In the absence of regulation, could the hourly wage ever be negative? Explain what that would mean and whether it would ever happen and make sense, in this production setting or others.

Further Reading

DeLeire, T. (2000). "The Wage and Employment Effects of the Americans with Disabilities Act." *The Journal of Human Resources*, 35, 693–715.

DeVaro, J. (2015). "In Sickness and in Health: Skipping Work Under the Cover of Law." *Compensation and Benefits Review*, 47, 121–26.

DeVaro, J. and Fung, S. (2014). "Public Bailouts, Executive Compensation and Retention." *Journal of Empirical Finance*, 26, 131–49.

Hanvey, C. (2018). *Wage and Hour Law: Guide to Methods and Analysis*, Springer.

MaCurdy, T. (2015). "How Effective Is the Minimum Wage at Supporting the Poor?" *Journal of Political Economy*, 123, 497–545.

MaCurdy, T. (2015). “The Minimum-Wage Stealth Tax on the Poor.” *The Wall Street Journal* (February 22, 2015).

Posthuma, R. A., Roehling, M. V., and Campion, M. A. (2006). “Applying U.S. Employment Discrimination Laws to International Employers: Advice for Scientists and Practitioners.” *Personnel Psychology*, 59, 705–39.

Shultz, K. S. and Adams, G. A. eds. (2019). *Aging and Work in the 21st Century*, 2nd ed., New York: Routledge.

5 Internal Constraints on Pay

When I joined the faculty of the CSUEB business school in 2008, the organization faced a problem. Faculty research productivity was low. The dean wanted to increase the quality and quantity of the faculty's research output, i.e., publications in professional journals. The reason for that desire is that a strong research profile increases a business school's chances of maintaining its accreditation, which affects the prestige of its programs and its ability to recruit students and place them in lucrative jobs after graduation.

The dean understood the forthcoming issues from Chapter 9 and introduced an aggressive pay-for-performance system that rewarded faculty with monetary bonuses when they achieved new research publications. Those bonuses were significantly more generous for publications in the most prestigious, competitive, top-ranking journals, as determined by a list of journal rankings that was circulated to everyone on the faculty and posted on the college website. The preconditions for a successful pay-for-performance plan, which we will examine at greater length in Chapter 9, were present; for example, output was clearly defined and easy to measure, and quality could be clearly defined and measured using well-known lists of journal rankings. The plan was off to a successful start, and the number of research articles that the business-school faculty published in prestigious journals increased.

5.1 Internal Constraints: an Example

The successful program that I just described was abruptly terminated shortly after it started. A disgruntled faculty member (who had not published and had, therefore, not received a bonus) complained to the California Faculty Association (CFA), i.e., the faculty union. Once the union caught wind of what was happening, the dean was forced to abort the program because of concerns that it violated the collective bargaining agreement (CBA), i.e., the contract agreed to by the union and the CSU system. The reason is that "research and publishing" is considered part of the regular job description of a professor, and additional compensation cannot be awarded for activities that are part of one's regular job description. This is an example of how rigidly faculty compensation is regulated by the CBA. The union's interference in the dean's management of the business school displeased her, but there was little she could do about it. She was frustrated with the union for two reasons.

First, the original organizational problem that the dean was attempting to solve (i.e., faculty were not producing enough published research) was, if not created by the union, certainly exacerbated by it. In a typical organization, the boss has the power to award raises to high performers. So a worker's strong performance throughout the year often leads to a good performance evaluation and a nice raise, which gives workers incentives to perform well. But the CBA at CSUEB prohibits such merit-based raises. Once professors are hired, the dean has no control over their base salaries and is, therefore, powerless to award raises based on research publications and other indicators of good performance. With no prospect of raises for good performance, professors have no monetary incentive to undertake the challenging and time-consuming task of producing high-quality publications. They also have no incentive to engage in "service" activities, which is a problem you are asked to confront in the "CSUEB CBE" case at the end of the chapter.

Second, the dean's proposed solution to the organizational problem (i.e., introducing pay-for-performance based on the number and quality of faculty publications) was blocked by the union.

In short, the union created or exacerbated an organizational problem by prohibiting merit-based raises in base compensation, and then prevented the compensation system from being used by management to solve that problem via pay-for-performance bonuses. It was a double whammy!

5.2 Internal Constraints and "the 3 Cs"

The preceding section offers an example of an *internal constraint* on pay. Internal constraints refer to those constraints that apply only within the organization (or, more accurately, within the "entity", which need not be an entire organization). In contrast, the external constraints (most of which are laws and regulations imposed by various levels of government) covered in Chapter 4 apply more broadly, both within and outside the "entity". The "entity" to which the word "internal" refers depends on the context. It could be a single firm, an individual establishment within a multi-establishment firm, a system of organizations (as in the example of the CSU system in the preceding section), an individual department, a division within a company, a team, etc.

As I mentioned in section 4.3, the "3 Cs of compensation constraints" that we discussed in Chapter 4 concerning external constraints also pertain to internal constraints. After all, constraints are constraints. Rules are rules. And from your standpoint as a manager, if there's a particular rule that you're supposed to follow, it doesn't usually matter much whether the origin of that rule is a law, a provision of a union contract, a mandate from the corporate office, or an instruction from your boss.

There's a caveat, however, to what I just said. In section 4.3, I argued that constraints generally impede your ability to maximize your company's bottom line. But if the internal constraints are being handed down by your boss or by the corporate office, then (presumably!) those constraints are actually helpful to the organization in the sense that they were designed specifically with the aim of maximizing the bottom line (see section 1.4). In that case you'd want to "skip the middle C" of the "3 Cs" and just "comprehend" the rules and "comply" with them, with no attempt to circumvent them.

In other cases the internal rule, even if designed by the company's senior management, might be at odds with the organization's objective, and in that case you'd apply the "3 Cs" as usual. How might this happen? In a multi-establishment firm, for example, the corporate office might design a benefits package that induces the desired sorting effects in most of its locations. However, at the particular site where you're a manager, the local labor market may be very different, and the benefits package doesn't generate the right sorting effects. In short, the corporate office's "one-size-fits-all" approach to compensation works well in most (but not all) of its locations, and the organization's bottom line could be improved to the extent that you can *circumvent* the rules in your location. The ideal circumvention, if possible, is to convince the corporate office that an exception to the rules in your location would benefit the company. Perhaps the original rule was designed because senior management lacked the detailed information that you have about local labor market conditions at your site, and once you convey that information, senior management realizes that its original rule should be adjusted.

It's unsurprising that the example of an internal constraint from the preceding section involved a union. Although unions are by no means the only source of internal constraints, they are an important source of them. So let's say a bit more about unions . . .

5.3 Unions and Union Contracts

Unions, on the behalf of groups of workers, negotiate contracts (i.e., CBAs) with an organization's management. The contracts regulate compensation, working conditions, hiring and layoff policies, and other aspects of the employment relationship. Contracts are typically renegotiated every three or so years. In the United States, unions are far more prevalent in the public sector than in the private sector. Unions have a greater presence in Europe than in the United States.

Exactly what unions care about, and do, varies from union to union and organization to organization. A good way to get a feel for what unions do is to read a CBA. And if you're a manager at a unionized firm, or seek to become one, then you definitely should read that firm's CBA. More than once. You should know it intimately. CBAs can be tedious reading. The one between the CFA and the CSU management for the contract period from November 2014 through June 2017 is 274 pages long and has eight appendices. But thoroughly understanding the internal regulations that apply to your organization is as important as understanding the relevant government regulations that apply and that we reviewed in Chapter 4.

5.4 Unions and Compensation Levels

One of the primary aims of most unions is to negotiate higher compensation levels for their workers. As part of the firm's management, this usually puts you at odds with the union. Your aim, as a good manager, is to pay the lowest possible compensation cost for a specified quantity and quality of labor services, whereas the union's objective is to extract the highest possible compensation cost for that same amount of labor. This fight over compensation levels is not just about whether the hourly wage will be $12 or $15, or whether the raises next year will be 2% or 5%. Compensation has multiple components other than wages and salaries, and in many cases unions are negotiating not just on the level of total compensation but also on how that compensation is allocated into wages, benefits, performance-based pay, etc. So negotiation is over the *design* of the compensation package as well as over its *level*.

Recall the breadth of the definition of *compensation* given in section 1.1: "A person's total compensation is properly understood as including *everything that the person likes about a job*." Unions negotiate over many aspects of the workplace environment of interest to workers. For example, Article 37 of the CBA referenced at the end of the section 5.3 is entitled "Safety" and stipulates a number of provisions for workplace safety beyond those that are mandated by federal and California state law. Those provisions, to the extent that workers value them, should be understood as part of total compensation. Thus, they are associated with compensating differentials, as we discussed in Chapter 3. That is, firms that offer such additional safety provisions can typically afford to pay less in other areas of compensation than competing firms that are otherwise identical but do not offer such provisions.

The battle over compensation levels between a union and management resembles the battle over compensation levels between an individual worker and the employer, which involves a bargaining process of the type that we will discuss in Chapter 14. So if your role as a manager requires you to engage in contract negotiations with a union, you will find the material in that chapter useful. But there are also some important differences between employer–worker negotiations and employer–union negotiations that you should be aware of as a manager. For example, in the case of an individual worker, there is no question about whose interests that worker represents. Each worker is his own best advocate and is out for himself. Each additional dollar of compensation extracted from the employer, whether it be in wages or benefits, represents a worker victory. But in the case of a union, the question of who the union represents and is fighting for can be more complicated. To illustrate that point, let's return to the CFA . . .

5.5 Diverse Preferences of Union Members

The CFA represents everyone who teaches in any of the 23 campuses of the California State University system. That's a lot of teachers, but they can be divided into two distinct groups that have very different interests and that are both represented by the same CFA. One group is the "tenure-line", or permanent, full-time faculty. These are the assistant professors, associate professors, and professors, who either have tenure (i.e., lifetime employment) within the CSU system or are on a six-year probationary "tenure track" with the goal of acquiring tenure. The other group contains temporary, or adjunct, faculty who do not have permanent jobs in the CSU system. They're hired semester by semester, on a short-term contract basis, to teach specific courses. They are paid by the course and receive no benefits (Chapter 11). Many of them have full-time jobs outside of the CSU system and only teach "on the side".

Historically, adjunct faculty are heavily represented in the CFA, and they dominate its actions. Following the global financial crisis of 2008, California entered an economic crisis that led to deep budgetary cuts to public higher education. The CSU's annual budget was slashed, and the CSU Chancellor's Office didn't have enough money to cover faculty compensation. In a typical organization confronted by such realities, large-scale layoffs occur. But because the CSU faculty are unionized, a different approach was taken. The union was asked to choose either "layoffs" or "furloughs", to be decided by a majority vote among its members. Layoffs would mean a large number of CSU teachers would lose their jobs. Furloughs would mean that all CSU employees would incur a 10% nominal cut in their wage and salary compensation for one full year, along with a 10% reduction in work hours.

The tenure-line faculty were strongly in favor of layoffs, and the adjunct faculty were strongly in favor of furloughs. The reasons for the difference in preferences between these two groups were obvious. The union contract is explicit about the order in which faculty layoffs will happen, should they ever be necessary. Article 38 of the aforementioned CBA is entitled "Layoffs" and stipulates these rules in great detail. The bottom line is that if layoffs are needed, the first people on the chopping block are the adjunct faculty . . . no tenure-line faculty can get laid off until all of the adjuncts have already been cut. Specifically, section 38.11 (entitled "Order of Layoff", page 156) of Article 38 stipulates:

> *The order of layoff within a unit of layoff designated by the President for a reduction in force shall be:*
>
> a. *first, less than full-time temporary faculty unit employees who do not hold a three-year (or longer) appointment;*
> b. *next, full-time temporary faculty unit employees who do not hold a three-year (or longer) appointment;*
> c. *next, less than full-time temporary faculty unit employees who hold a three-year (or longer) appointment;*
> d. *next, full-time temporary faculty unit employees who hold a three-year (or longer) appointment;*
> e. *next, faculty in the Faculty Early Retirement Program;*
> f. *next, probationary faculty unit employees;*
> g. *last, tenured faculty unit employees.*

Adjunct faculty are in items (a) through (d), and tenure-line faculty are in items (e) through (g). Thus, in the event of layoffs, the contract would require everyone in (a) through (d) to be terminated before anyone on the tenure-line faculty would suffer a job loss.

Each union member got one vote, but because adjunct faculty members dominated the union, furloughs won. Consequently, my pay, and everyone else's, was cut by 10% for a full year. You can imagine how disastrous that was for talent retention, and indeed some of the business school's top faculty members chose to take early retirement, while others left CSUEB to join other universities or private-sector firms. I discuss this further at the start of Chapter 12.

The example illustrates vividly that union members do not all speak with one voice, and sometimes the union negotiates towards outcomes that are strongly at odds with the preferences of large swaths of its membership. As a manager, you should be sensitive to this. A contract that is generally lauded as "successful" from the standpoint of the union may be quite distasteful to certain groups of workers, and that should worry you from the standpoint of talent management if those workers are the high-value stars you most want to retain.

5.6 Compensation Dispersion

I mentioned in section 1.10 that compensation compression (the opposite of compensation dispersion) is both a cause and a consequence of making employees' compensation levels public, as is done voluntarily as a company policy at the private-sector firm Buffer, Inc. (see the case discussion at the end of Chapter 1), or involuntarily as a matter of legal mandate within the CSU system. In the context of that discussion, I mentioned that although the public information about salaries was likely one reason for the salary compression within the CSU system, another important reason is the union. In addition to striving for higher compensation levels, unions generally try to reduce the dispersion of compensation among workers. That is, unions care about "pay equity" and reducing the compensation gap between the organization's highest-paid and lowest-paid workers.

There are various ways to achieve this. Sometimes the contract explicitly states provisions to reduce compensation dispersion. For example, hypothetically, a contract might specify across-the-board salary increases of 2% per year for all workers in each of the next three years, with an additional 2% awarded in the third year to all workers whose annual salaries in that year fall below $50,000. Alternatively, rather than specifying a percentage increase in salary, the contract might call for every worker to receive a $1000 increase in annual salary. For workers whose annual salaries are $30,000, that amounts to a raise of about 3.33%. But for workers whose annual salaries are $60,000, the raise is only about 1.67%.

Another way unions can reduce compensation dispersion is by negotiating over the *design* of compensation, as opposed to its *level*. Typically, the percentage of total compensation that is paid in the form of fringe benefits (see Chapter 11) is higher in unionized than in non-unionized firms. So unions tend to focus heavily on benefits (e.g., pension and health insurance) in their negotiations. One reason is that, in contrast to wages and salaries, benefits generally apply equally to all workers in the firm. So negotiating a generous health insurance plan helps the low-income workers in the firm just as much as the high-income workers, whereas a 3% salary increase would help the high-income workers more.

Similarly, as illustrated in section 5.1, unions tend to dislike pay-for-performance, because such pay plans usually increase compensation dispersion . . . the workers who are highly productive make lots of money, whereas those who have low productivity make far less. The general principle here is as follows. Most firms have considerable dispersion in talent . . . there's a big difference between the productivities of the top performers and the worst performers. If pay is tightly tied to performance, then a big dispersion in talent naturally implies a big dispersion in compensation. If unions want to reduce dispersion in compensation, the way to do that is to decouple pay from performance, weakening the link between pay and performance . . . and one way to achieve that is to eschew pay-for-performance. So it's no surprise that the dean of the business school at CSUEB encountered resistance from the union when she tried to reward professors financially, based on their research productivity.

In section 5.5, I mentioned that because union members do not all share the same preferences, the union's efforts may be welcomed by some members and frowned upon by others. The points about compensation dispersion offer another such example: the union's efforts to reduce pay dispersion are typically welcomed by the firm's low performers and loathed by the firm's high performers. This can really complicate your task, as a manager, when it comes to retaining your top talent. A good example is the one that opened this chapter . . . the dean wanted to use pay-for-performance as a way to reward top researchers. That would have increased pay dispersion in the business school, but it was blocked by the union, thereby reducing the dean's ability to retain star researchers.

Incidentally, shortly after the pay-for-performance plan at CSUEB was killed by the union, the dean creatively revived it in a partial (and imperfect) fashion by using a different form of payment . . . time rather than money. When a professor achieves a major publication, rather than receiving a cash award, the professor receives a course or two off from teaching. Being able to teach less is valuable to professors (no offense intended, if you're one of my students!), because it allows more of their time to be devoted to research and other desirable activities. So, recalling the broad definition of compensation from section 1.1, giving professors more time off from teaching should be understood as a form of compensation. And the union contract is more permissive of time-based awards than of money-based awards.

The contract also permits monetary funds to be awarded to professors if those funds are used strictly for research purposes (e.g., purchasing data, hiring research assistants, or traveling to conferences to present research) and not taken as personal monetary compensation, i.e., salary or bonuses. This modified pay-for-performance plan that abides by union regulations is further explored in the case discussion entitled "CSUEB CBE" that ends this chapter.

5.7 Compensation Floors and Ceilings

In Chapter 4, we discussed compensation floors and ceilings arising from external constraints like government regulation. Floors and ceilings can also arise from internal constraints, which may (though need not) be part of union contracts.

An example is offered by CSUEB. As discussed earlier in this chapter, professors at all 23 campuses of the CSU system are represented by the same union contract (CBA) that is negotiated every three years between the CFA and the CSU administration (i.e., "management"). Among other provisions, the collective bargaining agreement stipulates three caps on annual salaries. The lowest is for assistant professors; these are the professors of the lowest rank, and they are untenured, meaning that they do not enjoy guaranteed, lifetime employment. The highest cap is for professors (sometimes called "full professors"); these are the professors at the highest rank, and they are always tenured. The middle cap is for associate professors; these are mid-level faculty who are usually tenured but can sometimes be untenured.

The caps apply to all 23 campuses of the CSU system. They can create great difficulties for recruiting and retaining top talent, particularly in the highest-paid academic disciplines. They are even more damaging than the legislated caps on CEO pay that we discussed in Chapter 4, because there are no "back door" loopholes or "trapdoors" that allow the constraints to be partially evaded.

For example, suppose that CSUEB is trying to hire a new assistant professor of accounting, which is the highest-paid discipline in the entire university, as in many other universities. Accounting professors are in high demand, so it is likely that anyone CSUEB wishes to hire probably also has employment offers from other universities. A bidding war ensues, and CSUEB (like other CSU campuses) tends to lose to institutions that face weaker (or no) internal constraints and are able to offer salaries above CSU's cap. In the typical case, CSUEB can only win such a bidding war if the faculty member has such an intense love for living in the San Francisco Bay Area that he or she is willing to sacrifice a large amount of income for that opportunity, just like the workers at the extreme left end of the line I described in section 3.4.2 in the context of compensating differentials. The "CSUEB CBE" case discussion that ends this chapter invites you to grapple further with the managerial challenges implied by the CFA's restrictions on compensation.

5.8 Other Internal Constraints

Much of this chapter focuses on unions, which are an important source of internal constraints. But there are various other sources of internal constraints that do not involve unions. Often these are rules created by the company's corporate office. Salary ranges (section 12.1.1) and pay structures (section 13.2) are examples. If your company's internal salary range for senior software engineers caps out at $122,000, that constraint poses a problem if you're trying to retain one of your star engineers who just got an offer of $135,000 from a competing firm.

Benefits (see Chapter 11) are often designed at the firm level and apply throughout the entire organization, even when individual establishments (or departments, or divisions, etc.) retain discretion over other components of compensation, like salaries. This "one-size-fits-all" approach to benefits can be problematic, because it might work better in some locations of the company than in others. Earlier in the chapter I gave an example in which a particular location of the organization (where you happen to be a manager) faces local labor market conditions that are very different from those faced by other locations in the organization. So the sorting effects induced by the company-wide fringe benefit offerings might be desirable in most of your company's establishments but not in those with unique labor market conditions.

In section 4.3, I argued that constraints are generally undesirable from your standpoint as a manager, because they impede your efforts to enhance your company's bottom line. That's true when it comes to the external constraints of Chapter 4, or the union constraints that are the focus of much of this chapter, but things become more complicated in the case of internal constraints set by the corporate office rather than by a union. Presumably, if senior management knows what it's doing, those constraints should be consistent with the objective of profit maximization. After all, the company didn't have to impose those internal constraints but did so for some reason. Sometimes internal constraints that appear to impede profit maximization might actually support it. Even in the example I just gave involving a "one-size-fits-all" company-wide benefits system, it might be that the corporate office was fully aware that profits at your establishment would be lower as a consequence of this rule (due to less-than-ideal sorting effects) but that other considerations were even more important.

For example, imposing a health plan across the entire company and all of its locations may have allowed the corporate office to negotiate lower insurance rates for its workers, yielding cost savings that outweigh the cost of lower profits at your establishment. Sometimes companies impose internal constraints that seem almost certain to reduce profit, yet from the proper (long-term) perspective on profit, the constraints are fully consistent with profit maximization. To elaborate, suppose that a non-unionized firm that pays low wages and benefits is facing the prospect of its workers organizing to form a union. Senior management realizes that if a union successfully forms, compensation levels are likely to increase, and management will have less discretion over how the pay plan is designed. In such a situation, senior management might decide to voluntarily increase salaries and the generosity of fringe benefits, so as to reduce the workers' eagerness to unionize. If enough workers are reasonably satisfied with their compensation, the required minimum number of votes to formally unionize might not be obtained. In effect, a firm might voluntarily tie its own hands to prevent them from being tied even more tightly and permanently by someone else.

5.9 Lessons for Managers

The internal compensation constraints created by unions or policies from the company's corporate office can be as problematic for managers as external constraints. The "3 Cs" of compensation constraints generally apply for internal constraints just as they do for external constraints, with the following caveat. If the internal constraints are set by the company's senior management, they may well have been chosen to support the goal of profit maximization, and in that case you should "skip the middle C" (i.e., *circumvent*). In other cases, perhaps because senior management lacks important information that's specific to particular locations at which the firm operates, the internal constraints might hinder profit maximization. In that case, you can improve the bottom line to the extent that you can creatively "play the middle C" (a pun specially crafted to delight the pianists among my readers).

But . . . be extremely careful here! As I mentioned in Chapter 4, you should take care not to violate the law or to compromise your ethics. You should also not explicitly break company policy. In the ideal case, you would successfully convince the corporate office (by providing new information that corporate didn't have when imposing the internal constraint) that the constraint is problematic for the company and should be relaxed. It might happen that when you engage this discussion with corporate, you discover that internal constraints that appeared to hinder profit maximization are in fact consistent with it. You should think carefully before engaging such a discussion so that you are virtually certain that your analysis is correct and complete and that the new information you provide will be persuasive.

Remember that unions strive to increase worker pay and reduce compensation dispersion, and both of these things spell bad news for managers. The increased pay is bad (if it leads to compensation levels that are "above market") because it hurts the bottom line, and the reduced dispersion is bad because it thwarts efforts to retain top talent, while overpaying low performers and making them too comfortable and complacent in their jobs. Unions also make managers' lives difficult in other ways, e.g., by making it harder to discipline or fire poor performers and by regulating the workplace environment in a variety of ways that may be constraining and inconsistent with the pursuit of higher profit. For example, notice that the layoff rules described in section 5.5 are unrelated to job performance. If layoffs are necessary, the union contract forces you to follow the stipulated order for layoffs, which might mean that poor performers who have lots of seniority are protected, whereas high performers with low seniority lose their jobs. Unions present you with three managerial challenges:

(1) trying to prevent unions from forming in the first place
(2) negotiating union contract terms that are as favorable as possible to management
(3) managing the workforce in the best way possible once a contract has been signed.

Chances are you'll never have to deal with the first challenge. Most managers either work in firms where there's no union or in firms where a union is already in place . . . unionization is a rare event. But if you find yourself as a manager at a non-unionized firm where workers are murmuring about unionizing, pay attention, and do what you can to discourage that effort. One way is to make compensation concessions . . . the more satisfied workers are with their compensation, and the more fairly they think that they're being treated, the weaker their appetites will be for organizing into a union and paying union dues. Another effective approach is to appeal directly to workers, particularly those who might be adversely affected by a union (typically your stars). They can be educated and prevailed upon to oppose union formation, though often they are small in number.

Chances are good that you'll never have to deal with the second challenge either. Most managers in unionized firms aren't directly involved in contract negotiation. But if you find yourself in that position, the material in Chapter 14 offers you some advice. The third challenge is one you will definitely confront if you're a manager at a unionized firm. Once a contract is signed, what can you do as a manager? Unfortunately, not very much. Compensation should be understood as a powerful set of tools that managers can wield to advance organizational objectives, but union contracts heavily constrain managers' abilities to use those tools. The problem faced by the CSUEB dean, as described at the start of this chapter, is a case in point.

Even if you're a manager in a non-unionized company with no immediate threat of unionization, be aware that the National Labor Relations Act (NLRA) identifies many expressions of worker complaints and dissatisfaction as protected activities and possible first steps to unionizing. The National Labor Relations Board (NLRB) is an independent federal agency that protects the rights (established by the NLRA) of private-sector workers to organize to improve their wages and working conditions, *with or without* a union. So even if you don't directly face any of the three preceding challenges, you must pay attention to the internal constraints implied by the NLRA.

Virtually every written rule and policy, just like any law, is subject to interpretation. It is also your responsibility as a manager to understand the process by which adjudication occurs when issues of interpretation arise. In the example that opened this chapter, a disgruntled faculty member complained to the union. Formally, this process is called "filing a grievance". The procedural details about how this works are typically articulated in the CBA; for example, in the aforementioned CBA, Article 10 describes "Grievance Procedures".

Knowing the grievance procedures is only half the battle. You also need to know the key players involved. At CSUEB, for example, it is the associate provost's job to hear grievances and serve as mediator between the union member and the management. What are the historical precedents in your firm? Does the grievance panel tend to favor the union or the management? What issues, when grieved, are most likely to yield a victory for the union, or the management? Knowing the relevant decision-makers and historical precedents in your firm is important. Taking the associate provost, or the equivalent administrator in a non-academic setting, out for coffee to establish a rapport and collect some information would not be a bad investment! Everything I've mentioned in this paragraph, incidentally, is part of the "comprehend" component of the "3 Cs". Full *comprehension* of constraints means understanding the process by which those constraints are interpreted and enforced and how rigidly they are applied.

Finally, even if you're a manager at a non-unionized firm, reading this chapter's material on unions is worthwhile. One reason is that it sensitizes you to the desirability of discouraging union formation in your firm. A second is that your current or prospective workers might have offers from unionized employers, so you might have to compete against unionized employers even if you are not one yourself. The third and most important reason is that although the discussion focuses on unions, it is fundamentally about constraints on compensation systems that are beyond a manager's control. Such constraints – in this chapter and in Chapter 4 – can originate from unions, government regulation, or even internal rules and customs within firms or divisions of firms. Thus, the broad lessons in Chapters 4 and 5 are all about managing your workforce in environments in which the design and operation of compensation systems is heavily constrained by forces that lie largely beyond your control.

Case Discussion 6: CSUEB CBE

The College of Business and Economics (CBE) at CSU East Bay faces the problem of creating incentives for faculty, particularly tenured faculty, to serve on committees.

Background: There are three ranks of professors (assistant professors, associate professors, and professors). Assistants are the lowest, and professors are the highest, with associates in the middle. Associates and professors have tenure, which means they can never be fired (except for gross misconduct or negligence) and effectively have jobs for life. Assistants are on a six-year probationary contract. In the 6th year, they face an "up-or-out" tenure review, which means they either get promoted to associate professor (with tenure) or they get fired. Tenure decisions are based on research productivity (i.e., the number and quality of publications), teaching performance, and, to a lesser extent, service on various committees. If promoted to associate, they remain in that rank for five years, and then are reviewed for promotion to professor. If they fail to get promoted, they keep their job as an associate and can try again in a future year.

Assistants and associates typically get no annual raises (unless it's a year in which the faculty union has negotiated across-the-board raises), but upon promotion, they get an automatic raise of at least 7.5%, as stipulated by the collective bargaining agreement. Professors never get annual raises except for across-the-board ones negotiated by the union.

All faculty members, regardless of rank, do the same three tasks (teaching, research, and service). The standard teaching load is eight classes per year (four per semester). Research involves publishing articles in academic journals. Service involves serving on various committees at the Department, College, or University levels (e.g., curriculum committee; retention, tenure, and promotion evaluation committee; faculty affairs committee; university committee on research; academic senate; faculty recruitment committee). Figure 5.1 displays the breakdown of tasks:

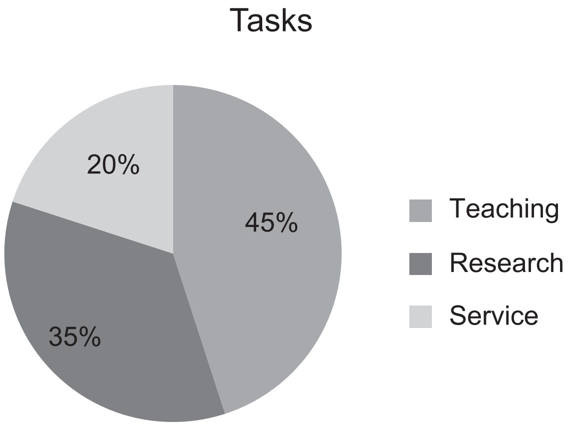


Figure 5.1 Task allocation for faculty in a business school.

Incentives: Assistants have the strongest incentives on all three tasks because they face an up-or-out tenure review in which they can be fired for low performance. Associates have moderate incentives because they face a promotion review, though the stakes are much lower than for assistants, because even if they are denied promotion, they get to keep their job. Professors face very weak incentives because there are no future promotion prospects and no performance-based salary increases (only the across-the-board pay increases negotiated by the union every several years).

Gross negligence in teaching (e.g., not showing up for class, refusing to grade exams, etc.) is one of the few things that can get a tenured faculty member fired. However, the teaching just has to be done; it doesn't have to be done well! **For service, there is virtually zero incentive to do it once you have tenure (plus, faculty members find it distasteful).** For research, three factors might induce tenured faculty to do it. One is intrinsic motivation (i.e., they simply enjoy doing it or feel they should). Another is external marketability (if they ever want to switch to a new university, they must have a strong, current publication record). Another is a publishing incentive program, designed in 2016 by the CBE leadership, which awards professors with reduced course loads (or, alternatively, money in their research accounts that can be used to attend conferences, purchase data, pay research assistants, etc.)

Publishing Incentive System: The CBE leadership constructed three lists of high-quality academic journals (A+, A, and A–) that cover all the business disciplines (accounting, finance, marketing, economics, management, entrepreneurship, etc.). If an article is published in an A+ journal, that faculty member gets three courses off the following year (i.e., they only have to teach five courses rather than eight). Similarly, a publication in an "A" journal earns two courses off, and an A– publication gives one course off.

Business Problem: Faculty at all three ranks (but particularly professors, who face no future promotions or raises) are shirking on service. Even assistant professors shirk because teaching and research are given much higher weight in tenure reviews, and because service is worthless from the standpoint of external marketability (i.e., if an assistant professor is denied tenure and must switch universities, the only things that will help them get a new job are a strong research record and evidence of good teaching performance from student evaluations, etc.).

What Can Be Done about This Incentive Problem?

Remember: The dean (whom you can think of as the "CEO" of the CBE) cannot give the faculty raises for excellent performance on service because this is prohibited by the collective bargaining agreement negotiated by the union. Monetary bonuses are also banned by the union. Moreover, "paying" by awarding further teaching reductions becomes very expensive and is already being used to encourage research. Each time a professor is awarded one course off, an adjunct lecturer must be hired as a replacement, which costs over $5000.

Question

What would happen if you got rid of tenure? How would it affect incentives and the sorting effect? How would it affect compensating differentials?

Further Reading

Bennett, J. T. and Kaufman, B. E. eds. (2017). *What Do Unions Do? A 20-Year Perspective*, New York: Routledge.

Brueckner, J. K. and Neumark, D. (2014). "Beaches, Sunshine, and Public Sector Pay: Theory and Evidence on Amenities and Rent Extraction by Government Workers." *American Economic Journal: Economic Policy*, 6, 198–230.

Farber, H. S. (2005). "Nonunion Wage Rates and the Threat of Unionization." *Industrial and Labor Relations Review*, 58, 335–52.

Hatton, E. (2014). "Temporary Weapons: Employers' Use of Temps Against Organized Labor." *Industrial and Labor Relations Review*, 67, 86–111.

Hirsch, B. T. and Macpherson, D. A. (2017). *Union Membership and Earnings Data Book: Compilations from the Current Population Survey*, Washington, DC: Bureau of National Affairs.

Lewin, D., Keefe, J., and Kochan, T. A. (2012). "The New Great Debate About Unionism and Collective Bargaining in U.S. State and Local Governments." *Industrial and Labor Relations Review*, 65, 749–78.

Pencavel, J. (2009). "How Successful Have Trade Unions Been? A Utility-Based Indicator of Union Well-Being." *Industrial and Labor Relations Review*, 62, 147–56.

Rosenfeld, J. and P. Denice, P. (2018). "What Do Government Unions Do? Public Sector Unions and Nonunion Wages, 1977–2015." *Social Science Research*, 78, 41–56.

Rosenfeld, J., Denice, P., and Laird, J. (2016). "*Union Decline Lowers Wages of Nonunion Workers*." Washington, DC: Economic Policy Institute.

Shin, T. (2014). "Explaining Pay Disparities Between Executives and Nonexecutive Employees: A Relative Bargaining Power Approach." *Social Forces*, 92, 1339–72.

6 Compensation Analytics I

Business decisions nowadays are driven by data analysis to a greater extent than in the past. There are two related reasons for this, both of which result from advances in technology in recent years. First, business data in electronic form are more plentiful than ever before, and they've become easier and faster to collect. Second, advances in hardware and software have facilitated analysis, particularly of large data sets (or "big data"), that was once infeasible. Jobs in "business analytics" are increasingly in demand, and degree programs in that area have become popular. Human resources (HR) is very much a part of these trends, particularly as personnel records have moved from hard copy to electronic records. Business analytics has various subfields corresponding to the core business functions, of which HR is one; "HR analytics", sometimes more affectionately called "people analytics", has become increasingly relevant. The part of HR analytics that concerns compensation data is called "compensation analytics", which basically means using compensation data in the pursuit of better business decisions. As we discussed in section 1.4, "better" business decisions are those that most enhance the organization's objective, which usually means maximizing profit.

Given how essential compensation analytics is for business managers, you might wonder why the standard textbooks in compensation – that are used to train the MBAs and undergraduate business majors who will be future managers – lack a chapter like this one and the next. I can suggest several reasons. Such chapters are hard to write, they are hard to read, and they are hard to teach, at least in the confines of a semester-long compensation course covering many other core topics. One reason why the material is hard to read, write, and teach, is that the students who enroll in compensation courses typically dislike it. Few people go into business, or specialize in compensation, because of a fiery passion for statistics. Many readers will, I fear, skip or gloss over this chapter and the next, which are the most technical ones of the book. But they and their companies risk being outperformed by readers who persevere! I don't have any special secret, or magic formula, for making this material riveting and fun, and that's not my goal. Rather than trying to put lipstick on a pig, I'm hoping to change your view of the pig, so that you can appreciate it as something that is, in the right hands, potentially very useful and not to be feared, while at the same time extremely dangerous and destructive when mishandled.

A challenge in writing Chapters 6 and 7 is that readers vary widely in their prior knowledge of statistics. Maybe you've never taken a statistics course. Or maybe you're a practicing manager who took statistics years ago but forgot a lot. Or maybe you've recently taken a bunch of statistics courses and are a stats whiz. No matter who you are, you'll find these chapters frustrating in some respects. I brush some stuff under the rug, and I don't explain some things that would be taught in depth in more technical courses. The whizzes will find the treatment too shallow, and those who are less statistically inclined will find the discussion too technical and won't understand it all. Nonetheless, I think both groups can extract value from these chapters.

6.1 What Types of Questions Can You Address Using Compensation Analytics?

Compensation analytics begins by defining concrete questions that can be informed by data analysis. Notice that I wrote "informed" rather than "answered". Usually the most important questions in business are complex, and a definitive answer is too ambitious a goal. For example, the question you might *really* want to answer is, "If I eliminate the pay-for-performance plan that has been used for the company's computer programmers during the last several years, and replace it with a more generous health insurance package, *what will happen* to the programmers' turnover rate?" The only way to definitively answer that question is to cut the pay-for-performance plan, beef up the health insurance, *continue doing everything else exactly as is*, and see what happens. Then you'll know the answer, but the information will have come too late to be of much use, because the decision will have been made already and, if it was a bad one, might have caused you to lose your star programmers. An alternative approach would be to use compensation analytics – perhaps using data from other firms in the industry that made similar changes to their pay plans – to help you reach a more informed prediction than you'd have reached without any data analysis, as to what will happen if you change the compensation system as described.

As the preceding example highlights, one set of questions that can be informed by compensation analytics concerns the effect of a change in the compensation system on some outcome that is of interest to your company. That outcome might be employee productivity (i.e., the incentive effect) or it might be turnover and the composition of workers who are attracted to the firm (i.e., the sorting effect), or something else. In all of these cases, compensation is an *independent variable* in the analysis. That is, it's not the outcome of primary interest (like turnover, employee productivity, etc.) but rather a variable that affects the outcome of interest. For other questions, compensation is the *dependent variable*, i.e., the outcome of interest. For example, you might want to know how various worker characteristics or credentials affect their compensation.

Another use for compensation analytics is *benchmarking*, which means comparing the compensation packages at your company with those of industry competitors or peer organizations. You can also do internal benchmarking where you assess how the compensation in one job (or rank) in your company compares to that in other jobs (or ranks) in the company. Yet another use is learning about your workers' preferences and tastes. For example, as we will discuss in Chapter 11, some years ago Walmart was interested in changing its benefit offerings and wanted to learn which benefits were most valued, and which were least valued, by different groups of Walmart employees.

Sometimes such information acquisition can be achieved by asking workers directly in survey questions, and other times it can be achieved by observing workers' actual choices. For example, as explored in the dual Merrill Lynch case discussions, during the recession of 2001, Merrill Lynch offered voluntary buyout packages to its workers . . . interested workers were asked to step forward to express their desire to take the buyout offer, though they were told that their request might not be granted. Tracking who steps forward to request a given buyout package provides useful information about employee preferences. Predicting future compensation costs is another area in which compensation analytics might be helpful. Such predictions need to correctly account for anticipated changes in worker behavior.

Compensation analytics can also facilitate your compliance with government laws and regulations, or union rules, of the type discussed in Chapters 4 and 5. For example, the Equal Pay Act of 1963 and Title VII of the Civil Rights Act of 1964 offer protections against gender discrimination in compensation. Compensation analytics can shed light on how vulnerable your company is to allegations of gender discrimination and how your company compares to your peers in the industry. Concerns about potential gender discrimination in pay might motivate compensation analytics of the type illustrated in Chapter 7's section 7.1.

6.2 Acquiring Data

Compensation data can take many forms. For example, Table 6.1 displays the first several rows of a hypothetical compensation *data set* from a particular company.

As you can see, the data set can be presented in a rectangular array in an EXCEL spreadsheet. The columns represent "*variables*" and the rows represent "*observations*" or "records". I have provided data for only the first seven observations, though I have presented an eighth (blank) row to remind you that the data set

Table 6.1 Compensation data from a company's HR records

Obs.	Last name	First name	Base salary in 2018	Total salary in 2018	Age	Sex	Years at company as of 2020
1	Skywalker	Luke	$55,785	$63,891	24	M	5
2	Organa	Leia	$79,831	$85,371	24	F	8
3	Kenobi	Obi-Wan	$90,781	$108,562	62	M	32
4	Vader	Darth	$81,371	$83,793	51	M	20
5	Fett	Boba	$63,471	$90,473	46	M	3
6	Hutt	Jabba	$193,731	$193,731	58	M	28
7	Lars	Beru	$48,931	$53,381	52	F	21
8	...	...	...	...	...	...	...

might continue for many more observations, sometimes hundreds or thousands, or tens of thousands, or more. But such "big data" can still be represented in a rectangular array ... just a much bigger one than the preceding, both in terms of rows (i.e., observations) and in terms of variables (i.e., columns).

In this simple example, one observation (or row of the data set) pertains to exactly one person, and likewise, each person has only one row. The first column, "*Obs.*", is short for "*Observation number*". For example, observation #5 in this data set describes a male worker named Boba Fett. As of 2020, Mr. Fett was 46 years old, had been with the company for three years, earned a base salary of $63,471 in 2020, and earned a total salary (which includes bonuses) of $90,473 in 2020.

In section 7.1, we will analyze a data set with a structure very similar to the preceding. It has 61 observations (i.e., workers). Although it does not include a variable for "Years at company" it does include some other variables, such as geographic location and rank in the organization. Other data sets might have more complex structures. For example, rather than having just one observation for each worker, there might be many. Boba Fett, in the preceding example, might have three observations (one for each year of his employment with the company). Or he might have many more observations, which could happen if he is paid on a piece rate (see Chapter 9) and we have weekly, or even daily, information on his productivity and pay. In this chapter and in the next, we will focus only on data sets with simple structures such as the preceding array.

Compensation analytics can draw on internal compensation data from your own company, or on external compensation data from other companies, or on a mix of the two. An example of using internal data is Walmart's project that I mentioned in section 6.1; the company administered an internal survey to assess workers' preferences for various benefits packages. Another example is the forthcoming analysis in section 7.1, where the Chancellor's Office of the California State University system uses data from three of its 23 campuses to investigate gender differences in pay in the sciences. An example for which external data would be relevant is a proposed change in the compensation plan in your company, where such a change has been implemented in other companies but not yet in yours. An example of compensation analytics that requires both internal and external data would be benchmarking.

Administrative records provide the source for some compensation data. In other words, someone is collecting and keeping records of compensation information . . . this could be the government collecting the information for tax collection purposes. Or it could be the firm itself . . . the HR Department maintains compensation information for the purpose of documenting regulatory compliance, such as information on California workers' paid sick leave, as we discussed in section 4.7.5. Sometimes such data from administrative records are described as "objective", as opposed to "subjective", because they don't reflect anyone's opinion. If there is a record in an HR file that Sally Smith, the most recently hired senior research associate at the firm, was hired on October 24, 2020, at a salary of $115,192, then unless someone keyed the salary number incorrectly into the data base, that is indeed her actual starting salary.

Surveys provide an alternative source of compensation data. Surveys involve asking people (e.g., workers, managers, HR directors, union representatives) something about compensation and recording their responses. Survey data are generally more prone to measurement errors than are administrative data. For example, if a manager is surveyed and asked about the salary of the most recently hired senior research associate at the firm just mentioned, the manager might not know the exact salary number. Perhaps he never knew it exactly. Or perhaps he once knew it and subsequently forgot. If he doesn't know the exact answer, he might track down someone in the firm who does, but that takes time that he'd probably rather not spend. He might respond to the survey question with his best guess, which could be close to the correct number of $115,192 or quite far away from it. Such "recall errors", associated with the imperfect knowledge or memories of survey respondents, are common problems with survey data.

Additionally, survey data often ask questions that draw on the respondent's opinion. For example, earlier in the chapter I mentioned that Walmart was interested in learning more about its workers' preferences for various types of fringe benefits. Walmart employees were asked to complete a survey and to rank (on a 5-point scale, where 5 is "very satisfied" and 1 is "unsatisfied") their satisfaction with each of a variety of fringe benefits. Questions about satisfaction are inherently subjective, and opinions may vary from one employee to the next. Moreover, respondents might not always truthfully express their opinions. For example, if you ask them how satisfied they are with their health insurance, the truth might be that they are very satisfied, based on what they know their friends are getting at other firms. But they might be afraid that if they check the "very satisfied" box on the survey, their employer might decide that further enhancements of the health plan are unnecessary or that it is too generous. Or their employer might decide to cut back a bit on some other components of the compensation package, knowing that workers who are "very satisfied" with their health insurance are unlikely to quit. In other situations, workers overstate their satisfaction. Suppose that they suspect that their employer is considering eliminating or cutting back on a particular fringe benefit. Even if it's not a benefit that's of enormous value to the worker, the worker might not wish to see it eliminated . . . and if enough workers respond that they love the benefit and care deeply about it, the employer is less likely to eliminate it. The bottom line is that "subjective responses" that reflect respondents' opinions are prone to various types of biases.

You should not infer from this discussion that survey data are useless or that they are always worse than administrative data. They have their problems, but they also have their strengths. For example, they often contain richer and more detailed information than is available in administrative records. Moreover, you should not automatically assume that "subjective" data are "wrong" and that administrative data are always "right". Even administrative data can contain errors (e.g., someone incorrectly typed a salary into a data base, or applied an incorrect formula when computing a change to compensation).

Often firms conduct their own internal surveys of workers, such as the one Walmart did. Additionally, compensation surveys are available on much larger scales that are more sophisticated and involve a lot more firms and a lot more data. For example, the Current Population Survey (CPS) is regularly conducted, is the primary source for labor-force statistics in the United States, and contains a lot of information about compensation. It is jointly sponsored by the US Census Bureau and the US Bureau of Labor Statistics (BLS). The CPS data are free and publicly available. The Survey of Income and Program Participation (SIPP), which surveys US households repeatedly over time and asks a number of questions about income and compensation, is another widely-used data source that is free and publicly available. Another example is Standard and Poor's *ExecuComp* data, which is the most widely-used data base on executive compensation. The National Compensation Survey (NCS), conducted by the BLS, provides measures of occupational earnings, compensation cost trends, benefit incidence, and detailed plan provisions.

6.3 Cleaning Data

Once you have acquired some compensation data, what next? You should think about data much like you think about fruit. Suppose that you buy an apple from the supermarket. You would probably not just bite right into it after the purchase. Who knows where that apple has been, what has been sprayed on it, and how many

Table 6.2 Annual salary data for administrative assistants

	Annual salary ($)
Secretary 1	41,352
Secretary 2	43,731
Secretary 3	45,371
Secretary 4	42,853
Secretary 5	449,988
Secretary 6	41,987
Secretary 7	44,582
Secretary 8	43,381
Secretary 9	42,717

people have touched it with dirty hands before you bought it? Washing the apple before you eat it is a critical step, and if you omit that step you might even get sick. Data analysis is the same. Once you acquire some data, you must thoroughly "clean" the data before analyzing or "consuming" them. Otherwise you might make your organization sick.

Data "cleaning" simply means preparing data for analysis and ensuring their integrity, which involves a number of steps and checks. One step is identifying obvious errors in the data. For example, suppose that you're working with data on the salaries of administrative assistants hired by your firm within the last 24 months. There are nine of them, and their recorded annual salaries appear in the database in Table 6.2.

Notice that Secretary 5 has a recorded annual salary of $449,988, an exorbitant number that might even rival the CEO's salary. An error in the data is strongly suggested. Perhaps if you investigate further, you learn that the correct salary for Secretary 5 is actually $44,998. What evidently happened is that when someone keyed the salary into the database, the "8" was accidentally typed twice, or there was a sticky key on the keyboard. It's just a typo, but it could render any data analysis severely inaccurate. The preceding example is a straightforward case. You can simply fix the typo and proceed. But often matters are less straightforward, and sometimes problems can be identified but not fixed. There were two distinct steps in the process that allowed you to correct the flawed data. Let's examine each of them more closely.

First, you had to discover something suspicious in the data. That first step is the most important, because if you don't notice a problem, you can't fix or address it, and you will unknowingly produce a flawed statistical analysis. Catching such problems isn't always easy. In this example with nine secretaries it was easy ...

the errant salary jumped out at us right away as soon as we glanced at the nine salaries. But what if you were analyzing "big data" involving many thousands of workers in many different years? It's impractical to eyeball each and every salary. The best practical approach is to examine "descriptive statistics" for every variable in your data. Common descriptive statistics are the arithmetic mean (i.e., average), mode, standard deviation, minimum, maximum, coefficient of skewness, and various percentiles such as the 10th, 25th, 50th (i.e., median), 75th, and 90th.

The extreme values (i.e., the minimum and maximum) are particularly useful for the purpose of data cleaning. In the example with nine secretaries, the minimum salary is $41,352, and the maximum is $449,988. The maximum immediately alerts you to a problem, and that would be true even if the database had 9000 secretaries (far too many to visually inspect one by one) rather than just nine. The minimum of $41,352 looks reasonable in this case, but if another data-entry error involved mistakenly cutting off the last digit from one of the salaries, examining the minimum would alert you to that.

Or suppose that the minimum salary in a data set you are working with is -99999. Such a finding isn't unusual if you're working with survey data. Often "missing values" (e.g., which occur when a survey respondent refuses to answer a question or was never asked it) are purposely recorded in survey data as some generic number that is chosen by the data-collection staff to stand out as very obviously wrong, so that a data analyst can easily catch it. The codebook or documentation that accompanies such survey data will usually indicate what entry is used to denote missing values. Data analysts usually drop from the sample those observations that are missing. So if you have a data base with 1000 workers, and 59 of them have missing values for salary, then you would compute the average salary and any other statistics using only the 941 workers for whom you have actual salary data. There are more sophisticated statistical approaches for handling missing data than simply dropping those observations from the sample (e.g., *multiple imputation*), but explicating those would take us too far afield.

Another useful check is to compare the mean to the median. If the mean exceeds the median, the data are typically "right-skewed", meaning that there are some particularly large numbers that inflate the average. If the median exceeds the mean, the data are typically "left-skewed", meaning that there are some particularly small numbers that drag down the average. An alternative, more precise, way to assess a variable's skewness is via the coefficient of skewness (see section 6.5 for an example). The coefficient of skewness is positive for right skewness, negative for left skewness, and zero for a perfectly symmetrical distribution with no skewness. Larger magnitudes of the coefficient indicate greater skew. For highly skewed distributions, an advantage of the median is that it is unaffected by extremely large or small numbers, since it is simply the "middle" value of the ordered salaries (i.e., the fifth highest salary in our sample of nine, or, if the number of observations is even, it is the average of the two middle salaries). In the preceding example, the median is $43,381, and the mean is $88,440. The mean is more than twice the median and seems like a high average salary for secretaries in your firm, which should alert you to a problem. But if the sample had 9000 workers rather than 9, the difference between the median and mean would be far smaller and might not sound any alarms. Examining the maximum gives you a clearer and more dramatic alert, whether the sample size is 9 or 9000.

The first step in cleaning the secretary data was detecting the problem. The second step involved conducting further investigation to find out the truth behind Secretary 5's exorbitant salary. Sometimes you can't find out the truth. Suppose that these are survey data involving hundreds or thousands of firms, and the firms and workers might even be anonymous to protect the confidentiality of the survey respondents. Then if you see a suspicious salary number, there is no way to double check it and no one to "call up" and ask. What should you do in that case? The safest approach is to drop (i.e., not include in the analysis) any data that you believe to be wrong to an extent that would materially affect your analysis. Judgment calls are required here, and reasonable people may disagree about the proper way to handle certain cases. If I were analyzing the secretary data, encountered Secretary 5, and was unable to find out the truth of the matter, I would drop that secretary from the analysis. Few people who have experience in data analysis would question that decision. But some cases are in a greyer area, and for those you should alert the readers of your analysis to any significant judgment calls you have made during data cleaning. Another good approach is to present your analyses multiple ways . . . you can present the results using all of the data, but then alert the reader to your concerns about some of the data points. Then present the analysis that drops the suspicious observations, and leave it up to the reader to decide which of the two sets of results are preferred.

The practice of dropping extremely large (or small) observations is referred to as "trimming" the data. Trimming can be very useful for ensuring the integrity of the data. But throwing away data is never a decision that should be taken lightly, because usually you do not wish to discard data that are extreme for legitimate reasons. Give careful thought to your decisions, and be forthright in alerting consumers of your analysis to the decisions that you've made. In short, when you can't find out the truth underlying some suspicious data, keep a record of all of the judgment calls that you make while cleaning the data, and be transparent about it.

In the secretary example, inspecting the maximum salary allowed you to detect and fix a problem. As I mentioned, if the data base had 9000 secretaries rather than just nine, the maximum would also work for catching the problem and would save you the significant burden of having to eyeball each of 9000 numbers. But in a sample as large as 9000, you're likely to have more than one extremely high value, and the maximum will only alert you to the highest value. If Secretary 5 appeared with the highest salary in your sample of 9000, you could drop that secretary and recompute the maximum (which perhaps is $396,987), and so on, repeatedly computing the maximum and dropping it until you reach a maximum that you believe is trustworthy. But that's tedious. A simpler approach is to look at some extreme percentiles of the data. For example, suppose that the 90th percentile of the salary data is $44,852, the 95th percentile is $45,082, and the 99th percentile is $398,275. Then you know that at least the top 1% of the data are problematic, but not the entire top 5%. You might then look at the 98th percentile and find that it looks reasonable, which tells you that the problem lies somewhere between the 98th and 99th percentiles. This information helps you to decide "where to draw the line" when trimming.

Sometimes in survey data the highest compensation numbers are not reported exactly but are truncated at some cap. For example, every worker whose annual salary exceeds $500,000 might be recorded in the data as having a salary of exactly $500,000. This practice is referred to as "top coding". It is usually done to protect the anonymity of workers (and firms). Because the number of workers earning extremely high salaries is small, if you knew the actual salary you might be able to figure out exactly who the worker is, which would raise a host of concerns about privacy. You should be aware of top coding and how it affects your analysis. For example, computing the average salary for a sample of workers that includes top coding will not be meaningful, because it will fall far short of the true average. A better approach is to focus on the median and other percentiles that are unaffected by top coding.

So far we've discussed two types of "incorrect data" that must be addressed when data cleaning. The first type is "accidentally incorrect" data (e.g., typos). The second type is "purposely incorrect" data. Purposely incorrect data can occur because of malicious intent, but more often they arise transparently and with a legitimate purpose in mind, such as top coding or the replacement of missing values with an obviously incorrect placeholder like -99999. Regardless of whether the errors were introduced accidentally or purposely, data cleaning involves more than catching and addressing incorrect data. Sometimes the data have no reason to be doubted, but they do not exist in a form that is amenable to immediate analysis. For example, suppose that in the HR records that produced the secretary data, you also have data on the highest educational degree completed, as in Table 6.3.

Note that only the salary variable is numerical. The education level is a "string variable", meaning that each value of the variable is a string of characters (including letters of the alphabet, apostrophes, and spaces). This is a problem, because statistical operations can only be performed on numerical data. For example, let's consider one of the simplest and most familiar statistical functions … computing an average. Computing the average of three numbers (say, 7, 15, and 26) is easy. You just add them up and divide by three, which yields 16. That computation is extremely easy using EXCEL, any statistical software package, a hand calculator, or

Table 6.3 Annual salary and educational attainment data for administrative assistants

	Annual salary ($)	Highest degree completed
Secretary 1	41,352	High school
Secretary 2	43,731	High school
Secretary 3	45,371	College
Secretary 4	42,853	Associate's degree
Secretary 5	449,988	College
Secretary 6	41,987	High school
Secretary 7	44,582	College
Secretary 8	43,381	Associate's degree
Secretary 9	42,717	Associate's degree

simply your own mind (at least if you've had your morning coffee!). But how would you compute the average of three character strings like "red", "green", and "blue"? You can't! Statistical analysis, like computing simple averages, involves performing computations on *numbers*, so string variables cannot be used in their raw forms.

The way to deal with string variables, such as the education level in the secretary data set, is to convert them into binary "dummy variables" (or just "dummies" for short) that assume only two values (usually zero and one). For example, you can create three dummy variables, one for each education level. The *High School* dummy variable will equal "1" for Secretaries 1, 2, and 6, and it will equal "0" for the other secretaries. The *Associate's Degree* dummy variable will equal "1" for Secretaries 4, 8, and 9, and it will equal "0" for the others. The *College* dummy variable will equal "1" for Secretaries 3, 5, and 7, and it will equal "0" for the others. Thus, the original string variable that can equal any of three different character strings gets converted into three new, numerical dummy variables, one for each education level.

Data cleaning is required not only for the original data but also for any new variables that you construct from those data. Once you have constructed the three dummy variables just described, you should examine descriptive statistics for each of them to check for any mistakes you may have made inadvertently when generating them. One standard check that you should always do is to verify that the sum of the averages of dummy variables within a given group (education in our example) equals one. The reason is that the average of a dummy variable equals the total number of times it equals "1", divided by the total sample size. In the example, because three of the nine workers have high school as their highest completed degree, the mean of *High School* is 3/9. Similarly, the means of *Associate's Degree* and *College* are each 3/9. So the sum of the means for the three dummy variables is 3/9 + 3/9 + 3/9, which equals one. Each of the three variables should have a minimum value of zero, a maximum value of one, and their means should add up to one. If any of those conditions are not met, it means that you made a mistake in the computer code that produced those variables, and examining the descriptive statistics should allow you to catch your mistake. In Chapter 7, we'll see an example in which the means of a set of dummies for various job titles add up to some number greater than 1, which alerts us to the fact that some of those job titles overlap (meaning that the same person can hold two titles simultaneously).

The preceding discussion about dummy variables reveals that there are two types of numerical (as opposed to "string") variables, namely "continuous" and "discrete" variables. An example of a continuous numerical variable is *Salary*, or anything else measured in dollars or units of another currency. A defining feature of a continuous variable is that differences between the values that the variable assumes are quantitatively meaningful. For example, if Secretary 8 has salary $43,381, and Secretary 9 has salary $42,717, the difference between them, i.e., $43,381 – $42,717 = $664, is a meaningful number that everyone understands. Everyone understands that the number is smaller (about 72% smaller, to be precise) than the difference between Secretaries 2 and 1, namely $43,731 – $41,352 = $2379. In contrast, with dummy variables, which are discrete variables, the quantitative values that the numbers assume are not meaningful . . . the variables only serve as indicators of the groups to which each worker belongs (e.g., if *High School* equals "1" then the worker belongs to the group of workers whose highest degree is a high school diploma). But we could have accomplished the same thing by making the dummy variables assume values of 3.551 and -1.34 (or any other two values) rather than zero and one. Zero and one are just labels and happen to be the standard choices that are used in most cases.

Another point concerning data cleaning is that with numerical variables you must pay attention to the *units* of all variables to ensure that you are making apples-to-apples comparisons instead of apples-to-oranges comparisons. For example, suppose that you have data on the compensation of both secretaries and janitors. The data for secretaries (who are salaried) are recorded in "dollars per year" whereas the data for janitors (who are hourly employees) are reported as hourly wages. Suppose that, instead of cleaning the data first, you jumped right into statistical analysis, constructing a multivariate statistical model of pay on the combined sample of secretaries and janitors. Mixing together two different types of workers whose compensation numbers are reported in different units (i.e., annual versus hourly) leads to meaningless statistical results. Secretary 6 has an annual pay of $41,987, whereas a particular janitor might have an hourly wage of $13.75. From the standpoint of the statistical model based on these flawed data, the janitor earns about 0.033% (i.e., far less than 1%) of what Secretary 6 earns, which is a grossly incorrect inference.

Before comparing the pay for these two types of workers, you have to render them comparable by measuring them both in the same units. For example, you can convert the janitors' pay from an hourly rate to an annual rate. Doing this accurately requires information about the annual work hours of janitors and also how many of those hours were overtime hours subject to time-and-a-half pay (see the discussion of FLSA in Chapter 4). For example, let's take a particular janitor who worked 2250 hours in the year, 250 of which were overtime hours . . . perhaps that's a 45-hour workweek for 50 weeks of the year, where the five hours of each workweek in excess of 40 are subject to overtime regulations. Then that worker's annual pay is

[(2000 hours) Â $13.75] + [(250 hours) Â $13.75 Â 1.5] = $27,500 + $5156.25 = $32,656.25.

That janitor earns about 78% of what Secretary 6 earns in a given year, a figure that differs wildly from the 0.033% figure we found before cleaning the data to place them in comparable units. Similar issues arise when very large salary numbers are scaled for convenience. For example, top executives may have their salaries or bonuses recorded in "thousands of dollars", as opposed to "dollars", simply to avoid reporting lengthy strings of numbers. In that case, if an executive has an annual compensation number of $10,332, that actually means compensation of $10,332 Â 1000, which is over $10 million. If the executives' pay is to be compared to that of other workers (whose pay numbers are not scaled by dividing by 1000) you must first rescale the data so that all compensation numbers involve apples-to-apples comparisons.

In summary, you must ensure the integrity of the data by cleaning them thoroughly before proceeding to analysis, just like you should wash your fruit before eating it. If you analyze flawed data that haven't been properly cleaned, your statistical results may be meaningless. A smart review of descriptive statistics can help you to identify potential problems in the data that require cleaning. But data cleaning and descriptive statistics are like chickens and eggs, in the following sense. Descriptive statistics on compensation data are interesting in their own right, and as a manager you should be keenly interested in descriptive statistics on compensation. However, the descriptive statistics aren't meaningful until the data have been properly cleaned, yet the data usually can't be properly cleaned without consulting the descriptive statistics. An iterative process is recommended, whereby you use descriptive statistics to identify problems, fix those, re-examine the descriptive statistics, and continue the cleaning until the descriptive statistics no longer suggest errors. Remember that a reasonable set of descriptive statistics does not mean that you have solved all of the data problems. There is no guarantee that you will catch every problem. But at a minimum, you should take care not to overlook any problem that could have been detected by examining descriptive statistics.

Another approach to identifying data problems is to directly examine the data themselves rather than their descriptive statistics. If the sample is small (e.g., the nine secretaries from earlier in this section), this is easy to do by simply looking directly the data. But if the sample size is large, it can be tedious inspecting each and every observation. In that case, a graph may be helpful. Even if there are thousands of observations, a graph such as a scatter diagram can quickly reveal any *outliers* in the data. I provide an example of this approach in section 7.1.

6.4 Regression and Data Analysis

After you have cleaned your data, you can proceed to data analysis. The type of analysis that you conduct is driven by the particular business question(s) you want to address. Analysis of compensation data usually involves specifying a *statistical model*, which is an equation that describes how a set of variables are related and that recognizes the randomness that's inherent in all business data. For example, you might want to construct a multivariate statistical model (i.e., a statistical model containing multiple variables) to examine how compensation relates to education levels.

Throughout this section, let's suppose that you have a data sample of secretaries, like in the examples of section 6.3. But suppose that instead of just nine secretaries you have data on a larger number, say 100. More precisely, in a given year (say, 2020), for each of the 100 secretaries, you know their annual salary, gender, age, and highest educational degree completed. The data are arranged in a rectangular array in an EXCEL spreadsheet, much like the one shown at the start of section 6.2. In the present case, there are 100 rows in the array (i.e., one for each secretary) and a separate column for each of the variables just mentioned.

Following the discussion of section 6.3, the education information originally existed in the database as a string variable, but let's suppose that you have converted it to series of numerical dummy variables (one for each education level).

In compensation analytics, the most common approach for specifying a multivariate statistical model is to use a *linear regression model*. Here's an example:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate\text{'}s\ Degree)_i + \beta_4 College_i + \varepsilon_i.$$

Like all regression models, this one is comprised of three different types of "things", namely *variables*, *parameters*, and an *error term*. Let's examine each of them, starting with the variables.

6.4.1 Variables (i.e., the Data)

There are five variables: $Salary_i$, Age_i, $Male_i$, $Associate\text{'}s\ Degree_i$, and $College_i$, and collectively these comprise "the data". Each variable has a subscript "i" attached to it. The "i" is called an "index", and it simply indicates that we are talking about 1 of the 100 secretaries, but we're not particular about which one. To elaborate, if our data sample has 100 secretaries, you can think of giving a “name” (well, actually a number) to each of them, so that we can keep track of who’s who. There’s Secretary 1, Secretary 2, Secretary 3, and so on, up to Secretary 100. The 2020 annual salary is available in the database for each of the 100 secretaries. For example, for Secretary 17, it is $Salary_{17}$, which might be $46,932, and for Secretary 93 it is $Salary_{93}$, which might be $43,833. Similarly, the age of Secretary 24 in 2020 is Age_{24}, which might be 51, and the gender of Secretary 67 is $Male_{67}$, which is a dummy variable equaling either zero or one. If it equals one, then Secretary 67 is male, and if it equals zero, then she is female. Workers who self-identify as neither male nor female would require special coding, but let’s ignore that possibility to simplify the discussion.

The highest educational degree attained by Secretary 88 is described by three variables, namely *(High School)*$_{88}$, *(Associate’s Degree)*$_{88}$, and $College_{88}$. Shortly, I’ll explain why one of these variables, namely *(High School)*$_{88}$, is excluded from the regression equation. All three of these variables are dummy variables that equal either zero or one, and recall that these numerical variables were created to replace the original string variable. Together, they form a “set” that describes the education level of Secretary 88. Suppose that the highest degree of Secretary 88 is an associate’s degree. Then *(Associate’s Degree)*$_{88}$ equals one. But then we immediately know that both *(High School)*$_{88}$ and $College_{88}$ must equal zero, because if the secretary’s highest educational degree is an associate’s degree, then obviously the highest degree cannot be a high-school degree or a college degree.

In the preceding example, knowing the value for one of the dummy variables in the set immediately tells you the values for the other two dummy variables in the set. But that’s only because the variable equaled one. Suppose instead that *(Associate’s Degree)*$_{88}$ equals zero. Then you know for sure that Secretary 88’s highest educational degree is either high school or college, but you don’t know which one it is. You need one more piece of information . . . either $College_{88}$ or *(High School)*$_{88}$. If you know either of those two – in addition to knowing that *(Associate’s Degree)*$_{88}$ equals 0 – then you can infer the third.

This illustrates a more general point . . . whenever you have a “set” of dummy variables that were all derived from the same underlying variable (in this case, highest degree obtained), you don’t need to know the values of all of them in the set. You only need to know the values of *all but one* of them in the set. The one that you don’t know can be inferred from what you know about the others. So in our example with three education dummy variables, one of them must be omitted from the regression equation because it represents redundant information. Once we know the values for any two of the education variables, we can immediately infer the value of the third one. Which variable you omit is arbitrary. I dropped (*High School)*$_{88}$, but I could have instead dropped either of the other two. These statements assume that your regression includes a constant term (denoted β_0 in the preceding equation), but that will almost always be the case.

The same principle applies in the case of gender. For gender there are only two categories (male and female) in light of our simplifying assumption that neglects other possibilities, and a dummy variable can be defined for each of them. $Male_{51}$ and $Female_{51}$ are the two dummy variables in the set that describes Secretary 51's gender. If he is male, then $Male_{51}$ equals 1 and $Female_{51}$ equals 0. If she is female, then $Male_{51} = 0$ and $Female_{51} = 1$. But once we know the value of one of these dummy variables, we can infer the value of the other, so only one variable is sufficient to capture all of the information. In the regression equation, I chose to include $Male_{51}$ and omit $Female_{51}$, but I could just as well have done the reverse. The general rule is that if there are K categories, the regression should include $K - 1$ dummy variables describing those categories. So with $K = 3$ education categories, the regression includes two education dummies. And with $K = 2$ genders, the regression includes one gender dummy.

The five variables in the regression model can be classified into *dependent variables*, *independent variables*, and *control variables*. The dependent variable is easy to identify. It's the one that appears on the left-hand side of the equation, so in the preceding example it's $Salary_i$. The dependent variable is sometimes called the *outcome variable*. It's the variable that you're trying to better understand, or explain, i.e., the variable of primary interest. The other variables, that you expect are related to salaries, appear on the right-hand side of the equation as either *independent variables* or *control variables*.

In the preceding regression, which of the four variables on the right-hand side of the equation are independent variables, and which are control variables? The answer depends on what business question you want to address. Suppose that your main objective is to understand how salary (the variable of primary interest) differs between men and women, as in section 7.1. Then $Male_i$ is the independent variable, whereas Age_i, $(Associate's\ Degree)_i$, and $College_i$ are control variables.

The purpose of control variables in a regression is easiest to understand with an example. Remember that your primary objective is to understand how salaries differ between men and women. A simple approach is just to compute the average salary for the men in the sample and compare it to the average salary of women in the sample. Suppose that you do that and find that the average salary of men exceeds that of women by 50%. Before you take that result to your boss in a panic, arguing that your firm is guilty of gender discrimination in pay, you should pause and consider the potential role of other influences.

What if the men in your sample are mostly older than 45, whereas the women are mostly younger than 35? People who are older, more experienced, and who are more advanced in their careers, naturally are paid more than younger people at the start of their careers, and that's true for both women and men. So it might be that the 50% pay gap that you found between men and women largely reflects the age difference between the two genders. What you really want to do is not to compare the average salaries of men and women but rather the average salaries of men and women *of the same age*. Then you are comparing apples-to-apples rather than apples-to-oranges, and you can achieve this by including Age_i as a control variable in the regression. Similarly, adding the education variables as control variables allows you to measure the difference in average salary between men and women *of the same age and education levels*. So including control variables in a regression allows you to better approximate apples-to-apples comparisons of how the dependent variable relates to the independent variable(s) of primary interest.

Shortly, I'll explain exactly how to use the regression to get an apples-to-apples answer to the question "How does the average annual salary differ between men and women?" But first I want to repeat, for emphasis, that the distinction between independent variables and control variables depends entirely on the business question that you're trying to answer. To illustrate that, let's take the preceding regression but change the business question.

Suppose that you want to understand how annual salaries in your firm differ between older and younger workers. Then $Salary_i$ remains the dependent variable, but Age_i now becomes the independent variable. Simply comparing the average annual salaries for older versus younger workers would be misleading because (as we have already discussed) the men in the sample tend to be older than the women. So you would be comparing apples-to-oranges unless you include $Male_i$ in the regression, now as a control variable. Similarly, you might be interested in understanding how workers' educational credentials relate to their compensation in your organization, and in that case $(Associate's\ Degree)_i$ and $College_i$ would be the independent variables, and $Male_i$ and Age_i would be the control variables.

Again assuming that we have a sample of 100 workers (i.e., $n = 100$), the data can be organized in a rectangular array in an EXCEL spreadsheet. There will be 100 rows (one for each worker) and five columns (one for the dependent variable, and the remaining four for the independent and control variables). Assuming that there is no missing information for any of the workers on any of the variables, the data set contains 100 Â 5 = 500 numbers, each of which we can look at in the spreadsheet.

6.4.2 Parameters (or Regression Coefficients)

The *parameters* of the statistical model are β_0, β_1, β_2, β_3, and β_4. These are also called *regression coefficients*. They are important, because they hold the answers to the business questions that you're pursuing. The first thing to notice about them is that there are only five of them, and, unlike the variables (i.e., data) that we just discussed, none of them has an "i" subscript. So when we compare different workers (say, worker 72 and worker 89) their values for each of the five *variables* may differ (e.g., worker 72 might be a 52-year-old college-educated male with an annual salary of \$89,832, whereas worker 89 might be a 41-year-old college-educated female with an annual salary of \$78,551) but the five *parameters* (β_0, β_1, β_2, β_3, and β_4) do not differ from one worker to the next.

The next thing to notice about the parameters is that we don't know their values. They are five unknown numbers, unlike the 500 "data" numbers that we can stare at in the data spreadsheet that we just described, based on a sample size of 100. The best we can do is to use statistical analysis to make educated guesses about the five unknown parameters, based on our data. Let's call those educated guesses b_0, b_1, b_2, b_2, b_3, and b_4, ignoring for the moment how we actually compute those educated guesses (or "estimates") of their unknown counterparts (β_0, β_1, β_2, β_3, and β_4). Our best estimate of β_0 is b_0, our best estimate of β_1 is b_1, and so on. Once we've produced the five estimates (b_0, b_1, b_2, b_3, and b_4) they're just numbers that we can look at in a spreadsheet just like the 500 numbers in our data array.

How do we interpret the estimates b_0, b_1, b_2, b_3, and b_4? To answer that, it's helpful to start with a simpler version of the preceding regression. Suppose that we have just two variables … one dependent variable ($Salary_i$) and one independent variable (Age_i). The regression equation is

$$Salary_i = \beta_0 + \beta_1 Age_i + \varepsilon_i.$$

Since the independent variable is Age_i, what we're interested in here is how workers' annual salary in the company varies with age. There are no control variables in this regression. Suppose that our best guesses, or estimates, of β_0 and β_1, based on our data on $Salary_i$ and Age_i, are $b_0 = 1017$ and $b_1 = 2308$. In regression analysis, usually you're primarily interested in β_1, i.e., on the coefficient on the independent variable(s) and, therefore, on your best estimate (namely b_1) of that unknown parameter.

What $b_1 = 2308$ tells us is that "each additional year of age is associated with \$2308 more in annual salary". Based on these results, if you compare two workers in the organization (one who is 45 years old, and the other who is 50), their annual salaries are likely to differ by (5 years) Â (\$2308 per year) = \$11,540. That does *not* mean that every pair of workers who are five years apart in age will have an annual salary difference of exactly \$11,540! Some will have more, some will have less, and it's even possible that in some rare cases the younger worker might be paid more than the older one. It simply means that, based on the sample of 100 workers, \$11,540 is roughly the pay gap that tends to exist between workers spaced five years apart in age.

What about b_0? What is the meaning of $b_0 = 1017$? In this example, b_0 isn't a particularly meaningful parameter, but it plays an important role in allowing us to predict the dependent variable, $Salary_i$. The two numbers, b_0 and b_1, are all that you need to predict any worker's annual salary, even for workers who are not in the sample. The predicted salary for worker *i* is

$$b_0 + b_1 Age_i, \text{ or}$$
$$1017 + 2308 Age_i.$$

You just plug in a worker's age, and out pops the worker's predicted salary. For example, if worker i is 36 years old, then $Age_i = 36$, and the predicted annual salary for that worker is

$$1017 + (2308\ \hat{A}\ 36) = \$84{,}105.$$

Now, that doesn't necessarily mean that the 36-year-old worker's salary is *exactly* \$84,105 . . . the \$84,105 is just our "best guess" based on the regression equation. Recognizing that $b_0 + b_1 Age_i$ is the predicted salary is helpful for understanding why b_1 tells us the "change in annual salary associated with an additional year of age". Pick any age you like, say, 42, and plug it into the prediction equation to get the predicted annual salary, i.e.,

$$b_0 + b_1 Age_i, \text{ or}$$
$$1017 + (2308\ \hat{A}\ 42) = \$97{,}953.$$

Now add a year of age, and compute the predicted salary again, i.e.,

$$b_0 + b_1 Age_i, \text{ or}$$
$$1017 + (2308\ \hat{A}\ 43) = \$100{,}261.$$

Now take the difference between these two predictions, i.e.,

$$\$100{,}261 - \$97{,}953 = \$2308,$$

which is just b_1! And you'd get the same answer if you had started with any age other than 42.

Salary predictions can also be made "out of sample". For example, suppose that the 100 workers in the sample range in age from 25 to 71, but none of them is exactly 57. That still doesn't stop you from predicting the annual salary of a 57-year-old, which would be

$$1017 + (2308\ \hat{A}\ 57) = \$132{,}573.$$

But don't get carried away. In principle, you could also predict the annual salary of a 116-year-old, which would be

$$1017 + (2308\ \hat{A}\ 116) = \$268{,}745.$$

Or you could predict the salary of a 6-year-old, which would be

$$1017 + (2308\ \hat{A}\ 6) = \$14{,}865.$$

Neither of those predictions would be very meaningful. Setting aside the fact that hiring a 6-year-old at the organization would be in violation of the child labor laws in the FLSA (see Chapter 4), both the 6-year-old and the 116-year-old are far outside the age range of the observed data, which is 25 to 71. To take the point even further, what would be the predicted annual salary of a newborn infant with Age_i = 0? It would be

$$1017 + (2308 \text{ Â } 0) = \$1017,$$

which is just b_0. That is, b_0 can be interpreted, literally, as the average annual salary for a worker of age zero, and now you understand why I said that “In this example, b_0 by itself isn’t a particularly meaningful parameter . . .”

The lesson here is that it’s dangerous to extrapolate regression results far outside the range of the observed data. You’re on safe ground predicting a 57-year-old’s salary, because 57 lies well within the range from 25 to 71. But the closer you get to the extremes of the data range, and even more so once you predict outside the data range, you’re skating on increasingly thin ice. Predictions are most reliable where the bulk of the data are concentrated.

Before discussing the full regression that includes all five variables, let’s take another simple example with only two variables, but this time let’s make the independent variable $Male_i$ rather than Age_i, so that the regression equation is

$$Salary_i = \beta_0 + \beta_1 Male_i + \varepsilon_i.$$

The difference between this example and the last one is that $Male_i$ is a dummy variable (equaling either zero or one) whereas Age_i is a continuous variable. When the sole independent variable in a regression (other than the constant term) is a dummy variable the estimates b_0 and b_1 have special interpretations and are particularly easy to compute, even without any knowledge of regressions. The estimate b_0 equals the average annual salary for females (i.e., the average value of the dependent variable for the worker group for which the independent variable, $Male_i$, equals zero). The estimate b_1 equals a difference in average salaries, in particular

$$(\textit{average annual salary for males}) - (\textit{average annual salaries for females}),$$

i.e., it is the difference in the average value of the dependent variable between the worker group for which the independent variable equals one and the worker group for which the independent variable equals zero. If the regression estimates, b_0 and b_1, can be computed as I’ve just described, as a simple average or difference between two averages, then you might wonder why regression techniques are needed at all. The answer is that computing b_0 and b_1 becomes more complicated when the regression equation contains more than two variables.

If there are multiple variables on the equation’s right-hand side, the preceding interpretations of the regression estimates b_0 and b_1 are similar, but they move closer to the ideal “apples-to-apples” comparisons that are desired. For example, suppose that the independent variable of interest is Age_i and that the full regression is estimated:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \varepsilon_i.$$

Now the estimated b_1 can be interpreted as the "change in annual salary associated with an additional year of age, *holding constant gender and educational attainment*". Sometimes "holding constant" is read as "controlling for"; again, the idea is that by comparing workers of the same gender and educational credentials, we have something closer to an apples-to-apples comparison that will give us a clear picture of how workers' ages relate to their compensation.

Similarly, in the preceding full regression model, suppose that the independent variable of interest is $Male_i$, so that we want to understand gender differences in compensation. Then the estimated b_1 represents the "difference in average annual salary between men and women, *holding constant [or controlling for] age and educational attainment*".

Suppose next that the independent variables of interest are the educational attainment ones, so that you're investigating how compensation varies with different levels of education. To get something closer to an apples-to-apples comparison of workers who differ in their education levels, you want to compare workers of the same gender and age. Now there are two estimated coefficients of interest (i.e., b_3 and b_4), namely the coefficients on the two independent variables. The estimated b_3 represents the difference in average annual salary between workers with an associate's degree and those with a high school degree (the omitted category in the set of education dummy variables) *holding gender and age constant*. Similarly, the estimated b_4 represents the difference in average annual salary between workers with a college degree and those with a high school degree (again, the omitted category in the set of education dummy variables) *holding gender and age constant*. Notice that the effect of a dummy variable is always interpreted relative to that of the *omitted* category!

In all of the examples that we have discussed so far, the relationships between salary and the variables on the regression's right-hand side are all *linear*, which means that estimated relationships do not depend on the values of certain variables. For example, when Age_i was the independent variable, the estimated relationship between salary and age was b_1, which is "just a number". It tells us how one additional year of age relates to annual salary, and it's the same amount of additional salary whether the worker is 30 or 55. This chapter's appendix deals with *nonlinear* estimated relationships that may differ according to the values of certain variables.

6.4.3 Error Term

We have discussed variables (i.e., the data) and parameters (both the unknown β's and their estimated counterparts, the *b*'s). The third ingredient in a regression equation is the "error term", which is denoted by ε_i. It has a subscript "i", just like all of the variables, meaning that each worker in the sample has his or her own error term. Unlike the variables, however, which can be seen as columns in your spreadsheet, the values of ε_i do not appear in a column of the spreadsheet . . . you cannot observe them. They represent all of the many factors that affect salaries that you do not observe in your data. So the regression equation's right-hand side decomposes salaries into a part that's explained by observed data (like age, gender, and educational attainment) and a part that is based on unobserved, "random" factors.

Let's rewrite the regression equation putting the error term on the left-hand side:

$$\varepsilon_i = Salary_i - [\beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate\text{'}s\ Degree)_i + \beta_4 College_i].$$

Writing it this way highlights that the error term depends both on observed data, i.e., $Salary_i$, Age_i, $Male_i$, $(Associate\text{'}s\ Degree)_i$, and $College_i$, and also the five unknown parameters. Notice that the term in square brackets looks a lot like the "predicted salary" that we discussed earlier. The only difference is that the estimated b_0, b_1, b_2, b_3, and b_4 are replaced by their (unknown) counterparts, β_0, β_1, β_2, β_3, and β_4. Now you can see why ε_i is called an "error term". It is just the difference between worker i's *actual* salary, i.e., $Salary_i$, and worker i's *predicted* salary, which is the expression in square brackets. Obviously, smaller prediction errors are better than larger ones, and that insight finally allows us to talk about how we can estimate the unknown parameters β_0, β_1, β_2, β_3, and β_4 to produce b_0, b_1, b_2, b_3, and b_4. There are various ways to do it, but the basic idea is the same in all cases, i.e., choose values for β_0, β_1, β_2, β_3, and β_4 that make the prediction errors "small" in some sense. But remember that there is a separate prediction error, ε_i, for each worker, and we'd ideally like all of these to be small.

The most common approach is to square the prediction error for each worker, then add up all of those squared prediction errors, and then choose the values of β_0, β_1, β_2, β_3, and β_4 that make that "sum of squared prediction errors" as small as possible. There are many software packages that you can use, including EXCEL (if the Data Analysis ToolPak is active), that can do this in the blink of an eye. The resulting estimates (b_0, b_1, b_2, b_3, and b_4) are called "ordinary least squares" (OLS) estimates, or simply "least squares" estimates, or often just "regression estimates". Once you have those, you can substitute them into the expression in square brackets to get

$$e_i = Salary_i - [b_0 + b_1 Age_i + b_2 Male_i + b_3 (Associate\text{'}s\ Degree)_i + b_4 College_i].$$

This is an expression for the *observed* prediction error, e_i, sometimes called the "regression residual". This e_i is an estimate of the (unknown) prediction error, ε_i, and it can be computed because everything on its right-hand side is known (i.e., salary, age, gender, education, and all five regression estimates). The e_i, or "residual", is the *observed* prediction error, i.e., the difference between a worker's actual salary and their salary predicted by the regression equation.

6.5 Levels or Logs?

I mentioned that compensation variables, like wages or salaries, can appear as the dependent variable or as independent variables, depending on the business question that you want to answer. If they are used as dependent variables, then frequently they are not used "as is" but are first transformed using the *natural logarithm*. So instead of using $Salary_i$ as the dependent variable, we use $ln(Salary_i)$, where "*ln*" denotes the natural logarithm. One reason is that we often are more interested in *percentage changes* in compensation levels than in *actual changes* in compensation levels, and logarithms are helpful for that purpose. I'll elaborate shortly.

EXCEL and any other software program that you'd use for statistical analysis, and indeed any scientific calculator, will allow you to compute natural logarithms. For example, on the iPhone scientific calculator you can type "100,000", and then press the "*ln*" button. What comes back is 11.513. It's actually a much longer number like 11.5129254 . . . and so on, but I rounded it to 11.513. So if $Salary_i$ = \$100,000 then $\ln(Salary_i)$ is approximately 11.513.

The symbol "*ln*" always means *natural* (i.e., "base *e*") logarithm, which distinguishes it from other logarithms, such as a "base-10 logarithm", which also appears on the iPhone scientific calculator with a "$\log_{10}$" button. Sometimes you will just see "log". Usually that means the natural logarithm, just like *ln*, but occasionally it means a different logarithm (usually base 10). If you're not sure which it is, a simple check is to compute the logarithm of 2.718. If what comes back is a number very close to 1, then you know you have the desired natural logarithm. For example, on the iPhone calculator, typing "2.718" followed by "*ln*" gives us a long number that rounds to 0.9999. In EXCEL, "*ln*" is the natural logarithm, and "log" is the base-10 logarithm.

The natural logarithm of 100,000 is about 11.513, so the logarithm induces a huge reduction in the scale of the number. But the natural logarithm of 20,000 is about 9.903. So whereas the original salary numbers (\$100,000 and \$20,000) are very far apart, differing by a factor of five, their natural logarithms are pretty close together. Transforming the compensation data by taking the natural logarithm has the effect of shrinking extremely large values and bringing them down much closer to the rest of the data, while still preserving the same ordering of the numbers. Salary and other compensation data tend to be highly right-skewed, meaning that the distribution has a "long right tail" because a small number of people receive exceptionally high compensation. In contrast, a *normal distribution*, which you have seen in your statistics class, has a perfectly symmetrical "bell-shaped" graph. When you take the natural logarithm of salary data, you get something that's less right-skewed and that more closely resembles a normal distribution.

Let's take a look at some data on real hourly starting wages (in 1990 dollars) for 3092 workers surveyed between 1992 and 1995 in four metropolitan areas of the

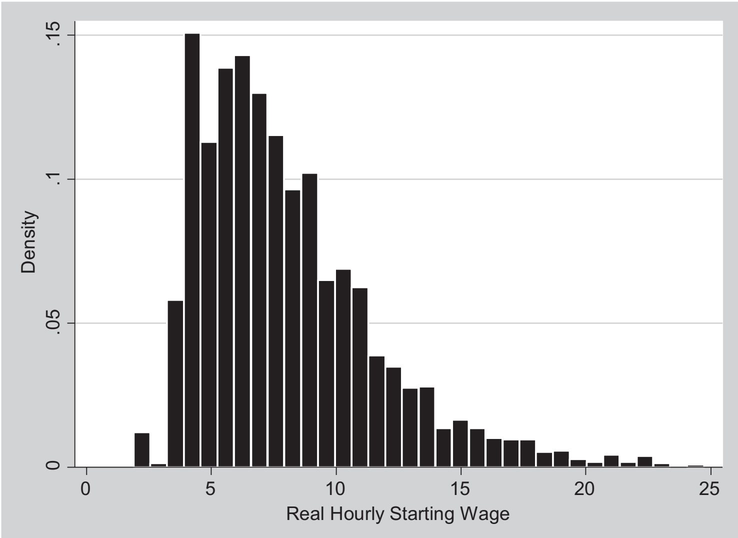


Figure 6.1 Histogram for real hourly starting wage (1990$) for n = 3071 workers.

United States (Atlanta, Boston, Detroit, and Los Angeles). Each worker was the most recently hired worker at his or her establishment at the time the survey was conducted. The "coefficient of skewness" for the real hourly starting wage is 5.87, which indicates that the wage data are highly right-skewed, as is common for wage data. A distribution that is perfectly symmetric (i.e., no skew at all), such as the normal distribution, would have a skewness coefficient of zero, and a distribution that is skewed to the left (i.e., has a "long left tail") would have a negative skewness coefficient. The natural logarithm of the wage variable, however, is much closer to symmetric, with a skewness coefficient of only 0.42. Figure 6.1 displays the histogram for the original wage variable, in levels.

Even though the right skew is quite evident, the histogram actually understates it considerably. The reason is that I dropped 21 workers with extremely high wages, because including them would have produced such an extreme right skew that the scaling of the graph would have made it hard to read. So I dropped the 21 workers whose real hourly starting wages exceeded $25. By the way, if the hourly wages seem low, remember that they're measured in 1990 dollars (see Chapter 1's appendix). Now let's take a look at Figure 6.2, which is the histogram for the natural logarithm of the same wage variable; for consistency I again drop the 21 highest-paid workers. Notice that it looks much closer to the perfectly symmetrical bell-shaped curve that describes the normal distribution.

Our discussion so far has focused on the compensation variable alone (either in levels or in logs). Now let's consider its relationship with other variables, in a regression context. In particular, let's revisit the regressions from section 6.4 and discuss how their interpretations change if they are estimated in logs rather than in levels.

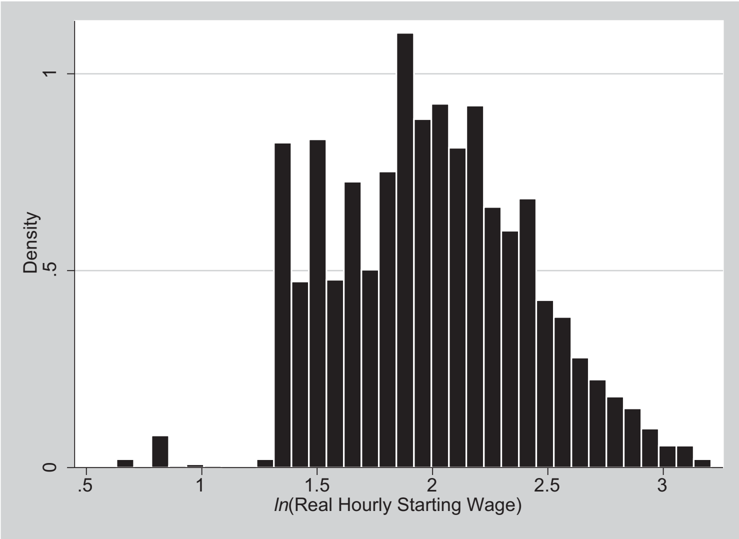


Figure 6.2 Histogram for *ln*(real hourly starting wage) for n = 3071 workers.

If $Salary_i$ appears as the dependent variable, we say that the regression is estimated "in levels", whereas if $ln(Salary_i)$ is used, we say that the regression is estimated "in logs". One advantage of estimating a regression in logs is that often what you care about isn't an actual change in compensation *levels* but rather the *percentage change* in compensation levels, and logs let us do that easily. When you hear workers talk about their raises, for example, they'll usually state a percentage change rather than a change in levels. It's more common to hear someone say, "I got a 7% raise!" than "I got a raise of $6285.51!"

Let's return to the following regression ("in levels") from section 6.4:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \varepsilon_i.$$

Recall how to interpret the parameters, or regression coefficients. For example, β_1 measures the change in salary that is associated with one additional year of age, holding gender and educational attainment constant. So if the regression estimate of β_1, which we call b_1, is 2500, that says that each extra year of age is expected to raise the worker's salary by $2500, holding gender and educational attainment constant, so that if we compare two workers (ages 42 and 47) we expect the 47-year-old to be making $12,500 more per year (i.e., 5 Â $2500) than her colleague who is five years younger. Similarly, β_2 measures the change in salary that is associated with being a male worker rather than a female one, holding age and educational attainment constant. If the regression estimate of β_2, called b_2, is $4300, this means that male workers, on average, enjoy annual salaries that are $4300 higher than those of female workers, holding age and educational attainment constant.

Now let's consider the same regression "in logs":

$$ln(Salary_i) = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \varepsilon_i.$$

Now β_1 measures the *percentage change* in salary that is associated with one additional year of age, holding gender and educational attainment constant. So if the regression estimate of β_1, which we call b_1, is 0.024, that says that each extra year of age is expected to raise the worker's salary by 2.4%, so that if we compare two workers (ages 42 and 47) we expect the 47-year-old to be making 12% more per year (i.e., 5 Â 2.4%) than her colleague who is five years younger. Similarly, β_2 measures the *percentage change* in salary that is associated with being a male worker rather than a female one, holding age and educational attainment constant. If the regression estimate of β_2, called b_2, is 0.131, this means that male workers, on average, enjoy annual salaries that are 13.1% higher than those of female workers.

One final word on the matter of logs versus levels . . . recall that you cannot take the logarithm of zero, or of a negative number. If you don't believe me and are feeling defiant, give it a try on your iPhone or scientific calculator, and see what happens! The requirement that the compensation numbers be positive to conduct analysis in logs isn't usually a problem, because unlike slaves and some prisoners (see section 1.6) all workers you'll be dealing with in a business setting have positive compensation levels. For example, all 3077 workers in the data set we used for the preceding histograms had positive real starting wages. Unemployed people have wages of zero, but you won't typically be studying them in a business setting.

There are some situations, though, in which zeros show up in compensation data, and you should remember that you cannot directly work with those data in logs. For example, your workers might be paid on commission, and on some days or weeks they might have zero sales, and therefore zero commission. In such cases, you cannot estimate a regression in logs using commission as the dependent variable. But sometimes it's possible to transform the data to eliminate the zeros, perhaps by aggregating the data. For example, even though some workers have days or weeks with no commission, perhaps all workers have positive sales and commissions within a month (otherwise they'd probably have been fired before having had a chance to appear in your data!). Then if you aggregate the daily data to the monthly level, an analysis on the monthly commission data could be conducted in logs.

6.6 Precision

Everything we've discussed so far concerns estimating the unknown regression parameters and interpreting them. An estimate is just an educated guess . . . the best guess that we can make using the data at hand. But how seriously should we take our estimates b_0, b_1, b_2, b_3, and b_4? How confident are we in them? In addition to computing those estimates, we want to form some assessment of how reliable they are. A complete discussion of precision in the regression context would take us too far afield in a book on compensation. So I will have to refer you to statistics books for that, though noting that the irony of offering only an imprecise discussion of precision is not lost on me!

The software that you use to produce the regression estimates b_0, b_1, b_2, b_3, and b_4 will also produce a "standard error" for each of these estimates, though I will omit the details of how those standard errors are computed. The standard errors describe how precise our estimates are. The lower the standard errors, the more precise the estimates, which is desirable. In the regression output of most software packages, the standard error for each estimate typically lies immediately below or to the right of that estimate, and we will see many examples of this in Chapter 7.

There is no magical threshold for "high" versus "low" precision, but a common rule of thumb that many researchers look for is that *the magnitude of the estimated regression coefficient is nearly double the size of its standard error*. So if we pick a particular parameter estimate, say, b_3, we would like for the ratio b_3/s_3 to be nearly two (and higher is even better), where s_3 is the standard error for b_3. Note that we ignore negative signs when making this evaluation. So if b_3 is -2.158 and s_3 is 0.891, then this meets our rule of thumb for taking the result very seriously, because 2.158 is more than twice 0.891. Less stringent rules of thumb that are easier to meet are also sometimes used, and we will occasionally rely on them in Chapter 7.

The sample size affects the precision of the regression estimates. If our estimates (b_0, b_1, b_2, b_3, and b_4) and their corresponding standard errors are generated using a sample of 100 workers, we could generally increase our precision (i.e., lower our standard errors) by instead using a sample of 500 or 1000 workers. The more data we have, the more reliably we can estimate β_0, β_1, β_2, β_3, and β_4, and the more closely our estimates of b_0, b_1, b_2, b_3, and b_4 will be to their unknown counterparts. However, this argument presumes that we still wish to estimate only five parameters. What matters is the sample size relative to the number of parameters being estimated. If you use a sample of 5000 to estimate five parameters, the parameters will be estimated with high precision. But if you use the same sample of 5000 to estimate 4950 parameters, precision will be much lower. For a fixed sample size, precision deteriorates as the number of parameters you try to estimate increases. Moreover, the sample size must exceed the number of parameters you wish to estimate; otherwise the regression coefficients cannot even be computed (i.e., the software package you're using will deliver an error message). If you don't believe me and are feeling defiant, give it a try and see what happens!

6.7 Lessons for Managers

Recent technological advances in hardware and software allow for more extensive data collection and more sophisticated data analysis than in the past. Careful data analysis often leads to better business decisions, including those concerning compensation, so good managers require competence in data analysis. Even if you are not personally analyzing the compensation data yourself, you need to be able to understand the underlying methodologies well enough so that you can correctly interpret the results that your subordinates or consultants produce, and use them to make better business decisions. You also need to be able to ask the right questions to be sure that your subordinates who conduct the statistical analysis aren't bamboozling you.

When contemplating any compensation-related decision, first clearly define the business question. Then ask yourself whether analyzing some data might clarify the picture and lead you to a better decision. If you conclude that data analysis would be helpful, what type of data would you need? Are such data readily available? If not, could you collect them? How long would that take, and can you afford to delay the business decision until after data collection, cleaning, and analysis? Compensation analytics can be conducted using data from your own organization or from other organizations. Be aware of what data are available, both internally and externally.

With the data in hand, the temptation to jump directly to analysis is strong. You probably want answers ASAP, and perhaps your boss is pressuring you to respond and deliver quickly. But you must always invest the necessary time and effort in meticulously cleaning the data to ensure their integrity. High-quality data analysis is challenging enough as it is, even under the best of circumstances, so you must not cripple yourself at the outset by pursuing analysis of flawed data. Even the most careful statistical analysis is a waste of time if the data are flawed. Your job during data cleaning is that of a detective, devising creative ways to detect subtle clues that would alert you to potential problems, and then pursuing and resolving those problems. In your reported analysis, be transparent about problems or doubts with the data that you can't resolve and also about the judgment calls you have made to resolve problems.

Data cleaning and analysis can be tedious and time-consuming. Sometimes many lines of computer code are required, with the need for lots of judgment calls along the way, both during data cleaning and data analysis. Even a single mistake or typo in your computer code can render the entire analysis meaningless and worthless. To ensure that you always conduct your data analysis with sufficient care, continuously remind yourself that the results of your analysis will serve as the basis for a business decision. So if you screw up the analysis, it's going to lead you to a bad, and possibly catastrophic, decision. If you can't trust your data analysis enough to base business decisions on it, then what is the purpose of the data analysis in the first place?

Appendix: Nonlinear Relationships among Variables in a Regression

In the main regression of section 6.4, i.e.,

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \varepsilon_i$$

the parameter of primary interest was β_2, the coefficient on $Male_i$, which measures the relationship between gender and salary, holding constant the other variables on the regression's right-hand side. That relationship can be described by "just one number", namely the unknown β_2 (or its known, estimated counterpart, b_2). But maybe the relationship between gender and salary isn't "just one number" that is the same for all workers. Maybe it depends on whether or not someone graduated from college.

One way to investigate this possibility is to estimate the regression from section 6.4 twice (with one slight difference both times). First, estimate it using only those workers *with* a college degree. Then estimate it again using only those workers *without* a college degree. What's the "one slight difference"? You would omit $College_i$ from the right-hand side of both regressions. Why? Just like you could never estimate a gender difference in pay using a sample of men only (or of women only), you cannot measure the effect of a college degree on salary if you only use data on college graduates, or if you only use data on non-graduates.

When you have estimated the preceding regression twice, you'll have two b_2 estimates, one from the sample of college graduates and the other from the sample of non-graduates. If they're pretty similar, then the relationship between gender and pay doesn't depend on whether or not the worker went to college, whereas if they're very different, it does. How similar is "pretty similar"? There's a statistical test that you can do to see if the two estimates of b_2 are basically identical in a statistical sense, but explaining it would take us too far afield. Plus, there's an alternative approach that involves an easier statistical test and that requires only one regression rather than two. Here's the regression:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \beta_5 \boldsymbol{(Male_i \times College_i)} + \varepsilon_i.$$

The only difference between this regression and the one from section 6.4 is that it includes a new independent variable, $\boldsymbol{Male_i \times College_i}$, indicated in boldface. So there are two independent variables. The first, which is $Male_i$, is called the "main effect", and its coefficient is β_2. The second, which is the product $Male_i$Â $College_i$, is called the "interaction term", and its coefficient is β_5. In the regression of section 6.4 that omitted the interaction term, the relationship between gender and salary was described by "just one number", namely β_2, so once you produced an estimate of that, namely b_2, you were home free. In the present regression that also includes the interaction term, the relationship between gender and salary is no longer “just one number” but rather

$$\beta_2 + \beta_5 College_i.$$

This relationship involves two unknown parameters (β_2 and β_5) and one variable, $College_i$. So once you produce estimates of β_2 and β_5, namely b_2 and b_5, you’re home free, because you can then compute an actual number:

$$b_2 + b_5 College_i.$$

Remember that $College_i$ is a dummy variable that either equals 1 or 0. If it’s 1, meaning that worker *i* holds a college degree, then the relationship between gender and salary is $b_2 + b_5$Â 1, or just $b_2 + b_5$. If instead it’s 0, meaning that worker *i* never graduated from college, then the relationship between gender and salary is $b_2 + b_5$Â0, or just b_2.

Suppose that you find that your estimate of β_5, namely b_5, has a very large standard error, so that the estimate is statistically insignificant. That means that, in a statistical sense, it’s basically zero. In that case, $b_2 + b_5 College_i$ becomes b_2 + 0Â $College_i$, or just b_2. In other words, it doesn’t matter whether or not someone went to college, i.e., it doesn’t matter whether $College_i$ equals 1 or 0, because the relationship between gender and pay is, once again, measured by “just one number”, b_2. In this case, the new interaction variable that you added to the regression turns out to be statistically irrelevant, which effectively brings you back to the situation from section 6.4.

We’ve now seen two ways to investigate whether the relationship between gender and pay depends on whether workers have a college degree. One way is to estimate the original regression from section 6.4 (dropping $College_i$) once for college graduates, then again for non-graduates, and then compare the estimates of b_2 from each regression. Another way is to estimate the original regression from section 6.4 (including the interaction $Male_i$Â$College_i$) and then determine whether b_5 is statistically significant. Neither approach is better or worse than the other. They just have different plusses and minuses.

The main advantage of the two-sample approach is also its main disadvantage. That is, when you use the two-sample approach you get two estimates not only of β_2 but of all other parameters as well. For example, β_1 (the effect of age) is also allowed to differ between college graduates and non-graduates, whereas in the interaction approach where you just estimate one regression (including an interaction term) using the entire sample, there is just one estimate of β_1. Allowing all of the right-hand side variables to have differential effects on pay according to whether the worker has a college degree might be desirable if you have a lot of data. But it also requires estimating a larger number of parameters, meaning your estimates will be less precise, which is a problem if you have a small sample. The two-sample approach requires you to estimate eight parameters in total (i.e., β_0, β_1, β_2, and β_3

from the college regression, and the same four parameters from the non-college regression), whereas the interaction approach only requires you to estimate six (i.e., β_0, β_1, β_2, β_3, β_4, and β_5). Moreover, that disparity between the two regressions grows when the number of control variables increases.

Another advantage of the interaction approach is that it works even if the variable that you suspect affects the relationship between gender and pay is continuous (like Age_i) rather than discrete (like $College_i$). When it was discrete, the two-sample approach simply involved estimating the regression once for the sample of workers for whom $College_i = 1$ and then again for the sample for whom $College_i = 0$. But think about how that approach would work using Age_i rather than $College_i$. It's impractical to estimate the regression separately for every age level. You might have only four workers, for example, who are age 45, only one or two who are 49, and maybe none who are 57. These samples aren't large enough to estimate the regression from section 6.4, and even if they were, estimating the regression separately for each age level would be too tedious. No one wants to look at all those regressions!

The interaction approach, however, works beautifully. Just estimate the following regression using the full sample of workers:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \beta_5 (\boldsymbol{Male_i \times Age_i}) + \varepsilon_i.$$

Now the relationship between gender and pay is

$$\beta_2 + \beta_5 Age_i,$$

and once you have the estimates b_2 and b_5 in hand, you have your desired estimate of the relationship between gender and pay, namely

$$b_2 + b_5 Age_i.$$

If b_5 is estimated with low precision (i.e., has a high standard error) then the desired effect is "just one number", i.e., b_2, as in section 6.4. But if it is estimated precisely, then the effect depends on the worker's age. For example, suppose that your estimates of β_2 and β_5 are $b_2 = 4513.35$ and $b_5 = 62.95$. Then the relationship between gender and pay is

$$4513.35 + 62.95 Age_i.$$

This says that, holding educational attainment constant, men tend to get paid more than women by about

$4513.35 + $62.95Â45 = $7346.10 for 45-year-olds and

$4513.35 + $62.95Â57 = $8101.50 for 57-year-olds.

Nothing stops you from computing the gender gap in annual pay for 116-year-olds, which is

$4513.35 + $62.95Â116 = $11,815.55,

but you should refrain from doing so, because it's virtually certain that none of the workers in your sample are anywhere near 116. Your predictions are most reliable within the range where the bulk of your data are concentrated.

Now suppose that Age_i, rather than $Male_i$, is the independent variable. That is, you're primarily interested in measuring the relationship between age and pay. In the regression from section 6.4, this relationship is "just one number", namely β_1, that you can estimate by b_1. But suppose you suspect that the relationship itself depends on the worker's age. In other words, the answer to the question "How much additional annual salary is associated with one extra year of age?" might not be the same for 30-year-olds as it is for 55-year-olds, i.e., it may not be "just one number" for workers of all ages. To explore this possibility, you can estimate the following regression:

$$Salary_i = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 (Associate's\ Degree)_i + \beta_4 College_i + \beta_5 \boldsymbol{Age_i^2} + \varepsilon_i.$$

This is the regression from section 6.4 but with an additional independent variable added in boldface, namely the square of the worker's age. So there are two independent variables in this regression, namely Age_i and Age_i^2. The relationship between age and salary is now

$$\beta_1 + 2\beta_5 Age_i,$$

and if you're wondering where that mysterious "2" came from, don't worry about it for now!

If your estimate of β_5, namely b_5, turns out to be statistically insignificant, then the estimated relationship between age and salary is "just one number", i.e., b_1, as in section 6.4. Why? Because if you plug $\beta_5 = 0$ into

$$\beta_1 + 2\beta_5 Age_i,$$

you get β_1, which you can estimate by b_1. But if b_5 is estimated precisely, then the relationship between salary and age itself depends on age. If your estimates of β_1 and β_5 are $b_1 = 2575.37$ and $b_5 = -28.67$, for example, then an additional year of age is associated with \$855.17 more in annual salary for a 30-year-old and \$578.33 *less* in annual salary for a 55-year-old. To obtain those numbers, just plug the ages 30 and 55 into

$$b_1 + 2b_5 Age_i.$$

For 30-year-olds, that is

$$\$2575.37 + 2(-28.67)(30) = \$855.17,$$

and for 55-year-olds it is

$$\$2575.37 + 2(-28.67)(55) = -\$578.33.$$

An additional year of age not only has differential effects on annual salary depending on age, but even the sign of the effect is reversed!

To summarize, in section 6.4 all of the estimated relationships were *linear*, meaning that they were described by "just one number", whereas in this appendix we considered some examples of *nonlinear* relationships that themselves depend on the values of certain variables. We considered three examples of nonlinear relationships: an interaction involving two discrete variables (i.e., $Male_i$ and $College_i$), an interaction involving one discrete variable and one continuous variable (i.e., $Male_i$ and Age_i), and the square of a continuous variable. You can even think of the third example as an interaction . . . it includes in the regression model a new interaction term that is the independent variable, Age_i, interacted with itself, because Age_i Â Age_i is just Age_i^2. Many other types of nonlinear relationships can be explored, but let's leave the discussion here.

Oh, one last quick thing! Are you still wondering where that mysterious "2" came from that showed up in $\beta_1 + 2\beta_5 Age_i$? If you don't care, or if you care but haven't had multivariate calculus, stop reading and skip to Chapter 7. If you care and have taken multivariate calculus, then notice that $\beta_1 + 2\beta_5 Age_i$ is just the partial derivative of the regression's right-hand side, with respect to Age_i. The right-hand side is just a function of a bunch of variables, and whenever you want to know the change in a function that's associated with a small change in one of the variables in that function (holding the function's other variables constant), you simply compute the partial derivative of the function with respect to the variable of interest. That insight is helpful for dealing with all sorts of nonlinear relationships beyond those we've discussed here.

Case Discussion 7: Wage–Insurance Tradeoff (Part A)[1]

You're a compensation consultant who is hired by a client firm that wants to gain a better, data-driven understanding of the relationship between wages and employer-provided health insurance for entry-level, low-skilled jobs in northern California, where the client operates.

The client's specific questions are:

(1) Can local firms that offer health insurance get away with offering lower wages?
(2) If so, by how much?
(3) How do these answers depend on whether or not firm size is held constant, if at all?

[1] Note: This case is based on "The Elusive Wage-Benefit Tradeoff: The Case of Employer-Provided Health Insurance", *International Journal of Industrial Organization*, 37, 2014, pages 23–37, by Jed DeVaro and Nan L. Maxwell.

Table 6.4 Regression results

	Dependent variable = Wage	
	Regression (1)	Regression (2)
Health insurance	0.824 (0.397)	0.809 (0.401)
Firm size	–	0.086 (0.345)
Sample size	n = 884	

You have data (collected in 2005 and 2006) on 884 establishments sampled randomly in northern California. A particular establishment is represented by the index "i". For each establishment, there is information on the average hourly starting wage for workers in the typical entry-level, low-skilled job. That variable is called $Wage_i$ for establishment i. A dummy variable called $HealthInsurance_i$ equals 1 if establishment i offers health insurance, and 0 if it does not.

All of the control variables are dummy variables. Of greatest interest is $FirmSize_i$, which equals 1 if the firm to which establishment i belongs employs at least 20 workers, and 0 if it employs fewer workers. The other control variables are indicators for whether the establishment is in the for-profit sector, whether it is unionized, whether it is in a non-metropolitan area, whether it is in the service sector, and whether it is in the manufacturing sector.

You estimate two regression equations:

$$Wage_i = \beta_0 + \beta_1 \hat{A}\ HealthInsurance_i + \text{"}other\ control\ valiables\text{"} + \varepsilon_i \quad (1)$$

$$Wage_i = \beta_0 + \beta_1 \hat{A}\ HealthInsurance_i + \beta_2 \hat{A}\ FirmSize_i + \text{"}other\ control\ valiables\text{"} + \varepsilon_i \quad (2)$$

The ε_i represents the error term of the regression. The "other control variables" are described above; to save space they are not written out individually, although they are all included in the regression. You estimate the regression without a control for firm size (Regression 1) and then with that control included (Regression 2). Comparing those two estimates of β_1 reveals how sensitive the relationship between wages and health insurance is to whether or not firm size is held constant.

Estimated coefficients can be found in Table 6.4, with standard errors in parentheses. For conciseness, the estimated coefficients for the "other control variables" are not displayed, and neither are the estimated constant terms.

Questions

1. In Regression (1), what is your estimate of β_1, and how precisely estimated is it?
2. Carefully explain the interpretation of that estimate. Do the same for your estimates of β_1 and β_2 in Regression (2).
3. What are the answers to the three specific questions that your client firm posed?

4. Taken at face value, are the regression results consistent with what the theory of compensating differentials would predict concerning the relationship between wages and health insurance?
5. What are the challenges associated with interpreting your estimates of β_1 as representing compensating differentials arising from employer-provided health insurance?
6. Suppose that your client suspects that the relationship between wages and health insurance likely varies by firm size. In particular, your client suspects that firms that employ at least 20 workers have a *positive* relationship between wages and health insurance (holding all other variables constant), whereas firms that employ fewer than 20 workers have a *negative* relationship (again holding all other variables constant).
   (a) Explain how you would evaluate your client's hunch by estimating a particular regression *twice*. Write down the regression equation that you would estimate (twice) and state which two samples you would use to estimate it. Explain how you would use the parameter estimates from both estimations to measure the relationship between wages and health insurance and how that relationship varies with firm size.
   (b) Explain how you would evaluate your client's hunch by estimating a particular regression that includes an interaction term. Write down the regression equation that you would estimate and state which sample you would use to estimate it. Explain how you would use the parameter estimates to measure the relationship between wages and health insurance and how that relationship varies with firm size.
   (c) Suppose that your client's hunch is correct. Could the estimates of β_1 that are reported in the preceding table still occur or do those estimates contradict your client's hunch?

Further Reading

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press.

Davenport, T. H. (2006). "Competing on Analytics." *Harvard Business Review*, 1–9.

Greene, W. H. (2018). *Econometric Analysis*, 8th ed., Pearson.

Moody, K. (2017). "Can Data Solve Employers' Compensation Headache? Quantifiable Compensation Decisions Go A Long Way Towards Addressing Pay Equity." *HR Dive*, August, 24.

Stock, J. H. and Watson, M. (2015). *Introduction to Econometrics*, 3rd ed., Pearson.

Studenmund. A. H. (2017). *Using Econometrics: A Practical Guide*, 7th ed., Pearson.

Weaver, D. (2015). "7 Ways to Use Compensation Analytics". LinkedIn. Published June 1, 2015.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*, 5th ed., South-Western Cengage Learning.

7 Compensation Analytics II

Suppose that the administrators in the Chancellor's Office of the California State University (CSU) system are interested in the experiences of female science professors within the 23-campus system, particularly with respect to compensation. The administrators are aware that there is a gender gap in compensation in the US labor market in general – with women tending to receive lower pay than men – and that academia is no exception. Moreover, they know that in the technical "STEM" fields (i.e., science, technology, engineering, and mathematics) the gender pay gap is a particular concern.

These facts lead the administrators to wonder whether there is a gender gap in pay for the science faculty within the system and, if so, how large it is. One factor that leads them to suspect only a small gap, if any, is that in the CSU system the faculty union exerts tight control over compensation (see Chapter 5), which limits the opportunities for supervisor discretion that could create or sustain gender discrimination. Nonetheless, the administrators want to conduct data analysis to investigate matters. You are hired as a compensation consultant to tackle this problem, using the analytical tools we developed in Chapter 6, and the administrators have asked you to start with the fields of chemistry and biochemistry. The next section walks the administrators through your analysis, step by step.

7.1 Application: Gender Differences in Pay in the Sciences

Recall from Chapter 6 that compensation analytics begins by defining concrete business questions that can be informed by data analysis. The concrete question that you have been asked to tackle is, "Are there gender differences in compensation within the CSU system in the disciplines of chemistry and biochemistry, and, if so, how large are they?"

Your next step is to identify the type of data that you would need to address this question, and then obtain those data or, if necessary, collect them from scratch. For this analysis you would ideally like data from the chemistry and biochemistry departments on all 23 campuses over a number of years. But the purpose of this chapter is just to provide an illustrative example, and we don't need a lot of data for that. So suppose that you obtain data from the departments of chemistry and biochemistry from three campuses in the CSU system, namely Los Angeles (CSULA), Long Beach (CSULB), and Fresno (CSUF). All data are for the year 2016 and pertain to the regular (i.e., tenured and tenure-track) faculty. Your sample covers 61 professors and contains the following ten variables:

$Last_i$: Worker i's last name (e.g., "Curie")
$First_i$: Worker i's first name (e.g., "Marie")
$Campus_i$: Worker i's CSU campus (i.e., "LA", "Long Beach", or "Fresno")
$TotalSal2016_i$: Worker i's total salary in 2016
$Base2016_i$: Worker i's base salary in 2016
$Male_i$: Dummy variable equaling 1 if worker i is male and 0 if female
$Assistant_i$: Dummy variable equaling 1 if worker i is an assistant professor, 0 otherwise
$Associate_i$: Dummy variable equaling 1 if worker i is an associate professor, 0 otherwise
$Professor_i$: Dummy variable equaling 1 if worker i is a professor, 0 otherwise
$Administrator_i$: Dummy variable equaling 1 if worker i is an administrator, 0 otherwise

Three of the variables are "string" variables (namely *Last*, *First*, and *Campus*). The remaining seven are numerical variables, two of which (namely the compensation variables) are continuous and the rest of which are discrete dummy variables equaling either zero or one.

After defining a concrete business problem that can be informed by data analysis, identifying the type of compensation data necessary to analyze that problem, and then obtaining those data, the next step in compensation analytics is cleaning those data. The cleaning step is critical, but many data analysts give it insufficient attention, or skip it entirely, resulting in garbage analysis! A good place to start the cleaning is by examining descriptive statistics for each variable.

The following analysis uses the STATA statistical software program, but there are many alternative programs available for this type of analysis (e.g., SAS, SPSS, R, and EXCEL if the Data Analysis ToolPak is active). The STATA command for computing summary statistics is "summarize", or just "sum" for short. Specifying the additional "detail" option, or just "det" for short, provides a more detailed set of descriptive statistics. The STATA command "sum, det" produces detailed descriptive statistics for every variable in the data set. To start, Table 7.1 displays the output for the first four of the ten variables.

The first three variables appear to have no observations, meaning that they do not exist in your data set. One possibility is that there was an error importing the data into the software package. In the present case, however, the issue is simply that

Table 7.1 Descriptive statistics for four variables

```
                            Last
-------------------------------------------------------------
no observations

                            First
-------------------------------------------------------------
no observations

                            Campus
-------------------------------------------------------------
no observations


                        TotalSal2016
-------------------------------------------------------------
      Percentiles      Smallest
 1%        5251          5251
 5%      19909.5        18744
10%        33122        18888       Obs                  60
25%      77923.5        20931       Sum of Wgt.          60

50%      92042.5                    Mean           86224.65
                        Largest     Std. Dev.      30162.52
75%     103883.5        122530
90%       116260        139477      Variance       9.10e+08
95%     131003.5        140945      Skewness      -.7979129
99%       147202        147202      Kurtosis       3.682792
```

these are string variables, whereas the statistical command "summarize" that produces descriptive statistics only works on numerical variables (see section 6.3). This makes sense. How would you compute the "average" of a list of last names?? It's a good idea to print all string variables to the screen and eyeball them to check that they were imported correctly. Workers' names aren't important for our purposes, and they are typically omitted for reasons of privacy, so we will ignore them henceforth. But each worker's campus is a potentially important variable in this analysis. It must be converted from a string to a numerical value so that it can be used in quantitative statistical analysis. $Campus_i$ is a "categorical" string variable in which there are three categories (i.e., LA, Long Beach, and Fresno), with each worker falling into one (and only one) of those categories. You create the following set of dummy variables . . . one for each category. These are:

$csula_i$: Dummy variable equaling 1 if worker i is at CSU Los Angeles, 0 otherwise

$csulb_i$: Dummy variable equaling 1 if worker i is at CSU Long Beach, 0 otherwise

$csuf_i$: Dummy variable equaling 1 if worker i is at CSU Fresno, 0 otherwise

Observe a few things about the fourth variable, *TotalSal2016*. First, check that the reported sample size matches what you expect it to be. In this case it is 60 (listed as

"Obs", for number of *obs*ervations). But you should have information on 61 professors, not 60. What's going on?

After investigating the matter, you discover that one of the professors at CSUF does not have a total salary recorded in 2016. Such "missing values" are automatically dropped by the computer program when it computes descriptive statistics and other statistical functions. If you have time for further investigation and can track down the missing salary number using the person's name, great. But because it's only one observation out of the 61, it's not worth losing a lot of sleep over this. Just make a note of it, and move on. If half the professors had missing salary data, that would be a different story and a more serious problem deserving more of your time.

The mean annual total salary is $86,224.65. Knowing what you know about salaries in chemistry and biochemistry within the CSU system, that number strikes you as suspiciously low. You would have expected something closer to $100,000. Investigating further, you notice that the median (i.e., the 50th percentile) is $92,042.50, which amplifies your suspicion. It's unusual in salary data (not just for professors of chemistry and biochemistry, but also in general) for the median to exceed the mean. As we discussed in section 6.5 and saw there in a histogram, usually salary data are "right-skewed", meaning that the salary distribution has a "long right tail" in which a small number of workers receive extremely high pay. These "compensation outliers" inflate the mean, whereas the median is insensitive to them, so typically the mean salary exceeds the median.

In the present case the data are actually left-skewed, meaning that the salary distribution has a "long *left* tail" in which a small number of workers receive extremely little pay. This can be seen by inspecting the skewness coefficient, which is about -0.8. As we discussed in sections 6.3 and 6.5, a negative skewness coefficient indicates that the data are left-skewed, a positive coefficient indicates that they are right-skewed, and a coefficient of zero indicates that they are symmetric, as in the case of a normal distribution or other "bell-shaped curve".

Looking at the extreme values of the data sharpens the picture further. The maximum value is $147,202. That's a high number, but not so high as to be alarming like in the example with secretaries in Chapter 6, i.e., you are aware that a small number of science professors in the CSU system receive such high total pay. But the minimum value is $5251, which is clearly a problem. No professor of chemistry and biochemistry in the CSU system (or anywhere else, for that matter) receives such a low annual total salary. The other three salaries in the lowest four are also too low to be realistic, and even the 10th percentile of the distribution (which is $33,122) is too low. The descriptive statistics have revealed a big problem with the data, in that (at least) 10% of the salary observations are too low to be believed.

Proceeding with statistical analysis when over 10% of the sample is seriously flawed would be a big mistake, and you could never trust the results. Unfortunately, that happens too often, because data analysts are often in a hurry, impatient to get on with the analysis, and therefore sloppy with their data cleaning. One solution to the problem you discovered is to omit the problematic observations from the analysis, which in this case means "left trimming" the data. If the data set is very large, and/or if there is no way to figure out what's going on with the problematic salary observations, then omitting them may be the safest and only practical option. But it is always better to retain data when you can do so credibly, particularly in a sample that is already so small. Examining each of the problematic cases more closely reveals the nature of the problems.

There are ten problematic observations. You suspect that many of these cases involve new hires. For example, when a new assistant professor joins the university at the start of the academic year, in August or September, they are only paid for the last three or four months of the calendar year. Because the salary information is recorded by calendar year, those professors who joined the university in fall 2016 will have artificially low annual salary data for 2016. To confirm this, for each of the professors with suspiciously low pay, you go back to the original database and look up their pay in the previous year, i.e., 2015. As expected, you cannot find these people in the 2015 database, because they joined the university in fall 2016.

So you conclude that anyone who was missing from the public database in 2015 and had a suspiciously low salary in 2016 must have joined the university late in 2016, at the start of the academic year. You know when the academic year started in 2016 at all three universities, so you can figure out how many months the (low) recorded salary figure corresponds to and then inflate it up to an appropriate 12-month figure. When you make those corrections, you immediately get sensible total salary numbers. For example, an assistant professor at CSULA had a recorded 2016 total salary of $18,888, and a "corrected" salary of $77,792, which very closely matches what you know to be the ballpark starting salary for assistant professors of chemistry and biochemistry in the system.

Half of the ten problematic observations could be fixed as just described. Of the remaining five professors, two appeared in the 2015 database (and therefore could not safely be assumed to be new hires in 2016). These appeared to be older professors who had probably retired at the end of the 2015–2016 academic year, so they were only paid for the first part of 2016. Their absence in the 2017 database is consistent with that conjecture. Checking the 2015 database for both professors, they were found to have reasonable-looking (i.e., full-year) salaries. For these observations, a defensible approach is to replace the 2016 figures with the 2015 figures (per the CSU collective bargaining agreement, no significant raises were awarded between 2015 and 2016).

The remaining three problematic observations required a bit more involved investigation, and the bottom line is that for two of them you were able to generate a "corrected" salary number with confidence. But for one of them there was

Table 7.2 Descriptive statistics for (cleaned) total salary

```
                        TotalSal2016_clean
-------------------------------------------------------------
      Percentiles      Smallest
 1%        72083          72083
 5%        74041          73224
10%        77792          74041       Obs                  59
25%        82430          74277       Sum of Wgt.          59

50%        94333                      Mean           97168.53
                        Largest       Std. Dev.      18633.23
75%       106731         139477
90%       121327         140945       Variance       3.47e+08
95%       140945         147202       Skewness       1.207786
99%       158584         158584       Kurtosis       4.538537
```

insufficient information to detect what was going on, so you dropped that professor from your analysis. You also followed the good practice of creating a dummy variable to "flag" any data observations to which you made adjustments. It equals one if you "corrected" the salary number, and zero otherwise. You call that variable $CleanFlag_i$, and it equals one for all ten problematic observations (including the observation that you dropped from the analysis by replacing the 2016 total salary figure with a "missing value"), and zero otherwise.

The new, "cleaned" total salary variable is called $TotalSal2016\_clean_i$, and its descriptive statistics appear in Table 7.2, based on the STATA command "sum TotalSal2016_clean, det".

The sample size has dropped to 59. Recall that the original data had 61 professors, but one of them (a professor from CSUF) had missing data for total salary, and a second one was assigned a missing value by you during data cleaning. The mean has increased dramatically, from $86,225 to $97,168.53 as a consequence of data cleaning, and it is now pretty close to your initial expectation of about $100,000. This change in the mean reveals how highly sensitive the mean is to extreme values in the data. In contrast, the median is insensitive. Any data modifications you make on the lowest (or highest) salary numbers should have no effect at all on the median. So why has the median (i.e., 50th percentile) increased from $92,042.50 to $94,333?

The reason is that the sample sized dropped from 60 to 59, and the professor you dropped was one with a low salary. When the sample size was 60, which is an even number, the median was the average of the salaries of the two "middle" professors, namely Professor 30 (say, "Jane") and Professor 31 (say, "John"), where the professors are ordered from 1 (the lowest-paid professor) to 60 (the highest-paid professor). After you dropped one of the lowest-paid professors, the sample size became an odd number, so the median is simply the middle professor's salary. That professor is John (who is now professor 30), who has a salary of $94,333. In the original sample of 60 the median, i.e., $92,042.50, was the average of John's salary of $94,333 and Jane's, which must be $89,752. You can infer Jane's salary by solving for X in the equation ($94,333 + X)/2 = $92,042.50 or simply by looking at it directly in your data.

The new median is below the new mean, suggesting the anticipated right skew of the salary distribution, and the positive skewness coefficient of 1.2 confirms that. The minimum salary number of $72,083 is a very reasonable number for an assistant professor in the CSU system in this field. The maximum salary is now $158,584. This exceeds the maximum salary in the uncleaned data (which is now the second highest salary), which means that one of the professors that you "cleaned" is now at the top. A professorial salary of $158,584 is high, but not totally unheard of within the CSU system. Finally, data cleaning has reduced the standard deviation (a measure of how spread out the salary distribution is) from $30,162.52 to $18,633.23. That occurred because an "artificially long left tail" in the uncleaned data was shortened during data cleaning, thereby reducing the overall dispersion in the salary data. The bottom line is that the descriptive statistics on the cleaned data all look reasonable and in harmony with the institutional knowledge you have about the CSU system. That variable, at least, appears ready for analysis.

The next variable in the data set is $Base2016_i$, which is worker *i*'s 2016 base salary. The base salary is what is used for the purpose of pension calculations in the CalPERS system (see Chapter 11). The base salary might also be lower than the total 2016 salary that we just examined. For example, professors might do overload teaching for extra compensation in a given year, which adds to their compensation that year even though the base salary remains unchanged. Table 7.3 displays descriptive statistics for $Base2016_i$.

This variable, not surprisingly, suffers from all of the same problems we identified for $TotalSal2016_i$, and the same steps that you used to clean that variable can also be used on the base salary. One encouraging point about the base salary, even in its uncleaned state, is that the mean and median base salaries both fall below their

Table 7.3 Descriptive statistics for (uncleaned) base salary

```
                              Base2016
-------------------------------------------------------------
      Percentiles      Smallest
 1%         5251           5251
 5%      19000.5          18744
10%        26818          18744       Obs                  60
25%      73261.5          19257       Sum of Wgt.          60

50%        83860                      Mean           79779.78
                        Largest       Std. Dev.      29176.86
75%      96295.5         115806
90%     103883.5         139192       Variance       8.51e+08
95%       127499         139477       Skewness      -.6757275
99%       140455         140455       Kurtosis       3.639382
```

Table 7.4 Descriptive statistics for (cleaned) base salary

Base2016_clean

	Percentiles	Smallest		
1%	62533	62533		
5%	70888	69793		
10%	73299	70888	Obs	59
25%	78988	72083	Sum of Wgt.	59
50%	89736		Mean	90731.61
		Largest	Std. Dev.	18288.96
75%	96478	139192		
90%	108761	139477	Variance	3.34e+08
95%	139477	140455	Skewness	1.70171
99%	158584	158584	Kurtosis	6.415586

counterparts in the (uncleaned) $TotalSal2016_i$. That must be so, because total salary can exceed or equal base salary but can never fall below it. If it were to fall below the base, that would have triggered a deeper investigation to detect and correct the underlying problem. You call the cleaned base salary $Base2016\_clean_i$, and Table 7.4 displays its descriptive statistics (from STATA's "sum base2016_clean, det" command).

Everything looks reasonable here, and the mean of the cleaned base salary falls appropriately below the mean of the cleaned total salary variable. As we saw before with the total salary, the median of the cleaned base salary variable has also dropped slightly because one low-paid professor was dropped from the sample. So the base salary now looks ready for analysis.

Remember that another approach (in addition to inspecting descriptive statistics) is to graph the data so that you can actually see the outliers. Panel A of Figure 7.1 displays a scatter diagram with the two original (unclean) salary variables, namely $TotalSal2016_i$ on the horizontal axis and $Base2016_i$ on the vertical. The graph reveals three high-paid outliers and ten low-paid outliers, though it looks like nine because two of the low-paid data points nearly perfectly overlap. For reasons I discussed earlier, the high-paid outliers are not cause for concern, but the low-paid ones are, and we have already explained how you fixed them. Panel B of Figure 7.1 displays a scatter diagram with the two (cleaned) salary variables, which no longer exhibit any low-paid outliers.

You next inspect descriptive statistics for the remaining variables, all of which are dummy variables. Five of them are in the original data, three you created from the original string variable, $Campus_i$, and one was the dummy variable you created to "flag" the workers whose compensation data you "corrected". For *dummy variables*, as opposed to *continuous variables* like the two compensation variables, it is not necessary to inspect the full set of descriptive statistics including the percentiles, skewness, standard deviation, etc. The information in Table 7.5 suffices.

Table 7.5 Descriptive statistics for dummy variables

	Sample size (n)	Minimum	Maximum	Mean
Male	61	0	1	0.672
Assistant	61	0	1	0.279
Associate	61	0	1	0.180
Professor	61	0	1	0.541
Administrator	61	0	1	0.098
CSULA	61	0	1	0.311
CSULB	61	0	1	0.377
CSUF	61	0	1	0.311
CleanFlag	61	0	1	0.164

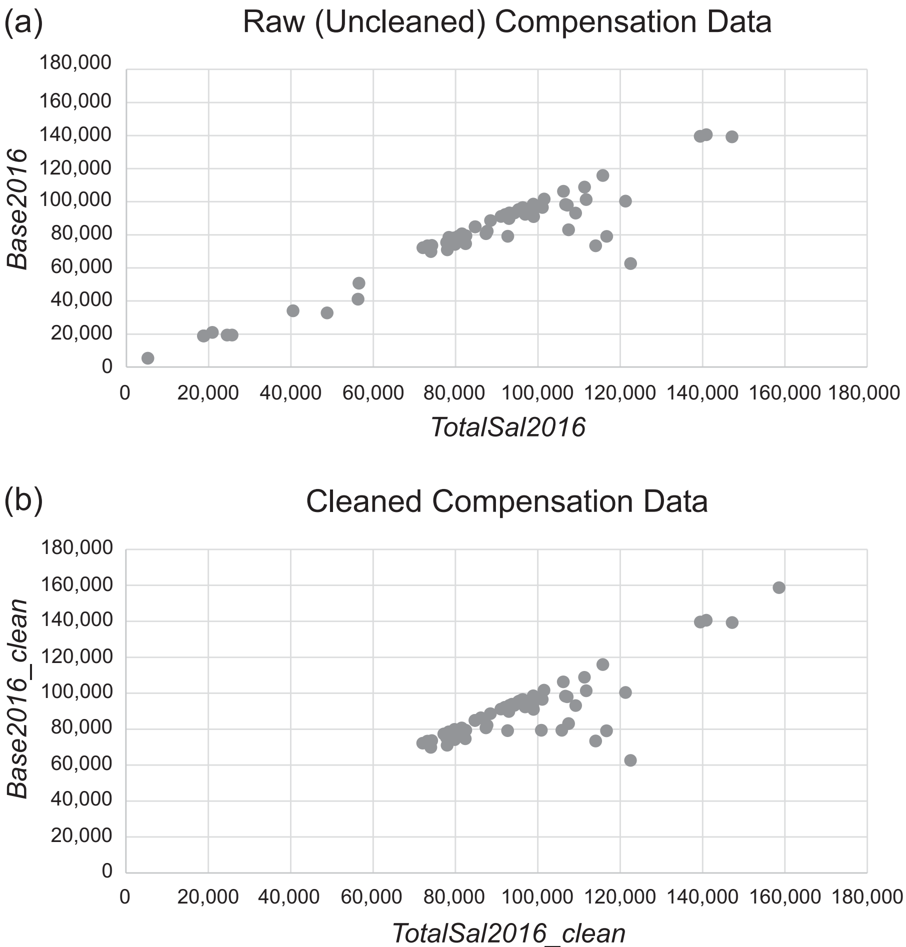


Figure 7.1 Scatter diagrams of the two compensation variables.
Panel A: original (uncleaned) data.
Panel B: cleaned data.

You have complete data for all variables for the full sample of 61. That means that in the subsequent statistical analysis you will lose no additional observations beyond the two that you lost because of the compensation variables. All of the minima are zero and the maxima are one, as expected. For a dummy variable, the mean is just its total number of ones divided by the sample size. So if you know the mean, you can infer the number of ones, and vice versa. For example, the mean of *Male* is 0.672, so you can infer that there are 41 males, i.e.,

0.672 Â 61 = 40.992,

and 20 females. The reason why you get 40.992 instead of exactly 41 is because of rounding error. You report the mean of *Male* as 0.672 for brevity, but the computer output actually reported several more digits, i.e., 0.6721311, and even that longer number is still truncated. If that truncated number is multiplied by 61, you get 40.9999971. Recall that you “corrected” the compensation variables for ten professors during data cleaning, so you expect *CleanFlag* to be one for exactly ten observations. You can check this by noticing that

0.164 Â 61 = 10.004.

The next thing to check is that all “sets” of dummy variables have means that sum to one. For example, the set of three dummy variables that you created for the CSU campuses should have means that sum to one, because every professor must belong to one (and only one) campus. Fortunately,

$$0.311 + 0.377 + 0.311 = 0.999,$$

which deviates from one only because of rounding errors in the means. Next check the means for the workers’ ranks (assistant professor, associate professor, professor, or administrator):

$$0.279 + 0.180 + 0.541 + 0.098 = 1.098.$$

This exceeds one by an amount (nearly 10%!) that is clearly not due simply to rounding error. Notice also that it exceeds one by an amount exactly equal to the mean of $Administrator_i$, which immediately reveals what’s going on. Every worker must be either an assistant professor, or an associate professor, or a professor, and can hold only one of those three ranks. Additionally, such workers may or may not be administrators. The most common administrative job is “Department Chair”. Usually department chairs are professors, but sometimes they are associate professors or even assistant professors. There are also administrative appointments other than department chair, and these can be held by faculty of any rank. The bottom line is that the three campus dummy variables form their own set, which means that they sum to one, whereas $Administrator_i$ belongs to its own set of two (i.e., administrators versus non-administrators), just like $Male_i$ belongs to its own set of

Table 7.6 Regression output using (cleaned) total salary as dependent variable

```
      Source |       SS       df       MS              Number of obs =      59
-------------+------------------------------           F(  1,    57) =    5.70
       Model |  1.8299e+09     1  1.8299e+09           Prob > F      =  0.0203
    Residual |  1.8307e+10    57   321184097           R-squared     =  0.0909
-------------+------------------------------           Adj R-squared =  0.0749
       Total |  2.0137e+10    58   347197275           Root MSE      =   17922
------------------------------------------------------------------------------
totalsal20~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   11765.17   4928.965     2.39   0.020     1895.086    21635.25
       _cons |   89391.55   4007.394    22.31   0.000     81366.88    97416.22
------------------------------------------------------------------------------
```

two (i.e., males and females). Table 7.5 raises no concerns, so all of the data appear to be clean and ready for analysis.

After defining a concrete business problem that can be informed by data analysis, identifying the type of compensation data necessary to analyze that problem, obtaining those data, and cleaning those data, the next step in compensation analytics is data analysis. Recall that you're interested in studying possible gender differences in compensation in chemistry and biochemistry on the three CSU campuses. You start by estimating a simple regression with total salary as the dependent variable and gender as the independent variable. The relevant STATA command is "regression", or just "reg" for short. Table 7.6 displays the output from "reg totalsal2016_clean".

The sample size in this regression is 59. This is because the regression contains the cleaned total salary variable, which is missing two observations (one was missing from the outset, and you inserted another missing value during data cleaning). Recall from section 6.4 how the two coefficient estimates are interpreted when the regression only contains a single variable on the right-hand side, which is a dummy variable. The constant (or "intercept") term is the mean value of the dependent variable for the worker group for which the independent variable equals zero. That is, the average 2016 total salary for women is $89,391.55. Its standard error appears to the right of the coefficient and is 4007.394. Recall that a rough rule of thumb for good precision is that the magnitude of the estimated coefficient be nearly twice as large as its standard error, and larger is even better. That is definitely true here, i.e., the regression constant term is estimated very precisely.

Next consider the regression "slope", i.e., the coefficient on the independent variable $Male_i$, which is the coefficient of greatest interest. It says that the average 2016 total salary is $11,765.17 higher for males than for females, a result that is precisely estimated because the estimate is more than twice its standard error. That's a pretty sizeable difference in pay and is worth investigating further, though remember that it's an apples-to-oranges comparison because there are no control variables on the regression's right-hand side.

Table 7.7 Regression output using (uncleaned) total salary as dependent variable

Source	SS	df	MS
Model	1.3504e+09	1	1.3504e+09
Residual	5.2326e+10	58	902180360
Total	5.3677e+10	59	909777660

Number of obs =	60
F(1, 58) =	1.50
Prob > F =	0.2261
R-squared =	0.0252
Adj R-squared =	0.0084
Root MSE =	30036

totalsal2016	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	10063.88	8225.784	1.22	0.226	-6401.821	26529.57
_cons	79515.4	6716.325	11.84	0.000	66071.22	92959.58

You next extend the regression to include some control variables, thereby bringing us closer to an apples-to-apples comparison of the pay between men and women. But before discussing your results, let's take a look at what would have happened if you had made a mistake that happens all too often, namely you didn't bother with the initial data cleaning and just jumped directly to data analysis. Using the original, uncleaned total salary variable as the dependent variable, the preceding regression (using the STATA command "reg totalsal2016") yields the results in Table 7.7.

The sample size is 60. It increased by one because it contains the professor whose salary you replaced with a missing value during data cleaning. The estimated regression constant suggests that the average 2016 total salary for women across the three CSU campuses is $79,515.40, and it is precisely estimated. That number is considerably smaller – in fact about $10,000 smaller – than the corresponding number in the "clean" regression. That is expected. Remember that the big problem in the original, uncleaned data is that there are some "annual" salaries that are artificially low, because the professors were only paid for part of the calendar year.

The estimated regression slope suggests that the average 2016 total salary was $10,063.88 higher for males than for females. That's a significantly smaller number than you found in the "clean" regression. But even more importantly, it is estimated with low precision, falling considerably short of the "nearly twice the standard error" rule of thumb for precision. The number even falls short of a more relaxed precision criterion that is sometimes used in statistics. The bottom line, using statistical jargon, is that the pay difference between men and women is "statistically insignificant". What that means is that it's so imprecisely estimated that it might as well be zero, meaning that there's a good chance that there's actually *no* gender difference in pay at all. Hopefully you now appreciate the importance of data cleaning before statistical analysis!

Returning to the clean data, let's discuss your results from the expanded regression that includes control variables for the professor's rank, CSU campus, and administrator status. Table 7.8 reports regression results from the STATA command

"reg totalsal2016_clean male associate professor csula csuf administrator".

Table 7.8 Regression output using (cleaned) total salary as dependent variable

```
      Source |       SS       df       MS              Number of obs =      59
-------------+------------------------------           F(  6,    52) =    9.30
       Model |  1.0424e+10     6  1.7373e+09           Prob > F      =  0.0000
    Residual |  9.7139e+09    52   186805570           R-squared     =  0.5176
-------------+------------------------------           Adj R-squared =  0.4620
       Total |  2.0137e+10    58   347197275           Root MSE      =   13668
------------------------------------------------------------------------------
totalsal201~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   5732.383   4064.717     1.41   0.164    -2424.068    13888.83
   associate |  -785.2518    5615.16    -0.14   0.889     -12052.9    10482.39
   professor |   19518.06   4709.995     4.14   0.000     10066.77    28969.36
       csula |   -2260.33   4365.328    -0.52   0.607       -11020    6499.343
        csuf |   823.4794   4551.476     0.18   0.857    -8309.726    9956.685
administrator |   18099.19   6234.887     2.90   0.005     5587.975    30610.41
       _cons |   81258.88   4814.742    16.88   0.000     71597.39    90920.36
------------------------------------------------------------------------------
```

The sample size drops back down to 59, because you are using the cleaned salary variable. The coefficient on $Male_i$ now suggests that the average 2016 total salary is \$5732.38 higher for men than for women *controlling for (or holding constant) rank, campus, and administrator status*. That number is considerably smaller than the \$11,765.17 that you found in the simple regression, and it is more accurate because it is closer to an apples-to-apples comparison. Precision of the estimate falls short of the "nearly twice the standard error" rule of thumb (i.e., it is only 1.41 times the standard error), but it meets a less stringent standard for precision that is sometimes used in statistics, so you should still pay some attention to this gender difference in pay.

For practice, let's examine the other estimated coefficients. Remember that there are three "rank" dummy variables, and you have omitted the lowest rank, i.e., $Assistant_i$, from the regression. The coefficient of $Associate_i$ (which is one rank higher than $Assistant_i$) is -785.2518. What that says, literally, is that the average 2016 total salary is actually \$785 *lower* for associate professors than for the omitted group (i.e., assistant professors), controlling for (or holding constant) gender, campus, and administrator status. That may seem puzzling, because you'd expect the higher-ranked job to be higher paying. But the standard error is 5615.16, which is far higher than the estimated coefficient. Precision is so low that in a statistical sense there's no difference in average total salary between assistant and associate professors. Even that result may surprise you. After all, the collective bargaining agreement that pertains to all CSU campuses requires that all assistant professors receive a (minimum) raise of 7.5% when promoted to the rank of associate professor. So what's going on?

One thing to keep in mind is that no assistant professors in the CSU system have tenure (i.e., a guaranteed job for life) whereas most associate professors have tenure.

So when people get promoted from assistant professor to associate professor, in addition to a raise of (at least) 7.5% they are typically granted a *job for life*. Even before reading Chapter 9, you can guess what that does to incentives! Some associate professors are ambitious and want to be promoted to the highest rank of professor (which comes with yet another raise of at least 7.5%), but others are content to "coast" and take a minimalist approach to their work, knowing that tenure prevents them from getting fired. Some remain for many years, even for the remainder of their careers, at the rank of associate professor. What this means is that they never get the pay increase that comes with promotion to the rank of professor . . . they get only the very meager cost-of-living increases that are occasionally negotiated in the collective bargaining agreement. But meanwhile, new assistant professors are hired at higher salaries that must remain competitive with the market. The collective bargaining agreement stipulates hard salary caps at each rank, but those caps are periodically increased when the contract is renegotiated, mainly so that the salary offers to assistant professors (the rank at which the vast majority of hiring occurs) do not fall too far below market levels.

In short, the assistant professor salaries are more market-driven than the associate professor salaries, so the regression results are unsurprising. The phenomenon is so common that it has a name . . . *salary inversion*, which has this Wikipedia definition:

> ***Salary inversion*** *refers to situations in which the starting salaries for new recruits to an organization increase faster than those for existing employees.*

As the holder of an administrative position in the CSU system, I can attest that higher-level administrators express concerns about salary inversion within the faculty ranks. It's a concern because it creates morale problems for the more senior faculty.

Returning to the regression, the coefficient on $Professor_i$ reveals that the average 2016 total salary is $19,518.06 higher for professors than for the omitted group (i.e., assistant professors). That's a big difference, and it is estimated very precisely. The coefficient is well over twice its standard error of 4709.995. In fact it is 4.14 times its standard error, as seen in the "t" column. So although the average salaries of assistant and associate professors are basically the same from a statistical standpoint, they are both considerably lower than those of professors.

Next consider the set of three campus variables, recalling that CSU Long Beach is the omitted group. The estimated coefficients suggest that faculty in chemistry and biochemistry at CSULA receive average 2016 total salaries that are $2260.33 less than those of the omitted group (CSULB). And salaries at CSUF are $823.48 higher than those at Long Beach. However, the standard errors are so high on both estimates that no statistically meaningful differences among them can be discerned. In a statistical sense, the three campuses don't differ in their pay, and that result is

Table 7.9 Regression output using (uncleaned) total salary as dependent variable

Source	SS	df	MS		
				Number of obs =	60
				F(6, 53) =	6.74
Model	2.3230e+10	6	3.8717e+09	Prob > F =	0.0000
Residual	3.0447e+10	53	574463915	R-squared =	0.4328
				Adj R-squared =	0.3686
Total	5.3677e+10	59	909777660	Root MSE =	23968

totalsal2016	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	-2382.351	6996.468	-0.34	0.735	-16415.49	11650.79
associate	19457.32	9642.723	2.02	0.049	116.4658	38798.17
professor	37591.12	8100.239	4.64	0.000	21344.1	53838.15
csula	-13226.39	7653.445	-1.73	0.090	-28577.25	2124.479
csuf	-2198.332	7841.814	-0.28	0.780	-17927.02	13530.36
administrator	26990.51	10926.91	2.47	0.017	5073.902	48907.13
_cons	65719.4	8437.448	7.79	0.000	48796.02	82642.77

unsurprising given that they are all subject to the same collective bargaining agreement. Finally, not surprisingly, administrators such as department chairs receive higher pay than non-administrators. The coefficient on $Administrator_i$ reveals that administrators have an average 2016 total salary that is $18,099.19 higher than that of non-administrators. The coefficient is 2.9 times its standard error, so it is estimated quite precisely.

To drive home my earlier warnings about the importance of data cleaning, let's take a look at the preceding regression using the "unclean" total salary dependent variable. Table 7.9 reports regression results from the STATA command

"reg totalsal2016 male associate professor csula csuf administrator".

Observe two differences between these results and the "clean" ones. First, the coefficient on $Male_i$ is actually negative, suggesting that females have higher average total salaries by $2382.35! Granted, the standard error corresponding to that estimate is extremely large, so from a statistical standpoint no gender difference in pay can be inferred, but the point is to observe the radical difference between this result and the earlier result from the "clean" regression. Second, the CSULA coefficient is now estimated with reasonably high precision, coming within firing range of the "nearly twice standard error" rule of thumb. In fact, it is 1.73 times its standard error, and it meets the bar for a less stringent criterion that is often used in statistics. This could lead you to the false conclusion that annual total salaries are $13,266.39 lower at CSULA than at the omitted campus (i.e., CSULB). But the earlier results from the cleaned data clearly reveal that from a statistical standpoint there are no pay differences across campuses.

All of the preceding results focus on total annual salary, which equals base salary plus various forms of extra pay. Sometimes the extra pay is compensation for

Table 7.10 Descriptive statistics for extra pay

```
                          extrapay
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs                  59
25%            0              0       Sum of Wgt.          59

50%         1287                      Mean           6436.915
                        Largest       Std. Dev.      11495.69
75%         7889          26557
90%        21557          37726       Variance       1.32e+08
95%        37726          40757       Skewness       2.735451
99%        59997          59997       Kurtosis       11.05117
```

"overload teaching", when faculty teach more than their required annual number of courses. Other times it may be compensation for additional duties like being a program director. Or it could reflect bonus payments with no additional work attached; for example, new faculty hires might be given "summer pay" for one or a few years, which is effectively a signing bonus. Does such "extra pay" differ by gender in chemistry and biochemistry within the CSU system? Let's see!

Start by creating a new "extra pay" variable, which is the difference between total and base salary, i.e., $ExtraPay_i = TotalSal2016\_clean_i - Base2016\_clean_i$. The relevant STATA command to produce the new variable is "generate", or either "gen" or "g" for short. The complete line of code is

```
g extrapay = totalsal2016_clean – base2016_clean.
```

Whenever defining a new variable, it is good practice to inspect its descriptive statistics. Table 7.10 displays results from the STATA command "sum extrapay, det".

There are 59 observations because the extra pay variable was created using the cleaned total compensation variable that has two missing observations. At least a quarter of the workers receive no extra pay, but one worker receives nearly $60,000 in annual extra pay. The average is $6436.92 in extra salary in 2016, but the median is much smaller at $1287. These numbers suggest that extra pay is strongly right-skewed (i.e., has a long right tail), which is confirmed by a large, positive coefficient of skewness (2.735).

Let's look at a regression that uses extra pay as the dependent variable and the same variables as before on the right-hand side. Table 7.11 displays results from the STATA command

"reg extrapay male male_administrator associate professor csula csuf administrator".

There is no statistically meaningful gender difference in the amount of annual extra pay received. The $Male_i$ coefficient's standard error (3367.35) is of far larger magnitude than the estimated coefficient itself (-248.341). In contrast, the

Table 7.11 Regression output using extra pay as dependent variable

Source	SS	df	MS		
Model	998081627	6	166346938	Number of obs =	59
Residual	6.6667e+09	52	128205256	F(6, 52) =	1.30
				Prob > F =	0.2748
				R-squared =	0.1302
				Adj R-squared =	0.0299
Total	7.6648e+09	58	132150947	Root MSE =	11323

extrapay	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	-248.341	3367.35	-0.07	0.941	-7005.423	6508.741
associate	825.9538	4651.791	0.18	0.860	-8508.547	10160.45
professor	560.4167	3901.921	0.14	0.886	-7269.361	8390.194
csula	4964.513	3616.387	1.37	0.176	-2292.299	12221.32
csuf	1333.879	3770.599	0.35	0.725	-6232.381	8900.138
administrator	11433.22	5165.193	2.21	0.031	1068.504	21797.94
_cons	2987.848	3988.697	0.75	0.457	-5016.058	10991.75

coefficient on $Administrator_i$ is estimated quite precisely, and it says that when the other variables on the right-hand side of the equation are held constant, having the status of administrator is associated with $11,433.22 in extra pay per year.

So it seems that $Administrator_i$ is a pretty important variable that relates to extra pay. Could it be that the gender gap in extra pay depends on whether the person is an administrator? To answer that question, we need to add an "interaction term" ($Male_i$ Â $Administrator_i$) to the right-hand side of the regression, as explained in the appendix to Chapter 6. This interaction term, which is a product of two dummy variables that equal zero or one, is itself a dummy variable equaling either zero or one. It equals one when both $Male_i$ and $Administrator_i$ are equal to one (i.e., when the worker is a male administrator), and otherwise it equals zero. To create the new interaction variable in STATA, which we'll call $Male\_Administrator_i$, you again use the "generate" command, or either "gen" or "g" for short:

"g male_administrator = male*administrator".

It's hard to imagine what could possibly go wrong with such a simple command. Nonetheless, it's always good practice to inspect the descriptive statistics for each variable that you create or modify. So let's do that, both for the interaction term and for both of the original variables of which it's comprised. You've already checked the latter two variables, but it's helpful to see their descriptive statistics alongside those of the new variable. Table 7.12 displays results from the STATA command

"sum male administrator male_administrator".

As expected, the new variable has 61 observations, a minimum of zero, and a maximum of one. Moreover, it has a mean of 0.0819672, which seems reasonable. After all, only 9.84% of the sample are administrators, so the percentage of *male*

Table 7.12 Descriptive statistics for three variables

Variable	Obs	Mean	Std. Dev.	Min	Max
male	61	.6721311	.4733326	0	1
administra~r	61	.0983607	.3002731	0	1
male_admin~r	61	.0819672	.2765913	0	1

Table 7.13 Regression output using extra pay as dependent variable

Source	SS	df	MS
Model	1.5846e+09	7	226373064
Residual	6.0801e+09	51	119218500
Total	7.6648e+09	58	132150947

Number of obs =	59
F(7, 51) =	1.90
Prob > F =	0.0890
R-squared =	0.2067
Adj R-squared =	0.0979
Root MSE =	10919

extrapay	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	1747.694	3369.576	0.52	0.606	-5017.011	8512.4
male_administrator	-28183.64	12706.44	-2.22	0.031	-53692.89	-2674.399
associate	-71.8307	4504.016	-0.02	0.987	-9114.019	8970.358
professor	259.0073	3765.134	0.07	0.945	-7299.815	7817.829
csula	3780.482	3527.956	1.07	0.289	-3302.185	10863.15
csuf	2001.838	3648.494	0.55	0.586	-5322.819	9326.496
administrator	34479.52	11522.47	2.99	0.004	11347.19	57611.85
_cons	2237.992	3861.188	0.58	0.565	-5513.668	9989.651

administrators must be even smaller, which (at 8.20%) it is. Looking at the means for the three variables, you might be tempted to conclude that because 9.84% of the sample are administrators, and 67.21% of the sample is male, then the percentage of male administers should be 9.84% Â 67.21%. But that equals 0.066, which is considerably below the actual mean (0.082) of the interaction term. This is because the mean of a product is *not* generally the product of the means!

Incidentally, multiplying the means by the sample size of 61, you see that there are 41 males, six administrators, and five male administrators. So of the 55 non-administrators, 19 are female, or nearly 35%, whereas of the six administrators, only one is female, or less than 17%. This pattern, whereby the representation of women diminishes at higher levels of organizations, is sometimes called a "glass ceiling".

Let's get back to the task at hand, which is adding the interaction term to the "extra pay" regression. Table 7.13 displays results from the STATA command

"reg extrapay male male_administrator associate professor csula csuf administrator".

As explained in the appendix to Chapter 6, the presence of the interaction term in the regression allows the relationship between $Male_i$ and $ExtraPay_i$ to differ between administrators and non-administrators. Holding the other right-hand side variables constant, the relationship between $Male_i$ and $ExtraPay_i$ is given by

$$\$1747.69 + (-\$28{,}183.64\ \hat{A}\ \mathit{Administrator}_i).$$

Because of the minus sign preceding the \$28,183.64, which comes from the regression output, the preceding relationship is

$$\$1747.69 - (\$28{,}183.64\ \hat{A}\ \mathit{Administrator}_i).$$

Remember that $\mathit{Administrator}_i$ must be either one or zero. If it is one, then the preceding relationship is

$$\$1747.69 - (\$28{,}183.64\ \hat{A}\ 1), \text{ or } -\$26{,}435.95.$$

If it is zero, then the preceding relationship is

$$\$1747.69 - (\$28{,}183.64\ \hat{A}\ 0), \text{ or } \$1747.69.$$

So the relationship between gender and one's annual extra pay critically depends on whether one is an administrator! For administrators, being male is associated with considerably *less* extra pay than being female. For non-administrators, there's a small gender difference (of less than \$2000 per year) favoring males, but it's estimated so imprecisely that from a statistical standpoint there's no gender difference.

The result that, among administrators, males get considerably less extra pay than females is a bit surprising and warrants further investigation. Remember that $\mathit{ExtraPay}_i$ is defined as the difference between two variables, i.e., total salary and base salary. Let's consider those two variables separately, as dependent variables, including the $\mathit{Male}_i\ \hat{A}\ \mathit{Administrator}_i$ interaction variable on the right-hand sides of those regressions. Starting with total salary as the dependent variable, Table 7.14 displays the output from the STATA command

"reg totalsal2016_clean male male_administrator associate professor csula csuf administrator".

Table 7.14 Regression output using (cleaned) total salary as dependent variable

```
      Source |       SS       df       MS              Number of obs =      59
-------------+------------------------------           F(  7,    51) =    7.83
       Model |  1.0432e+10     7  1.4903e+09           Prob > F      =  0.0000
    Residual |  9.7050e+09    51   190294425           R-squared     =  0.5181
-------------+------------------------------           Adj R-squared =  0.4519
       Total |  2.0137e+10    58   347197275           Root MSE      =   13795
------------------------------------------------------------------------------
totalsal2016_clean |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
              male |   5486.866   4257.125     1.29   0.203     -3059.67     14033.4
male_administrator |   3466.654   16053.33     0.22   0.830    -28761.75    35695.06
         associate |  -674.8223   5690.378    -0.12   0.906    -12098.73    10749.09
         professor |   19555.14   4756.874     4.11   0.000     10005.31    29104.96
             csula |  -2114.691   4457.223    -0.47   0.637    -11062.94    6833.557
              csuf |   741.3188   4609.511     0.16   0.873    -8512.661    9995.299
     administrator |   15264.44    14557.5     1.05   0.299    -13960.96    44489.85
             _cons |   81351.11   4878.229    16.68   0.000     71557.66    91144.56
------------------------------------------------------------------------------------
```

Table 7.15 Regression output using (cleaned) base salary as dependent variable

Source	SS	df	MS
Model	8.3374e+09	7	1.1911e+09
Residual	1.1063e+10	51	216917744
Total	1.9400e+10	58	334485964

Number of obs	= 59
F(7, 51)	= 5.49
Prob > F	= 0.0001
R-squared	= 0.4298
Adj R-squared	= 0.3515
Root MSE	= 14728

base2016_clean	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
male	3739.171	4545.179	0.82	0.415	-5385.655	12864
male_administrator	31650.3	17139.56	1.85	0.071	-2758.802	66059.4
associate	-602.9916	6075.41	-0.10	0.921	-12799.89	11593.9
professor	19296.13	5078.742	3.80	0.000	9100.13	29492.13
csula	-5895.173	4758.815	-1.24	0.221	-15448.89	3658.547
csuf	-1260.519	4921.408	-0.26	0.799	-11140.66	8619.619
administrator	-19215.08	15542.52	-1.24	0.222	-50417.98	11987.83
_cons	79113.12	5208.309	15.19	0.000	68657	89569.23

The coefficient on $Male_i$ Â $Administrator_i$ is positive with a high standard error, so clearly there is no evidence that, among administrators, males fare worse in terms of total salary. But you just learned that, among administrators, males fare considerably worse in terms of extra pay. If they fare worse in terms of extra pay but no worse in terms of total salary, we would expect that they fare better than female administrators in terms of base pay. That conclusion is confirmed in Table 7.15, which reports regression output from the STATA command

"reg base2016_clean male male_administrator associate professor csula csuf administrator".

The coefficient on the interaction term is \$31,650.30, with a standard error small enough to be taken seriously. Holding the other right-hand side variables constant, the relationship between $Male_i$ and base pay is given by

$$\$3739.17 + (\$31{,}650.30 \text{ Â } Administrator_i).$$

For administrators (i.e., $Administrator_i = 1$) this relationship is \$35,389.47, and for non-administrators (i.e., $Administrator_i = 0$) it is only \$3739.17.

So what's the bottom line? The offset isn't perfect, but let's suppose hypothetically that it were, meaning that every additional dollar of *base pay* received by men is exactly counterbalanced by an additional dollar of *extra pay* received by women. In that case, you might think that the difference doesn't matter. After all, a dollar of pay is a dollar of pay, whether we call it "base salary" or "extra pay", right?

Wrong! There are three reasons why receiving a dollar of base pay is more valuable than receiving a dollar of extra pay. First, increases in base pay are permanent and cumulative ... once you get a raise, you rarely lose it, whereas extra pay may vary from year to year, and getting it one year is no guarantee that it will continue in the future. Second, extra pay is often associated with extra work. For example, if a professor chooses to teach an additional course, beyond the required teaching load, that "overload teaching" is compensated with extra salary that semester. But it involves a workload increase, whereas increases in base salary generally do not. Third, as we will discuss in Chapter 11, the CSU's defined-benefit pension formula is tied to base salary, not to extra salary or total salary, so an additional dollar of base pay has a multiplier effect in that it raises the future value of the pension, whereas that multiplier effect is absent for an additional dollar of extra pay.

So the bottom line is that your statistical analysis reveals a rosier compensation picture for males than for females in chemistry and biochemistry on the three campuses, and the difference is driven by administrative positions. There are various holes that one could poke in your analysis, but exploring those would take us too far afield. My objective here was not to provide an airtight analysis of gender differences in pay in a particular discipline in the CSU system. It was simply to illustrate the mechanics of regression analysis in a concrete context, and to convince you that deeper study of this material is well worth your time. Nonetheless, despite its limitations, even the simple analysis you have just done is informative and way better than no analysis at all.

7.2 Exploratory Data Mining, Causality, and Experiments

In the preceding analysis, we knew which variables to study and where we were interested in looking for correlations. We started with a specific research question concerning gender differences in pay in a particular academic discipline within the CSU system, and that question steered our selection of variables and how they were organized in a statistical model. We were most interested in the coefficient of $Male_i$, and when the model included an interaction term, i.e., $Male_i$ Â $Administrator_i$, we were also interested in that variable's coefficient. But sometimes when approaching a compensation data set we do not have a clear idea of how the statistical model should look or even what variables it should contain. In such cases, we may be interested in exploring the data in a search for interesting patterns.

Extracting interesting patterns can be difficult, particularly in large data sets, i.e., so-called *big data*. Technological advances in hardware and software in recent years have facilitated such exploration of big data. How big are "big data"? There's no magical sample size beyond which mere "data" become "big data". The analysis in the preceding section obviously doesn't qualify. On the other hand, many of the issues that arose in that analysis also apply to big-data analysis.

Exploratory data analysis poses various challenges. The first and most obvious I've already mentioned, i.e., you don't know exactly what you're looking for.

Second, particularly if you're analyzing big data, you might need special programming techniques and software to cope with the high volume and heavy computational costs. Third, even if interesting patterns are found, it may be hard to interpret them. *Why* does the interesting pattern occur? What's behind it? Even if you discover an interesting pattern in the data, you wouldn't necessarily want to make business decisions based on it. Finding interesting correlations among variables doesn't shed any light on what caused the correlations. Good business decisions should account for how employer and worker behaviors are likely to *change* in response to a stimulus (e.g., recall the incentive and sorting effect from section 1.8), such as a change in a company's compensation system. But despite these concerns, data mining can still help uncover patterns that would otherwise have been missed and that merit further investigation.

Establishing *causality* when using compensation data is often extremely difficult. Suppose that you're interested in knowing how a particular change in the compensation system will affect worker productivity. One way to establish causality is to conduct experiments within the company. The basic idea mimics the approach often used in medical research. Suppose that a team of medical researchers claims to have developed a pill that can cure the common cold. How can its efficacy be tested?

A first reaction might be to assemble a sample of people who began suffering from a cold within the last 24 hours, give them all the pill, and check back with them a number of days later to see if they're cured. There are several problems with that approach. First, how many days is "a number of days later"? That number has to be chosen carefully. If the test is done after 15 days, then mostly likely everyone in the sample will have recovered whether the pill works or not, because most colds go away on their own within two weeks. But if the test is done after only two days, then it's possible that everyone will be found to still be sick even if the pill works. For example, the pill might reduce the average length of a cold from ten days to five, but checking the subjects after only two days would be too soon to detect this.

One way to improve the preceding experiment would be to check on the subjects every day for 15 days (or ask them all to keep a daily diary recording their symptoms). Then, after 15 days, when presumably everyone has recovered, we will know exactly how long the cold lasted for each person, and we can compare those cold durations to the "average length of a cold" from the medical literature. If the average length of the cold for the experimental subjects is shorter than the average length from the medical literature, it might seem that the pill works.

Another problem with the experiment is that the "average length of a cold" from the medical literature isn't an ideal benchmark for comparison. The reason is that the characteristics of the experimental subjects may differ, in systematic ways, from those on which the medical literature is based. In fact, we already know of one very important way in which the two groups differ . . . the subjects in the experimental group received the pill, whereas those on which the medical literature is based did not. This is important, because there may be a "placebo effect" in which even a "fake pill", or placebo, produces positive results.

That is, if patients aren't given the actual medicinal pill but rather an identical-looking pill made of some neutral substance (like gelatin), and if the patients recover from the cold just as quickly as if they'd taken the medicinal pill, then it can't be the medicine that's curing the cold. The only way to rule out such a placebo effect is to divide the experiment's participants into two groups . . . randomly assigning some subjects to the "experimental group" and others to the "control group". The people in the experimental group are given the actual medicinal pill, the people in the control group are given the placebo, and no one in either group knows which group they're in. Then if the people in the experimental group experience faster recoveries than those in the control group, it can safely be concluded that the medicine actually works and that the positive result is not just a placebo effect.

These concepts of experimental and control groups are very important, and it's particularly important that the control group be chosen so that its members are virtually identical to those in the experimental group in all ways *except* for the fact that they receive the treatment. Randomly assigning participants to the experimental and control groups should accomplish this "apples-to-apples" situation. So the bottom line is that the experiment, as I first described it, could be significantly improved by randomly assigning subjects into experimental and control groups.

There are also other ways in which the experiment could be improved. We can randomly assign subjects to experimental and control groups, as just discussed, but before that even happens let's back up a step and ask if the original sample was drawn in an ideal manner. What we did was to "assemble a sample of people who began suffering from a cold within the last 24 hours". The problem with that approach is that sickly people will be overrepresented. To see the problem, suppose that there are two types of people . . . "typical" people and "sickly" people. A typical person has about three colds per year, each lasting about a week, so those people are sick about 21 days per year. A sickly person has twice as many colds per year, and each cold lasts for two weeks, so those people are sick about 84 days per year. Now, if we ask who came down with a cold "within the last 24 hours" then sickly people are far more likely to meet that criterion than typical people, so the sample will end up being heavily laden with sickly people.

The problem with this is that the pill might have different efficacies for sickly people versus typical ones. Perhaps it doesn't help sickly people at all but it helps typical people a lot. Then a sample that is heavily laden with sickly people will give a misleadingly bleak picture of the pill's efficacy. The problem arises because, by focusing on people who came down with a cold during the last 24 hours, we have sampled *colds* rather than *people*. An alternative approach is to take a random sample of *people* and to follow each person until (and throughout) their next cold.

Such an experiment, though superior, is more expensive to conduct, because it will take a lot longer to complete . . . the original experiment that sampled colds could be wrapped up in a couple of weeks, whereas if people are sampled, time must be spent waiting for them to get sick with a cold. For some super healthy people, that might involve a lot of waiting. But the point is that if the population is comprised of, say, 10% super healthy people, 70% "typical" people, and 20% sickly people, then if we randomly sample people and track them until they catch a cold, about 10% of our sample is likely to be super healthy, about 70% is likely to be typical, and about 20% is likely to be sickly, so we will not be oversampling sickly people as happened when we sampled colds rather than people.

This is probably more detail than you wanted to know about medical experiments. But there are some important lessons here that carry over directly to experimental research on compensation in organizations. The most important lesson, which I hope by now is clear, is that you have to think very carefully about how to design any experiment. That means drawing a sample that is representative of the population that you want to study, and it means randomly assigning members of the sample to experimental and control groups. The cost of running the experiment is also a consideration. In the preceding example, sampling people is preferable to sampling colds, but sampling colds has the advantage of being a cheaper and faster experiment. Evidence from a less-than-ideal experiment may be better than no evidence at all, but when interpreting the results of the experiment it's important to understand how the results are likely to be affected by the limitations inherent in the design of the experiment.

How do these concepts apply to experiments involving compensation in organizations? I provide one example at the start of Chapter 9, which investigates worker productivity (in planting trees) on both fixed wages and piece rates. As another example, suppose that you're a manager in the sales industry, and a significant component of your employees' compensation is a 10% commission on individual sales. Your boss asks you to make a projection about what would happen to individual productivity if the commission rate were to double to 20%. Doubling the commission rate could be a great idea or a terrible one, so before implementing such a policy company-wide, it would be nice to get an idea about whether it's great or terrible. Following the medical discussion, you could begin by randomly assigning the company's employees (i.e., those who are currently getting a 10% commission rate) to experimental and control groups. Then, after some time, you can take individual productivity measurements to identify the productivity change in the experimental group (i.e., those getting 20%) versus the control group (i.e., those still getting 10%).

One practical problem you're likely to encounter if you run the experiment this way is that members of the experimental and control groups who work in close vicinity (or who are in communication) will discover that some people are getting double the commission rate that others are getting for doing the same work. This may upset those getting 10%, and it may make those getting 20% feel awkward and uncomfortable. Morale problems associated with pay equity can arise, as we've discussed in sections 1.10 and 5.6. For that reason, it might be wise to separate the treatment and control groups to the greatest extent possible. One way might be to select a particular establishment (in a multi-establishment firm), try out the change to the compensation system discreetly, and then implement the change more broadly within the organization if it turns out successfully.

Another thing to be aware of when running experiments in organizations is the so-called *Hawthorne effect*, which means that experimental subjects behave differently than they otherwise would, simply because they know that they are being observed in an experiment. For example, if workers are aware that you are testing out a piece-rate compensation system to gauge its effects on productivity, they might work harder than they would if the piece rate were permanently installed, simply because they know they are being watched in an experiment and that you're focusing on their productivity.

Hawthorne effects can lead to misleading conclusions. For example, a big positive productivity effect from introducing piece-rate pay might lead you to adopt such pay company wide, but if the productivity bump was due entirely to a Hawthorne effect, then you will be disappointed with the results once the plan goes into effect company-wide. If workers are unaware that they are participating in an experiment, then Hawthorne effects do not arise. Ethical issues arise with conducting experiments on people without their knowledge and consent, which is why any type of "human subjects research" in academia requires pre-approval by an institutional review board.

7.3 Lessons for Managers

Experience and good intuition are key to success as a manager. But even successful, experienced managers can sometimes be led astray by their intuition, and dispassionate data analysis is a powerful tool for avoiding such mistakes. Careful data analysis can confirm, support, and sometimes alter or even reverse your intuition. Managerial intuition and solid data analysis are complements, not substitutes. If a highly experienced and successful manager has strong intuition concerning a particular business problem, then the bar is going to be quite high for that intuition to be reversed by data analysis, though any good and intellectually curious manager should take such data analysis seriously and try to understand why its conclusions are counterintuitive. Similarly, the more convincing and airtight a data analysis is, the higher the bar is for its conclusions to be reversed by appealing to the intuition of an experienced manager.

Use your organization as an experimental laboratory for generating your own data and testing out various changes to the compensation system. Careful thought must go into the design of such experiments, particularly in selecting the treatment and control groups.

Dramatic increases in the availability of business data in recent years, and advances in hardware and software, have increased the importance of the role that compensation analytics plays in business decisions. Overall, that's a positive development. But don't fall into the trap of thinking that a "data-driven" decision is necessarily a good decision. Plenty of terrible business decisions are made on the basis of flawed or incomplete data analysis. Sometimes shoddy analysis is obscured behind a thick veil of complex, sophisticated statistical methods that are beyond the understanding of most managers. After all, it's hard to question and criticize something that you don't fully understand, and there's often a tendency to think that if a methodology is complex enough, then the people executing it "must be smart" and know what they're doing.

Such thinking creates the conditions for big mistakes. As a manager, don't allow yourself to be intimidated and browbeaten by a sophisticated statistical analysis. Ask basic questions, even those that reveal your ignorance. Consult with those you trust whose backgrounds in quantitative analysis exceed yours. Learn enough about statistical methods on your own so that you become more independent in your assessments and less dependent upon consultations with others.

Even the most careful and competent quantitative analysis will fail to influence business decisions if it is poorly presented. Whether you are writing a formal report, an email to your boss, or a PowerPoint presentation, you should invest effort in communicating the results effectively so that they are persuasive and easily understood. Remember the following "4 Cs" of communicating the results of compensation analytics (not to be confused with the "3 Cs" of compensation constraints from Chapters 4 and 5!); your presentation must be:

Correct, Clear, Concise, Captivating.

Correctness goes without saying. Clarity is important because your results must be understood to be evaluated and, hopefully, to influence business decisions. Concision is important because most managerial consumers of quantitative analysis are busy, impatient, and don't have time to pore over reams of numerical output. Captivating the audience is essential for obvious reasons; you need to excite the consumers of your analysis and convince them that it is important and should affect business policy. When preparing a presentation and when communicating your results, know your audience well . . . in most business settings there is a lot of variation in the quantitative backgrounds of the managers, just like the readers of this book vary widely in their knowledge of statistics, as I mentioned at the start of Chapter 6.

Speaking of Chapter 6, I recommend that you now reread it. Chapter 7 gave you a detailed, concrete example that illustrates the concepts and methods presented in Chapter 6. Now that you have seen that example and have a feel for how these methods work in practice, you will get more out of Chapter 6 than you extracted on your first reading. And don't forget the appendix!

Case Discussion 8: Wage–Insurance Tradeoff (Part B)[1]

This is the second part of a two-part case. Part A provides background and should be read first (see Case Discussion 7).

Your client is a multi-establishment firm but competes in the same market with several single-establishment firms, so analysis for both types of firms is of interest. Recall from Part A that the client firm was originally interested in three questions:

(1) *Can local firms that offer health insurance get away with offering lower wages?*
(2) *If so, by how much?*
(3) *How do these answers depend on whether or not firm size is held constant, if at all?*

Now the client would also like to understand how the answers to those questions differ, if at all, between multi-establishment firms and single-establishment firms.

In Part A of the case, you estimated Regressions (1) and (2) for the full sample, and those results are reproduced below in the top panel of Table 1.16. The middle panel displays new results from estimating Regressions (1) and (2) on the sample of 446 multi-establishment firms, and the bottom panel displays new results for the 438 single-establishment firms. As in Table 6.4 from Part A, for conciseness, Table 7.16 does not show the regression estimates for the constant term and other control variables even though they are included in every estimated model.

Questions

1. Focusing only on the results for *multi-establishment firms*, in Regression (1), what is your estimate of β_1, and how precisely estimated is it? Carefully explain the interpretation of that estimate. Do the same for your estimates of β_1 and β_2 in Regression (2).

[1] Note: This case is based on "The Elusive Wage-Benefit Tradeoff: The Case of Employer-Provided Health Insurance", *International Journal of Industrial Organization*, 37, 2014, pages 23–37, by Jed DeVaro and Nan L. Maxwell.

Table 7.16 Regression results

Dependent variable = Wage			
Full sample	Regression (1)		Regression (2)
Health insurance	0.824 (0.397)		0.809 (0.401)
Firm size	–		0.086 (0.345)
Sample size		n = 884	
Multi-establishment firms			
Health insurance	-2.008 (0.701)		-2.381 (0.707)
Firm size	–		2.131 (0.577)
Sample size		n = 446	
Single-establishment firms			
Health insurance	1.755 (0.492)		1.921 (0.490)
Firm size	–		-1.547 (0.488)
Sample size		n = 438	

2. Focusing only on the results for *multi-establishment firms*, what are the answers to the three specific questions that your client firm posed? Taken at face value, are the regression results for multi-establishment firms consistent with what the theory of compensating differentials would predict concerning the relationship between wages and health insurance?
3. Focusing only on the results for *single-establishment firms*, in Regression (1), what is your estimate of β_1, and how precisely estimated is it? Carefully explain the interpretation of that estimate. Do the same for your estimates of β_1 and β_2 in Regression (2).
4. Focusing only on the results for *single-establishment firms*, what are the answers to the three specific questions that your client firm posed? Taken at face value, are the regression results for single-establishment firms consistent with what the theory of compensating differentials would predict concerning the relationship between wages and health insurance?
5. What do you think explains the difference between multi-establishment firms and single-establishment firms? Why might the results differ as they do between the two subsamples?
6. Rather than comparing multi-establishment and single-establishment firms by estimating Regression 2 separately for the two subsamples of firms, suppose instead that you estimate a version of Regression 2 (that includes an interaction term) only once, using the full sample. Write down the regression equation you would estimate, and explain how you would use the parameter estimates to measure the relationship between wages and health insurance.

Further Reading

Agarwal, D., Bersin, J., Lahiri, G., Schwartz, J., and Volini, E. (2018). "People Data: How Far is Too Far? 2018 Global Human Capital Trends." *Deloitte Insights*. March 28, 2018.

Aral, S., Brynjolfsson, E., and Wu, L. (2012). "Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology." *Management Science*, 58.

Davenport, T. H. and Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*, Boston: Harvard Business School Press.

Davenport, T. H., Harris, J., and Shapiro, J. (2010). "Competing on Talent Analytics." *Harvard Business Review*, 2–6.

Fitz-Enz, J. (2010). *The New HR Analytics: Predicting the Economic Value of Your Company's Human Capital Investments*, New York: American Management Association.

Murray, M. P. (2006). *Econometrics: A Modern Introduction*, Addison-Wesley.

Pemmaraju, S. (2007). "Converting HR Data to Business Intelligence." *Employment Relations*, 34, 13–16.

Verbeek, M. (2017). *A Guide to Modern Econometrics*, 5th ed., Wiley.

8 Training

Seattle Genetics – based in Bothell, Washington – is a biotechnology company with around 800 employees. Its focus is developing cancer treatments based on antibody-based therapies. On its website, within a list of offered employee benefits, Seattle Genetics advertises “tuition reimbursement for college courses” and “professional development opportunities”. The reimbursement for college courses is available to workers after only six months on the job, and it also reimburses employees for course-related materials such as textbooks. Such education and training offerings are not unusual fringe benefits, and we will discuss some other examples in Chapter 11. On the Seattle Genetics website, these training opportunities appear under a subheading called “Get Smarter”. Indeed, training makes your workers more knowledgeable and, therefore, more productive. But making your workers more knowledgeable and productive also makes them more marketable, and some of them might take that knowledge and go work for another firm, after you’ve invested resources in training them.

An example is Scott Cook, who in 1983 founded Intuit, a business and financial software company headquartered in Mountain View, California. Cook began his career at Procter & Gamble, which was known at the time for providing intensive training to its marketing experts, leading former employees to refer to it as the “Marines”. Cook was among those Procter & Gamble employees who was trained in product development, market research, and marketing. More recently, Cook has been quoted as saying that “Intuit would not be here today were it not for what I learned at Procter & Gamble.” An interesting twist to the story is that currently, as of 2019, Cook serves as a director at Procter & Gamble. But it might well have been even better for Procter & Gamble if Cook had built his entire career there.

8.1 What Is Training?

Training is an investment in your workers’ future productivity. It includes anything that you do, directly or indirectly, that augments your workers’ knowledge and skills in a way that is expected to result in a sustained increase in their future productivity. Like any investment, training involves paying a cost today to reap an expected benefit in the future. The cost can be either a direct monetary cost of providing the training, or an indirect cost in the form of reduced productivity during the training period (e.g., if your workers spend a week taking an off-site leadership course, then that means one week when they are not in the office working), or both. Training can be formal (e.g., courses, programs, and workshops) or informal (e.g., experienced employees mentoring and helping less experienced ones). It can be mandatory or optional.

Training is closely connected to compensation, for two related reasons. First, training increases *future* compensation by enhancing the “skills” leg of the three-legged compensation stool that we discussed in section 1.6. Second, training is a component of *current* compensation. To see why, recall that compensation includes everything that a worker likes about a job. And workers typically like training. Why? Precisely because of the first reason! Workers like anything that gives them reason to expect higher future compensation. By making workers more productive in the future, training increases the future compensation workers can expect to receive.

8.2 Portability of Training

When deciding whether to train your workers, how much to train them, and what types of training to provide, an important issue to consider is the *portability* of the training. That is, how transferrable and applicable would the training be if your trained workers were to move to other (competing) firms? Some types of training are highly portable, meaning that they equip workers with skills that are valued at a wide range of potential employers. Other types of training are more narrow and specialized and are valued by a smaller set of employers . . . in the extreme case, the training might be so unique that it provides value only within your company and is of no value to other employers.

Training that is highly portable is sometimes called *general training*, and the skills that it imparts are called *general skills*. Training that is not very portable is sometimes called *specific training*, and the skills that it imparts are called *specific skills*. Talking about specific skills requires stating the domain or sphere in which the skills are relevant. For example, *firm*-specific training imparts *firm*-specific skills, which are relevant only to the particular firm in question. So if you provide your workers with firm-specific training, you don’t have to worry that, like Scott Cook, they will take those skills and move to another firm, because another firm would not value those skills. Similarly, if you provide your workers with department-specific training, then the skills imparted are useful only within your department, and even if one of your trained workers moves to a new job within the same firm (but outside of your department) those skills will not be valued. In addition to firm-specific or department-specific, skills could be occupation-specific, industry-specific, region-specific, establishment-specific, team-specific, and so on. In all cases, the key issue is the extent to which the skills are portable. True specificity of skills means that the skills are portable only within the domain in question, and not outside it.

Even when particular skills are largely general, when they are packaged together into a portfolio, the portfolio as a whole might be largely firm-specific. That is, even if multiple firms are all interested in hiring workers with the same general skills, the firms may differ in the particular ways in which those individual general skills are used on the job. So even if the skills are individually general, the "mix" of those skills, i.e., the proportions in which they are blended together in a given job, might be firm-specific. Again, the greater the degree of specificity (or, equivalently, the lesser the generality) the less portable is the portfolio.

Don't confuse *specificity* of training with *uniqueness* of training. To elaborate, suppose that your firm does certain activities in a unique way that differs from what all other firms are doing. Even though no other firms are doing things that way, that doesn't mean that they wouldn't like to, particularly if what you're doing uniquely is profitable. For example, in section 15.7 we will discuss Kamdesh Afghan Kabab House, and the challenges faced by that small business concerning turnover of its cooks. In the restaurant industry, recipes are important trade secrets, and it is difficult to hide them from the people (i.e., the cooks) who execute them on the job. When those cooks are poached by competing restaurants, you run the risk that your secret recipes also end up "in enemy hands". Even if your successful recipes are unique to your firm, that doesn't mean that they will remain so. Specificity, as opposed to uniqueness, refers to portability . . . it means that the training is only of value within the sphere to which the specificity pertains (e.g., a firm, a department, an industry).

8.3 Who Pays for Training?

Training is an investment in workers' future productivity. Like all investments, training involves paying something today in exchange for some expected rewards in the future. Who actually pays? Is it you (the manager) or the workers you're training? In most cases, the party who is expected to reap the future rewards is the party who bears the upfront training cost. Which party is expected to reap the future rewards of the training? The answer to that question hinges on the portability of the training. The more portable the training is (i.e., the more general are the skills), the more the training cost will be borne by the workers. As a manager, you should be hesitant to pay for your workers' general training because, once they're trained, your workers are more marketable elsewhere and may quit (in which case you incurred a training cost without reaping any future reward) or demand a large raise to prevent them from quitting (in which case you incurred a training cost without reaping any future reward, because the worker's future increase in productivity is cancelled out by the higher compensation that you must pay to retain them).

There's a similarity here to our wage-theft discussion from Chapter 2. I mentioned there that it's rare for new hires to show up on their first day of work to find a paycheck waiting for them on their desk. The reason employers don't like to make compensation payments upfront is because of a fear that workers might "take the money and run". Similarly, in the present context of training, the reason why employers don't like to incur (general) training costs upfront is because of a fear that their workers, once trained, will "take their higher productivity [which translates into higher pay] and run to a new firm". The problem in the wage-theft case arose because the work and the pay weren't simultaneously exchanged, so that the party who put something on the table first was at risk that the other party would take it and run. One solution to the wage-theft problem is to move closer to simultaneous exchange, and if the exchange is perfectly simultaneous, then the wage-theft problem disappears.

By the same logic, in the training context, anything that can bring the situation closer to simultaneous exchange should make the employer less fearful of paying for the worker's general training. For example, sometimes workers have to sign a contract that requires them to pay their employer back for the general training costs if they quit within a specified time period after the training ends. Such a contract prevents the worker from "taking the training and running", just like, in the case of true simultaneous exchange, the worker is unable to "take the training and run". Replicating, or even approximating, simultaneous exchange, is usually not easy in a training context ... even contracts like I just mentioned don't always solve the problem, because they can be costly to enforce. So in many cases workers end up paying for their general training.

Be aware that when workers "pay" for their general training, the payment may be, and typically is, indirect rather than direct. Indirect payments occur in the form of lower monetary compensation than the workers would otherwise receive. Because workers know that general training will enhance their future productivity (and, therefore, their future compensation) they are willing to accept lower compensation in the short-term, thereby implicitly paying for their training. Recalling the discussion of Chapter 3, training is a job characteristic that workers value and that, therefore, gives rise to a compensating differential; jobs that offer training can be expected to offer lower compensation than *otherwise identical* jobs (to ensure an "apple-to-apples" comparison) that offer no training.

Who pays for the training if it is firm-specific? In that case, both the employer and the worker benefit from the future productivity gains arising from training. So they both can be expected to share the training cost. "Share" does not necessarily mean a 50/50 split. The relative costs that both parties bear will depend on their relative bargaining power and how well they negotiate (see Chapter 14).

The Google case discussion at the end of the chapter reveals an interesting strategy of a company providing training that is quite general (i.e., an online IT certificate program) at a very low price that is heavily subsidized by the employer, even though most individuals who complete the training will go on to join other companies. Grappling with the case allows you to explore the ways in which this strategy may be profitable to Google.

8.4 Should You Train Your Workers?

Deciding whether or not you should train your workers involves comparing the *current* cost of training to the *future* benefits of training, which are defined in terms of enhanced worker productivity. Typically the costs are incurred upfront, and all at once, whereas the benefits are spread out over a span of time in the future. The problem with costs (or payments) occurring at different points in time is that an apples-to-oranges situation arises. As in many other situations throughout this book, what we need are "apples-to-apples" comparisons. An apples-to-oranges problem exists for two reasons. The first is inflation. As we saw in Chapter 1's appendix, in an inflationary environment, the same nominal monetary amount (whether a benefit or a cost) is worth less in real terms if it occurs in the future rather than today. To simplify our discussion, I'm going to assume for the rest of the chapter that the inflation rate is zero, so that there's no difference between nominal and real values, i.e., all dollar figures can be interpreted in real terms. This will allow us to focus all of our attention on the second reason for the apples-to-oranges problem . . .

Whenever costs or payments accrue over time, it is necessary to properly *discount* them to produce a correct investment decision. The method for accomplishing that will also be applied later in the book in other contexts (e.g., Chapter 12). Costs (or payments) that occur in the future – even in real, constant dollars – are less important than the same costs (or payments) that occur in the present. And the further into the future they occur, the less important they become. So when you have to incur costs, it is always best to defer those costs to the future, whereas if you're receiving benefits, it's best to receive them as soon as possible.

The reason for this is quite intuitive. If you're going to receive a payment of \$100,000, you're better off receiving it now than one year from now. If you get the

Table 8.1 Discounted value of costs incurred for next 5 years

Future year (t =)	Actual real payment (or cost)	Discounted real payment (or cost)
0 (i.e., the present)	\$100,000	$\$100{,}000/(1.05)^0 = \$100{,}000$
1	\$100,000	$\$100{,}000/(1.05)^1 \approx \$95{,}238$
2	\$100,000	$\$100{,}000/(1.05)^2 \approx \$90{,}703$
3	\$100,000	$\$100{,}000/(1.05)^3 \approx \$86{,}384$
4	\$100,000	$\$100{,}000/(1.05)^4 \approx \$82{,}270$
5	\$100,000	$\$100{,}000/(1.05)^5 \approx \$78{,}353$

payment a year from now, all you get is \$100,000. But if you get it now, you could invest it somewhere for a year (say, at 5% interest), and then in a year you'd have \$105,000. Similarly, if you have to incur a cost of \$100,000 it's better to incur the cost a year from now than right away. If you incur it right away, it costs you a full \$100,000. But if you incur it a year from now, it costs you only about \$95,238, assuming again a 5% annual interest rate. Why? Because if you were to invest just \$95,238 today, in a year it would be worth 1.05 × \$95,238, which is about \$100,000 . . . just enough to pay your bill. So to pay off a bill of \$100,000 a year from now, all you need is \$95,238 today, whereas to pay off a bill of \$100,000 today, you need a full \$100,000 today.

The rule for discounting is simple. You simply divide a payment (or cost) by $(1+r)^t$, where r is the one-period discount rate and the exponent, t, is the number of periods in the future. In the preceding examples, the periods are years, and the annual discount rate is 5%, i.e., $r = 0.05$. Because we considered payments or costs one year into the future, t was 1, and we compared those numbers to payments or costs that are incurred today, i.e., $t = 0$. Notice that when $t = 0$, the term $(1+r)^t$ becomes $(1+r)^0$, which is just 1. And if you divide any number by 1, obviously the number remains unchanged. Thus, payments or costs that happen *today* are not discounted. Notice also that when t becomes larger (i.e., we consider years that are further into the future) the discounting gets heavier and heavier, so that the payments or costs shrink considerably. Table 8.1 gives the discounted values of a \$100,000 cost that is incurred today and in each of the next five years, again assuming a 5% annual interest rate.

The numbers are rounded slightly, so I use the "≈" symbol rather than the "=" symbol to remind you that they are not exact. If we were to extend the table with many more rows, the discounted amounts would become smaller and smaller. After 30 years, the discounted payment would be about \$24,295, and after 300 years it would be just shy of five cents. As you can see, if you go far enough into the future the discounted values become essentially zero. You can easily verify these statements using EXCEL or a financial or scientific calculator, and you should do so if you're not already experienced with such calculations.

How exactly do you interpret a number like \$82,270, as appears in the preceding table when $t = 4$? That number says that a payment of \$100,000 four years from now is equivalent to a payment of \$82,270 today, because if you were to invest \$82,270 today at an annual interest rate of 5%, you'd end up with about \$100,000 in four years.

What does all of this have to do with training? Most of the time the training cost is an upfront cost that is paid today, when $t = 0$, so it is not discounted. But the future benefits (in terms of enhanced worker productivity) accrue over a number of years. So you must compare the present cost to the present discounted value of the future benefits to reach the correct conclusion about whether or not to train. Let's consider an example.

Example 8.1

Suppose that it would cost \$45,000 to train your worker today, and, if you do so, you anticipate that she will generate \$12,000 extra in sales revenue per year, starting next year. To keep things simple, assume that her compensation will remain the same after training and that you expect her to remain with your company for five full years after the current year (i.e., the training year).

a. If the annual interest rate is 5%, should you train your worker?
b. If the annual interest rate is 12%, should you train your worker?
c. At approximately what annual interest rate would you be indifferent between training and not training your worker?

Answer:

a. *You need to compare the discounted value of the training cost to the discounted value of the training benefits. The costs are easy, because there's only a one-time cost of \$45,000 that you'd pay upfront in the present year. As we saw, $\$45{,}000/(1.05)^0$ is just \$45,000. In other words, there's no need to discount the \$45,000, and its discounted value is identical to its actual value.*

   The discounted benefit is the sum of the additional revenue over the five years, properly discounted. Next year (i.e., when $t = 1$), the discounted sales revenue is \$12,000/1.05. In years 2, 3, 4, and 5, the discounted sales revenues are $\$12{,}000/(1.05)^2$, $\$12{,}000/(1.05)^3$, $\$12{,}000/(1.05)^4$, and $\$12{,}000/(1.05)^5$, respectively. Adding those five numbers is approximately \$51,954. Because \$51,954 > \$45,000, it is profitable to train your worker.

b. *The discounted benefits over the five years following the training are:*

\$12,000/1.12, $\$12{,}000/(1.12)^2$, $\$12{,}000/(1.12)^3$, $\$12{,}000/(1.12)^4$, $\$12{,}000/(1.12)^5$.

Adding those five numbers is approximately \$43,257. Because \$43,257 < \$45,000, it is not profitable to train your worker.

c. *If you're indifferent, then the discounted values of costs and benefits must be equal. So you have to find the (approximate) value of the annual interest rate that would equate \$45,000 with*

$$\$12{,}000/(1+\rho) + \$12{,}000/(1+\rho)^2 + \$12{,}000/(1+\rho)^3 + \$12{,}000/(1+\rho)^4 + \$12{,}000/(1+\rho)^5.$$

You can infer that ρ must lie somewhere between 0.05 and 0.12, because when the interest rate was 5% you found (in part a) that training was profitable, whereas when the interest rate was 12% you found (in part b) that training was unprofitable, so the breakeven interest rate at which training is neither profitable nor unprofitable must lie somewhere in between those percentages.

You can also infer that ρ is closer to 12% than it is to 5%, because at 12% interest the discounted benefits (of \$43,257) were only slightly below the discounted costs (of \$45,000), whereas at a 5% interest there was a larger discrepancy between the discounted benefits and costs. A reasonable guess might be 10%. This is close but a bit too low, as it yields a discounted value of benefits of \$45,489.

A better approximation is 10.4248% (i.e., $\rho \approx 0.104248$), in which case the discounted benefits are about \$45,000.05, which is extremely close to \$45,000. You can use a financial calculator or EXCEL to approximate ρ with great accuracy, or you could even use trial and error as I just did, plugging in different values for ρ until you find one that makes the discounted value of training benefits approximately \$45,000.

The "breakeven" interest rate, ρ, from part (c) of Example 8.1, which was about 10.42%, is called the ***internal rate of return***. The concept is used for all sorts of investment decisions, not just training decisions. In the training context, whenever the actual real interest rate, r, exceeds the internal rate of return, ρ, it is unprofitable to train the worker, whereas whenever the actual interest rate, r, falls below ρ, training is profitable.

Figure 8.1 graphically depicts the training problem described in Example 8.1. The horizontal axis represents the actual annual real interest rate, r, that is available to your company on the money you invest. The vertical axis represents the present discounted value of training benefits and also the present discounted value of training costs, so that both graphs are displayed on the same set of axes. The graph for the ***present discounted value*** (PDV) of training costs is a horizontal line at \$45,000, because training costs are always \$45,000, regardless of the value of the

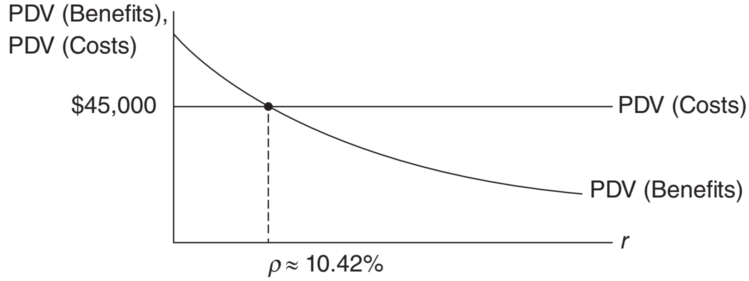


Figure 8.1 Internal rate of return, ρ, for a worker training problem.

interest rate, r. The reason that the training costs don't depend on the interest rate is that they're incurred entirely upfront, in "year 0". In contrast, the PDV of training benefits depends on the interest rate, r, and the higher the interest rate, the lower the PDV of training benefits. The intersection of the two PDV graphs reveals the internal rate of return, ρ, which part (c) of Example 8.1 revealed to be approximately 10.42%. Note that for interest rates, r, that are below ρ, the graph for the PDV of benefits lies above the graph for the PDV of costs, so training is profitable, whereas when r exceeds ρ, the reverse is true.

Changing Example 8.1 around a bit, and redoing it, will deepen your understanding of it. First, suppose that training costs increase from \$45,000 to \$55,000. Remember that the PDV of benefits was found to be \$51,954 in part (a). That number falls below the new training cost of \$55,000. Therefore, at an interest rate of $r = 0.05$, training is unprofitable. If training is unprofitable when the interest rate is 0.05, it will be even less profitable for higher values of r, which means in part (b) it definitely makes no sense to train. For part (c), the internal rate of return, ρ, must lie below 0.05. You know this because at $r = 0.05$ (and at all higher values of r) training is unprofitable, and any interest rates at which training is profitable must lie below ρ. To find the new (approximate) internal rate of return, identify a value of ρ that comes close to satisfying the following equation, which equates the PDV of training benefits to the PDV of training costs:

$$\$12,000/(1+\rho) + \$12,000/(1+\rho)^2 + \$12,000/(1+\rho)^3 + \$12,000/(1+\rho)^4 \\ + \$12,000/(1+\rho)^5 = \$55,000.$$

If ρ is 0.02972, the left-hand side of the preceding equation is approximately \$55,000.45, so the approximate internal rate of return is slightly less than 3%. The internal rate of return has decreased considerably from the value of about 10.42% in part (c) of Example 8.1, which was based on a training cost of \$45,000. This can also be seen graphically. If the horizontal line in Figure 8.1 that represents the PDV of training costs is shifted upwards vertically, to a vertical intercept of \$55,000, it will intersect the PDV of training benefits graph at an interest rate slightly less than 3%. This makes sense. Because the increase in training costs has made training less attractive to the employer than it originally was, the training should only be pursued for exceptionally attractive (i.e., low) interest rates.

To take an extreme case of the point I just made, suppose that the training costs increases to \$60,000. Then you would expect the internal rate of return to decrease further still. It turns out that when training costs are \$60,000 the internal rate of return is $\rho = 0$, meaning that it is never profitable to train, regardless of the interest rate. You can figure this out immediately, just by noticing that the "best case scenario" for training benefits is that you reap the full \$12,000 in additional sales revenue for each of the five post-training years. That scenario only happens if you face an interest rate of $r = 0$, so that there is no discounting, and future monetary amounts can be treated exactly the same way as current monetary amounts. But even in that ideal, best case scenario, your PDV of training benefits (i.e., \$60,000) exactly equals your PDV of training costs, so the best you can do is break even. And if the interest rate were to increase even an iota above zero, you'd be incurring a loss by training. So the internal rate of return, ρ, is zero, and it never makes sense to train the worker, regardless of the value of the interest rate, r.

The preceding discussion considered raising the training cost beyond the \$45,000 level covered in Example 8.1. We saw that increasing the training cost lowered the internal rate of return, and if the training cost reaches as high as \$60,000 (the amount of the total, undiscounted, training benefits over the five-year post-training period) the internal rate of return drops all the way to zero. Similarly, if we drop the training costs below \$45,000 the internal rate of return increases beyond the 10.42% approximation that was found in part (c) of Example 8.1. For example, if training costs drop to \$40,000, then the approximate internal rate of return is 0.15235, which yields a PDV of training benefits of about \$40,003.06, which is very close to the PDV of training costs. You should verify this for practice.

What happens if the training costs drop so low that training actually becomes free? Without doing any computations, the answers to parts (a) and (b) of Example 8.1 should be immediately obvious. Whether the interest rate is 5% or 12% doesn't matter; training is obviously a good idea because it increases future productivity at no current cost. You might expect that the internal rate of return would be enormous in this case, and that would be an understatement! Remember that the internal rate of return, ρ, is the interest rate that equates the PDVs of benefits and costs, so it would need to satisfy this equation:

$$\$12,000/(1+\rho) + \$12,000/(1+\rho)^2 + \$12,000/(1+\rho)^3 + \$12,000/(1+\rho)^4 + \$12,000/(1+\rho)^5 = \$0.$$

There's no actual solution to this equation, because the left-hand side is always positive, no matter what number you insert for ρ. The larger the value of ρ that you insert, the smaller the left-hand side becomes. So it approaches zero but no number you plug in can actually make it *equal* zero. The internal rate of return in this case is infinite! You can also see this graphically in Figure 8.1. The horizontal line that represents the PDV of training costs shifts down vertically from a vertical intercept of $45,000 to a vertical intercept of $0 (i.e., it coincides with the horizontal axis). But that graph will never be intersected by the graph for the PDV of training benefits, because the latter never actually reaches the horizontal axis, though it gets closer and closer to it as the interest rate, r, gets larger and larger. The bottom line, again, is that you'd always train the worker in this situation, regardless of the value of the interest rate, r.

In a standard training problem, you would train your worker if the interest rate lies below the internal rate of return, and you would not train your worker if the interest rate lies above it. If the interest rate coincides with the internal rate of return, you would be indifferent between training and not training. But this interpretation, in which you make one decision if the interest rate is below the internal rate of return, and another if it is above it, requires that the internal rate of return not be too extreme. If it is too low (i.e., if it is zero) then you would never train your worker, regardless of the interest rate. And if it is extremely high, then practically speaking you'd always train your worker.

8.5 Practical Applications

The preceding section lays out how to think about training investment decisions. You need four pieces of information to make an informed decision about whether to train your worker:

(1) per-period amount by which training is expected to increase your worker's post-training productivity
(2) training cost
(3) expected number of periods your worker will remain employed with you after being trained
(4) real interest rate, r.

In Example 8.1., I gave you all four pieces of information. The training cost was $45,000, the amount by which training was expected to increase productivity was $12,000 in each year following the training year, the expected number of periods your worker would remain after the training period was five, and the real interest rate was 5% (in part *a*). But how would all of this work in a real business setting, and where would you obtain these four key pieces of information?

Rarely can you expect to know any of the four pieces of information with complete confidence. In most situations you will have to make educated guesses based on all available information. In some cases you might be able to use tools from compensation analytics (see Chapters 6 and 7) to arrive at better guesses than you would have made in the absence of data analysis. The following four subsections consider each of the key pieces of information.

8.5.1 Post-Training Increases in Worker Productivity

It's difficult to quantify the amount by which training can be expected to increase a given worker's per-period productivity. It's a bit like predicting what grade a particular student will get in a statistics course that they're just about to start. Sometimes even a very good student underperforms and earns a lower course grade than expected, and other times even a historically weak student does surprisingly well and aces the course. To make an educated guess about a given student's performance, you might look at their current grade point average (GPA), which is a cumulative measure of their past academic performance. You might predict that a student with a 3.9 GPA will get an A in the course, whereas a student with a 2.1 GPA will get a C.

If you have more detailed information on their course histories, you might be able to make more accurate predictions. For example, suppose that the student with the 3.9 GPA has taken mostly art and music classes and has received As in all of them but has received B+s in the few math classes that were taken. Similarly, the student with the 2.1 GPA has received Cs and Ds in a number of art and history classes but has taken several math classes, receiving a B in all of them. Then, relying on the assumption that performance in prior math classes is the best predictor of success in a statistics class, you might predict that the first student will get a B+ and that the second student will get a B. Such predictions could easily be wrong, but they're likely to be the best you can do, and better than using overall GPAs with no attention to the courses taken.

The situation is likely to be even more challenging in a training context, because you're unlikely to have data on the workers' past training histories. Training is given most often to new hires, and your new hires probably haven't been around long enough to have generated much performance data. You might be able to use the tools from compensation analytics, however, to develop a forecast based on historical data from your other workers. Recalling Chapters 6 and 7, you would build a regression model in which the dependent variable is a worker's sales revenue, perhaps measured by the month. The independent variable would be a dummy variable, equaling 1 if the worker has undergone the training program as of the month in question, and 0 if they have not. The control variables would include various worker characteristics and job characteristics. The estimated regression coefficient on the training variable would then measure the additional monthly sales revenue that is associated with having been trained, analogous to how, in Chapters 6 and 7, the estimated regression coefficient on *Male* measured the additional annual salary that is associated with being male.

You can make the forecast more sophisticated by including interaction terms (see the appendix to Chapter 6) on the right-hand side of the preceding regression equation. In particular, you would interact the training variable with various worker and job characteristics that are included as control variables. Doing so would allow the measured effect of interest (i.e., the additional monthly sales revenue that is associated with having been trained) to differ by worker and job characteristics, just like, in the appendix to Chapter 6, the effect of gender on annual salary was allowed to differ by worker characteristics such as prior education levels.

I've given you just a brief sketch of an approach to take. We could have a much longer discussion at this point about how to specify the regression properly and about various complications that arise. But I'm going to leave the discussion here. The main take-away is that the compensation analytics methods that we developed in the preceding two chapters (including the appendix to Chapter 6) can be used to produce a more sophisticated "educated guess" about your workers' post-training productivities than you could produce in the absence of any data analysis. Hopefully you're now sold even more on the value and wide applicability of the methods in Chapters 6 (including the appendix!) and 7. And maybe if you skipped or glossed over those chapters, you'll now think twice about that decision!

8.5.2 Training Costs

There are two main types of training costs, and I only addressed one of them in Example 8.1 of section 8.4. The first is the direct cost of the training. In Example 8.1, that was a one-time, current cost of $45,000. For example, if you send your workers away to a one-week training seminar, the $45,000 reflects the fees or tuition that you must pay to enroll your worker in the seminar, along with any travel reimbursement or other expenses that you must pay. It should also account for the lost productivity that occurs during the training period; if your workers had not gone to the training seminar, they would have spent that entire week working at the company and generating sales revenue. So whatever sales revenue is foregone during that week, due to their absence, must be included in the $45,000. All of these costs, however, can be thought of as *one-time costs*. They are paid upfront, once, in the current period, and they are not ongoing.

The second type of costs are *ongoing costs* that are incurred in each period after training has occurred. The most important ongoing cost is compensation. Once your workers are trained, they are more productive, and workers who are more productive need to be paid more, or you risk losing them to competitors. Training, therefore, leads to an increase in expected compensation costs. How much does it drive up compensation costs?

The answer hinges heavily on how portable the training is, i.e., how general or specific it is, as we discussed in sections 8.2 and 8.3. The more general the training, the more leverage your workers have in commanding higher pay, because the training has equipped them with skills that competitors highly value. Even in the case of purely firm-specific training, however, you can expect to incur some increase in your employees' compensation once they are trained. Firm-specific training is a shared investment that creates value for both the employer and the workers, and the workers typically have bargaining power to extract some extra compensation as a consequence, because you would lose a lot if they were to quit.

Example 8.1 didn't consider the ongoing costs of higher compensation that result from training. Instead, it assumed for simplicity that compensation remained the same after training. That omission from the example isn't quite as big a deal as it may seem, because a small change to the example would allow it to be reinterpreted in a way that incorporates ongoing costs. More precisely, the $12,000 that is accrued in each of the five post-training periods can be interpreted not as "annual sales revenue" but rather as "annual sales revenue net of additional compensation costs". So it might be that the trained worker is actually generating $17,000 in extra annual sales revenue, but, because their compensation also increases by $5000 annually as a consequence of their higher, post-training productivity, their *effective* annual sales revenue is only $12,000.

The example is conceptually easy to fix so that it properly accounts for ongoing training costs arising from higher post-training compensation. But in practice, it can be difficult forecasting those ongoing costs. In principle, you can apply the tools of compensation analytics to forecast increases in compensation costs, much like we sketched (in section 8.5.1) how those tools can be used to forecast post-training worker productivity. But there are some challenging issues that arise for such an analysis, one of which concerns turnover, the subject to which we now turn.

8.5.3 Expected Post-Training Worker Tenures

When your workers quit, you typically don't have complete data on their compensation packages at their new employer. That means any data analysis that you do using your own company's data must be based on the workers who stay with you. For those workers who have left, you typically have information on them only up until the time they left. This makes it hard to get a complete picture of how training affects compensation, because your workers who leave are likely to do so shortly after the training ends; they leave precisely because the training has made them more productive and allows them to command higher pay elsewhere.

Of course, their likelihood of leaving for higher pay is influenced by how willing you are to increase their compensation in the post-training periods. To illustrate the problem that turnover creates for your analysis in section 8.5.2, let's consider the extreme case in which your company follows a "no offer matching" policy (see section 12.7). In that case, when a competitor offers one of your workers more compensation you simply allow the worker to be poached. Many of your workers will then leave after training, to take more lucrative offers elsewhere. But some will stay even though you haven't increased their compensation, perhaps because family considerations limit their mobility (recall the three-legged stool of compensation from section 1.6). Since the workers who stay are the only ones for whom you have compensation data before and after training, they are the ones who would appear in a regression analysis if you tried to forecast post-training compensation much like we discussed forecasting post-training worker productivity in section 8.5.1. Such a regression analysis would predict that training is associated with no increase in pay, but that's only because the regression is (necessarily) restricted only to your workers who stay and not to those who leave for higher pay elsewhere.

An important take-away here is that there are close connections among the type of training (i.e., general or specific), the expected post-training increase in compensation, and the amount of time you can expect your workers to remain with you after they're trained. These issues must be considered simultaneously. You will later see that section 12.7, on raiding and offer matching, relates closely to the preceding discussion; it deals with the situation that arises when one of your workers threatens to quit after receiving a more lucrative offer from a competing firm. Sometimes you win the resulting bidding war and retain your worker. Other times you lose your worker to a competitor, as happened when Procter & Gamble lost Scott Cook as I mentioned at the start of this chapter.

8.5.4 Interest Rates

A key point of this chapter, which extends to all investment decisions, including those that we will address in Chapter 12, is that monetary amounts (whether revenues or costs) that accrue in the future must be properly discounted. But what interest rate should you use for this purpose? The choice matters a lot, as we saw in Example 8.1 from section 8.4. There, the difference between interest rates of 10% and 11% is the difference between training and not training your worker.

Echoing the discussion from Chapter 1's appendix on nominal versus real compensation, it's important to understand the distinction between *nominal* and *real* interest rates. The real interest rate, r, adjusts for inflation, i, whereas the nominal interest rate, n, does not. These three rates are connected in the following way:

$$r \approx n - i$$

That is, the real interest rate is approximately the nominal interest rate minus the inflation rate. In an environment with no inflation (i.e., $i = 0$) the real and nominal interest rates coincide, i.e., $r = n$. In an environment with deflation (i.e., $i < 0$) the real interest rate exceeds the nominal interest rate, i.e., $r > n$. That situation is unusual, so we're not going to discuss it. Henceforth, let's focus on the most common situation, which is an inflationary environment (i.e., $i > 0$) in which the nominal interest rate exceeds the real interest rate, i.e., $r < n$.

For training investment decisions, the real interest rate is the one that you should use, and that's why I used the "*r*" notation in Example 8.1. The best way to account for the effect of inflation is to remove it from the analysis. This is done by converting all nominal values (e.g., revenues, salaries, bonuses, training costs, etc.) to current values. Then use the real interest rate, r, for the purpose of discounting future streams of revenues and costs. In the examples of this chapter, we implicitly assumed that this had already been done, i.e., the monetary values are in current dollars, and the interest rate used for discounting, r, is the real interest rate that has inflation removed from it via the approximation $r \approx n - i$.

8.6 Lessons for Managers

The most important lesson for managers is to think of training as an investment and to analyze training problems as you would any other investment, which includes considering the risk profile of the future returns associated with the investment. You incur training costs today in exchange for a more productive worker tomorrow. The challenge is to figure out whether the additional future productivity is attractive enough (and safe enough, from a risk standpoint) to outweigh the short-run training costs. Because the benefits accrue in the future, the present discounted value of the stream of future revenues must be computed and compared to the present discounted value of training costs that you must pay.

What complicates the problem is that in most cases training is at least somewhat portable, which means that, once trained, your worker will have a higher market value. This will necessitate paying your worker higher compensation to avoid losing them to a competitor. That higher compensation level eats into whatever profit you might otherwise make from having trained the worker. It also raises the possibility that other firms might free ride off of your training investment, meaning that you spend money upfront to train the worker, and then the worker (who is now more productive) moves to another firm for a higher compensation level.

You should be wary of this prospect, and plan accordingly. This might involve depressing your employee's initial compensation for a while after the training ends, which has the effect of implicitly charging your worker for the training. Then if your employee leaves, you don't end up spending a lot of money on training for nothing. Another approach that is sometimes used in the public sector is to have employees charged for the training if they leave before the end of a specified time period. For example, police and fire employees often sign an agreement that they will reimburse a portion of their (expensive) training if they leave within five years of completing the training. The bottom line is that you want to avoid your workers "taking the training and running", much like in Chapter 2 we worried that they might "take the money and run" if paid before doing their work.

You need to find some way to effectively tie your worker to the firm, at least for a while, to prevent them from doing this, to make training investments worthwhile for you. Basically, you are looking for ways to insure yourself against the possibility of post-training turnover, which is the main risk associated with a training investment. An obvious but expensive way to do this is to increase employees' compensation immediately after any significant training that enhances their (portable) skills or knowledge. For example, in the public sector, certified employees are given a specific monetary incentive (anywhere from $25 to $100 per month) in extra pay. Alternatively, they are given a percentage of their pay (say, from 2.5% to 10% extra) per month depending on the department. Police officers who earn a degree can receive, say, 5% extra per paycheck.

The post-training turnover problem is exacerbated even further if the training that you have given your workers involves trade secrets or internal processes and procedures that give you a competitive advantage and that you would not want made public and copied by your competitors. Then the challenge becomes training your workers so that they can make productive use of this proprietary information, while simultaneously safeguarding the information even in the event that they separate to join competing firms.

You should think about training as compensation from the standpoint of your workers, because it's something that they like about their job (see section 1.1). The main reason why they like it is that it can be expected to increase their future compensation, possibly through enhanced promotion prospects (see Chapter 13). So training is an *investment*, both from your standpoint and from your workers' standpoints, and it is also *compensation* from your workers' standpoints. The extent to which it is financed by you or by your workers hinges, in large part, on the portability of the training.

Case Discussion 9: Google[1]

In January 2018, Google began publicly offering its internal information technology (IT) training course online, hosted by Coursera, an online learning website. The introductory course is designed for, and is open to, anyone without an IT education. Those who complete it receive the Google IT Support Professional

[1] Note: This case is based on the article on Next.gov by Michael J. Coren (January 17, 2018) entitled "Google's Latest Hiring Tactic is Training Other Companies' Employees".

Certificate, which should prepare them for entry-level IT jobs in any company, including Google. The curriculum includes labs, guidance on "soft skills", interactive evaluations, and 64 hours of video lessons. The certificate program is designed to prepare students for basic IT support functions such as troubleshooting, customer service, networking, operating systems, system administration, automation, and security.

The program costs $49 per month. Estimated time to completion is eight months for a student who is willing to invest eight to ten hours per week in the courses. Total expected cost for one student is about $400 to $600. Google offered a full scholarship for 10,000 qualified applicants who either applied by February 20, 2018 or were selected by nonprofit partners offering in-person support. The total expected cost to Google of providing the course is in the range of $4 million to $6 million.

The scholarship program that offers free tuition to 10,000 people can be interpreted, at least in part, as a hiring strategy, because Google will ultimately hire some of the people who complete the certificate program. However, many people who complete the program will end up joining other companies, in which case Google will have provided free training to those workers without reaping any returns on the investment.

Questions

1. Firms are often reluctant to provide "general training" because, once trained, the worker might get poached by a competing firm that effectively "free rides" on the initial firm's training investment. Why would Google spend millions of dollars knowing that, in many cases, it will ultimately end up training another firm's workers for free?
2. Review section 7.2. Suppose you are a manager at Google and you're in charge of implementing the scholarship training program. How would you design an experiment to analyze how well the policy works in terms of delivering new hires and publicity? How would you quantify the benefits of "publicity"?
3. What sorting effects (both positive and negative) is this training program likely to generate, both for Google and for other firms in the industry?
4. If you were the Google manager in charge of designing and administering the program, would you do it precisely like Google is doing it, or would you make any changes? Explain.
5. In what other industries and types of jobs would this type of training program be a good idea? In what settings would it be a bad idea? Use specific characteristics of industries and jobs to support your answers.

Further Reading

Almeida-Santos, F. and Mumford, K. (2005). "Employee Training and Wage Compression in Britain." *The Manchester School*, 73, 321–42.

DeVaro, J., Maxwell, N. L., and Morita, H. (2017). "Training and Intrinsic Motivation in Nonprofit and For-Profit Organizations." *Journal of Economic Behavior & Organization*, 139, 196–213.

Heslin, P. A., Vandewalle, D., and Latham, G. P. (2006). "Keen to Help? Managers' Implicit Person Theories and Their Subsequent Employee Coaching." *Personnel Psychology*, 59, 871–902.

Lazear, E. P. (2009). "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy*, 117, 914–40.

Muehlemann, S. and Wolter, S. C. (2011). "Firm-Sponsored Training and Poaching Externalities in Regional Labor Markets." *Regional Science and Urban Economics*, 41, 560–70.

Prasad, S. and Tran, H. (2013). "Work Practices, Incentives for Skills, and Training." *Labour Economics*, 23, 66–76.

Santos, A. and Stuart, M. (2006). "Employee Perceptions and Their Influence on Training Effectiveness." *Human Resource Management Journal*, 13, 27–45.

Skule, S. (2004). "Learning Conditions at Work: A Framework to Understand and Assess Informal Learning in the Workplace." *International Journal of Training and Development*, 8, 8–20.

Tharenou, P., Saks, A. M., and Moore, C. (2007). "A Review and Critique of Research on Training and Organizational-Level Outcomes." *Human Resource Management Review*, 17, 251–73.

9 Pay for Performance

One “lesson for managers” in section 7.3 was to use your organization as an experimental laboratory for generating your own data and testing out various changes to the compensation system. An interesting example of such an experiment was conducted some years ago in a tree-planting firm in the British Columbia province of Canada.[1] You might want to review section 7.2, where I discussed the design of such experiments, before you continue reading.

Before the experiment, the tree-planting firm in British Columbia typically paid its workers in proportion to the number of trees that they each planted per day. So if you’re working as a planter at the firm, and you plant N trees on a given day, your pay for that day is $\beta \times N$, where β (i.e., the amount of compensation you receive per tree planted) is known as the *piece rate*. Such a compensation system is known as a *piece-rate pay plan*.

To conduct the experiment, nine male workers were randomly selected from the firm, and each of them was observed planting trees under two different pay plans . . . the piece-rate plan just described, and a fixed wage where the planters were paid a fixed hourly rate regardless of how many trees they planted. This design of the experiment allows us to measure the performance of the same workers, doing the same work, in the same firm, at the same time, under two alternative modes of compensation.

The bottom line is that productivity was considerably higher . . . to the tune of about 20% . . . under the piece-rate pay plan than under fixed wages. This is a classic example of a performance-pay success story. However, such plans can also backfire in various ways.

For example . . . when you open a new bank account at Wells Fargo, it is likely that somebody there directly benefits via higher pay. Compensation bonuses at the bank are – or at least *were* at the time I wrote this chapter – tied to sales targets based on the number of new accounts opened (e.g., six per day), so employees have strong incentives to convince you to open a new account. But if you decide not to open an account, there’s nothing they can do, right?

---

1 See Shearer, B. (2004). “Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment.” *The Review of Economic Studies*, 71, 513–34.

Wrong! Wells Fargo's employees created a whole bunch of new, false accounts for people who didn't request them. Some employees got rich from the bonuses. Then news of the scandal broke explosively, tarnishing the bank's reputation. In April 2017, the bank reported the results of a six-month-long investigation into the scandal and announced that it would claw back an additional $75 million from the two executives it claims deserved the most blame.

9.1 Hourly Sales Quotas at ProDirect

What happened at Wells Fargo is hardly unprecedented. When I first read the headlines, I recalled an epoch from my own job history. During my senior year in high school, I held a telemarketing job at a company called ProDirect. The job paid $10 per hour. That was very high pay at the time for someone in high school; it was triple the minimum wage of $3.35 that I had earned at Taco Bell the previous summer, and it would be more than $20 in 2020 dollars.

At ProDirect, we sold credit cards and credit-card insurance, over the phone, to customers who were, typically, extremely reluctant to buy them. We had hourly quotas (e.g., selling five cards per hour, much like the sales targets for Wells Fargo employees). Managers stood in the front of the room, where there was a whiteboard with all of our names written on it. When you'd sell a credit card, you'd raise your index finger high in the air, indicating "one" new sale. Then the manager would use a dry-erase marker to add one stroke next to your name on the white board. So everyone could observe everyone else's performance, in real time.

The hourly quota varied from shift to shift and depended on what products we were selling, but five was a typical number. If you managed to meet that quota in a given hour of your shift, you were allowed to stay for the next hour. Otherwise, you were sent home early for the day, and paid only for the hours you had worked, which was devastating for those workers who really needed the money from a full shift. If you sold seven cards in the first hour rather than the required five, then you only needed to sell three more in the second hour to be allowed to stay for a third hour. On very rare occasions, you'd have enough "carryovers" from previous hours that you'd meet the quota for the final hour of your shift before that hour even started.

Imagine having just ten minutes left in an hour of a particular shift and knowing that unless you make one more sale within ten minutes, you're going to get sent home. A new call connects, and you know it's your last chance to meet the quota. A kind, elderly woman answers the phone, and a miracle occurs. She agrees to take the card! You begin completing the sales application with pen and paper. With three minutes to go, the application is complete, and you're reading her the final disclosures that will complete the sale. Then at the very end of the verbal disclosures she says, “Actually . . . now that I think about it . . . I’m terribly sorry, but I’ve changed my mind. I’m not interested in the card.”

The call ends. There’s one minute left of the hour. You stare down at the completed credit-card application sitting on your desk. All you have to do is raise your index finger, and you will earn $10 (plus the prospect of remaining even longer and making more money if things go well during the next hour) while a manager comes over to collect the completed application. Morality aside, what’s the cost of doing that? Well, supposedly there was a manager sitting “upstairs” who would randomly select completed applications, and call the customers back to ensure that the recent sale was legitimate. Ostensibly, you could get in trouble if you submitted a fake application. But that never seemed to happen, and moreover, everyone knew that if you got caught you could always claim that the customer must have changed their mind after the call ended, and it would be hard for the company to know otherwise (the sales calls were randomly monitored but not recorded, and if you were paying close attention you could always tell when the call was being monitored because of a very subtle sound in the background).

At ProDirect, I can attest that index fingers sometimes went up without justification, additional hours of pay were improperly collected, and some customers were unfortunately signed up for credit cards against their will, just like the fake accounts at Wells Fargo. Clearly, performance pay has its downsides. Oh, and just in case you’re wondering . . . no, I never lied about a sale!

9.2 Pay for Performance: Some Basics

The examples of the Canadian tree-planting firm, Wells Fargo, and ProDirect dramatically illustrate the power of performance pay to shape workers’ behaviors. Such pay can induce greater effort and productivity as it did in the tree-planting firm, but it can sometimes backfire magnificently by encouraging behaviors that hurt the organization. The Wells Fargo scandal is a prime example, where the ethical failures induced by performance pay tarnished the bank’s reputation and imposed ongoing financial costs.

Performance pay goes by many different names: pay for performance, performance-related pay, performance-based pay, incentive pay, output-based pay, output-contingent pay, pay by results, etc. It directly ties pay to some performance measure, *P*, and it is based on a worker’s “output” (i.e., what they actually produce) rather than the worker’s “inputs” (i.e., how long they work and how much effort they invest). Virtually everyone’s pay is tied to performance, at least indirectly. For example, even a salaried worker’s pay depends on performance, because if performance is sufficiently poor, the worker will be fired (and, therefore, receive zero future pay). But when we refer to performance pay, we have in mind a more direct connection between pay and a performance measure, like we saw at Wells Fargo and ProDirect. Examples of such pay include piece rates, bonuses, commissions, stock options, tips, etc.

Sometimes when thinking about performance pay, and distinguishing it from input-based pay such as wages and salaries that do not directly depend on performance, it is helpful to visualize the compensation contract in a graph, with a performance measure (perhaps on a scale from 0 to 100, where the units might represent items sold within a certain time period) on the horizontal axis and pay on the vertical. Someone who earns an annual salary of $60,000 would have a pay graph that is a horizontal line starting at a vertical intercept of $60,000. The fact that the line is flat means that, regardless of the worker's performance measure, *P*, the same pay of $60,000 is received. See panel *a* of Figure 9.1.

Alternatively, in lieu of a salary, suppose that a worker is paid $1000 per unit sold (i.e., $1000 per performance point). Pay for that worker would start at $0 and would be an upward-sloping straight line with a slope of $1000. See panel *b* of Figure 9.1. This is called a "piece-rate" pay system because workers are paid by the "piece" (or by the performance point in this example). The performance measure, *P*, in the case of piece rates is "units of output produced" (e.g., the number of apples picked, or the number of pounds of crab meat picked, as in the Lindy's Seafood case discussion at the end of the chapter).

Now consider both pay plans simultaneously, in panel *c* of Figure 9.1, which simply overlays the two graphs from panels *a* and *b*. The upward-sloping "piece-rate" line intersects the flat "salary" line at a performance level of 60. So if there are two workers, one who is paid an annual salary of $60,000 and the other who is paid by the piece at a rate of $1000 per performance point, they earn the same annual pay if they both have performance levels of 60. For performance levels below 60, the salaried worker is paid more, and for performance levels above 60, the piece-rate worker is paid more.

Comparing the two graphs (or "pay contracts") illustrates the difference between performance pay and non-performance pay. The salary graph in panel *a* has a flat (or zero) slope, whereas the performance-pay graph in panel *b* has at least some regions where it is not flat . . . in fact, it is not flat anywhere.

Graphs like Figure 9.1 express pay as function of a performance measure, *P*. For example,

$$Pay = f(P) = \alpha + \beta P,$$

where f is a *linear* function (of P), β is a positive number called the "piece rate", and α represents "base pay" or fixed pay that does not depend on performance. The pay contract is called a *linear* contract, because its graph is represented by a straight *line*, as in the upward-sloping one in panel *b*. In fact, a salary or hourly wage, such

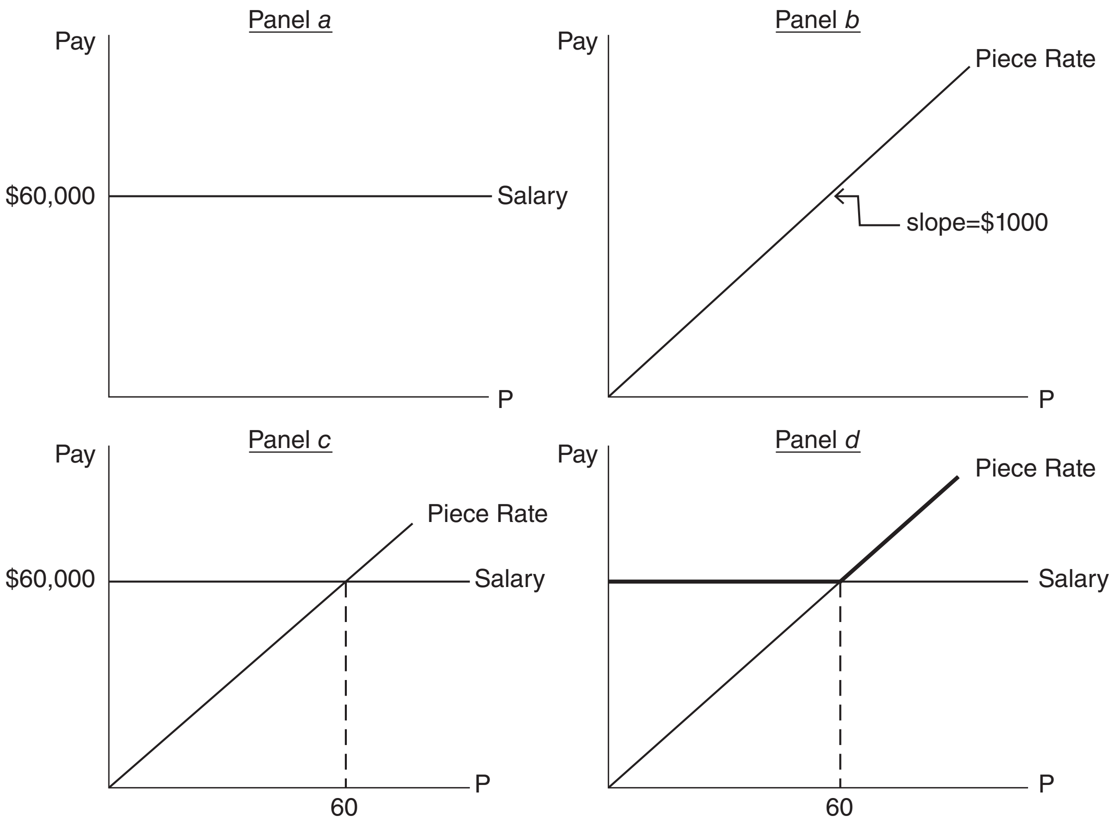


Figure 9.1 Different types of compensation contracts.

as the \$60,000 annual salary shown in panel *a*, is also a linear pay contract, because its graph is a straight line. But it's a rather uninteresting one because it has no slope (i.e., $\beta = 0$), and *worker incentives are created only by non-zero slopes.*

There are other forms of performance pay in which, unlike those shown in panels *a* and *b*, the "*f*" function is *nonlinear*, meaning that its graph is not a simple straight line, even if some portions of its graph can be represented by a straight line. Examples include stock options and executive compensation bonuses, as discussed in Chapter 10. As another example, let's consider a third contract to accompany the two linear ones in panel *c*.

This (nonlinear) pay contract offers your worker a guaranteed annual salary of \$60,000 and, in addition, your worker gets \$1000 per performance point for any performance points that are achieved beyond 60. So if your worker only obtains 17, or 57, or 60 performance points, he would just collect the base salary of \$60,000. But if your worker obtains 63 performance points, he collects the \$60,000 plus an additional \$3000 in commission (i.e., \$1000 $\times$ 3 performance points in excess of 60), for a total of \$63,000. This nonlinear pay contract is shown in the boldface graph in panel *d* of Figure 9.1, which also includes as overlays the two linear contracts from panel *c*. Note that parts of the boldface graph overlap with the other two pay contracts. Compensation plans like those depicted in the boldface graph are known as *draw schemes*.

Using the preceding general function for performance pay, i.e., $Pay = f(P) = \alpha + \beta P$, we can express the equation for all three pay contracts (two of them linear, and one of them nonlinear) that are graphed in panel *d*. These are as follows:

Flat line (panels *a* and *c* of Figure 9.1):

$60,000 annual salary:

$$\text{Pay} = f(P) = \alpha + \beta P = \$60,000 + \$0P, \text{ so } \alpha = \$60,000 \text{ and } \beta = \$0.$$

Upward-sloping line (panels *b* and *c* of Figure 9.1):

Piece rate of $1000 per performance point:

$$\text{Pay} = f(P) = \alpha + \beta P = \$0 + \$1000P, \text{ so } \alpha = \$0 \text{ and } \beta = \$1000.$$

Kinked (nonlinear) "draw scheme" (panel *d* of Figure 9.1, shown in bold):

$60,000 annual salary plus piece rate of $1000 per performance point for units in excess of 60:

$$\begin{aligned}\text{Pay} = f(P) &= \alpha \text{ if } P \leq P_0 \\ &= \alpha + \beta(P - P_0) \text{ if } P > P_0,\end{aligned}$$

where α = $60,000, β = $1000, and P_O = 60. Plugging these three numbers into the preceding nonlinear pay contract lets us write it more simply like this:

$$\begin{aligned}\text{Pay} = f(P) &= \$60,000 && \text{if } P \leq 60 \\ &= \$60,000 + \$1000(P - 60) && \text{if } P > 60.\end{aligned}$$

The last function is expressed in two pieces (just like its boldface graph appears in two pieces, one of which is flat and pertaining to performance levels not exceeding 60, and another that is upward sloping and pertaining to performance levels in excess of 60). The function shows that if performance is 60 or less, only the $60,000 annual salary is relevant, whereas if performance exceeds 60, there is an additional component that rewards performance in excess of some performance target, P_O, which is set at 60. The number of performance units that are obtained in excess of the target is $P - P_O$, so if the worker obtains 63 units (i.e., P = 63), then

$$P - P_0 = 63 - 60 = 3,$$

and extra compensation of $1000 × 3 = $3000 is added to the base salary of $60,000.

I mentioned that worker incentives are created by non-zero slopes of pay contracts. Let's return to this point, with reference to the three alternative pay contracts shown in panel *d* of Figure 9.1.

Start with the linear contract that has zero slope, namely the straight annual salary of $60,000 that is shown by itself in panel *a*. Your worker faces no incentives to perform better. Whether the performance is 5, 17, 69, or 83, pay is simply $60,000, and if your worker goes the extra mile and obtains one more performance point, their pay still remains stuck at $60,000.

Next consider the piece-rate contract shown by itself in panel *b*. Because it is upward sloping over the entire range of performance, it always creates incentives. Whether a worker's performance is 5, 17, 69, or 83 (or even zero!), the worker can always increase pay by $1000 simply by obtaining one additional performance point.

Finally, consider the draw scheme (a nonlinear contract appearing in bold in panel *d*). No incentives exist for performance levels of 60 or below, because this simply replicates the zero-slope "flat" picture from panel *a*. But once performance hits 60, incentives kick in at the "***kink point***" in the boldface graph, because the slope switches from zero to positive (i.e., from $\beta = \$0$ to $\beta = \$1000$). Starting at $P =$ 60, one additional unit of performance will earn your worker $1000.

With a draw scheme, often the relevant part of the graph is the upward-sloping part which actually generates incentives, and in that situation there is no difference between a draw scheme (as in panel *d*) and a straight piece rate (as in panel *b*), because for performance levels of at least 60, the two graphs are identical. What about workers on a draw scheme (panel *d* of Figure 9.1) whose performances fall below 60? Those workers are likely to be out of a job, particularly if it happens repeatedly, and one of my MBA students provided a nice illustrative example . . .

He used to work in the restaurant industry in Alabama. The hourly minimum wage was $7.25, but servers were paid only $3.12 (recall from section 4.7.1 that waiters and waitresses are exempt from the minimum wage because of the expectation that income from tips will allow them to match or exceed the minimum wage). The majority of the pay was from tips, and at the end of every shift the workers had to declare their total tip income. Many workers listed "$0" so as to avoid paying income taxes on the cash income. Management was unhappy about this, because the workers then needed to be paid enough extra money to bring their hourly wage to $7.25, as required by law. One worker was written up multiple times for failing to report tips, and eventually the servers were all summoned to a meeting and told that they had to start claiming their tips. Management told them that they didn't have to claim all of their tips, but just enough so that the hourly wage came out to $7.25, so that the restaurant wouldn't have to incur extra compensation costs. One worker asked what if they didn't make enough in tips to bring their pay to $7.25. And the manager replied that if a worker can't make about $4 in tips, then maybe they shouldn't be working as a server!

Question: Management didn't care if the servers reported any of their tip income beyond the minimum amount (i.e., $7.25 – $3.12 = $4.13) required to get the servers' hourly pay up to $7.25. Setting aside the ethical problems with failure to report cash income for tax purposes, why might management actually *prefer* that workers underreport their tip income, as long as they report the necessary $4.13?

Hints for answering this question are scattered all over the book (e.g., reread section 4.7.5).

Before turning to a different question in the next section, let's take a look at another example of a nonlinear compensation contract that resembles a draw scheme in that its graph is flat up until a certain threshold level of performance is met, and then upward-sloping beyond that kink point.

Example 9.1: Daily Performance Bonuses for Technical Support

A private company in the internet industry, headquartered in New York City, provides social media marketing and other services to large corporate clients. In 2018, a bonus system was instituted to incentivize the company's technical support specialists. The bonus system established a daily individual productivity target, P_O, which was the number of service tickets that a technical support specialist solved in one day. This target was set at five. Bonus payments of $10 per ticket were awarded for every additional ticket that was solved beyond the target of five. For example, if a specialist solved nine tickets in a day, they solved four in excess the target, so they would get a $40 bonus for that day. The daily bonus payment was then:

$$\begin{aligned} \textit{Daily Bonus} &= \$0 && \text{if } P \leq 5 \\ &= \$10(P-5) && \text{if } P > 5 \end{aligned}$$

where P is the number of service tickets a technical support specialist solves in one day. See Figure 9.2 for the graph of this daily bonus contract, which is shown in boldface. Notice that a kink occurs at $P = P_O$, where $P_O = 5$.

The company sweetened the preceding bonus deal as follows, with a further twist beyond what is displayed in Figure 9.2. The top five technical support specialists, based on ticket closures, would earn double the reward amount. For example, if a specialist was in the top five and earned $250 in daily bonuses for that month, they would be paid $500.

There is no easy way to show this additional twist in a graph, because the total bonus payment doesn't just depend on a specialist's own individual performance but also on the performances of the other specialists. This twist is an example of *relative performance rewards*. The daily bonus system is a (nonlinear) contract that blends aspects of both *absolute* performance pay (because specialists get paid $10 for each ticket solved beyond five per day, regardless of what any of the other specialists do) and *relative* performance pay (because specialists get extra money if they are among the top five performers, an outcome which obviously depends on the performances of other specialists).

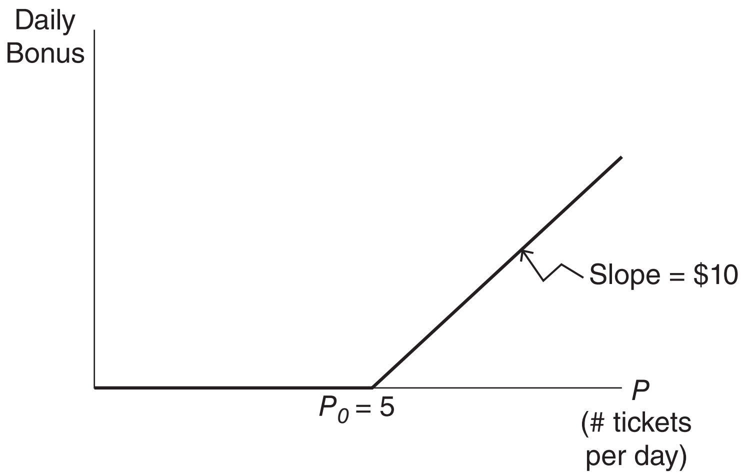


Figure 9.2 Daily bonus for technical support specialists.

Remember . . . slopes are what create incentives, and changes in slope (as occur at the kink point in the bold graph) tend to change worker behavior! We will see this again in the discussions of stock options and executive bonuses in Chapter 10.

9.3 What’s the Purpose of Performance Pay?

Employers use performance pay in the hope that it will influence worker behavior in ways that benefit the firm. As illustrated in the preceding section in the discussion concerning positively sloped regions of $f(P)$, the primary purpose of using performance pay is to create *incentive effects* . . . inducing higher levels of effort and, therefore, productivity. I say “therefore” with hesitation, because greater effort doesn’t always lead to higher productivity, and in fact it could worsen productivity. It all depends on how the effort is spent. There are various ways in which workers can expend effort to increase the performance measure, P, on which their performance pay is based, while harming the organization. The Wells Fargo scandal is a perfect example.

A second reason for using performance pay is the *sorting effect* (section 1.8). Generally, the high performers tend to be more excited than the low performers about performance pay, and they are attracted to firms that offer it. The reason is that the high performers know that they are likely to make a lot of money from such a plan. Thus, the sorting effect of performance pay has important implications for talent management. In short, there are two reasons to use performance pay . . . *incentive effects* and *sorting effects*. Section 9.2 focused on incentive effects, and that entire discussion assumed that you have access to a good, accurate measure of performance, P. When the performance measure is flawed, then problems arise, as discussed in section 9.1 (and the paragraph preceding it) and section 9.7. Performance measurement is critical to any successful performance-pay plan, and it's covered in section 9.8.

9.4 How Prevalent Is Performance Pay?

Performance pay gets a lot of attention in textbooks, academic research, and the popular press. One reason is that in the fairly dry terrain of compensation, performance pay is the source of many colorful stories, such as those from Wells Fargo and ProDirect. But how common is performance pay in practice? Less so than you would guess based on the attention it receives. Surely in some jobs (e.g., sales and CEOs) it's extremely important. But in most jobs, if it's present at all, it's by no means the centerpiece of the compensation plan and represents a relatively small portion of total compensation.

For example, consider my own job history. I've held six main jobs since high school: as a worker at Taco Bell, a telemarketer at ProDirect, a research assistant at the Federal Reserve Bank of Richmond, a research assistant at Stanford, a faculty member at Cornell, and a faculty member at CSUEB. Four of those jobs (Taco Bell, Richmond Fed, Stanford, Cornell) involved absolutely no performance pay . . . compensation was strictly wage and salary, plus (except at Taco Bell) benefits. ProDirect, as discussed in section 9.1, was based on performance pay.

Compensation at CSUEB, as at Cornell, is virtually entirely salary plus benefits. However, there is a small component of performance pay, designed to encourage professors to produce and publish research. CSUEB's business school has produced three lists of academic journals. The most prestigious, competitive journals are in the "A+" list. Those just outside of that list are the "A" journals. And those that are one notch below, though still quite good, are on the "A-" list. The performance measures for the pay plan are the *numbers of articles published annually* in each of the three lists.

The pay isn't in monetary terms but rather in terms of reduced teaching loads. The standard teaching load at CSUEB requires faculty to teach eight courses per year. If a faculty member publishes a research article on the A+ list, then the following year they get "three courses off", so they only have to teach five that year. If they publish an article on the A list, they get two courses off, and if they publish an article on the A- list, they get one course off. In lieu of the course reductions, faculty members can choose to take equivalent cash payments that go into their "research accounts", which can be spent on things like conference travel, hiring research assistants, purchasing research equipment, etc. This performance-pay plan, which is featured in a case discussion at the end of Chapter 5, represents an extremely small sliver of total compensation at CSUEB. So for the most part, of the six main jobs I've held, only one had a serious performance-pay component.

The same holds true more generally. Good data on this are available from the 2011 British Workplace Employment Relations Study. A nationally representative, random sample of 2680 British workplaces was surveyed in 2011. Of the 2680, less than a third (28% to be precise) report using output-based "pay by results". And not all of the 28% are using it very extensively. For example, as I just described, only a tiny sliver of compensation at CSUEB comes from performance pay, and similar jobs in the UK would be included in the 28%. In fact, if even just a single worker at the surveyed establishment is getting paid "by results", and even if that one worker is only receiving a tiny portion of her pay in that form, that would be enough to include that establishment in the 28%.

The language of the survey that produced that statistic is very clear about the meaning of the term "pay by results", so that all respondents interpret it consistently. The definition is:

> *Payment by results includes any method of payment where the pay is determined by the amount done or its value rather than just the number of hours worked. It includes commission, and bonuses that are determined by individual, workplace or organization productivity or performance. It does not include profit-related pay schemes.*

9.5 Risk and Workers' Attitudes Concerning Risk

Understanding how performance pay works requires understanding risk and workers' attitudes towards risk. So let's develop those concepts in this section, and then I'll connect them to performance pay in the next section. Before getting started, let me make clear that by "risk" I'm talking about *income risk*, i.e., the risk of losing money from one's pay. I'm not talking about other types of risk, like the risk of getting injured on the job, as explored in the "Deadliest Catch" case discussion at the end of Chapter 3. When talking about performance pay, "risk" almost always refers to income risk. It turns out that performance pay also has a connection to injury risk, but let's defer that unrelated topic until the end of section 9.7.

Suppose that I offer you the following choice. One option is that I give you $50 in cash, which we can interpret as your compensation for the pain and suffering of having dutifully read this much of my book! Alternatively, we can flip a fair coin. If heads comes up, I will give you $100 in cash; if tails comes up, you get nothing. Which option would you prefer? When I teach my compensation classes, I pose that question to my students, asking them to answer by show of hands. A clear majority of them choose the guaranteed \$50 in cash, but more than a handful choose to gamble on the coin toss, or claim that they're indifferent. I then repeat the question, but this time I offer \$50 million in cash for sure, or we can toss a coin, which gives you \$100 million if heads appears and nothing if tails appears. Except for the occasional, attention-craving fanfaron, everyone chooses the guaranteed \$50 million.

I'd like to highlight a few points about this.

First, the two gambles have the same *expected values* as their accompanying certain cash payments. For example, you have a 50/50 chance of getting \$100 (versus \$0) if you take the first gamble. Suppose that you take the gamble and everyone else in your class takes the \$50. Then one of two things will happen. Either you end up with \$100, and everyone else gets \$50, or you end up with \$0, and everyone else gets \$50. Either way, what you get is way different from what everyone else gets. But now suppose we play this game not once, but every day of the year. Again suppose that every day you pick the coin toss, whereas everyone else in your class takes the sure \$50. Some days, you'll end up with \$100, and the rest of the class gets \$50. Other days, you'll end up with \$0, and the rest of the class gets \$50. But at the end of the year, if we tally up everyone's earnings, yours will be virtually the same as theirs. Everyone other than you will get exactly \$50 × (365 days) = \$18,250. As for you, we can't say with certainty exactly what you'd get . . . to do that, we'd need to toss a fair coin 365 times, once for each day of the year, and I have no interest in doing that! But if the coin is fair, we can expect that about half the time it will come up heads, and half the time tails. So on about half of the days (i.e., about 365/2 days) you'll earn nothing, and on the other days you'll earn \$100. So we expect you to earn about

$$[\$100 \times (365/2\text{ days})] + [\$0 \times (365/2\text{ days})] = \$18{,}250.$$

Now, you might earn a bit more than this, or a bit less than this, but we expect you to be pretty close to \$18,250. And if we were to play the game for 3650 days rather than for just 365, you could expect to get even closer to what everyone else is earning. You get the idea.

Second, even though the gamble offers the same *expected* payment as the sure \$50, most students prefer the sure \$50. In other words, people are *risk averse* – they don't like risk – at least when it comes to their compensation.

Third, when the stakes of the gamble are high relative to a person's wealth, people's distaste for risk becomes even stronger. In the first gamble, the expected payment is only \$50. So if you gamble and the coin comes up tails, you're disappointed that you got nothing, but it's not the end of the world. But in the second gamble, the expected payment is \$50 million. If you gamble and the coin comes up tails, you'll be devastated and spend the rest of your life thinking about what you could have done with the guaranteed \$50 million that you foolishly chose not to take.

Fourth, if you were indifferent between the $50 and the gamble, we would call you *risk neutral*. People generally are not risk neutral when it comes to their incomes. They're at least somewhat *risk averse*. Though again, the smaller the gamble is relative to a person's wealth, the bolder they can afford to be, and the more risk neutral they may appear.

Fifth, an implication of risk aversion is that people would be willing to pay something to avoid risk. Suppose that the second gamble is exactly as I described, but if you skip the gamble your cash payment is only $49 million instead of $50 million. In this case, the gamble gives you a higher *expected* payment than the cash . . . if you were to play the game for a large number of days, you'd end up accumulating more money by gambling every day than by taking the cash every day. Nonetheless, if I gave you the choice between the coin toss or the $49 million with only one chance to play the game, then – unless you're an attention-craving fanfaron – I'm confident that you'd choose the $49 million. How much would someone be willing to pay to avoid risk? Or, stated another way, how much would they have to be paid to induce them to take the gamble? We call this amount their *risk premium*. We have seen the concept before, in the context of compensation contract failure and wage theft (see Chapter 2), which also concerns income risk. The concept will also come up a lot in Chapter 10 when we discuss executive stock options, because options are an extremely risky form of pay for performance.

Returning to the low-stakes gamble, which has an expected value of $50, recall that most of my students prefer to take $50 with certainty. This is unsurprising, because the gamble pays no risk premium, i.e., its expected payment is $50, which exactly matches what you could get by incurring no risk. Suppose that we change the gamble, however, so that if heads appears, you get $120 rather than $100. Now we have sweetened the pot by giving the gamble a higher expected value (i.e., $60) than the certain cash payment of $50. This change might be enough to entice some students who originally preferred the cash payment to take the gamble instead.

Suppose that there's a student who originally preferred the certain cash payment of $50 but who, after we modify the gamble to sweeten the pot, is now *indifferent* between the gamble and the cash payment. That student's risk premium is $10. We know this because there is a $10 difference in expected value (i.e., $60 – $50) between the gamble and the certain cash payment, and the student is indifferent between these options. The $10 can be interpreted as compensation for bearing the risk of the gamble, and at that level of compensation this particular student is indifferent. Those students whose risk premia are higher than $10 would still prefer the certain cash payment of $50 even after the gamble is modified, and those whose risk premia are lower than $10 would definitely prefer the modified gamble to the certain $50.

9.6 Risk and Performance Pay

Often a worker's performance is affected, positively or negatively, by random events. When pay is tied directly to performance, this also means that *pay* is affected by random events. So performance pay exposes workers to a gamble. Things might turn out well, or poorly, regardless of the worker's effort investment. Luck matters!

The problem with this from your standpoint as a manager is that, as I explained in the preceding section, workers generally dislike income risk, and they must be paid a risk premium to accept a job that exhibits income risk. This echoes our discussion from Chapter 2 on wage theft, which is a form of income risk. In section 2.4, I gave an example in which an undocumented immigrant worker on a one-day job faced a 10% chance of wage theft, a random event. In that context, the worker feared that his employer might randomly steal some of his compensation. In the present pay-for-performance context, workers fear that random "bad luck" will lower their performance (and therefore their compensation). But these two situations share the similarity that in either case there is income risk caused by random events, and workers expect to be compensated for that risk via a risk premium.

My telemarketing job at ProDirect offers an example of the relationship between risk and performance pay. There the performance measure, *P*, was "number of credit cards sold in an hour". Several factors affected *P*. Certainly effort mattered . . . e.g., talking fast enough to not waste valuable time but slowly enough to persuasively engage with the customer and be clearly understood. But other factors also affected the performance measure that, at least from my perspective, seemed random. One such factor was the geographic location of the customers. The geographic locations varied from shift to shift, and sometimes within a shift, and I found that some areas (e.g., rural areas, and the South) tended to be easier for sales than others (e.g., Manhattan). Calls were clustered within regions. So if the first segment of the shift was from Georgia, you could expect that at least the next full hour, if not several, would be from Georgia. Which state(s) you happened to get during your shift was pure luck, but it clearly affected the performance measure. Even within a state, there's luck involved with whether the targeted customer answers the phone and whether they can be persuaded to get a credit card.

When I worked at ProDirect, I knew that even if I invested a lot of effort, I still might have bad luck and get sent home after just one hour of my shift. I needed to be compensated for this income risk, i.e., I needed to be paid a risk premium. If ProDirect had paid the minimum wage of $3.35 per hour rather than $10 per hour, I would definitely have preferred to have worked elsewhere, at an alternative minimum-wage job with more certain weekly work hours. The reason why ProDirect was able to attract workers, even with such extreme income risk, is that the high hourly wage of $10 incorporated a hefty risk premium. An alternative way to say this is that if ProDirect had been willing to commit to offering standard, full-time work of 40 weekly hours, then it wouldn't have had to pay anything close to $10 per hour. I, and many other workers, would have gladly taken the job for minimum wage, or close to it, because without the pressure of the threat of being sent home early, it would have been more comfortable than working at Taco Bell.

Paying your risk-averse workers high risk premia is an expensive proposition and diminishes the appeal of compensating your workers using performance pay. People also vary in their degrees of risk aversion, and if your workers have particularly intense risk aversion, then you'll have to pay them very high risk premia to compensate them for assuming the income risks of performance pay. The bottom line is that when there's a lot of income risk associated with the job, and/or when your workers are extremely risk averse, you might want to take your foot off of the performance-pay gas pedal.

But not so fast! We might actually reach the reverse conclusion if we consider an additional factor, namely the *extent to which your workers are sufficiently informed to respond productively to risk*. That's a mouthful, but the idea is very simple. The people who know a job the best are the workers who actually hold that job. As a manager, you probably know something about the jobs that your subordinates hold, but your subordinates are likely to know even more, because they are the ones who actually do the job day in and day out. The *unique information* they hold, and that you lack, can either be used to help your organization or to hurt it (or it can be ignored and not used for either purpose). You want to give your workers an incentive to use their unique information to *help* the organization (i.e., to improve performance), but this requires tying their pay to performance. If you're going to give your workers the discretion to make decisions that leverage their unique information, you need to hold them accountable by tying their pay to performance, so that they don't misuse that discretion by taking actions that may benefit themselves at the company's expense.

When is this unique information held by your workers likely to be the most valuable? When there's a lot of risk in the production process! For example, if really bad luck happens, your workers may be able to use their unique information to soften the blow and mitigate the damage to the organization. Or, if really good luck happens, your workers might be able to rely on their unique information to recognize a once-in-a-lifetime opportunity and seize it, for the benefit of the organization. The greater the importance of luck (both good and bad) in the production process, the greater the importance of leveraging the unique information that your workers possess so that it can be productively channeled to benefit the organization. The bottom line is that when your workers possess unique and valuable information that you lack and that can be used to affect organizational performance in response to bad (or good) luck, then you should *press harder* on the performance-pay gas pedal rather than taking your foot off it.

So there are two competing considerations that you must weigh when designing your workers' performance-pay plans. When there's a lot of risk, the preceding paragraph says that you should be *more* inclined to use performance pay (because in such settings your workers are likely to possess unique and valuable information that you want to induce them to use to help your organization in the wake of good or bad luck), whereas the paragraph before it says that you should be *less* inclined to use performance pay (because your workers are exposed to undesirable income risk and will require you to pay them an expensive risk premium).

9.7 Drawbacks to Performance Pay

The need to pay your workers a risk premium is one drawback of performance pay. But even ignoring risk considerations, performance pay can backfire, as we've seen. When it does, usually the problem lies in how the performance measure, *P*, on which it is based, is defined and measured. This is what went wrong at Wells Fargo. The performance measure used there (i.e., number of new accounts opened per day) didn't account for the *quality* of the new accounts, in particular whether the customers actually wanted them. There are two types of new accounts at Wells Fargo . . . those that customers want, and those that customers don't want. Only the former type should be included in a good performance measure, *P*. But at Wells Fargo, both types were included in *P* without any distinction, and the presence of the wrong type is what led to problems.

The Wells Fargo scandal is an example of *distortion* caused by a flawed performance measure. Distortion means neglecting some important aspects of performance (like quality) that are excluded from the performance measure, while focusing only on those aspects that are captured in the measure. The Lindy's Seafood case discussion at the end of the chapter offers another example of distortion. Workers who are paid per pound of crab meat picked have an incentive to distort their efforts across the different dimensions of performance, focusing too heavily on quantity (i.e., number of pounds of crab meat picked, which is the narrow performance measure used to pay them) and not enough on quality (as evidenced by too many fragments of shells showing up in the finished product).

Sometimes the performance measure is subject to *manipulation*, which gives rise to another potential drawback of performance pay. Manipulation means taking deliberate steps to actually falsify the performance rating, lying to make it look better than it actually is. Manipulation occurs when your workers are able to falsely report (i.e., inflate) the performance measure without being detected. The distinction between distortion and manipulation is subtle. Suppose that a Wells Fargo employee opens seven new accounts in a day, five of which were desired by customers and two of which were undesired. This is an example of distortion, because the performance measure, P, i.e., "number of new accounts opened in a day" is accurately reported as seven, even though two of those accounts are of low quality in that they were not desired by customers . . . so the employee distorted his effort by focusing too much on quantity and not enough on quality. An example of manipulation would be if the Wells Fargo employee reported having opened seven new accounts in a day, when in fact only five new accounts were opened.

The closely related concepts of distortion and manipulation are easy to confuse, so let's discuss another example to highlight the difference. Consider the CSUEB CBE case discussion that ends Chapter 5. As I mentioned earlier, professors in the CSUEB business school face a pay-for-performance plan to create incentives for high-quality research. The performance measure, P, is the number of publications in journals from "A+", "A", and "A-" lists that the business school devised and currently posts on its website. The problem is that the lists are quite imperfect. They are political compromises that resulted from highly tense and acrimonious debates in which professors from different disciplines (who have very different perspectives on what constitutes a prestigious journal) couldn't fully agree on rankings. Because rewards are only given for publications from these lists, professors sometimes purposely avoid submitting their articles to journals that failed to make the A- list but that are actually of higher quality than journals on the A- list. Or they submit to a journal on the A list, avoiding an even better journal that was relegated to the A-list only for purely political reasons, and they do this because the financial rewards from publishing an A paper exceed those from publishing an A- paper.

This is a classic example of *distortion*. Professors spend too much time and effort focusing on certain journals, while avoiding certain others that may be of even higher quality but that failed to make the lists for political reasons. Although distortion can happen in this setting, *manipulation* cannot. I cannot falsely claim that I published a paper in an A journal when in fact I did not, because my claim can very quickly and easily be verified, and indeed professors must furnish their published articles as proof before they can collect their reward.

Another drawback of performance pay is that it can sometimes condition your workers to expect compensation for things that they would otherwise have done willingly for free. For example, consider again the performance-pay plan at CSUEB to encourage professors to produce and publish research. Although the plan explicitly rewards research productivity, the fact is that producing and publishing research is already part of our job descriptions as university professors. It's expected that we do it, along with teaching our classes. The difficulty is that once a professor has tenure (making it nearly impossible to fire them) there is no consequence to shirking on research. If you stop doing research, nothing much happens, whereas if you don't show up to teach your classes, you'll get fired even if you have tenure. Many professors do research, not for compensation, but simply because they are *intrinsically motivated* (see section 15.2 for more on intrinsic motivation). They are motivated either by passion and genuine interest in the subjects they are researching, or they feel a sense of professional obligation. But once you start rewarding them for doing research using a performance-pay system, then if you ever remove the system they might become unhappy and stop doing research, even if they would have done research out of intrinsic motivation if the performance-pay system had never been implemented. In this case, performance pay "*crowds out*", or displaces, intrinsic motivation.

Yet another drawback of performance pay is that the fast pace of work that it encourages can increase the incidence of injuries and illnesses, both of the workers who are paid on such plans and of their co-workers whose work is either interdependent or co-located. One reason is that mistakes become more likely when workers are exhausted and moving quickly; this could result in misuse of machines or equipment. Sitting at a desk for long periods, or working intensely with no breaks, can cause fatigue and repetitive stress injuries, joint and muscle problems, and back problems. All of these health issues can hinder productivity.

Such health problems can also increase the likelihood of absenteeism, which itself hurts organizational productivity. The effect of performance pay on absenteeism, however, is unclear, and the empirical studies on this subject have reported mixed evidence. On one hand, the fatigue and health problems potentially caused by performance pay can make absenteeism more likely. On the other hand, workers who are paid for performance have an added incentive to show up for work, because the pay plan potentially allows them to make a lot of money.

The potential negative worker health consequences of performance pay are often somewhat forgotten among the set of well-known pitfalls associated with performance pay. But there's plenty of evidence of these costs. For example, employees at Wells Fargo during the scandal reported a range of symptoms including stress, anxiety, depression, and in one case a female worker reported regularly drinking the alcohol-based hand sanitizer in the bathroom, because she found that it relieved the anxiety symptoms induced by performance pay!

9.8 Performance Measurement

Good performance measures are needed for the compensation tools developed in this chapter to work *in the desired way*. The phrase "in the desired way" is an important qualification of "work", because performance-pay plans virtually always "work" in the sense of affecting worker behavior. Anecdotal evidence and academic research has shown, time and again, that people respond to incentives. Using compensation to induce your workers to change their behavior is pretty easy; what's often harder is getting them to change their behavior in the desired ways that will benefit your organization. Choosing or developing a good performance measure is the centerpiece of a successful performance-pay plan.

A good place to start is deciding whether the measure will be *subjective* or *objective*. A subjective measure of a worker's performance is an appraisal of that worker's performance based on the judgment of his supervisor. Consider, for example, the performance measurement system that was used for district managers at Borders Group, Inc., which was an international book and music retailer that went out of business in 2011. Borders produced a set of "competency scores" to rate the attributes of its district managers. The scores were based on evaluations performed by the supervising regional director. District managers were rated on each of the following ten different dimensions of performance:

> *creative, results-driven, leadership, persuasive, conflict management, informing, feedback, optimistic, persistent,* and *strategic*.

On each of those ten dimensions, each district manager was periodically evaluated on a five-point scale (known as a "Likert" scale) with the following levels: 1 = *unsatisfactory*; 2 = *needs development*; 3 = *meets expectations*; 4 = *exceeds expectations*; 5 = *far exceeds expectations*.

Imagine that you are the supervising regional director at Borders, and on the dimension of "conflict management" you give one of your district managers a rating of "3 = *meets expectations*". This is *subjective* in that it reflects your opinion (hopefully based on and supported by evidence) about that district manager's conflict management skills during the evaluation period. But a different supervising regional director, someone other than yourself, might have had a different opinion and given that district manager a "4 = *exceeds expectations*" or a "2 = *needs development*".

Alternatively, suppose that a particular district manager's performance is measured by the *monthly sales revenue of that district*. That's a number, and whatever it is, it is. That number is the same whether you are the supervising regional director who evaluates the district manager, or whether some other supervising regional director does the evaluation. That is an example of an *objective* performance measure.

The term "objective" is a bit of a misnomer, however, because it misleadingly implies that the performance measure is immune from subjective influence. But sometimes even "objective" measures can be manipulated or fudged. For example, suppose that district managers at Borders are paid a flat monthly bonus if their district manages to hit a monthly sales target. If district sales revenue exceeds that target in April, the district manager might be able to delay to May the reporting of some of the April revenue. This manipulation of the performance rating depresses the April revenue number and inflates the May revenue number, which allows the district manager to collect the April bonus while getting a head start on May sales revenue, thereby increasing the likelihood of collecting a May bonus. In the discussion of executive bonuses in Chapter 10, we will return to such tricks of shifting profits forward or backwards in time in response to the design of an executive bonus plan. This is a classic example of perverse behavior in the neighborhood of the “kinks” of a performance-pay plan, and you’ll fully understand what I mean by that after you’ve read section 10.2.

You must also decide at what level to measure performance. You can measure performance for individual workers, or for teams, departments, divisions, establishments, districts (as at Borders), or at the level of the entire company. See The Cheese Board Collective case discussion for an example of a worker-owned cooperative in which performance is measured at the level of the entire business, and all of the profit goes to its worker-owners. A disadvantage of group measures is *free riding*, which you’ve probably experienced if you’ve ever worked in a group. Even though everyone in the group gets the same rating, their individual contributions often vary substantially, and the lower performers free ride on the great performance of the stars. But an advantage of group measures is that they are less prone to manipulation by individual workers; they are also often easier to measure than individual performances. In the following example, the employer switched from individual to group-based incentive pay.

Example 9.2: Switch from Individual to Team-Based Pay for Performance at Koret Company

An analysis was done on the personnel records of workers employed between 1995 and 1997 at a garment factory operated by the Koret Company in California’s Napa Valley. The facility produced women’s “lowers” (including pants, skirts, shorts, etc.), and the sewers who produced them were paid individual piece rates. Over a three-year period the factory switched to team-based piece rates in which sewers worked in teams of six or seven members, arranged in a U-shaped work space approximately 12 feet by 24 feet that allowed them to observe each other. Compensation would be rewarded on the basis of the team’s total output and would be shared equally among the team members. At the start of the transition to group-based piece rates, teams were formed on a voluntary basis, so workers could either choose to join them or remain working alone in individual production. Eventually, because of the success of the switch, team production was required of all workers.

The switch from individual piece rates to group piece rates improved worker productivity by 14%, on average. The productivity improvement was greatest for the earliest teams (which were allowed to voluntarily form) and diminished as the remaining workers were eventually switched to team production. One possible mechanism for the improvement is that the U-shaped arrangement facilitated learning about "best practices", so everyone could watch the top sewer in the group and attempt to mimic her tricks and techniques.

Interestingly and somewhat surprisingly, the sewers who had the highest productivity under the original individual piece-rate system tended to be the first to volunteer to join teams, even though many of them ended up suffering a reduction in income as a consequence (because *average* productivity for their teams were less than what they had been producing alone, given that the teams included some low performers who brought down the average). One possible reason for this is that working in teams is more fun than working alone, and the opportunity to socialize while working is a non-monetary component of compensation that sewers value and are willing to pay for by accepting less monetary compensation (see Chapter 3 on compensating differentials, and the broad definition of compensation in section 1.1).

Source: Hamilton, B., Nickerson, J., and Owan, H. (2003). "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy*, 111, 465–97.

A measure of individual performance is *narrower* (i.e., less inclusive) than a measure of group performance, which is *broader*. There are tradeoffs associated with choosing broader or narrower measures. Generally speaking, the broader the measure, the greater the worker's exposure to income risk, because broad performance measures are affected by lots of influences that are beyond a worker's control. For example, a firm's stock price is the broadest possible measure of performance and is often used in performance-pay plans for CEOs, as we'll see in Chapter 10. But a firm's stock price is determined by many factors that are beyond the CEO's influence. Including more risk in the performance measure is a double-edged sword for the reasons we've discussed. On one hand, it exposes your worker to greater income risk, which requires you to pay a heftier risk premium. On the other hand, it tends to increase the value of the unique information that your workers possess, because a greater exposure to risk (i.e., instances of good or bad luck) means a greater chance that a situation will arise in which your worker – if he has the proper incentives! – can apply his unique information to benefit your organization.

The choice between narrower and broader performance measures hinges on more than just risk. Other considerations are that narrower measures increase the threats that your workers will either *distort* their efforts across different dimensions of performance, or outright *manipulate* the performance measures. Distortion and manipulation are both more likely to arise when the performance measure is narrow.

The problem of defining a good performance measure that completely captures all of the relevant dimensions of performance (correctly weighting them in the way that most benefits the organization) and that excludes everything else (i.e., all of the things on which the organization would rather the worker not focus) is a daunting task. The difficulty of devising good performance measures, and the potentially disastrous consequences of implementing performance-pay systems based on inferior performance measures, is the main reason why performance pay isn't usually the centerpiece of most compensation plans.

9.9 Designing the Performance-Pay Contract

If your organization decides to use performance pay and has identified a satisfactory performance measure, you must decide how to design the pay contract. For example, if piece-rate pay is being used to compensate crab pickers, and the performance measure is "number of pounds of crab meat picked", the key question is what piece rate to choose, i.e., how much to pay per pound of picked crab meat. In other words, what should the slope of the graph be in panel *b* of Figure 9.1? There's no easy answer to this. Draw on whatever information you have. For example, you can observe what industry competitors are doing, though you shouldn't assume that competitors always have the best solutions. Trial-and-error is often the way to go. Change the piece rate, up or down, and then, taking stock of all changes in worker behavior (positive and negative) assess whether the organization's bottom line was helped or hurt. If it was helped, the new rate is better than the original one. And you might be able to improve the bottom line even further. Experiment to find out what works best.

This is what you want to do in principle. But in practice, experimentation can be difficult. A big problem is that workers often respond negatively to changes in the pay contract. This happens in particular when they anticipate that you're going to raise standards in the future. If they perform really well, they might fear that you'll respond by raising performance standards (which effectively reduces their pay). This is such a concern in organizations that it has a name . . . the *ratchet effect*. The name captures workers' fears that employers will "ratchet up expectations" once they observe how productive the workers are on a performance-pay plan. Being able to communicate to workers that the performance-pay contract is stable and will not change (at least for the foreseeable future) will eliminate the ratchet effect. But that's difficult to do convincingly, plus it eliminates the benefits from experimentation.

So if you're going to experiment, try to do it either discreetly or in a way that will minimize the severity of the ratchet effect. For example, if your firm has multiple establishments, or multiple divisions or departments within the main office, with little worker interaction across them, you might experiment with different piece rates in different locations to see which ones are the most profitable. Or try to communicate the change to the pay plan in such a way that workers believe that the change was prompted by some reason unrelated to their productivity. It's when workers suspect that the pay plan is being changed in response to their past productivity that the ratchet effect looms.

Although trial-and-error might be needed to converge on the correct design of the performance-pay plan in your firm, some general considerations to keep in mind are how profitable it is to produce your company's product or service, how much risk or variability there is in the output the worker produces, to what degree workers are concerned about income risk, and how distasteful workers find it to do a bit of extra work. When production becomes more profitable, it generally makes sense to intensify performance pay to urge workers to higher levels of effort and (more profitable) performance. If there is significant risk or variability in workers' output, or if workers are very concerned about income risk, it's better to take your foot off the performance-pay pedal a bit; doing so will protect your workers somewhat from the income risk that they fear. And if your workers find a bit of extra work highly distasteful (e.g., if they are already being pushed to near exhaustion) then you should probably not intensify performance pay, and you might consider easing up on it.

Two closely linked questions are: (1) How much performance pay should your workers get? (2) How much decision-making authority should you grant your workers?

The Netflix case discussion explores the connection between these two questions and also how they both relate to performance evaluation. The general rule is that you should delegate more decision-making authority to your workers in situations where they are likely to be better informed than you (the manager) about what decisions would be best for the organization. Let's take an example. Suppose that you are managing some salespeople. Those salespeople are the ones who are interacting, one-on-one, with customers, and in some cases those interactions are repeated over time, which gives the salespeople a chance to form relationships with the customers. This means that the salespeople will have knowledge of what types of arguments work particularly well (or poorly) with certain customers. So rather than you stepping in as a manager and trying to micro-manage each sale by taking the lead on interacting with the customers, it's better to allow the better informed parties (i.e., your salespeople) to conduct the sales. So you delegate the decision-making authority to them.

However, employees who enjoy a lot of decision-making authority can sometimes misuse that authority, making decisions that benefit themselves rather than your company. A good way to "keep them honest" and to ensure that their incentives are aligned with those of your company is to tie their pay to performance. If the performance measure is appropriately chosen, this will keep your workers' eyes on the ball, and you have less to fear about them misusing their decision-making authority to benefit themselves. The bottom line is that, typically, performance pay goes hand-in-hand with workers being granted authority over decision-making and how to do their jobs.

9.10 Lessons for Managers

Much of this chapter covers problems with performance-pay plans. The reason for all of this focus on bad news is to make a lasting impression on you, so that you approach the design and management of performance-pay plans with appropriate caution. You need to understand and be aware of potential problems so that you can avoid or mitigate them. But you should not infer that performance pay is more bad than good! There are countless success stories involving performance pay; when it works correctly, it is an extremely powerful tool for increasing productivity through both incentive and sorting effects. The experiment from the tree-planting firm that opened this chapter is a case in point.

Performance pay always "works", meaning that it always affects workers' behaviors, and usually significantly. The key question is whether, and to what extent, it works as desired, i.e., by inducing behaviors (and attracting worker-types) that improve the bottom line. Understanding the drawbacks of performance pay (e.g., expensive risk premia, distortion, manipulation, crowding out of intrinsic motivation, and increased risks of occupational health problems) can allow you, as a manager, to avoid them or to minimize their occurrences and severity.

You should consider whether to use performance pay at all. If you use it, then your main problem is how to design it. A critically important piece of that problem concerns which performance measure(s) to use. Most of the pitfalls associated with performance pay can be traced to problems with how the performance measure is defined. Experiment judiciously to identify the best design of performance pay for your setting.

Finally, be sensitive to how the performance-pay plan fits with other components of the compensation plan and with other aspects of the HR system and organizational culture.

Case Discussion 10: Lindy's Seafood[2]

Lindy's Seafood is a company based in Woolford, Maryland. In 2016, Lindy's aimed to hire 75 seasonal workers to help extract and package crab meat. The company provided the following information about the new positions.

The positions are full time (in terms of weekly hours), but the jobs only last the duration of the season (April 1, 2016 to December 1, 2016). The job title is "seafood processor / crab picker". The work is supervised, and no education or experience is required. Workers must have good work ethics, follow directions, and be able to

[2] Note: This case is based on a LinkedIn job ad posted in July 2016.

perform manual labor. They must have good health, with no infections or diseases, not be allergic to handling seafood, and be physically fit to meet health department requirements. Job tasks require the worker to: "steam, clean and pick whole crabs; weigh, sort, ice, debone, pack, cook, seafood preparation for wholesale/retail sale; general plant labor such as cleaning and preparing work areas, load/unload, dump, clean, move, dock work, grade, freeze, prepare, seal, lift/carry up to 50 pounds; and any other activities related to seafood processing/crab picking."

Wages and Hours: Two work shifts run consecutively from 2am to 8pm, Monday to Friday. Overtime hours are sometimes worked when there is an oversupply of crabs. Piece-rate pay is used, at a rate of $2.80 per pound of picked meat, or the prevailing wage rate of $8.61 per hour (whichever is higher). When overtime work is assigned, the rate is $12.92 for hours in excess of 40 hours per week. Discretionary raises and bonuses are also paid. The (minimum expected) production standard is 3 pounds of picked crab meat per hour after a two-week training period. A single workweek is used to compute the wage due, and payment occurs every Friday.

Benefits: Shared housing is available to only seasonal full-time workers as a condition of employment at a rental fee of $45/worker per week, which includes utilities. Workers are not required to live in employer-arranged housing. If workers decide upon arrival or anytime thereafter that they want to live elsewhere, they are free to make their own arrangements. Workers who live outside a reasonable commuting distance are eligible for transportation cost reimbursement (and meals during travel) subject to certain restrictions; they must work half the period for reimbursement of inbound transportation, and the entire eight-month period (or get let go) for reimbursement of outbound transportation. The employer reimburses the H-2B (temporary, non-agricultural) worker in the first workweek for all visa, visa processing, border crossing, and other related fees (including those mandated by the government). Reimbursements do not include passport expenses or other charges primarily for the benefit of the worker.

Questions

1. Evaluate this compensation system. What are its strengths and weaknesses?
2. Suppose that customers have been complaining about too many pieces of shell appearing in their purchased crab meat. Why might this be happening? How would you address this problem?
3. Is "pounds of crab meat" the best output measure for piece-rate pay? Can it be improved?
4. Is this compensation system a good fit for this particular industry and job type? Why or why not? What kind of worker would be attracted by this compensation package?
5. How would things change at Lindy's if overtime and prevailing wage laws were abolished?

Further Reading

Adler, S. Campion, M., Colquitt, A., and Grubb, A. (2016). “Getting Rid of Performance Ratings: Genius or Folly? A Debate.” *Industrial and Organizational Psychology*, 9, 219–52.

Azar, O. H. (2018). “Do Fixed Payments Affect Effort? Examining Relative Thinking in Mixed Compensation Schemes.” *Journal of Economic Psychology*.

Cadsby, C. B., Song, F., and Tapon, F. (2017). “Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation.” *Academy of Management Journal*, 50, 387–405.

Conroy, S. A. and Gupta, N. (2018). “Disentangling Horizontal Pay Dispersion: Experimental Evidence.” *Journal of Organizational Behavior*, 40, 248–63.

DeVaro, J. and Heywood, J. (2017). “Performance Pay and Work-Related Health Problems: A Longitudinal Study of Establishments.” *Industrial and Labor Relations Review*, 70, 670–703.

DeVaro, J. and Kurtulus, F. A. (2010). “An Empirical Analysis of Risk, Incentives, and the Delegation of Worker Authority.” *Industrial and Labor Relations Review*, 63, 637–57.

Ederer, F. and Manso, G. (2013). “Is Pay for Performance Detrimental to Innovation?” *Management Science*, 59.

Kauhanen, A. and Napari, S. (2012). “Performance Measurement and Incentive Plans.” *Industrial Relations*, 51, 645–69.

Larkin, I., Pierce, L., and Gino, F. (2012). “The Psychological Costs of Pay-for-Performance: Implications for the Strategic Compensation of Employees.” *Strategic Management Journal*, 33, 1194–1214.

Lazear, E. P. (2000). “Performance Pay and Productivity.” *American Economic Review*, 90, 1346–61.

Manthei, K., Sliwka, D., and Vogelsang, T. (2018). “Performance Pay and Prior Learning: Evidence from a Field Experiment.” *IZA Discussion Paper* No. 11859.

Pencavel, J. (2001). *Worker Participation: Lessons from the Worker Co-ops of the Pacific Northwest*, New York: Russell Sage Foundation.

10 Executive Compensation and Stock Options

In 2012, the board of directors of Tesla Motors Inc. (TSLA) set an executive compensation contract for the CEO, Elon Musk, who also served as chairman of the board until the fall of 2018. The plan was aggressively focused on performance pay, and particularly stock options. It was designed to generously reward major successes years into the company's future. According to the 2016 proxy statement, "The 2012 CEO Grant was designed to be entirely an incentive for future performance that would take many years, if at all, to be achieved." The plan would grant Musk up to \$1.6 billion worth of stock options by 2022 if he could achieve challenging targets that would benefit Tesla's shareholders and workers. Musk has to meet dual targets to get paid. For every target related to the stock price, another requires that products be brought to market that are better than any before them. His base salary of \$37,584 is negligible as a fraction of his total compensation, and he has never even bothered accepting it.

Musk's compensation contract, which is explored further in the Tesla Motors case discussion, is a particularly extreme version of what is generally true of executive compensation contracts. They involve extraordinarily high pay, as I mentioned in section 1.9. They are also heavily focused on performance pay, using broad measures of organizational performance (like stock prices). For this reason, they are extremely risky. CEO pay contracts are heavily laden with stock options, which could earn the CEO a fortune if the firm does well but which could end up being worthless if the firm does poorly. Such risky compensation plans require paying the CEO a hefty risk premium at the company's great expense.

Millions of people are employed by firms. But each firm has only one CEO. So, statistically speaking, it's unlikely that you are, or will ever be, a CEO. Why, then, should you bother reading a chapter on executive compensation? Even if you're never a CEO, you might find yourself as a director on a firm's board, and it's the job of the board of directors to design the CEO's compensation contract. Also, such boards often rely on the advice of compensation consultants, so you might find yourself in the role of compensation consultant. But even if you're never a CEO, a director on a board, or a compensation consultant, I'll give you two good reasons to read this chapter.

First, studying executive compensation contracts deepens your understanding of compensation contracts in general. For example, understanding the incentives that are created by executive bonuses will help you to better understand how incentives work in non-executive performance-pay plans, even when they are designed differently from executive bonuses. Second, this chapter isn't just about executive compensation. It is also about stock options. Although stock options are a significant, and often the most important, component of CEO pay (Elon Musk's contract is a great example) they are also commonly granted to non-executive workers in so-called "*broad based stock options*" plans, so it is important to understand how they work and fit into a worker's total compensation package. Employee stock options that apply to non-executives are particularly common in industries or companies that focus on "upside" returns, such as technology companies, firms with investments in R&D, startups, and venture capital firms.

10.1 CEO Pay

In this chapter, I'll focus on the *design* of executive compensation rather than on its *level*. There are four components of most executive compensation packages: (1) base salary, (2) benefits, (3) bonuses, (4) stock options. This chapter focuses on items (3) and (4) rather than on (1) and (2). One reason for omitting (1) and (2) is that these are the least interesting of the four components from the standpoint of creating CEO incentives. Recalling the $f(P)$ function from Chapter 9, the graphs of components (1) and (2) are both flat lines, like in panel *a* of Figure 9.1. In contrast, bonuses and stock options have positively sloped graphs for at least some values of P, and for that reason they create incentives. A second reason for omitting items (1) and (2) from our discussion is that both of those components aren't much different for CEOs than for other workers (apart from having significantly higher levels for CEOs). So the same considerations that we've discussed elsewhere about base salaries and benefits also apply to CEOs. For example, everything in the discussion of benefits in Chapter 11 also applies to executive benefits.

At large public companies, CEO compensation is set by the board of directors. The fact that the CEOs themselves often sit on these boards raises concerns that CEOs have too much influence over their own compensation. For example, as I mentioned at the start of the chapter, until the fall of 2018 Elon Musk was both the CEO of Tesla and the chairman of its board, a situation known as "CEO duality" which is quite common in US companies. Many of these boards have interlocking or overlapping directorates, meaning that two different boards share some of the same directors. There might even be overlap in the executive compensation committee across these boards of directors, which raises further questions about how independently and objectively executive compensation is determined. Boards also overlap in the compensation consultants they use (i.e., two different boards may share the same compensation consultant). These overlaps across different boards, both in their directors and in their compensation consultants, result in similarity in the design of executive pay plans across different companies.

10.2 Executive Bonuses

Executive bonuses are standard in for-profit companies and apply to CEOs and other top executives. They are typically paid annually based on the preceding year's performance. Bonus contracts are explicit or "formal", meaning that their terms are clearly written down on paper, with limited room for discretion on the part of the board of directors. Discretion is limited by basing the bonuses on "objective" performance measures (see Chapter 9), though recall that even objective measures can be subject to influence. Section 10.5 discusses the distinction between formal and informal contracts in the context of CEO bonuses. Even if there is a formal bonus contract, any room for discretion to enter that contract gives it the flavor of an informal contract (i.e., one that is not written down and, therefore, is harder to enforce in court) which raises the possibility of the company reneging, in whole or in part, on the contract.

In fact, there are three ways in which discretion on the part of the board of directors might creep into the bonus system even with a formal contract. First, boards can make discretionary adjustments to reported earnings numbers on which the bonuses are based. Second, a modest fraction of the bonus (up to about 25%) is sometimes based on individual performance, as determined by the board, and this can be subjective in the same way that the performance ratings of Borders' district managers (see Chapter 9) were subjective. Third, boards of directors might have discretion in allocating a fixed bonus pool among the CEO and other top executives.

Executive bonuses are sometimes called "80/120" plans, for reasons that will soon be clear. These plans have three basic components (performance measures, performance standards, and the structure of the relationship between pay and performance). Let's consider a particular example, starting with some notation. As in Chapter 9, P represents the CEO's performance measure. Suppose that P_O represents the *performance standard*, which you can think of as a performance target that the CEO is expected to achieve. Also, suppose that $P_{80\%}$ is the performance level equaling 80% of the target performance. That is, $P_{80\%} = 0.80P_O$. Similarly, suppose that $P_{120\%}$ is the performance level equaling 120% of the target performance. That is, $P_{120\%} = 1.2P_O$.

The annual bonus varies according to the CEO's performance level. If the CEO fails to hit at least 80% of the target performance level, P_O, then no bonus is paid. But if the CEO manages to achieve exactly the magic level of 80% of the performance target, then a bonus is paid, though it is the minimum possible bonus that could ever be paid out under the plan. Let's call that minimum bonus B_{min}. If the CEO manages to exceed 80% of the performance target, then a bonus higher than B_{min} is paid. The amount of the bonus increases with the performance level. But that is only true up to a certain point. If the CEO manages to achieve a performance equal to 120% of the performance target, then the maximum possible bonus is paid. Call that maximum bonus B_{max}. Performance levels that exceed 120% of the performance target do not earn the CEO any additional bonus payment . . . the bonus is capped at B_{max}. If the CEO exactly achieves the performance target of P_O, then the target bonus is paid. Call it B_O.

Recalling the $f(P)$ function from Chapter 9, the annual bonus function (which is nonlinear) is expressed as follows:

$$\begin{aligned} f(P) &= 0 && \text{if } P < 0.8P_0 \\ &= B_{min} + \beta(P - 0.8P_0) && \text{if } 0.8P_0 \leq P < 1.2P_0 \\ &= B_{max} && \text{if } P \geq 1.2P_0. \end{aligned}$$

See Figure 10.1 for a graphical representation of $f(P)$. The graph has three segments, corresponding to the three "=" signs in the preceding expression for $f(P)$, and notice that the first segment coincides with the horizontal axis. The first and last segments are flat (meaning $\beta = 0$, so that no incentives are created) whereas the middle segment is upward sloping (meaning $\beta > 0$, so that incentives are created). The middle segment, which is just an upward-sloping straight line, resembles panel *b* of Figure 9.1, which showed a piece-rate compensation contract. You can think of the middle region of the executive bonus in Figure 10.1 much like a piece rate, in which β represents the piece rate (i.e., the bonus paid per unit of performance) and the "pieces" are units of performance, P. The middle region of Figure 10.1 is called the "incentive zone" because it is the range of performance values for which incentives are created for the CEO. Within this region, an additional unit of performance earns the CEO β extra bonus dollars.

Figure 10.1 is just one example of an 80/120 plan. The graphs for other 80/120 plans might look somewhat different. For example, the middle segment (i.e., the incentive zone) need not necessarily be a straight line, though it must have a positive slope. It could be nonlinear and have various segments within it. And, in fact, the shape of the graph in the incentive zone tends to vary across industries. Also, even though these plans are called "80/120" plans, the performance cutoffs don't always have to be 80% and 120%. For example, they could be 90% and 110%, or 85% and 115%. They don't even have to be symmetric around 100. For example, nothing stops the board from setting the incentive zone between 80% and 140%.

There are two "kinks" in the graph in Figure 10.1, meaning two points at which the slope of the graph changes. They occur at $P = 0.8P_O$ and at $P = 1.2P_O$. At the first kink, where $P = 0.8P_O$, the slope changes radically from its original level of zero to

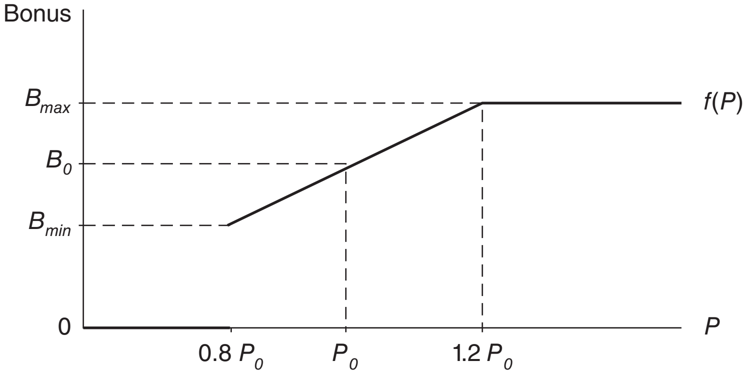


Figure 10.1 Annual executive compensation "80/120" bonus.

infinity because the $f(P)$ graph suddenly leaps up from a bonus of zero to a bonus of B_{min}. The change in slope that occurs at the second kink (where $P = 1.2P_0$) is less dramatic though still significant; it decreases from β to zero.

The reason that the kink points are worth your serious attention is that, as we saw in Chapter 9, positively sloped regions of $f(P)$ create incentives, and steeper positive slopes create even stronger incentives. So when the slope of $f(P)$ changes, so do the incentives faced by executives. At the first kink, where the slope changes from zero to infinity, the change in incentives is intense. For performance levels well below $0.8P_0$ the CEO gains no reward from performing a little better, because the bonus remains at zero. But when performance is on the verge of achieving the kink point at $0.8P_0$, performing just a little bit better generates a huge reward, i.e., the bonus jumps from zero to B_{min}. Similarly, consider the second kink point. If performance is at or above the kink point of $1.2P_0$, then performing a little better won't generate any additional payout for the CEO, because the bonus is B_{max}, regardless. But at performance levels slightly below (or clearly below) the kink point, performing a little better raises the CEO's bonus; more precisely, one additional unit of performance, P, increases the CEO's bonus by β.

Incentives are particularly powerful near the kink points. Imagine that the CEO of your organization is paid according to the bonus plan in Figure 10.1 and that her annual performance falls slightly below the kink at $0.8P_0$. The CEO might go to great lengths to try to manipulate the performance measure, P, higher so that it hits or surpasses $0.8P_0$. Assuming that P is measured as company profit or a related variable, one way to achieve this is to shift revenues or costs across calendar years. This might be done, for example, by front-loading capital expenditures or by delaying them until after the calendar turns to the next year.

I discussed this type of manipulation in section 9.8 when I mentioned the possibility that district managers at Borders might shift monthly sales revenues for their district either forward in time or backward in time according to how the bonus plan was set. If the CEO of your firm has hit and exceeded $1.2P_O$, then the CEO has incentives to manipulate the performance measure downward, by shifting some profit from this year into next year. This strategy has the advantage of not "wasting profit" (i.e., generating profit that won't add to the CEO's bonus), and it allows the CEO to get a head start on next year's profit, so that enough is generated to qualify the CEO for a nice bonus next year. Academic research has shown evidence consistent with executives manipulating performance measures as just described, in the neighborhoods of the kink points of their bonus contracts.

One of my MBA students who previously worked in the retail sector described an annual store-wide bonus schedule based on "comp sales", which is a common metric in the retail sector that just refers to current sales compared to last year's sales. Each store in the retail firm got a bonus if it beat its previous year's sales by at least 2%. This creates a discontinuity (or "jump") in the bonus graph, at the 2% mark, much like the one that occurs at a performance level of $0.8P_0$ in the executive compensation bonus schedule depicted in Figure 10.1. Suppose that sales are unusually bad for the retail store in a particular year, and that they drop significantly compared to the previous year. How should the store's management behave over the next several years? The bonus schedule creates perverse incentives for managers to benefit at the expense of shareholders. The shareholders would want the store to recover as quickly as possible from the anomalous bad year, perhaps posting double-digit sales growth over the next couple of years to return to the "normal" level that prevailed before the bad year. But that would quickly reset the manager's "bar" to a higher level, making it harder to achieve a 2% gain over that level. An alternative approach would be to post modest sales growth of only 2 to 3% per year over the next five to ten years, taking a longer time to return to the "normal" level of sales but ensuring that the annual bonus is collected for a larger number of years. The bonus schedule rewards high current performance with a bonus, while at the same time punishing high current performance with the prospect of lower future bonuses due to a higher bar. This built-in punishment (which savvy employees are pretty good at anticipating) resembles the ratchet effect (see section 9.9). Just like the executives faced with the bonus schedule in Figure 10.1 have no incentive to exceed a performance level of $1.2P_0$, the managers at the retail establishment have no incentive to post annual sales growth greater than 2%, even if double-digit sales growth is feasible.

The incentives created by executive bonuses, such as the one depicted in Figure 10.1, tend to be rather short-term in nature. The performance measure, P, tends to focus on short-term performance, and a portion of it may even include a subjective assessment (by the board of directors) of the CEO's performance during the past year. Longer-term incentives are provided through the equity-based compensation components discussed in the next couple of sections.

10.3 Equity-Based Compensation

Equity is an ownership share in a company. When some components of a compensation package are equity based, the worker's financial fate is tied to that of the organization. When the company does well, or poorly, the same will be true of the value of the employee's compensation package. This gives workers a strong reason to hope for the company's success and to do everything they can to realize that success. But it also exposes workers to a lot of risk and necessitates paying them hefty risk premia, which can get expensive for the company. Equity-based compensation plans are performance-pay plans (Chapter 9) in which the performance measure, P, is the broadest possible measure of firm performance, namely the stock price. Equity-based compensation plans such as profit sharing shift an increasing share of the company's ownership into the hands of its workers. See the comparison of Amazon and Sears in the case discussion. The most extreme form of profit sharing occurs when the firm is entirely owned by its workers, in which case all of the firm's profit goes into its worker-owners' hands, as in The Cheese Board Collective case discussion.

One form of equity-based compensation is *restricted stock grants*. These are grants of stock that are offered to the worker at a reduced price but with some strings attached that restrict the worker's ability to sell. For example, the worker might not be allowed to sell the shares until a specific date. In that case there is effectively a vesting schedule. During the global financial crisis of 2008, when caps were placed on executive compensation by the ARRA (see sections 4.4 and 4.7.4) there was an exception made for restricted stock grants. That is, even though the other components of executive compensation packages were capped for companies in receipt of public bailout assistance, those companies could award as much compensation in restricted stock grants as they wished. The loophole allowed those companies to make generous awards of restricted stock grants to high-performing executives who might otherwise quit.

The most prevalent form of equity-based compensation is stock options. In the next section I explain how options work. Options can be awarded as part of compensation packages both in privately owned and in publicly owned companies. Most of this chapter will deal with publicly traded companies. Private companies, such as startups, often award equity as a percentage of ownership in the company. If those companies "go public" (meaning start selling stock to the public) then an individual employee's ownership share may or may not change depending on how the initial contract was written. Another possibility is that a privately owned company does not go public but rather is bought, and in that case the workers who have ownership shares get a piece of the sales price in proportion to their ownership.

What your workers should worry about, when paid in the form of ownership shares in a privately owned company, is dilution, which occurs when you issue a lot of new shares, so that even if a particular worker had a 1% ownership share when the equity-based compensation was awarded, that percentage would diminish as you add new shares. A formal contract that protects the worker against dilution would avoid this situation. That is, the contract would state that the worker is guaranteed a 1% ownership share when the company goes public, regardless of how many new shares are issued. Employment offer letters become long and involved in such situations, to clearly detail exactly what is and is not being promised.

In the case of publicly traded companies the situation is more straightforward. Workers can simply purchase stocks at a discount, directly from their paychecks through payroll deductions. Public companies face regulation from the Securities Exchange Commission (SEC) and, among those regulations, companies are not allowed to sell shares to different workers at different prices.

Equity-based compensation often has a vesting schedule attached to it, as I mentioned earlier concerning restricted stock grants, meaning that workers don't have access to the full amount of stocks until a certain amount of time has elapsed. Such vesting serves as a retention device, because if your worker leaves the firm before the equity-based components of compensation vest, then the worker loses them. If your worker receives a grant of equity-based compensation (e.g., stock options) when starting at your firm, the shares typically vest over a period of time (say, four years) so that the closer the worker makes it to the four-year mark before quitting, the more of the shares that vest. And if the worker makes it the full four years, then all of the shares vest and the worker would lose none of them by leaving your firm. Often by that point the firm has granted the worker some new shares that have not yet vested, so that the worker always has to leave some money on the table when quitting the firm. By the way, the same concept of vesting also applies to defined-benefit pension plans (Chapter 11).

"Cliff vesting" means a sharp change in entitlement to the shares once a certain time threshold has been crossed; for example, the worker might lose all of the shares if they quit before the one-year mark. But if they hit that one-year milestone, then perhaps 25% of the shares automatically vest, and then each month thereafter that percentage increases until it hits 100% at the end of four years. This would be an example of a one-year cliff, because of the sharp change that occurs at the one-year mark. Typically as long as the shares are vested, your worker can hang onto them even after leaving the company.

10.4 Stock Options

A big fraction of CEO pay, particularly in the United States, is equity-based, as is the case with Tesla's CEO Elon Musk (see the case discussion at the end of this chapter). The primary compensation vehicle that is used is *stock options*, which are an extremely risky form of compensation. Therefore, recalling Chapter 9, we should expect the CEO's compensation to incorporate a hefty risk premium if much of the pay comes in the form of stock options, as it does in Elon Musk's case. The use of stock options in executive compensation contracts has increased over time, particularly in the United States. Many aspects of stock options compensation could be explored, but in this section I will focus primarily on the role of options in creating worker incentives, consistent with the incentive and sorting effect themes that permeate this book (see section 1.8). Except for some remarks in section 10.4.1 on private companies, I'll focus on publicly traded firms throughout the discussion.

10.4.1 Definitions

Stock options are of many types. Those that are commonly traded are called ***vanilla options*** (or ***standard options***), and those that are more complex are called ***exotic options*** (or ***non-standard options***). The options that are used in compensation packages are exotic, for reasons I'll soon explain. Nonetheless, much of the following discussion, particularly in section 10.4.2, focuses on vanilla options. The reason is that "simple" formulae have been developed that are useful for valuing vanilla options, but things get more complicated in the case of exotic options. Fortunately, most of the general intuition from studying vanilla options also extends to exotic options.

Options can be either ***call*** or ***put*** options, but I'll only discuss call options because put options are typically not used in compensation plans. A person who owns a call option for a particular stock has the right (though no obligation!) to buy a share of that stock at a specified price (called the ***exercise price***, ***strike price***, or ***striking price***) on or before a predetermined future ***maturity date*** (or ***expiration date***). I say "on or before" the maturity date, but some types of options can only be exercised exactly on the maturity date, as I'll explain soon. If an option holder chooses to exercise their right, this is known as ***exercising the option*** (or ***exercising the options***, if they hold more than one option and choose to exercise a bunch of them all at once, as is typically the case). A typical option contract has a ***contract multiplier*** of 100, i.e., the contract controls 100 shares of underlying stock, which implies that the price of the option must be multiplied by 100 to get the value of the entire option contract (don't fret, we'll get to the concept of "value" in section 10.4.2). For simplicity, let's assume throughout our discussion that the contract multiplier is simply 1, as I implicitly did earlier in this paragraph, so that the option contract controls only one share of stock.

Again, a call option is simply the right to buy a share of stock at a specified exercise price at some specified future date(s). Why would someone want to do that? So that they could then sell that share at an even higher price! For that strategy to be profitable, however, the stock price that prevails on the future date on which the option is exercised must exceed the predetermined exercise price, which leads us to our next definition. The ***moneyness*** of an option refers to the position of the current stock price relative to the exercise price. If those two prices are equal, the option is said to be "***at the money***". If the strike price is below the current stock price, the option is said to be "***in the money***". If the current stock price is above the exercise price, the option is said to be "***out of the money***". An option's moneyness determines whether any money can be immediately made from the option. In short, an option will only be exercised when it's in the money.

Vanilla options come in two different flavors: American and European. ***American-style*** options can be exercised on or before the maturity date. ***European-style*** options cannot be exercised early and can only be exercised on the maturity date, though they can be sold to another buyer prior to the maturity date. Most options traded on exchanges are American style. If two options are identical in all respects *except* that one is American style and the other is European style, the American-style one is worth more. The reason is intuitive and echoes our discussion of compensation constraints in section 4.3. A European-style option is just an American-style option with an additional constraint, namely that the only valid date on which the option can be exercised is its maturity date. As I noted in section 4.3, constraints make you worse off. Nothing stops the holder of an American-style option from waiting until the maturity date to exercise it, in which case the profit obtained would exactly match that of a European-style option. But the holder of the American-style option has a range of additional possibilities (i.e., earlier dates on which to exercise) that the holder of the European-style option lacks, and those have value.

The exotic options that are used in compensation plans are more complex than the preceding ones, because of the additional complication of ***vesting***. A person who receives options as part of their compensation package is not allowed to immediately exercise the options . . . they are required to hold the options until a certain amount of time has passed, perhaps three to five years, which is called the ***vesting period*** or ***vesting schedule***. Once the vesting period is completed the options are said to be ***vested***, which means they are eligible to be exercised. Such options that are granted in compensation packages are similar to ***forward start options*** with an early exercise possibility. The phrase "forward start" signals that a vesting period must be completed before the option can be exercised, and the phrase "with an early exercise possibility" means that the option can be exercised (once it is vested!) at any time on or before its maturity date, just like American-style options. Typically the options granted in pay packages are issued with exercise prices that are "at the money" at the time the contract is granted, so even though the stock price may fluctuate throughout the vesting period, the worker knows what the exercise price will be when the vesting period ends and the option becomes active.

The vesting schedule of an option has an important effect on the value of that option to the person who holds it. If one of your workers holds options that will vest

19 months from now, but that worker quits today to take a job at another firm, that worker forfeits those options, meaning that the options become worthless. The same thing happens if the person who holds the options allows them to expire, by forgetting to exercise them on or before the maturity date. Vesting schedules can be a powerful way to deter turnover. Workers are reluctant to quit before their options vest, since doing so means leaving money on the table. If you periodically give your workers new options grants (with new expiration dates), then even when some of their earlier grants vest, some of their more recent grants will remain unvested, so that your workers are always forced to forfeit some options if they quit. Often an options grant vests in a staggered way over time. Perhaps 25% vest after one year (which is called a ***one-year cliff***, echoing the cliff vesting for pensions that I'll describe in Chapter 11) and then an additional percentage vests in each subsequent month until 100% of the initial grant vests after four years. Again, this means that a worker who quits before working four years will leave money on the table, though the amount that is left on the table continuously shrinks as the four-year mark approaches.

The purpose of stock options as part of a compensation package is to tie your employees' compensation to company performance, as measured by the stock price. This induces your workers to behave in the interests of the firm's shareholders, i.e., by maximizing shareholder value. This approach works best for CEOs and other high-level executives whose decisions can be expected to have a substantial influence on the success or failure of the organization, whereas lower-level workers can't individually affect firm value very much. For those folks, giving them options is akin to giving them a lottery ticket tied to the employer's financial fate. They will certainly hope for the success of the company, but they have little reason to believe that their own behavior has any significant effect on the company's success or failure.

The performance measure, P, in a stock options compensation plan is the stock price. The P stands for ***p***erformance measure, as in Chapter 9, but also for stock ***p***rice. As I mentioned in Chapter 9, the stock price is a very broad measure of performance. Sometimes narrower measures (e.g., earnings or accounting profit) are used in other components of executive compensation contracts, such as the bonuses described in section 10.2. Let's use the notation "E" as shorthand for ***exercise price***. If the stock price exceeds the exercise price, i.e., if $P > E$, then the person who holds the option can make money by exercising the option. How? Remember that a (vested) option allows its holder to buy a share of stock at a price of E. They can then sell it at the higher price, P, which is the current stock price, thereby earning a monetary profit of $P - E$.

Exercising a vested option and converting it to cash is a quick and easy process if the company that issues the stock (and therefore the options) is a publicly traded company. In this case, the ownership of the company is dispersed among the general public, and shares of the company stock are freely traded on a stock exchange. This means that there are lots of buyers and sellers of these shares of stock, a situation sometimes called a "thick market" (or liquid market) for the stock, so when the option is exercised (i.e., a share of stock is bought at a price of E) it is easy to find a buyer to whom the share can be sold at a price of P. Moreover, the price P (though it fluctuates from day to day, and even minute to minute, like most stock prices on an exchange) can be readily observed and identified by everyone at every moment in time.

The situation is quite different if the company is not publicly traded (i.e., has not "gone public" and is therefore private). In that case, the ownership shares are held by a smaller number of people rather than by the general public, and they are not actively traded on an exchange. So when an option is exercised (i.e., a share of stock is bought at a price of E) it can be rather difficult to find a suitable buyer. There are far fewer buyers and sellers of shares than is the case for a publicly traded company, and this situation is sometimes called a "thin market" (or illiquid market) for the stock. In the case of a private firm, the stock price, P, represents the maximum amount that a buyer would pay for a share of the company's stock. If such a buyer can be found, and if $P > E$, then the person who is exercising the option in a private company can make a profit of $P - E$, just like in a publicly traded company.

But again, finding such a buyer isn't always easy. I have a friend at a private company who planned to exercise some of his options, which would have involved purchasing a bunch of shares of his company's stock at exercise prices ranging from \$0.25 to \$1, for a total cost of \$5000. He had to hunt to find prospective buyers of the shares, eventually talking to four prospective buyers, only one of whom was potentially interested. That potential buyer was willing to pay \$4 per share (i.e., $P =$ \$4). However, there were strings attached. My friend was asked by that buyer's lawyers to sign a contract that insured the buyer against the following negative outcome. If my friend's employment were to be terminated by his company, then the company reserved the right to buy back the ownership shares that he had held (and sold). This could mean, hypothetically, that the potential buyer (who wanted to buy the shares from my friend at a price of \$4 per share) might have to sell those shares back to the company at a price that could potentially be below \$4 per share, which would create a loss for the potential buyer. The potential buyer therefore requested that my friend sign a contract agreeing to reimburse the potential buyer to cover the amount of that loss, should it ever happen. My friend was unwilling to sign that contract and therefore decided not to exercise his options.

The bottom line is that converting vested options into cash is typically harder in a private company than in a public one. Workers often need to locate potential buyers themselves, and then the sale price, P, is determined by negotiation and bargaining (see Chapter 14) rather than by the active market competition that occurs when a public company's stock is traded on a high-volume exchange. As a consequence of P being harder to determine in the case of private companies than in publicly traded ones, the value of stock options is also harder to determine. Henceforth, unless I note otherwise, this chapter focuses on publicly traded companies.

10.4.2 Value of Stock Options

An option's total value is called the option premium, and it has two parts:

$$\textit{Option premium} = \textit{Intrinsic Value} + \textit{Extrinsic Value}.$$

The first part, ***intrinsic value***, we discussed in section 10.4.1. It is the immediate monetary profit that can be made today if the option is exercised. Another way to think about this is, suppose that the maturity date has been reached so that it's your last chance to exercise the option before it expires. How much money could you make? If P exceeds E, then you can buy at a price of E and sell at a price of P to make profit (or intrinsic value) of $P - E$. But if P is less than or equal to E, then the option's intrinsic value is zero. The option will not be exercised in that case, because buying at a price of E and then selling at a lower price of P would mean taking a loss rather than making a profit. So an option's intrinsic value can be expressed using the following formula:

$$\textit{Intrinsic value} = \max(P - E, 0).$$

The "max" is a function that simply tells you to choose the larger of the two items in parentheses. That will be $P - E$ whenever P is larger than E, and it will be zero otherwise. An intrinsic value of zero means that the option offers the option holder zero monetary profit at this moment in time. Always remember that the intrinsic value can never be negative! The reason is that the option holder would never voluntarily exercise an option that would yield a negative profit. They'd just hang on to the option, hoping that in the future P would increase by enough that they could make some money by exercising it, which brings us to our next point . . .

An option's ***extrinsic value*** (or ***time value***) captures the value that the option holder enjoys from the possibility that the stock price will increase between now and the maturity date, which would increase the option's intrinsic value and therefore the profit that can be made from exercising. To elaborate, suppose that you hold an option that expires one year from now and that currently it is "out of the money", i.e., $P < E$. The option's intrinsic value is zero. But that doesn't mean that the option is worthless and that you should throw it away. Why? Because the stock price might increase during the next year! Maybe seven months from now, when the option is still five months to maturity, P will exceed E, and in that case your option will suddenly be "in the money" and you can exercise it to make a profit. That "hope for the future", i.e., that the stock price will grow in the future and increase the intrinsic value of the option, is captured by the extrinsic value.

Of course, there's only "hope for the future" when there actually ***is*** a future! Once the maturity date is reached, an option's extrinsic value shrinks to zero, and its total value (i.e., option premium) coincides with its intrinsic value. By the same token, the longer into the future the option's maturity date lies, the greater the extrinsic value will be. As long as there's still time left on the clock before the maturity date hits, then extrinsic value is positive. That must mean that the option premium (which is the sum of intrinsic and extrinsic value) must also be positive, because intrinsic value can never be less than zero. To summarize, the option premium is always positive unless the maturity date is reached and intrinsic value is zero (in which case the option premium must also be zero, given that it coincides with intrinsic value once maturity is reached).

The Black–Scholes–Merton (BSM) formula was developed to explain the market value of European-style, exchange-traded options rather than those that are part of compensation packages. The formula is sometimes used by firms to value employee options, but it considerably underestimates the value that workers place on the options in their compensation packages. The central issue is risk. Risk-averse workers who accept options as compensation require risk premia of at least 30% above-and-beyond the BSM valuation. This is familiar logic that we've now seen a number of times, such as in Chapter 2 where employees needed to be paid a risk premium to compensate them for the risk of wage theft. I also had a lot to say about this in Chapter 9, because income risk is an inherent part of any performance-based pay plan, including stock options compensation. You might be wondering why I'm wasting your time on a formula that underestimates employees' valuations of their stock options to such a substantial degree. The reason is that the formula provides some helpful intuition for thinking about employee valuations. I will not present the actual BSM formula, which would involve a lot of new technical notation, but if you're interested you can find it online or in any finance textbook.

To get a feel for employees' options valuations, let's investigate the role of four individual factors that influence valuation. The first is P, the stock price. The second is E, the exercise price. The third is the maturity date, and the fourth and most important is volatility. Much of this discussion is based on intuition derived from the BSM model and its variants.

First, when the stock price, P, increases, the option premium increases, meaning that the worker values the option more. This is obvious if the option is either "in the money" or "at the money" prior to the increase in P. In both of those cases the intrinsic value of the option is $P - E$, which obviously increases as P increases. Moreover, if the intrinsic value increases, then so must the option premium, which is just the sum of intrinsic value and extrinsic value. Next, suppose that the option is "out of the money" rather than "at" or "in" the money prior to the increase in P. If the increase in P is big enough for P to surpass E, then the option will switch from being "out of the money" to being "in the money", and intrinsic value will increase from 0 to $P - E$. But suppose instead that the option remains "out of the money" even after P increases, albeit by a lesser amount. Even then the option premium increases! Clearly, this must be due to the extrinsic value, because intrinsic value is zero before and after the increase in P. Remember that extrinsic value captures “hope for the future”, and what you’re hoping is that ***eventually*** P will increase beyond E. The prospects of that happening are brighter after P has increased than before it increased. In other words, being slightly out of the money is better than being far out of the money. This argument requires sufficient stock price volatility (more on that below!). It also requires that there be some “time left on the clock”, because remember, extrinsic value hinges entirely on there being some time left before the maturity date, to allow a chance for good luck to strike. The bottom line? Whether the option is “in the money”, “at the money”, or “out of the money”, it’s worth more to your employee when the stock price is higher.

Second, consider the exercise price, E. The higher that number is, the lower the option premium will be. After all, if your worker exercises the option, the profit made will be the option’s intrinsic value, namely $P - E$, and that obviously shrinks when E grows. Another way to state the same thing is in terms of a drop in E rather than an increase in E. Your worker’s valuation of the option increases when E shrinks, and if E shrinks all the way to zero, then the option is exactly equivalent to a share of stock. Why? Remember the definition given in 10.4.1. A call option allows its holder to purchase a share of stock at a price of E, at which point the share could be sold at its market price of P. But buying a share at a price of zero is effectively the same as being given a share for free . . . in either case you can sell the share, if you wish, at the market price of P, making a profit of $P - 0$, or simply P. If a worker has an option with $E = 0$ (which is the right to hold one share of stock at no cost) the worker should exercise the option immediately so that money can be made off of the stock’s dividends. Buying the share of stock at the exercise price of zero simply means holding the share. The bottom line is that options with lower exercise prices are worth more, holding all else equal, and the most valuable option of all is one with an exercise of zero (which is identical to simply a share of stock).

Third, let’s imagine pushing the option’s maturity date further into the future, and remember that with employee stock options (just like with American-style options) the holder is permitted to exercise early and need not wait until the maturity date. The more distant the maturity date, the greater the option premium. The reason is simply that the longer the time that exists between the present date and the expiration date, the more time there is for P to grow and to push the option in the money (or even *further* in the money if it’s already in the money). In contrast, if the maturity date is at hand, so that P no longer has a chance to grow, then the market value of the option coincides with its intrinsic value, because the holder must immediately exercise the option if it is in the money (making profit of $P - E$). So the further into the future the maturity date, the greater the amount by which the option premium exceeds the intrinsic value, i.e., the greater the option’s extrinsic value. Another way to think about this is that extrinsic value is all about the value of "time on the clock", so the more time there is on the clock, the more your employees will value their options.

Fourth, let's consider the ***volatility*** of the stock price on which the option is based. Volatility measures how much dispersion there is in the stock price, P. When you think of dispersion, standard deviations quickly come to mind, and indeed that concept is relevant here. Technically, however, volatility in the BSM model is not the standard deviation of the distribution of stock prices, P. Rather, it is the standard deviation of the distribution of the *natural logarithm* of ***stock returns***, i.e., $ln(P_t / P_{t-1})$. Here the subscripts represent time periods, so P_t is the stock price in period t, and P_{t-1} is the stock price in period $t - 1$. Note that $ln(P_t / P_{t-1}) = ln(P_t) - ln(P_{t-1})$, which is the ***percentage change*** in the stock price between periods $t - 1$ and t. This rendering of percentage changes using logarithms echoes our discussion from section 6.5 on whether to analyze compensation variables in "levels" or in "logs". I mentioned there that when the dependent variable in a compensation regression is measured in logarithms (as opposed to levels) the regression coefficient describes the *percentage change* in the dependent variable when the right-hand side variable to which the regression coefficient pertains increases by one unit.

In any event, volatility is technically the standard deviation of $ln(P_t / P_{t-1})$, but for our present discussion it's okay to just think of it as an indicator of how much dispersion or day-to-day fluctuation there is in stock prices. If stock prices were to never change day to day, then volatility would be zero. In practice, of course, stock prices go up, and they go down, so volatility is positive. But that volatility is a larger positive number for some stocks than for others. The key point to remember, and possibly the most important point of our entire discussion of options, is that *options have higher total value when their volatility is higher.* The reason is that if there are big swings in P, there's a greater shot of winning the jackpot by getting an extremely large value of P (in which case the option can be exercised to make a hefty profit of $P - E$).

You might think that this advantage is cancelled out by the fact that a big swing in P could also lead to a crash in the stock price (i.e., a very low P that puts the option far out of the money). But that situation isn't disastrous, because the option holder simply wouldn't exercise the option in that case. Whether the option is slightly out of the money, or far out of the money, the option holder won't exercise it, and the intrinsic value is just zero in both cases. So there's an inherent asymmetry that the option holder benefits from, in that large values of P guarantee big payoffs, but crashes in P have limited downside because options are simply never exercised when they're out of the money. Remember the intrinsic value formula, i.e., $\max(P - E, 0)$, which automatically ensures against downside risk because it can never drop below zero. The bottom line is . . . *when it comes to the value of options, the more volatility the better*!

Figure 10.2 illustrates the preceding point. Suppose there are two different stocks, one with low volatility and the other with high volatility. Both have the same

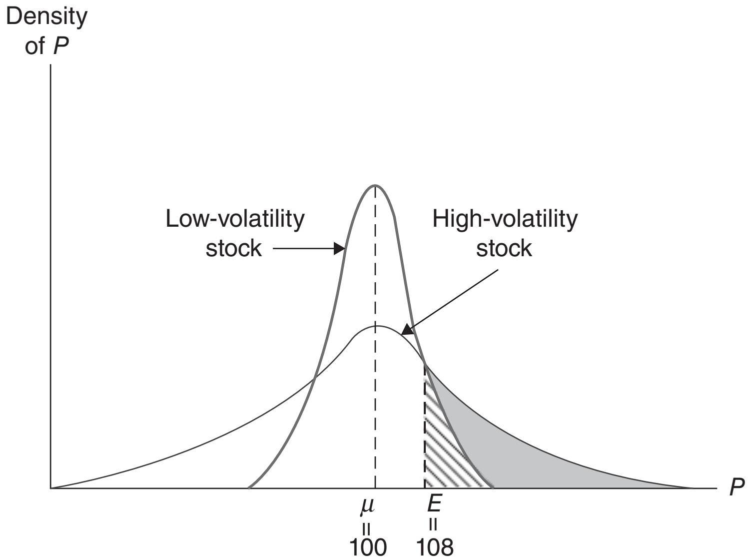


Figure 10.2 Stock options for 2 different stocks with the same expected price, μ.

expected stock price, μ, and let's say that's 100. Even the low-volatility stock has *some* volatility, meaning that there's uncertainty about what its future price will be even though the *expected* price is 100. This uncertainty about the future value of the low-volatility stock is represented in Figure 10.2 by the boldface graph, which is a bell-shaped curve centered about the expected stock price of 100. This boldface graph is called the ***density function*** for the low-volatility stock's price. If you're unfamiliar with the concept of a density function, it's like a smooth version of a histogram (such as those displayed in section 6.5 for wage levels and log-wages). As the boldface graph shows, the stock price, P, for this low-volatility stock might end up below 100, or above 100, but it won't deviate wildly from 100 because the density function surrounds the value 100 fairly tightly. In contrast, the density function for the high-volatility stock (depicted as a "regular", as opposed to boldface, graph) exhibits a lot more dispersion. Like its low-volatility counterpart, it has an expected stock price of 100, but unlike its low-volatility counterpart, it could deviate from 100 significantly.

Now let's consider stock options that are based on these two different stocks, and let's suppose that the options have the same exercise price, E, of 108. Even though both options have the same exercise price of 108, and even though the stocks that underlie both options have the same expected price of 100, the option based on the high-volatility stock is the more valuable of the two, because it's more likely to be "in the money". How much more likely? The probability that the high-volatility stock is "in the money" equals the area under the high-volatility density function that lies above the exercise price of 108. That is, it's the sum of the areas of the crosshatch and solid grey shaded regions. This is considerably higher than the probability that the low-volatility stock is “in the money”, which is the area under the low-volatility (boldface) density function that lies above the exercise price of 108, i.e., the crosshatch region. So volatility is good when it comes to options, because there’s a greater chance of a very high stock price that will put the option far “in the money”. Of course, there’s also a greater chance of an extremely low stock price, but in that case the option holder simply wouldn’t exercise the option. Once an option is out of the money, it doesn’t matter if it’s out of the money by a small amount or by a huge amount; in either case the intrinsic value is zero, and the option holder will not exercise it.

The BSM options pricing formula and its variants develop the points I’ve just made verbally with much greater precision. In practice, formulae like BSM or its variants are often used despite their problems in valuing the options in compensation packages. But it’s important to remember that the formula significantly overestimates the value that risk-averse workers place on the options in their compensation packages and, consequently, it is necessary to pay these workers hefty risk premia when their pay includes options.

10.4.3 Stock Options as Nonlinear Pay Contracts

Stock options are nonlinear pay contracts, meaning that a graph of the value of the option as a function of P is not a straight line. We have seen examples of other nonlinear (i.e., piecewise-linear) pay contracts in sections 9.2 (draw schemes) and 10.2 (executive bonuses). Chapter 9 uses the notation $f(P)$ to denote the compensation contract, which is a function, f, of performance, P. In the case of stock options, we would ideally like to focus on the graph of the option’s *market value* (as opposed to its *intrinsic value*) because that’s what’s of greatest importance to the worker who holds the option as a form of compensation. To see why, remember that options that are slightly out of the money have an intrinsic value of zero, but they are still valuable to their holder, because of the prospect that in the future they will be in the money. The value of that prospect is reflected in the option’s market value but not in its intrinsic value.

Graphing the market value of an option is challenging for a few reasons. Again, there’s no universal formula for expressing the market value of options in a compensation package, and we’d need a formula to produce a precise graph. Alternatively we could use the Black–Scholes–Merton formula, but it’s flawed for valuing the options of interest to us, and moreover I haven’t presented the formula (I refer you to finance texts for that) so we have no business drawing graphs based on it. So I’m instead going to draw graphs based on the option’s intrinsic value, and those graphs will also be nonlinear and will provide insights.

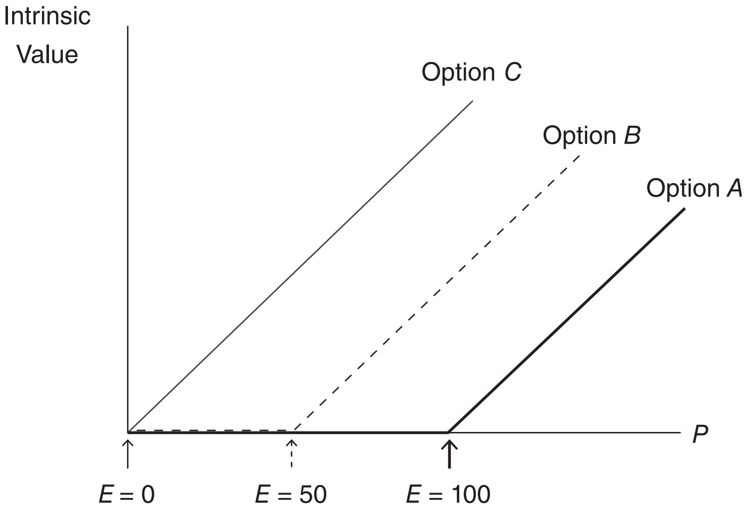


Figure 10.3 Stock options with 3 different exercise prices.

The graph, $f(P)$, of the intrinsic value of a stock option is "piecewise linear", which means that it is comprised of multiple linear segments that have different slopes. The graph of $f(P)$ for three stock options compensation contracts (with three different exercise prices, E) is shown in Figure 10.3. Consider first Option A, graphed in boldface, which is the one with the highest exercise price ($E = 100$). Notice that when the Option A is out of the money (i.e., $P < 100$) the boldface graph coincides with the horizontal axis, which implies an intrinsic value of zero. When Option A is at the money (i.e., $P = 100$) a kink occurs, at which point the graph changes slope (from zero to one). When the option is in the money (i.e., $P > 100$) the graph of the intrinsic value has a positive slope (equal to one). The slope equals one because for every additional dollar that P increases (beyond the exercise price of 100) the profit from exercising the option also increases by exactly a dollar. I'm ignoring taxes here, as I do throughout the chapter, to keep the discussion simple. But in practice workers must pay taxes on the income earned from exercising stock options, and these taxes erode the options' value.

Next consider Option B, which has a lower exercise price of $E = 50$. The shape is the same as Option A, but the flat portion of the graph (for which the option is out of the money) is shorter. Finally, Option C has an exercise price of $E = 0$ and is, therefore, always in the money. An option with an exercise price of zero is equivalent to a share of stock. In contrast to Options A and B (both of which are nonlinear because they exhibit kink points located at the exercise price), Option C is a linear contract just like a piece rate (Chapter 9).

Focusing on Options A and B, as is true of executive bonuses, the most interesting region of those nonlinear graphs from the standpoint of incentives surrounds the kink points, where the options switch from being out of the money to being in the money, and the slopes change from zero to one. Holders of the options have incentives to behave in ways that will increase the chance that their options will be in the money in the future. Workers may be tempted to manipulate the incentive system (e.g., executives may engage in illegal or unethical conduct to try to get their options in the money). Realistically, changes in behavior (both good and bad) in the neighborhood of kinks in the options contracts are less likely to occur for non-executives than for CEOs and other top executives who have significant influence over the company's main strategic decisions. Moreover, such behavior wouldn't happen if the workers were paid in stock rather than in options, since the relationship between pay and performance is "smoother" (in fact, it is linear) for stocks than for options, meaning there is no abrupt change in the slope of the contract.

10.4.4 Incentives from Stock Options

Stock options connect pay to (company) performance. So do shares of stock. Both compensation contracts have "intrinsic value" graphs with identical slopes of one, at least when the options are "in the money". However, options and stock don't generate exactly the same incentives, for two reasons, both of which relate to the fact that for options the exercise price, E, is positive. This should come as no surprise, because if E were zero, as we've discussed, the options would be identical to shares of stock, so (obviously!) they must generate exactly the same worker incentives as stocks.

The first reason is that options encourage the workers who hold them to make riskier decisions that might lead to great success for the company, or that might lead to colossal failure. The reason is simple. For a (vested) option to actually generate any cash for a worker who holds it, the option must be "in the money". In other words, the stock price, P, needs to clear a hurdle, E, before any money can be made. Riskier strategies that could yield big payoffs (or big losses) have a higher chance of landing the worker's options in the money (hopefully far into the money!), and the worker is insured against the downside risk because even if the stock price plummets the worker can simply choose not to exercise the options; remember, an option can never have negative value.

You might say that the preceding argument about options also applies to stocks, but the difference is that stocks (which are options with exercise prices of zero) are *always* in the money with positive intrinsic value, so there is no need to pursue risky strategies in an effort to get them in the money. Encouraging workers to take greater risks is sometimes a good thing. Facebook's CEO, Mark Zuckerberg, was quoted as saying that "The biggest risk is not taking any risk . . . In a world that is changing really quickly, the only strategy that is guaranteed to fail is not taking risks."

Managers often behave too conservatively and don’t want to be responsible for a bad decision that visibly hurts the company and, therefore, their career prospects, so encouraging them via options compensation to think more boldly can have a big upside for the company.

The second reason, which again results from a positive exercise price, is that if a worker’s options fall too far out of the money then they no longer provide much incentive because the worker believes that it’s essentially hopeless that they’ll ever generate any cash. In such cases, companies sometimes “reprice” the options, lowering the exercise price so that the options are closer to being in the money. To execute such repricing, workers exchange their existing options (which are far underwater) for a smaller number of options that are issued at the money. Shareholders often aren’t thrilled about such options repricing. The practice is interpreted as a bailout. In other words, when the options were originally granted, everyone understood that there was a chance that they might end up far out of the money, so if that bad event happens then why should the company step in to bail out the workers, particularly if the reason why the stock price tanked is because the company’s workers weren’t performing well enough?

If you’re part of the management team that’s considering repricing your workers’ options, you have to make the case to your shareholders convincingly so that it isn’t perceived as just a blatant bailout that rewards workers’ poor performance. One common argument is that repricing is needed to retain top talent. Often the top executives have a lot of their compensation in options, and if those options have very low market value (and zero intrinsic value) as a consequence of being far out of the money, then it may be necessary to reprice the options to retain those workers. Of course, a counterargument is that there are other ways to increase those employees’ compensation that might be less costly than options repricing. Another argument is that the options are underwater not because the company and its workers made bad decisions but because of factors beyond its control (e.g., a business cycle downturn that hurt the entire industry), in which case repricing workers’ options can reinvigorate their incentives and would not be perceived as a reward for past poor performance. Perhaps the most essential thing to remember is that you should communicate options repricing, both to your workers and to your shareholders, with carefully justified reasoning that makes the event seem rare and special and that won’t appear to create a precedent for the future.

10.5 “Pay for Luck”

The performance measure, P, on which executive compensation is based is the broadest possible performance measure for the organization, namely its stock price.

This is a very risky measure. Lots of factors affect the stock price that are beyond the CEO's control. P could be very high, not because the CEO did a great job but rather because the CEO (i.e., the company) got lucky. In that case, the CEO would be paid for being "lucky" rather than for performing well. The flip side of that coin is that P could be very low not because the CEO mismanaged the company but simply because the CEO (i.e., the company) got unlucky. In that case, the CEO would be penalized with low pay only for being unlucky rather than for poor performance. This doesn't seem like a great way to pay CEOs . . . you'd rather pay them not for luck but for performance (i.e., the outcomes they can actually control through their decisions), because doing so gives them incentives to make the good decisions that shareholders expect.

One situation, however, in which it may be profitable to pay CEOs (and other workers) for luck is when luck is *persistent* over time, meaning that good luck today is likely to be followed by good luck tomorrow, and similarly for bad luck. In an environment with persistent luck, a company that makes high profits today due to luck is likely to enjoy the same situation in the future. For example, if a government regulation is unexpectedly relaxed or eliminated (see Chapter 4), the company may enjoy higher profit this year but also in future years. To understand why the presence of persistent luck might make it profitable to pay CEOs for luck, keep in mind that there are two types of pay contracts (for CEOs, and for all other workers for that matter).

One is a ***formal contract***, which is written down on paper and explicates the terms of the compensation. Formal contracts tie a company's hands by creating a legal obligation to respect the contract. If the contract promises the CEO a bonus under certain conditions, then if the company tries to withhold that bonus the CEO could take the company to court for breach of contract. Many CEOs have formal contracts . . . this is why, during the global financial crisis of 2008, when a number of failing firms were receiving public bailout assistance, the CEOs of those failing firms were raking in large bonuses. The public was outraged by these "rewards for failure", to such an extent that the government enacted caps on executive pay for firms in receipt of bailout assistance, as discussed in section 4.7.4. Companies receiving bailouts, however, were often legally obligated to pay large executive bonuses, because of pre-existing, multi-year formal contracts that had been signed during the more prosperous economic times that prevailed before the financial crisis hit.

Another is an ***informal contract***, which is an unwritten understanding between the two parties (i.e., worker and employer) about the terms of compensation. Because an informal contract is unwritten, a company could renege on it without obvious legal consequences. In practice, sometimes even informal contracts are upheld in court. But even in those cases, it's a lot easier for workers to enforce a formal contract than an informal one in the event that the employer tries to renege.

Reneging, by the way, typically means not paying the worker some compensation that was promised (see Chapter 2 on wage theft).

CEOs and other executives often have contracts of both types – formal and informal. Let's focus on the informal ones, which permit the possibility that the company may renege and engage in wage theft. Recall from Chapter 2 that, in the wake of wage theft, workers often quit, and that threat is often enough to prevent the employer from stealing wages in the first place. A company that promises to pay the CEO a large bonus for good performance might later renege and withhold some (or all) of this bonus, even if the CEO has performed well, but the downside of this wage theft is that the CEO might quit, which hurts future profits.

Losing a good CEO would be particularly damaging if the company is doing really well and profits are running high. So a company that expects high future profits (even due mostly to luck) will be particularly keen on retaining its CEO, which in turn means that it will be less likely to engage in wage theft by reneging on the informal pay contract. Such a company can promise to pay a high bonus when the CEO has performed well (and the company has enjoyed good luck). That promise will encourage the CEO to achieve higher performance, because he is confident that the company will actually want to honor its commitment to pay a high bonus when luck is good. This explains why companies may actually purposely pay executives (and other workers) for luck even though, on the face of it, this seems like a bad idea. Paying for luck makes it easier for companies to uphold their informal promises to CEOs concerning bonuses . . . in an environment with persistent luck, everyone sees that good luck today is likely to bring good luck tomorrow (and, therefore, a desire on the company's part to retain the CEO, which means not antagonizing the CEO by engaging in wage theft).

The preceding arguments only pertain to *persistent* luck. But not all luck is persistent. If an earthquake damages the company's headquarters and requires the construction of a new building, that's bad luck, but there's no reason to think that bad luck in the future will be more likely as a consequence of today's earthquake. A good CEO, however, might be able to seize opportunities presented by good luck (even good luck that appears to be idiosyncratic rather than persistent), and leverage those into even more good luck in the future, thereby *creating* some persistence in luck. That's exactly the idea behind the adage "You make your own luck." Talented and creative executives are able to recognize unexpected opportunities, seizing them to create the conditions for even more good opportunities in the future. Such individuals can create the persistence in luck that actually justifies "pay for luck" from the standpoint of company profits. This amplifies the benefits of using pay for luck in CEO compensation, and it means that using "pay for luck" gives CEOs incentives to leverage unexpected opportunities so as to create fertile conditions for more such opportunities to arise in the future.

10.6 Lessons for Managers

Nonlinear pay contracts (e.g., executive bonuses and stock options) are common in executive compensation. Performance levels near the “kink points” (i.e., where the contracts change slope) are of particular interest, because CEOs face powerful and sometimes perverse incentives to push the performance measure into the upward-sloping region of the pay contract where they’re actually getting paid more money for higher performance. Keep your eyes open for unethical or even illegal behavior on the parts of your company executives and other workers!

The performance measure used in the main components of executive compensation contracts is the stock price, which is a very broad measure of organizational performance but also one that is very risky, which requires companies to pay hefty risk premia to compensate risk-averse executives. That can get expensive.

Stock options can be an effective way to create incentives for workers, particularly CEOs and other executives whose decisions materially impact the strategic direction of the company (and, therefore, the stock price, which is the performance measure on which this type of performance pay is based). Options can also embolden overly conservative managers to pursue riskier (but potentially higher return) strategies that the shareholders may favor, because doing so increases the chances that the options in their compensation portfolios will be in (and hopefully *far* in!) the money. Stock options, like other forms of pay for performance, can also generate sorting effects that benefit the company, by attracting and retaining those workers who are most optimistic about the company’s future outlook and about their own ability to move the company forward.

The biggest downside of paying workers with stock options is that they’re an extremely risky form of compensation, even riskier than stocks, which are themselves options with exercise prices of zero. This means that workers will demand a hefty risk premium to compensate them for accepting part of their compensation in such a risky form, and those risk premia are expensive for the company. If your workers’ options fall too far out of the money, you might want to consider repricing them, subject to the caveats discussed earlier.

Paying your CEO for “persistent luck” makes it easier to honor your informal promises to pay executive bonuses, which helps the bottom line. Creative CEOs can sometimes seize unexpected profit opportunities that create fertile conditions for even more unexpected profit opportunities in the future, thereby creating the very persistence in luck that justifies paying them for luck. These are the CEOs who can literally “make their own luck”. Whatever you can do to encourage executives to “make their own luck” (recall the incentive effect) or to hire and retain executives who have that talent (recall the sorting effect) will benefit your company and please your shareholders.

Case Discussion 11: Tesla Motors[1]

Tesla Motors Inc.'s TSLA compensation plan for CEO Elon Musk is aggressively focused on incentive pay. Musk stands to get $1.6 billion worth of stock options by 2022. The plan requires him to hit challenging targets that will benefit Tesla's shareholders and workers. For every target related to the stock price, another requires that products be brought to market that are better than any before them. Musk has to do both to get paid. He receives a base salary of $37,584 (which he has never accepted) and no cash bonus.

Musk gets 10% of his options every time Tesla adds (and sustains) $4 billion in new stock market value up to $43.2 billion, as long as the company also meets an operational goal that accompanies the milestone. Those include getting the Model X crossover through development and onto the market, then doing the same for the less expensive Model 3 sedan whose 2017 debut made Tesla a mass-market producer.

There are milestones for selling 100,000, 200,000 and 300,000 cars. At the end of 2015, Tesla had sold about 107,000. When the program was introduced, the first Model S had not yet been delivered. Also, Tesla must have gross profit margins of 30% for four straight quarters.

The situation as of April 2016 was as follows. Tesla had met seven of the market-cap milestones and five of the operating milestones, Tesla's proxy said. Only half of Musk's options had vested, since seeing the share price rise is not enough. The company believed it could hit the 30% margin on the Model S by the end of 2016, and that its 400,000 Model 3 orders in hand would meet the sales targets. To reach $43.2 billion in market value, Tesla's $250 stock would need to sustain a six-month moving average of $327.

Musk will receive significant financial rewards for major successes. But the bar is high. The 2016 proxy statement section on CEO compensation states (in underlined, boldface text): "**The 2012 CEO Grant was designed to be entirely an incentive for future performance that would take many years, if at all, to be achieved. Further, many of the requisite milestones were viewed as very difficult to achieve when the 2012 CEO Grant was made**."

Questions

1. Consider the "sorting effect". What type of CEO is attracted by this type of pay plan?
2. From the perspective of the firm and its shareholders, what are the advantages and disadvantages of this type of CEO compensation plan?

---

[1] Note: This case is based on an article in *MarketWatch* on April 25, 2016 – "Opinion: Tesla's Elon Musk is doing for executive compensation what he did for electric cars".

3. Is this executive compensation plan a good fit for Tesla? What specific characteristics of the company and industry have a bearing (good or bad) on the appropriateness of this CEO pay plan?
4. What kind of message does this CEO plan send to lower-level workers at Tesla?

Further Reading

Bebchuk, L. M. and Fried, J. M. (2003). "Executive Compensation as an Agency Problem." *Journal of Economic Perspectives*, 17, 71–92.

Bebchuk, L. and Fried, J. (2004). *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*, Cambridge: Harvard University Press.

Core, J. E., Guay, W., and Larcker, D. F. (2008). "The Power of the Pen and Executive Compensation." *Journal of Financial Economics*, 88, 1–25.

Cornett, M. M., Marcus, A. J., and Tehranian, H. (2008). "Corporate Governance and Pay for Performance: The Impact of Earnings Management." *Journal of Financial Economics*, 87, 357–73.

DeVaro, J., Kim, J.-H., and Vikander, N. (2017). "Non-Performance Pay and Relational Contracting: Evidence from CEO Compensation." *The Economic Journal*, 128, 1923–51.

Devers, C. E., Cannella, Jr., A., Reilly, G. P., and Yoder, M. E. (2007). "Executive Compensation: A Multidisciplinary Review of Recent Developments." *Journal of Management*, 33, 1016–72.

Hartzell, J. C. and Starks, L. T. (2003). "Institutional Investors and Executive Compensation." *Journal of Finance*, 58, 2351–74.

Jarrow, R. and Chatterjea, A. (2019). *An Introduction to Derivative Securities, Financial Markets, and Risk Management*, 2nd ed., World Scientific Publishing.

Murphy, K. J. (1999). "Executive Compensation." In O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, vol. 3B, North-Holland, 2485–2563.

Shi, W., Connelly, B. L., and Sanders, W. G. (2016) "Buying Bad Behavior: Tournament Incentives and Securities Class Action Lawsuits." *Strategic Management Journal*, 37, 1354–78.

11 Benefits

My first job after college was as a research assistant at the Federal Reserve Bank of Richmond. Once per year, all employees at the Fed were distributed a personalized folder (marked "confidential") with their name written on the front cover. Inside the folder, on the first page, was a dollar figure . . . their "annual compensation". When I opened mine for the first time, I was shocked. The number was significantly higher than my annual salary of $30,000 in 1995. Was this a mistake? For a split second of sheer joy I wondered, could I really be receiving a raise of about 40%? Are they really that impressed by my performance?

Alas, the reported number was not my salary but rather the total cost to the Fed of employing me. The value of my health insurance, retirement benefits, and so on, were monetized and incorporated into that figure. So the figure represented the cost to the Fed of my "salary plus benefits". In fact, that number, as substantial as it seemed to me at the time, was an underestimate. Recall the definition of compensation from section 1.1 . . . "everything a worker likes about a job". There were many aspects of my compensation that I valued highly but that were not incorporated in the number inside my confidential folder.

For example, the hours were flexible. If you wanted to work until 7pm, no one would complain if you didn't show up until 10am. More importantly, the research department where I worked was filled with fantastic researchers who were also great mentors and very generous with their time. They all had a significant positive influence on my future career. I highly valued those aspects of the work environment and am still grateful for them today. Recalling our discussion of compensating differentials in Chapter 3, I was willing to sacrifice a lot of cash compensation (i.e., annual salary) for the opportunity to work in such an intellectually stimulating environment.

11.1 Benefits and Value

"Benefits" typically mean all of the non-cash forms of compensation that an organization offers at some direct cost. For example, the SAS Institute (an American multinational developer of analytics software based in Cary, North Carolina) offers free on-site access to healthcare facilities, recreation areas, swimming, and a gym.

All of those amenities come at a direct monetary cost to SAS. In contrast, a "supportive work environment" is hard to quantify and monetize and, unlike SAS's other benefits that I just mentioned, would not typically be listed by the HR Department as part of the company's benefits package. But the best way to think about benefits, consistent with the broad definition of compensation in section 1.1, is "everything (*other than monetary payments*) that a person likes about a job".

As we learned in Chapter 3, any benefit of a job can give rise to a compensating differential, meaning that a job that offers that benefit can get away with offering less compensation in other areas compared with an *identical job* that does not provide that benefit. I've italicized "identical job" to highlight, yet again, the importance of "apples-to-apples" comparisons. Doing true apples-to-apples comparisons is really hard, so when you hunt for compensating differentials in the data using regression techniques of the type described in Chapters 6 and 7, you often don't find them. For example, consider the "Wage-Insurance Tradeoff" case discussions (Parts A and B) at the ends of Chapters 6 and 7. The Part-A case illustrates a failure to detect a compensating differential attached to employer-provided health insurance. The Part-B case successfully detects a wage-insurance tradeoff but only in multi-establishment firms. If apples (i.e., multi-establishment firms) are compared to apples, the expected compensating differential emerges, but not when those apples are compared to oranges (i.e., single-establishment firms) as in Part A.

Let's review the familiar logic of compensating differentials, in the context of employer-provided health insurance. Imagine that one of your competitors, let's say Firm A, wants to pay workers \$70,000 per year, with health insurance. You, as a manager at Firm B, want to pay workers \$70,000 per year to do the *exact same job* but with no health insurance. Everyone would want to work for your competitor in that case, because even though you both offer the same jobs at the same salaries, your competitor offers higher *total compensation*. To hire any workers at all, you would have to offer more than \$70,000 to compensate workers for the lack of health insurance in your firm. Let's suppose that offering the health plan costs firms (including both you and your competitor) \$3500 per worker per year. There are two alternative compensation systems that you and your competitor might consider. One offers workers a \$70,000 annual salary plus the health plan, and another offers workers a \$73,500 annual salary but no health plan. The monetary cost (from the employers' perspectives) of both plans is identical, but which one should you provide? The answer depends on how much the workers ***value*** the health plan.

What exactly do I mean by "value"? You probably have in mind the common English-language definition of value, which basically means how much workers "care about" the health plan. That's in the ballpark but a bit too vague. What exactly do we mean by "care about"? Our definition of "value" is the *maximum amount that a worker would be willing to pay* for the health plan. The value principle applies to everything that's potentially for sale, not just health plans. For example, when you're shopping for houses, the value that you place on a particular house is the maximum amount that you'd be willing to pay for that house. If you value a given house at $500,000, you'd be willing to pay (up to) $500,000, but if the price were any higher you'd walk away. If the price were $500,000 or less, you'd definitely buy. I want to emphasize that value is not the amount you *want* to pay for the house (which should be zero!), nor is it the amount that you think is a *fair price* to pay for the house, nor is it the amount that you *expect to pay* for the house. It is the *maximum amount that you would be willing to pay.*

11.2 Worker Value versus Employer Cost

In the preceding two-firm example, a worker's value of the health plan need not match the employer's cost of $3500. It could be either higher or lower. Moreover, workers vary in their valuations of the plan. Older workers care more about health plans than do younger workers, because they are more likely to need healthcare. Similarly, families care more about health plans than do single workers, and women care more than men. Employers in this example would love to offer the health plan only to those workers whose valuations exceed $3500 and to reduce each of their salaries by exactly their individual valuations, whereas workers with valuations less than $3500 would not be offered the plan. For example, if one of your workers valued the health plan at $5000 (i.e., they would be willing to pay up to $5000 for the plan, but no more) then you would like to offer that employee a compensation package that includes the health plan and a salary of $68,500 (i.e., $73,500 – $5000). And if another of your workers valued the plan at $7000, you would like to offer that worker a compensation package that includes the health plan and a salary of $66,500. These workers always have the option to work for another employer that pays $73,500 and no health plan, but there's no reason for them to quit and work for such an employer, because doing so wouldn't make them better off.

If, however, you try to be greedy and cut your workers' salaries by even more, then they would quit. For example, if you offer your second worker only $66,000 rather than $66,500, that worker would be better off quitting and working for a firm that offers $73,500 and no health plan. The reason is that by switching firms the worker gets compensation valued at $73,500 whereas by remaining employed by you the worker gets a salary of $66,000 plus a health plan that the worker values at $7000, for a total compensation valued at only $73,000. So you can increase your company's profit the most by offering the health plan and reducing each employee's compensation by exactly the worker's valuation of the plan. Again, for your workers who value the plan at less than your $3500 cost of providing the plan, you're better off offering them no health plan and a salary of $73,500.

There are three problems you face in behaving like this. First, it's usually impossible to observe each worker's valuation of the health plan, and your workers don't have an incentive to truthfully report this. For example, consider the two workers just mentioned. The one with the higher valuation got stuck with the lower salary. So why would the worker disclose that high valuation to you in the first place? Second, even if you could perfectly observe all your workers' valuations, paying certain individuals (or groups of workers) different rates could expose your company to a lawsuit. For example, suppose that all of the women at your firm ended up getting paid lower salaries, simply because of their higher valuations for the health plan. This could subject your firm to a gender-discrimination lawsuit (see Chapter 4). Third, frequently you must offer a health plan to all of your workers or to none of them. Some features of the tax law permit a benefit to be tax deductible to workers only if the plan is provided to a specified (large) fraction of workers.

The fact that workers can't be trusted to accurately report their true valuations means that conducting a survey of workers is not the best approach. Of course, that doesn't stop some firms from doing it. Walmart issued an internal memo some years ago (see the case discussion at the end of the chapter) that surveys workers to elicit their preference for a range of different benefit offerings, including health plans. How should a firm proceed that wants to get a feel for how much its workers value the health plan, without asking them directly?

Drawing on the compensation analytics skills that you developed in Chapters 6 and 7, a data-driven approach can be used, which relies on workers' *actual choices* rather than on choices they *claim* (perhaps falsely) that they will make in the context of a survey. A compensation consultant may have access to data on what's happening at a bunch of similar firms that employ similar workforces and that can be used to shed light on how workers trade off wages for health benefits. Such an example is explored in the case discussions called "Wage-Insurance Tradeoff" (Parts A and B).

11.3 One (Big) Problem with Benefits Compensation

You have surely received, at some point in your life, an unwanted gift. I used to get them once per year, at Christmas, from an aunt. Her typical gift to me would be a sweater, several sizes too large, that I'd never wear even once. Her gifts were by no means cheap. She would spend a lot of money on them, even though they offered me no value. I would have been much happier if she had instead just given me the money she spent on the sweaters, so that I could have bought whatever would have made me happiest at the time.

What I've just described is a fundamental problem that employers face when using benefits to compensate workers. Benefits are often quite expensive for employers to provide, but it's unclear that a particular worker will value those benefits at anything close to what the employer paid for them. And there's another consideration that makes matters even worse than in the sweaters example. In the case of sweaters, my aunt's gift was *chosen just for me*, by someone who knew me well, and it was *still* way off the mark, whereas in an organization, you have to choose a benefits package to please your entire workforce. What one worker finds worthless, another may love. For example, employer-provided health insurance for dependents is great if you have six kids, but it's less valuable if you're childless and don't plan to have kids. Similar arguments apply to tuition subsidies, on-site childcare, vision benefits, and even personal health insurance (because some workers are fully covered on their spouse's plan).

Imagine buying one gift that will be distributed to a large group of people, and your goal is to please them all. It's nearly impossible. You might think that the solution to that problem is not to buy everyone the same gift but rather to customize the gifts to each person. That occurs to some degree with benefits offerings, and it can help to mitigate the problem I've just described, but for the most part benefits are not tailored to the tastes of individual workers and are instead of a "one-size-fits-all" nature. Moreover, even if an organization tried to perfectly customize the benefits to each individual worker, there still would be a lot of misfires . . . again, my aunt was choosing a perfectly customized gift just for me and still got it wrong every year!

Let's take a closer look at the losses imposed by an ill-chosen gift. Suppose that my aunt spends \$100 on a sweater for me that I value at \$0. Then my aunt loses \$100 (or "gains" -\$100), I gain \$0, and if you add up the gains for both of us, you get -\$100 + \$0 = -\$100, which is a net loss. Even if I didn't find the gift completely worthless, there would still be a loss, albeit a smaller one. Suppose that I valued the gift at \$50. Then our "total gains" would be -\$100 + \$50 = -\$50, which is still a net loss. Giving me \$100 in cash would have been much better than buying me a sweater. In that case, our "total gains" calculation would have been -\$100 + \$100 = \$0, which is a wash rather than a net loss.

I've found there are two types of readers. The first type is satisfied with the preceding discussion. The second type, however, is bothered by the fact that the discussion focuses only on me and my aunt and ignores the vendor who profited from selling my aunt the sweater. If you're the first type of reader, skipping now to section 11.4 will allow you to avoid a small risk of becoming confused. If you're the second type of reader, continue reading to learn why bringing the vendor into the discussion doesn't change the conclusion and just makes the argument more involved.

The clothing vendor who sold my aunt the sweater must have valued that sweater at less than \$100 (otherwise my aunt would never have been able to buy it from them for \$100) and, therefore, made some profit on the sale. Suppose the clothing vendor values the sweater at $67 and sells it to my aunt for $100, to make a profit of $33. Then if I value the sweater at $0, the "total gains" calculation when the clothing vendor is included is -$100 + $0 + $33 = -$67. Fair enough. But if we're going to include a vendor in the "total gains" calculation for sweaters, then to be fair and consistent we also have to include a vendor in the "total gains" calculation for cash. If my aunt had given me a cash gift, I would have spent it on a sweater that I chose myself and that actually fit (or on a new pair of shoes, or electronics, or toys, or whatever else I wanted to buy), and the "generic" vendor who sold me that merchandise would also have made a profit.

Without further information, there's no reason to anticipate that the profit my generic vendor makes on a $100 purchase is any less than the profit my aunt's clothing vendor makes on a $100 purchase, so the profits for those two vendors are basically a wash. For that reason, nothing is lost by ignoring vendor profits in the preceding "total gains" calculations. To make that point more vividly, if we include vendor profits of $33 in our calculations, then the "total gains" are

-$100 (aunt) + $0 (me) + $33 (clothing vendor) = -$67 (net) for a sweaters gift and
-$100 (aunt) + $100 (me) + $33 ("generic" vendor) = $33 (net) for a cash gift,

with a difference of -$67 – $33 = -$100 between those two total gains. If we instead exclude the vendor profits, then the "total gains" are -$100 (aunt) + $0 (me) = -$100 (net) for a sweaters gift and -$100 (aunt) + $100 (me) = $0 (net) for a cash gift, with a difference of -$100 – $0 = -$100 between those two total gains, which is exactly what we found if we included vendor profits!

Hopefully you're now convinced that cash is the way to go when it comes to gifts. But that raises the following question . . .

11.4 Why Do Employers Offer Benefits?

Why don't employers just use cash to compensate workers? That would completely eliminate the "sweaters problem" described in section 11.3. Recalling the third recurring theme of section 1.8, your first thought might be that employers offer benefits simply because workers like them. Employers, after all, are forced by labor-market competition to care about workers' preferences. But that argument wouldn't really answer the question, because if workers have cash they can buy whatever they want on their own. If they love certain benefits, they can buy those themselves, just like if I happened to love grossly oversized sweaters, I could have used a cash gift from my aunt to buy one for myself. The key point is that I wouldn't be *forced* to . . . I could spend the money however would make me happiest. There are five good reasons why employers offer benefits even despite the big disadvantage of the "sweaters problem".

11.4.1 Legal Mandates

Some benefits are required by law. An example is workplace safety. The Occupational Safety and Health Act of 1970 created OSHA (Occupational Safety and Health Administration), which sets and enforces health and safety standards in the workplace. Other examples are California's 2014 HWHFA, as discussed in section 4.7.5, or the federal FMLA, as discussed in section 4.6, both of which mandate paid work leave. So, one reason why you might give an oversized sweater instead of cash as a gift is that you're forced to by law. As we discussed in section 4.3, from your perspective as a manager, such regulations are unwelcome, because if gifting an oversized sweater were truly a winning strategy from the standpoint of raising your company's profit, there was nothing stopping you from doing that anyway even without being forced to by law. The other four reasons for offering benefits, described in the following four subsections, are voluntary and do not require that you are forced.

11.4.2 Bulk Discounts on Employer-Purchased Benefits

The fact that you can buy in bulk for lots of workers means that you can purchase some benefits at lower prices than your workers would face if they individually bought the same benefits on their own. Health insurance is a great example. If one of your workers purchases health insurance on their own, the insurer is likely to worry that the worker is unhealthy and will end up costing the insurer money in the long run. But if you purchase a health plan for all of your workers, the insurer knows that some of your workers will be unhealthy, but many will be healthy, so the overall risk is modest because lots of healthy people counterbalance those who are unhealthy. The larger the firm, the more powerful that argument, and the deeper the discount the insurer will provide you. Although health insurance is the most important example of this second reason for offering benefits to your workers, there are other examples. For example, you can usually purchase gym memberships at a bulk rate that's lower than what your workers would be charged to buy their own memberships.

11.4.3 Tax Considerations

Tax law creates other reasons to voluntarily provide benefits to your workers. From a talent management standpoint, your goal as a manager is to compensate your stars generously enough to retain them but to do this as cheaply as you can, because every dollar that you save the company adds to the bottom line. Suppose that you need to pay one of your stars an extra \$1000 annually to retain her. One approach is to increase her salary by \$1000, but another approach is to offer \$1000 worth of tuition reimbursement. Such education benefits are provided by the Federal Reserve Bank of Richmond, as explored in an online case discussion. Starbucks goes even further by providing its workers with free tuition for a four-year undergraduate online degree program at Arizona State University.

A downside to offering education benefits is the "sweaters problem" from section 11.3. But let's set that aside for the moment and assume that your star worker values education benefits basically the same as cash, because she plans to take some courses anyway, regardless of who pays for them. An advantage of paying her via tuition reimbursement is that workers are often not taxed on benefits, whereas they always pay income tax on wages and salaries. So if you give her \$1000 in cash compensation, that only gives her $\$(1\text{-}t)\times 1000$ in actual "take-home" value, where t represents her marginal income tax rate, whereas if you give her \$1000 in benefits compensation, she gets the whole \$1000 in "take-home" value. That means that you could give her only $\$(1\text{-}t)\times 1000$ worth of benefits compensation, which would give her exactly the same value as if you were to take the more expensive (for you) route of paying her \$1000 in salary.

In fact, around the time I worked at the Richmond Fed, the Virginia state government made more than one change to its tax treatment of tuition-reimbursement benefits, which further underscores the importance of staying abreast of the laws. The first of the "3 Cs" from section 4.3 is to *comprehend* the law, which is a particular challenge when the law frequently changes. The tax argument for offering benefits also comes into play for health insurance, because health benefits are tax deductible for workers. That wasn't always the case, and once the tax code was changed to allow those deductions, employer-provided health insurance in the United States became more widespread. Another example is pension contributions. US tax law allows workers to invest pre-tax income and not pay taxes until the pension is paid out during retirement. This has the dual advantage, for workers, of allowing them to enjoy tax-deferred earnings growth and potentially face lower marginal tax rates when the income is eventually taxed (because incomes are generally lower during the retirement years).

11.4.4 Benefits Can Increase Worker Productivity

Sometimes providing workers with benefits makes them more productive. Airbnb is a company in the hospitality sector that facilitates an online marketplace for short-term rentals of lodging. It has caused the hotel industry headaches in much the same way that Uber and Lyft have caused headaches for taxi-cab drivers (for more on that, see the online Uber case discussion). At its headquarters in San Francisco, consistent with its mission of hospitality, Airbnb offers free meals to its employees and their guests . . . breakfast, lunch, and dinner, cooked every day, right on the premises, including a full bar with a variety of high-quality wines on tap for self-service, as well as beer and spirits. Employees can go back for seconds, and even thirds, and the fare is of restaurant quality. I can personally confirm this, because I've eaten dinner there twice and am hoping to return soon! Offering that amount of food, at that level of quality, every day, for all of those employees and guests is quite expensive. But there are at least three reasons why Airbnb workers' productivity might be enhanced to an extent that justifies the cost.

First, the practice helps to inculcate workers with the company's mission and values, simply expressed in its tagline "Belong Anywhere" (see Chapter 15 for more on a company's mission, particularly as a source of intrinsic motivation for employees). Home-cooked meals in a casual, relaxed environment help to make employees and guests feel at home and like they belong. This can enhance worker productivity by strengthening a sense of common purpose and understanding of the organization's objectives. Second, and more directly, workers are likely to spend more time at the office when there is free, delicious food there. They will show up earlier, to take advantage of breakfast, and stay later, to take advantage of dinner. More time at the office means more potential for work to be done. Third, the dining area is purposefully arranged to facilitate social interaction. There aren't many small tables for two, and it's difficult to sit alone. There are long tables, with long benches, and even if you sit down alone, you never know who might subsequently join you. This arrangement facilitates communication of all types . . . across functional areas, across hierarchical ranks, etc. I witnessed such productive communication firsthand during my first dinner there. I was there for a social dinner with one employee, but before and during our dinner he was briefly interrupted by co-workers for quick conversations that resulted in actual business decisions being made. These communications would have been less likely to happen if they had required knocking on his office door and interrupting his work.

There are other examples of ways in which benefits can improve worker productivity. In a sales context, if the workers who are selling the products are given discounts on those products, their increased exposure to the products, and familiarity with them, can make them better and more credible salespeople. For example, Starbucks offers all of its workers a pound of free coffee or tea each week, in addition to as much free coffee, tea, or milk-based beverages that they can drink during breaks on a shift. This means that when customers are thinking of trying a product for the first time and have questions about it, the baristas are more likely to have informed answers based on their own experiences. On-site parking is another good example. Rather than having employees waste time circling around looking for parking, they can quickly park and get to work. Flex-time is another benefit that has the potential to increase worker productivity, by allowing workers to do their work during the hours in which they're least distracted and most productive. For parents this might be after their children have gone to sleep.

11.4.5 Sorting Effects

A fifth reason to offer benefits is perhaps the most powerful, and that's the sorting effect (see section 1.8). Workers differ a lot in terms of which benefits they find attractive, and for that reason, probably no component of a compensation system generates more powerful sorting effects than benefits. What goes hand-in-hand with a powerful sorting effect is a greater likelihood of tighter matches between employers and their workers, which in turn reduces turnover. Employers that desire low turnover rates, perhaps because recruitment costs are particularly high, can achieve that by designing the benefits package to induce the right sorting. Another implication is that the company's workers who do not highly value the benefits package heavily "cross-subsidize" the workers who do. For example, if the health plan involves generous benefits for dependent children, the single workers are subsidizing those who have families with children.

Another example of a benefit that can induce sorting effects is a tuition reimbursement program, such as those offered by Starbucks or the Richmond Fed. At the Fed, the benefit was highly valued by entry-level workers who planned to go to graduate school, such as myself, but it was worthless to most middle-aged and older workers who had no intention of going back to school. Why might the Fed want to attract workers who plan to take classes? One reason is that workers who want more education might be particularly energetic, and they might tend to have high ability. These desired qualities are often difficult for employers to observe during the hiring stage, so it helps to set up the compensation system to make the desired workers more eager to apply. A downside to selecting such workers and paying for their education is that, once educated, they might leave the firm. Even if they stay, you might need to pay them more, because the new skills that you paid for increased their market value. Chapter 8 elaborates.

The particular array of benefits that you offer has an important influence on the types of workers that you can attract and retain. This means that learning about the preferences of the types of workers that you'd like to attract (or to avoid) is helpful. Also be aware that the design of benefits offerings can be used as a way to achieve *de facto* discrimination of the type prohibited by the laws covered in section 4.5. For example, posting a job ad that says, "Childless, single workers aren't wanted and need not apply" would surely invite legal problems. But offering a compensation package involving below-market salaries coupled with unusually generous health benefits for dependent children, along with on-site childcare, could legally accomplish much the same thing. Employers might want to induce such sorting if they believe that workers with families, and who have children in the local school system, are less likely to quit than single workers who are completely unattached to the geographic area.

Sorting effects induced by benefits can sometimes be damaging to productivity. An unusually generous health plan, for example, might disproportionately attract workers with serious health problems or whose children have serious health problems. If such workers use the healthcare more intensively, then the employer will have to pay a higher price for the same plan when negotiating with insurers in the future. On top of that, such workers are probably more distracted and less productive than workers in perfect health.

11.5 Cafeteria Plans

Your workers' diverse preferences make a one-size-fits-all benefits package problematic, because the workers who dislike the package may begrudge the fact that they must cross-subsidize those who like it. The disgruntled workers then become flight risks, which is a problem if they're your stars. A cafeteria plan attempts to mitigate this problem by offering all of your workers more choices. In a cafeteria plan, there is a "menu" of benefits that the firm offers, along with a "price" associated with each. Workers are each given some fixed amount of dollars and can pick and choose from the menu until all of their dollars are spent.

Different workers within the same firm might choose very different packages from the menu. Sometimes the prices of the menu items are not stated in dollar amounts. Each item might have some number of points attached to it, and each worker is given a fixed number of points to allocate across the items. The flexibility of a cafeteria plan helps to alleviate the "sweaters problem" from section 11.3. But it comes at the cost of diluting the advantages of benefits compensation that we discussed in section 11.4.

11.6 Pensions

One thing I noticed shortly after moving from Cornell to CSUEB was that professors' retirement ages are younger at CSUEB than at Cornell. There are probably multiple reasons for this difference, but I'm sure that one reason is the difference in the design of the pension plans between the two institutions. One of my former CSUEB colleagues retired shortly after I moved there. He was in his 50s and seemed young to be retiring. He also seemed to love his job and to enjoy teaching. One day I asked him about his decision, and he replied, "Yes, I do love my job, and I do enjoy teaching. But not enough that I'm willing to pay to do it. The idea of getting paid a negative wage just doesn't appeal to me!" I'll explain what he meant later in the chapter.

Some of the material that follows may seem a bit complex and overly detailed, and you might question how relevant it is for you. Maybe the company where you're a manager, or where you hope to be a manager in the future, doesn't even offer any pension plans, so your employees' compensation won't involve the features I'm about to describe. You might then be tempted to skip reading this material. Don't! You should continue reading for at least two reasons.

First, even if your company doesn't offer a pension plan, a competing firm that could potentially poach your workers might. Good talent management, in particular retaining your top talent, requires understanding the compensation packages of competing companies and how they compare to those you offer. After all, your workers have mobility and will be choosing among them. If a competitor that offers a pension is attempting to poach one of your workers, you need to understand what that means if you're to compete effectively in the bidding war (see Chapters 12, 13, and 14 for more on poaching, counteroffers in a bidding war, and how to negotiate over pay when one of your workers gets a competing offer).

Second, the primary purpose of this book is to teach you how to think in a more sophisticated way about how workers respond to the incentives created by compensation plans. Pensions, with all of their complexities and idiosyncratic features, offer a fantastic opportunity to develop these skills and intuitions. Those skills will equip you to evaluate other aspects of compensation systems (unrelated to pensions), including unique features of compensation plans that may arise in your company but that are not even covered in this book.

11.6.1 Defined-Benefit Pensions

A *defined-benefit* (DB) pension plan uses a formula to determine the pension benefits of the retiree. The typical formula is based on three important variables: the employee's age at retirement, their number of years of service, and "some number" that depends on their salary. The first two variables are straightforward, but the third is a bit vague because it varies from plan to plan. In some plans that "number" is the average of the worker's salary over their final few years (often three years) of employment. Other times it's the average of the three highest salaries that the employee has received, which are usually (but not always) the salaries in their final three years. Or it might be their largest annual salary ever received. And there are other possibilities. How that number is defined has a big effect on how workers behave, particularly as the retirement year gets close.

CSUEB's pension plan is a DB plan. Specifically, it is the CalPERS plan, which applies to California state employees. Details of the plan can be found on this website:

www.csueastbay.edu/af/departments/hr/benefits/retirement/Retirement-pers.html

CSUEB's plan is based on three variables: *age at retirement*, *number of years of service*, and the *average of the three highest consecutive annual salaries*. Prior to 2013 (so, for example, when I joined CSUEB in 2008) the third of these variables was not the average of the highest three consecutive salaries but, more simply, the highest income from salary that was received over a period of 12 consecutive months, even if those months split two calendar years.

A numerical example that I'll discuss shortly is a bit simpler to illustrate using the older (pre-2013) definition of the third variable that doesn't involve averaging over three years. So I'm going to use the pension formula and numbers that applied when I joined in 2008, but the essential points conveyed by the example also apply to the current (post-2013) definition.

The 2008 pension formula is:

$$Annual\ Pension = (c\ /\ 100) \times Largest\ Annual\ Salary$$

where c is a percentage (a number between 0 and 100) that increases with the worker's years of service and with their retirement age. In other words, c is the *replacement rate of the pension participant's largest annual salary* (meaning the percentage of the largest annual salary that the pension participant will receive annually throughout the retirement years).

If an employee retires at 68 or at some other age, "*Annual Pension*" is the amount that they are paid every year from retirement until death. Someone who works for CSUEB for many years and waits until an old age to retire will have a very high value of c, close to 100, or perhaps even equaling 100. In contrast, someone who has only a small number of years of service will have a small value of c, and it will be even smaller if they retire young than if they retire old. Values of c cannot exceed 100, meaning that the annual pension payout cannot exceed the worker's largest annual salary. However, a more favorable tax treatment of pension benefits versus employment income (e.g., the latter is subject to Social Security taxes, whereas the former is not) means that even if c is equal to, or less than, 100, the worker's annual "take-home" pension pay after taxes might exceed their largest annual after-tax salary.

The exact value of c is determined by a simple formula. To make things easy for workers, CalPERS presents it in the form of a table, or grid, where the rows are "years of service" and the columns are "age at retirement". If you run your finger along the appropriate row, and down the appropriate column, the cell where they intersect gives the value of c. The values for c get larger towards the lower-right corner of the grid. Table 11.1 displays an abbreviated version of the table that applies to the 2008 pension formula (i.e., pre-2013):

Table 11.1 Percentage of final compensation based on retirement age and years of service

Age	51	53	55	57	59	61	63+
Benefit factor	1.280	1.640	2.000	2.126	2.250	2.376	2.500
Years of service	Percentage of final compensation, c						
5	6.40	8.20	10.00	10.63	11.25	11.88	12.50
10	12.80	16.40	20.00	21.26	22.50	23.76	25.00
15	19.20	24.60	30.00	31.89	33.75	35.64	37.50
20	25.60	32.80	40.00	42.52	45.00	47.52	50.00
25	32.00	41.00	50.00	53.15	56.25	59.40	62.50
30	38.40	49.20	60.00	63.78	67.50	71.28	75.00
35	–	57.40	70.00	74.41	78.75	83.16	87.5
40	–	–	–	85.04	90.00	95.04	100.00

For example, a worker with 25 years of service who retires at age 61 would have $c = 59.40$. If a worker has fewer than five years of service, then their value of c is zero no matter when they retire. In other words, to be eligible for any pension payment at all when they retire, a worker must provide at least five years of service. Once they have done that, they are *vested* in the pension.

The retirement age isn't necessarily the age at which the person resigned from CSUEB. "Retirement age", from the standpoint of CalPERS, is the age at which an employee chooses to begin collecting pension benefits, which may be later than the age at which they stopped working at CSUEB. For example, suppose that a professor is 55 years old, has 20 years of service, and decides to leave CSUEB to relax and pursue various hobbies. The former professor might choose to delay the receipt of pension benefits for seven years, until age 62. The benefit of doing so would be a higher value of c in the pension formula; even though the row of the retirement grid would be fixed at 20 years of service, delaying retirement means stepping to columns that are further to the right on the grid, which raises the c. The obvious cost of this strategy, however, is the delay in the receipt of the pension payouts.

This kind of pension plan may sound very expensive to provide, and indeed it is. Such plans are not as common as they once were. And they exist almost entirely in the public sector, pertaining to employees of states, cities, or municipalities. The pensions are financed partially by the employees themselves, through payroll deductions, and partially by tax revenues extracted from the general public. For example, at CSUEB, 5% of our pay is deducted from each paycheck and sent to CalPERS, but those contributions are insufficient to fully cover all the pension obligations. I mentioned that employees require five years of service before they are vested. If someone quits after four years, what happens to the 5% per year that they contributed in each of those years? They receive a one-time, lump-sum payment reimbursing them for their contributions.

11.6.2 Defined-Contribution Pensions

A *defined-contribution* (DC) pension plan is a personal account that is set up for each worker, to which both the employer and the worker contribute according to some rules that the employer decides. There are several types of DC plans in the United States, and they vary according to their rules and by who can participate. I will give some examples.

401(k) plans are the most well-known DC plans. They apply to employees of public corporations and businesses. Earnings accrue on a tax-deferred basis. Employees make contributions to the plan via payroll deductions (post-tax and/or pre-tax). Employers may offer matching contributions.

403(b) plans apply to workers in public education organizations, some nonprofits (in particular 501(c)3s, as discussed in Chapter 15), cooperative hospital service organizations, and self-employed ministers. They are similar to 401(k)s in their tax treatment; in particular, earnings growth is tax deferred. Employee contributions to 403(b)s are made pre-tax. The plans are also known as tax-sheltered annuities (TSAs), or equivalently tax-deferred annuities (TDAs) even though the plans are not restricted to annuity forms.

457(b) plans (or simply 457 plans) are available for government employers (e.g., state and municipal employers) and some types of non-government employers. Employees make contributions to the plan via payroll deductions on either a pre-tax or post-tax basis. These plans are similar to 401(k)s and 403(b)s, but the rules for 457s are less restrictive with respect to early withdrawals and eligibility. In particular, there is a 10% penalty for withdrawals from 401(k)s and 403(b)s before the age of 59.5, but that restriction doesn't apply to 457s. Moreover, independent contractors are allowed to participate in 457s (whether offered by governmental or non-governmental employers) but not in 401(k)s or 403(b)s. On the other hand, 457s are more restrictive concerning early withdrawals for hardship reasons, which must be unforeseeable emergencies (e.g., buying a house would not qualify, whereas insufficient homeowners' insurance to cover damage to a house from a hurricane would qualify).

Simplified Employee Pension (SEP) plans are retirement plans that an employer or self-employed person can establish and that allow employers to make contributions to an employee's SEP Individual Retirement Account (IRA), which is a traditional (as opposed to Roth) IRA. Employers get a tax deduction for contributions made to their employees' SEP IRA plans. The contributions are immediately 100% vested, and the employee controls how the funds are invested. As with other IRAs (whether traditional or Roth), early (i.e., before age 59.5) withdrawals are subject to a 10% penalty.

Let's consider a concrete example of one of these types of DC pensions. When I worked at Cornell, I participated in a tax-deferred annuity (TDA) plan that had the following features. For every dollar that I put into my account via a payroll deduction, Cornell would contribute another dollar. Cornell would continue matching in this fashion, dollar for dollar, up to 15% of my income. If I wished to contribute more than 15% to my account, that was fine. I could contribute as much as I wanted, but Cornell would only match up to 15%. Other DC plans are structured differently. For example, maybe they match only 50 cents on the dollar, and maybe up to 20% of income, or perhaps only up to 10%.

What happens to the money once it's in the account? It's invested in some way, and usually the worker has at least some control over how it's invested. For example, the plan offered by Cornell, which was TIAA-CREF (now simply TIAA, as of February 2016), provides a menu of stock and bond mutual funds, and workers can allocate their investments in various proportions across these funds. They can also change those allocations over time.

The assets in a DC account grow over time for two reasons. One reason is the infusion of new contributions by the worker and the employer. For example, at

Cornell, a percentage of each of my paychecks went directly into the account, along with Cornell's matching. A second reason is investment returns, assuming that the assets are invested in mutual funds and other instruments that are performing well. The assets in a DC account can also diminish over time for two reasons. One is investment losses. A second is employee withdrawals. In most cases, employees must pay a 10% penalty on any funds that they withdraw before they reach the normal retirement age. However, they can take short-term loans from the retirement account, before retiring, without paying the penalty.

When a worker leaves a firm that has a DC pension plan, the worker continues to own that account even after leaving. I still own my DC account from Cornell, even though I haven't worked there for over a decade. The only difference is that I no longer contribute funds to it, and obviously Cornell doesn't either because I no longer work there. But the value in the account fluctuates from day to day according to the returns of the mutual funds in which it is invested.

11.6.3 Risk and Pensions

Retirement planning is pretty easy for workers on a DB plan. They know in advance exactly what annual payment they will receive each year in retirement, and they know this many years in advance of their actual retirement date. They face no market risk; even if the stock market crashes, they are still owed a pension determined by the formula in their DB plan. Of course, there is a risk that the entity that owes the worker the pension will renege or go bankrupt. But that is rare. Reneging is more likely to happen with "smaller" entities (e.g., it is more likely to happen with municipal pension plans than with state plans). Detroit reneged on its pension obligations, but it's less likely (I hope!) that the State of California will renege, particularly because of pension reforms in recent years that have enhanced the pension's long-run viability, some of which I discuss later in the chapter.

In contrast, retirement planning is more difficult under a DC plan, because the value of the DC plan fluctuates with the market, and the worker is fully exposed to that risk. A worker might plan to retire in five years, but after the third year the stock market might crash, and the DC account could suddenly lose a third of its value. As workers get closer to their desired retirement ages, they can shift the assets in their DC accounts more and more towards lower-risk investments, but those also have lower expected returns.

Inflation also poses risks to the value of pensions, though this is mitigated by annual cost-of-living adjustments (COLAs) in the case of DB pensions. When the inflation rate is high, each dollar of a worker's pension is worth less because the prices of goods and services are higher. So a dollar doesn't go as far. There are two periods in the "life" of a pension . . . a "pre-retirement" period during which money is being contributed to the pension, and a "post-retirement" period in which money is being drawn out of the pension. Inflation can occur in either or both periods, but it is particularly damaging to workers in the post-retirement period. I'm about to explain why, but before reading on, you might want to review Chapter 1's appendix on the subject of nominal versus real compensation.

Consider a DB plan. Recall that one of the three main variables in the formula for a DB plan is "some number" that is a measure of a worker's salary. In a high-inflation environment, salaries will often rise quickly, just like other prices. If all of the prices of goods and services that a worker wishes to buy double, that's not so bad as long as the worker's salary also doubles. And if the salary doubles, the pension amount increases accordingly. But what happens once a worker retires? Then the pension formula dictates an amount that the worker will receive every year until death. Suppose that's \$50,000. Inflation erodes the real value of that \$50,000 over time. For example, suppose that the prices of goods and services double in the ten years after a worker retires with a \$50,000-per-year pension. Then the purchasing power of that annual pension ten years after retirement is half of what it was at the start of retirement. The CalPERS DB plan indexes annual pension payments, increasing them by a small percentage each year to mitigate the damaging effects of post-retirement inflation.

The preceding issues also affect DC plans, i.e., they are hurt by post-retirement inflation but not by pre-retirement inflation. During the pre-retirement period the pension is being infused with new contributions from the worker and the employer. In an inflationary environment, those contributions typically increase along with all other prices. But once workers retire, the contributions stop, and the withdrawals begin. In that post-retirement phase, inflation is very damaging to the value in the pension account. Just like post-retirement inflation eroded the value of the fixed annual pension payout under the DB plan, it erodes the value of the fluctuating amount in the DC account in the years after retirement.

Apart from inflation risk (which affects both DB and DC plans in the post-retirement stage) another concern is market risk, which only affects DC plans. If a worker is on a DC plan and has accumulated a large chunk of money in the account by the time a normal retirement age is reached, that worker might choose to retire. Then the worker draws from the account throughout retirement. But suppose that the stock market crashes shortly after retirement, and the value of the worker's account takes a big hit. The worker is out of luck. Usually workers invest pension funds more conservatively as they age, to soften the hit of such an event. But still there will be some hit as long as the worker has retained some exposure to equity markets. In contrast, on a DB plan, retirees are immune from fluctuations in the market, because their annual payments are based on a formula that doesn't vary with market conditions. In short, workers on DC plans are exposed to market risk, whereas workers on DB plans are not.

It is possible, however, for workers on DC plans to eliminate market risk by using the DC account, upon retirement, to purchase an annuity. In exchange for a large lump-sum payment, the annuity will pay out a fixed amount each year, indefinitely, similar to what happens for retirees on a DB plan. In this fashion, purchasing an annuity mimics a DB plan and allows retirees to mitigate market risk. Workers who are more pessimistic or uncertain about future market returns (or who expect to live a really long time, beating the life-expectancy statistics) would find the annuity appealing, whereas those who are more confident and optimistic about future market returns would prefer to retain their full DC account in hopes of enjoying future investment returns on a larger chunk of cash.

One practice that increases the risk that workers bear from a DC pension is when the pension contributions are not invested in a broadly diversified fashion. For example, if Google employees have a large chunk of their pension contributions invested in technology stocks, that's a bad idea. And it's an even worse idea if the investments are in shares of Google stock. Google employees will already suffer if the technology sector, and particularly Google itself, takes a hit. There's no reason to make that hit even worse by having the retirement funds positively correlated with Google's fate. The more diversification in a portfolio the better when it comes to investing, and the same is true for pension contributions.

11.6.4 Pensions and Retirement Ages

The two different types of pension plans (DB and DC) have very different effects on workers' retirement ages. Remember the discussion at the start of section 11.6 concerning my former colleague at CSUEB, where the pension plan is DB. He retired at a relatively young age, citing his desire not to be paid a "negative wage". Once the years of service become very high, the *c* parameter in the pension formula nears 100. If it equals 100, then a worker's annual pension payment each year between retirement and death is (using the older CalPERS formula I discussed earlier) the same as their largest annual salary received during the career. If that "largest salary" is the worker's current salary in the last year on the job, then continuing to work means getting paid compensation of $0. The reason is that the worker could choose to retire, which means doing no work, and still get paid a full salary. So a full salary is received whether the worker chooses to work or not, and working provides no further income.

But suppose in the preceding example that the worker's highest salary received so far is not the current salary but rather the annual salary received four years ago. Perhaps that salary was the highest because the worker temporarily changed job roles within the organization, assuming a higher-paid role (such as becoming the interim associate dean in the business school at CSUEB). In that case, the worker would make a negative wage by continuing to work rather than retiring. If *c* nears 100, then the worker can receive the highest salary in the work history every year until death (simply by not working), or alternatively the worker can continue working and receive the current (lower) annual salary. Continuing to work effectively means a negative wage, because each year the worker loses the difference between those two numbers.

The likelihood of making a negative wage by continuing to work is even higher than I just described, and you can guess why if you remember my earlier point about how pension income and wage-and-salary income are taxed differently. Pension benefits are exempt from certain payroll taxes that must be paid on wage-and-salary income. Therefore, c doesn't need to reach as high as 100 before a worker, upon retirement, is entitled to full replacement of the maximum after-tax income received in the work history. Full replacement of after-tax income actually occurs at a number well under 100 . . . more like 85.

So that's the situation my CSUEB colleague faced that caused him to retire at a young age while still in his 50s. He spent his entire career at CSUEB, starting at age 27, so he had a lot of years of service. Moreover, he had exploited a special deal that CalPERS used to offer, which allowed workers to purchase "additional years" of service, at an attractive price. Up to five additional years of service could be purchased. For example, suppose that a worker buys five extra years, and works for a 26-year career before retiring. In the worker's pension calculation, "years of service" would be counted as 31 (i.e., 26 + 5) even though the worker only actually worked for 26 years. Purchasing five extra years allows someone to get their c close to 100 a full five years earlier. It was the purchase of five extra years that faced my colleague with the prospect of a negative wage if he had continued to work, even though he was still fairly young.

In 2011, CalPERS made some major changes to render the DB plan less generous for future employees. I was grandfathered, because I joined in 2008. However, one of the changes was to eliminate the offer to purchase up to five extra years of service. I believe the last year in which workers were allowed to do this was 2012. I was ineligible to purchase years, however, because I wasn't yet vested in the pension; I had only four years of service instead of the required five. Consequently, I won't be faced with the prospect of a negative wage for a long time!

The key point to notice is that on a DB plan, once years of service and age get sufficiently high, it strongly pays to retire. And lots of time that happens well before the point at which a negative wage would be reached. Suppose that c is high enough that a worker is entitled to 85 or 90% of their (highest ever) after-tax salary in a period of 12 consecutive months, simply by not working. Even if that's a bit less than they'd make by continuing to work, it's an attractive deal because they're being paid to do nothing! If they want, they can go take another job somewhere else, and effectively collect two salaries simultaneously while working only one job (the new one). Many professors at CSUEB find such prospects appealing when c gets high, and they retire early.

Now switch gears and think about what happens on a DC plan, which is very different. On that type of plan, there is no penalty from continuing to work. The worker can keep going beyond age 70 and 80, and although there are market fluctuations, over time the DC account can be expected to keep growing, both from worker contributions and from employer contributions (and, hopefully, investment returns). It's true that the longer the employee works, and the longer retirement is delayed, the less time there is to spend the money before death. But upon death, whatever money that remains in the account goes to the worker's estate, so the employer doesn't benefit if the worker dies before retiring or shortly after retiring. In contrast, under a DB plan, the employer benefits greatly and receives a windfall if the worker dies before or shortly after retirement. When workers are on a DB plan, their employers (or, more specifically, those who manage their DB plan) hope that they die quickly upon retiring!

Workers on a DB plan who retire young and live a really long time cost the pension a lot of money, and what they get out of it far exceeds what they put into it. In contrast, if workers are on a DC plan, the employer doesn't care how long they live, because even if they live until 105 it doesn't cost the employer anything more . . . the employer's contributions to the DC account stop the moment the worker leaves the organization. DB plans become extremely expensive to maintain when workers live for many years after retiring.

Recently, one of my two departments (Economics) hired a new assistant professor. After he accepted CSUEB's job offer, but before his first day of employment, he emailed me some HR-related questions. One of those questions was motivated by his reading of the URL I provided in Section 11.6.1, with details on the CalPERS plan. He asked,

> *Why is it called an account if it looks like I would be entitled to 2%*#Years_Service*Highest_Annual_Pay_Amount? i.e. what if I retire and live to be 105, wouldn't my withdrawals far exceed the amount I put into the system?*

The answer to his question is, "Yes!" His question insinuates surprise that the system would allow a worker to collect far more in benefits than the worker actually put into the system. But that's exactly the point we've been discussing. And it's the reason DB plans are often hard to sustain financially and why the plans sometimes become insolvent, leading to reneging on contracts with workers (a form of wage theft, as discussed in Chapter 2). That's what happened in Detroit and other cities. To avert such collapses of the system, sometimes the managers of a pension fund will take steps to lower future costs, by reducing benefits. Often current workers are grandfathered so that only new workers lose out. That's what happened with CalPERS several years ago. All professors who joined CSUEB in 2011 or later face a less generous pension formula than those who joined earlier.

The fact that the older CalPERS pension formula on which I've focused for illustrative purposes is based on the maximum annual salary that the worker achieved during all years of service creates incentives for workers to "spike".

Spiking means purposely increasing one's work hours or accepting a different job assignment, for a short duration, with the primary purpose of increasing the maximum annual salary for the pension calculation. Successfully doing so creates a spike upward in a worker's pension income. Hence the name, "spiking". In other DB pension systems, in some occupations, this can be achieved by working overtime hours. At CSUEB, this is not possible. As I mentioned near the end of section 7.1, only a professor's base salary counts in the pension calculation, and all other payments (e.g., one-time bonus payments for undertaking additional tasks) are not included.

But back to those workers who are on DC plans, such as my former colleagues at Cornell . . . none of them have an incentive to retire young. If they keep working, abstracting from short-run market fluctuations, their DC account just grows larger and larger with continuous employer and worker contributions. They will own the entire account when they eventually leave the university, and any of it that remains after their death transfers fully to their estate.

11.6.5 Sorting and Turnover

To an even greater extent than other benefits, pensions have a powerful effect on sorting and turnover. One reason concerns vesting schedules. Workers often must wait several years before becoming vested in the pension. In the CalPERS DB plan, it is five years. In the US military, until recently, it was 20 years. These vesting schedules are referred to as "cliff vesting" . . . a worker abruptly goes from having *no* pension eligibility to *full* pension eligibility as soon as the required time period for vesting is met. The name "cliff vesting" is based on the graph of pension benefits, which resembles a sharp cliff that occurs at the end of the vesting period.

Figure 11.1 displays the graph of pension benefits for a pension with cliff vesting on a five-year vesting schedule, such as CalPERS. The analogous graph for the pension system previously used in the US military would be even more dramatic; it would coincide with the horizontal axis for the full 20 years of the vesting period and then would have an enormous jump up at the 20-year mark, rising linearly thereafter as in Figure 11.1. Particularly on such a long vesting schedule with a huge cliff, virtually no turnover is seen in the period shortly before vesting, but turnover spikes at the moment of vesting or shortly thereafter.

This worker behavior in the neighborhood of the vesting period's termination is reminiscent of the worker behavior that we have seen in the neighborhood of the "kinks" in pay-for-performance plans (e.g., section 9.8 on district managers at Borders, and section 10.2 on executive bonuses). In both cases, workers' incentives are particularly strong in the region of their compensation graph where it dramatically changes slope. The difference is that in the case of performance-pay plans we are thinking of pay as a function of *worker performance*, P, whereas in the case of pensions we are thinking of pay (in particular pension compensation) as a function

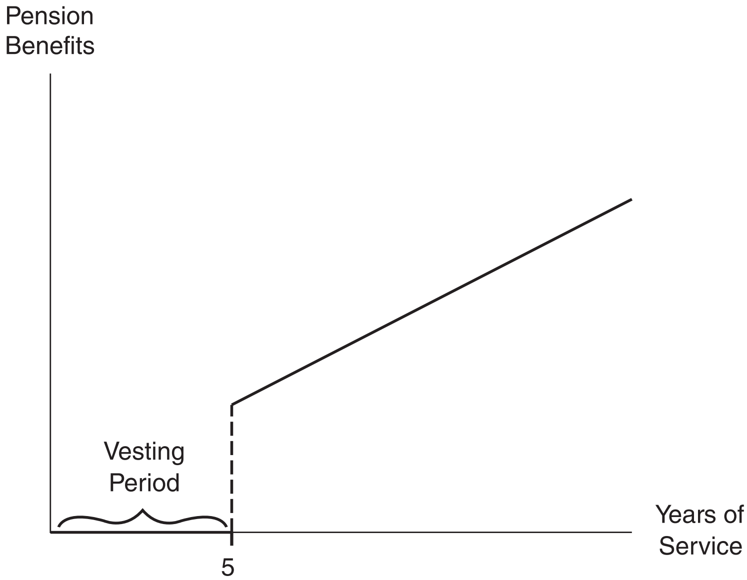


Figure 11.1 "Cliff vesting" of pension with a five-year vesting schedule.

of *years of service*. But the logic is very similar, because *P* and *years of service* are both variables that your workers control, and in making their choices about those variables your workers are keenly aware of how their compensation will be affected.

Even after vesting occurs, other features of the pension may limit mobility. For example, in the CalPERS DB system, the State of California pays a portion of retirees' health insurance, in addition to the pension payments. How much the State contributes to health insurance depends on a worker's years of service, and the vesting schedule for health coverage is longer than the five-year vesting schedule for the pension. The rules have become progressively less generous to workers over time. Currently, after a worker hits 15 years of service, the State covers 50% of their health insurance. Each additional year of service adds another 5 percentage points, so that workers who have 25 or more years of service get 100% of the cost covered by the State. This vesting schedule for health coverage looks similar to cliff vesting for the first 15 years, because the State's coverage jumps from 0% to 50% once the 15-year mark is crossed. After that major cliff, each subsequent year of service (up to 25) represents a smaller cliff in which the State's coverage increases by increments of 5 percentage points.

DB plans are more likely than DC plans to reduce worker mobility and tie workers down. Once a worker has accumulated a bunch of years of service in a DB plan, it is costly to switch employers unless the plan is "portable", meaning it transfers, undiminished, to the next employer. DC pensions are fully portable, because they are accounts owned entirely by the worker, so when the worker leaves, their account goes with them. The DB CalPERS plan is partially portable. If I were to move to one of the 22 other campuses in the CSU system, my pension would remain fully intact. So within the CSU system, the pension is fully portable. If, instead, I were to move to a university in another state, the pension would not transfer. I would have to start over from scratch in the new university's pension system. Even if the new university has an *identical* pension system, starting over from scratch (and having two pensions, one from the new university and the earlier one from CSUEB) is worse than staying at CSUEB. The following example illustrates:

Example 11.1

Suppose that a professor joins CSUEB at age 31, on a defined-benefit pension plan with a five-year vesting schedule, where the annual pension at retirement is given by the following formula that applied in 2008:

$$\textit{Annual Pension} = (c\ /\ 100) \times \textit{Largest Annual Salary},$$

and c is given by table presented in section 11.6.1. Suppose that the professor starts at an annual salary of \$90,000 and receives a 2% raise every year throughout the career. The professor plans to work for 30 years, retiring from CSUEB at age 61, and starting pension payments immediately thereafter. For 30 years of service and a retirement age of 61, the value of c is 71.28%.

a. What will the professor's annual pension income be upon retirement?
b. Suppose that after 20 years at CSUEB, the professor moves to another university that has an identical pension and system of annual raises. However, the CSUEB plan is not portable, so the professor must start over on the new pension system for the last 10 years of the career, collecting two pensions at age 61. What will the professor's annual pension income be upon retirement?
c. Suppose that the professor switches to the new university after 10 years rather than after 20. What will the annual pension income be upon retirement?
d. Explain why the answers differ across the preceding three cases.

Answer:

a. *Because the salary increases each year by 2%, the professor's largest annual salary is the salary received at age 61, just before retirement. That is \$90,000 × $(1.02)^{30}$, which is about \$163,022. The annual pension is then (71.28 / 100) × \$163,022, which is about \$116,202.*
b. *The professor's final CSUEB salary, which is also the starting salary at the new university, is \$90,000 × (1.02)20, which is approximately \$133,735. The professor's final salary upon retirement, after ten years of service at the new university, is about \$163,022, just as it was if all 30 years had been spent at*

CSUEB. The annual pension is then (47.52 / 100) × \$133,735 + (23.76 / 100) × \$163,022, which is about \$102,285.

c. *The professor's final CSUEB salary is \$90,000 × (1.02)10, which is about \$109,709. The annual pension is then (23.76 / 100) × \$109,709 + (47.52 / 100) × \$163,022, which is about \$103,535.*

d. *Even though the salaries, raises, and pension systems are identical between the two institutions, the pension payments are lower if the worker switches institutions mid-career. In all cases, the worker ends the career with a salary of \$163,022. If the entire career is spent at CSUEB then the entire 30 years of service are applied to that number, whereas if the worker has moved, then a smaller number of service years is applied to that maximum salary (and the other service years are applied to a lower salary number). But if the worker is going to switch universities, it's less costly to do it after 10 years than after 20, because after 10 years the worker can have 20 service years at the new university (as opposed to 10) applied to the maximum salary of \$163,022. The difference between parts (b) and (c) reveals that the longer a worker is in the pension system the more costly it is to drop out of it and start over somewhere else.*

The preceding example illustrates how a DB plan can form a powerful bond that attaches workers to an organization. The reason that CSUEB's pension is partially portable is that it is a plan that pertains not to one university but to all 23 campuses in the CSU system, and the terms of the pension are negotiated by the faculty union every three years and appear in the collective bargaining agreement (see Chapter 5). This is commonly the case, i.e., DB plans often are not tied to a specific institution but rather apply to a broader set of institutions (e.g., all city government jobs, or all jobs within a state university system). In the preceding example, if the new university were within the CSU system (e.g., California State University–Long Beach) then there would be no cost to the worker to switching institutions, because the entire 30-year career would be applied to the maximum salary of \$163,022, regardless of how those 30 years were spread across the two universities. So the chemistry and biochemistry professors at the three CSU campuses in Chapter 7 could freely move to any of the other CSU campuses without harm to their pension benefits.

A DB plan that is fully portable is the US Social Security system. That system has a vesting schedule of ten years, though the years need not be worked consecutively. So someone who works for three years, then takes five years off to go to school, then works another six years, then is unemployed for a year, then works for another year (the tenth), would become vested. In this example, it doesn't matter if the three periods of employment are with different employers. All that matters for vesting is the total years of service. The plan is a DB plan because there is a formula that determines what Social Security benefits a worker will receive each year, once the worker starts collecting benefits. As with any DB plan, if the worker lives for a long time, it's possible that he will collect far more Social Security payments than he paid into the system. In that case the government, i.e., the US taxpayers, would have to finance the shortfall.

11.7 Lessons for Managers

Benefits comprise a significant share of your organization's total compensation costs. About 30% of your employees' compensation may come from benefits, and this number could even be as high as 40%, particularly in the public sector. What I dubbed the "sweaters problem" is a considerable downside to offering benefits compensation, but there are several positive features of benefits that must be weighed against that big negative. Sorting effects are particularly powerful in the benefits component of compensation, because workers have different preferences for non-monetary job characteristics. The sorting effects can either be positive or negative, and as a manager you should be keenly sensitive to them when evaluating your current benefits offerings and considering changes to them.

Cafeteria plans mitigate the "sweaters problem" but at the expense of eroding some of the positive features of benefits compensation. As a manager, you should remain alert to creative opportunities for mitigating the "sweaters problem" without significantly undermining the positive aspects of the benefits package.

DB pensions are an extremely powerful tool for achieving employee retention. The problem is that those retention properties apply just as strongly to your weaker workers as to your top talent. DB pensions apply across the board to all employees, or to large groups of employees within the organization, and they cannot generally be tailored differently for high-value and low-value workers, unlike the buyout packages to be discussed in Chapter 12. Other components of compensation are better for achieving such differentiation (e.g., stock options with vesting schedules, which can be awarded more generously to your high-value employees). Vesting schedules and other characteristics of pensions, both DB and DC, also induce strong sorting effects; employees who intend to hold the job for some time will not be deterred by a vesting period. Workers who plan to leave in the near future would be put off by a vesting schedule and would prefer to have a higher salary with no pension at all. Again, you should be cognizant of the likely effects of the design of your company's pension plan(s) on the composition of employees that you can attract and retain, i.e., the sorting effect.

Pensions, and indeed benefits more generally, tend to have a "one-size-fits-all" flavor, meaning that they apply to all employees, or to large groups of employees, within the organization. As a manager, this is a part of the compensation plan over which you're likely to have little or no control, because such pension plans are typically run at higher levels than an individual organization (e.g., CalPERS applies broadly to California State employees in many different organizations). If you are a manager at one establishment in a multi-establishment firm, it is likely that the benefits package, including the pension, is designed at the corporate headquarters for the firm and applies to all of its locations, as in the multi-establishment firms in the case discussion at the end of Chapter 7. Your job as a manager, in that case, is to treat the pension system as given and to adjust the components of compensation that you *can* control to achieve the desired objectives (e.g., talent management). Typically those are monetary components like base salaries and bonuses.

Case Discussion 17: Walmart[1]

An internal memo sent to Walmart Inc.'s board of directors in 2005, and written by Ms. Susan Chambers (Walmart's executive VP for benefits), proposed ways to hold down spending on healthcare and other benefits while seeking to minimize damage to the retailer's reputation (because critics had attacked Walmart for being stingy on wage and health coverage). Recommendations included hiring more part-time workers and discouraging unhealthy people from working at Walmart. The memo also recommended reducing 401(k) contributions and taking other steps to attract younger and healthier workers. "It will be far easier to attract and retain a healthier work force than it will be to change behavior in an existing one," the memo said. "These moves would also dissuade unhealthy people from coming to work at Walmart." Note that these quotes directly pertain to incentive and sorting effects.

The memo voiced concern that workers were staying with Walmart longer, pushing up wage costs. Although it stopped short of calling for efforts to push out more senior workers, stating that "the cost of an associate with seven years of tenure is almost 55% more than the cost of an associate with one year of tenure, yet there is no difference in his or her productivity. Moreover, because we pay an associate more in salary and benefits as his or her tenure increases, we are pricing that associate out of the labor market, increasing the likelihood that he or she will stay with Walmart". The memo noted, "The least healthy, least productive associates are more satisfied with their benefits than other segments and are interested in longer careers with Walmart."

[1] Note: This case is based on a *New York Times* article (October 26, 2005) by Steven Greenhouse and Michael Barbaro entitled "Wal-Mart Memo Suggests Ways to Cut Employee Benefit Costs".

Questions

1. What are some steps Walmart can take to achieve the desired sorting effects?
   (a) What are the cost implications of the proposed steps in #1?
   (b) What are the public relations implications of the proposed steps in #1?
2. The theme throughout the memo was how to slow the increase in benefit costs without giving more ammunition to critics who contend that Walmart's wages and benefits are dragging down those of other American workers. Eager to burnish Walmart's image, CEO H. Lee Scott Jr. said that Walmart supported raising the minimum wage to help Walmart's customers. What are the advantages and disadvantages (to Walmart) of an increase in the minimum wage?
3. Ms. Chambers also said that she made her recommendations after surveying employees about how they felt about the benefits plan. "This is not about cutting," she said. "This is about redirecting savings to another part of their benefit plans."
   (a) Write down three questions that you would have included in this survey of employees to elicit employee preferences. Explain any biases you might expect in the answers to such questions.
4. The memo noted that Walmart workers "are getting sicker than the national population, particularly in obesity-related diseases", including diabetes and coronary artery disease. The memo said Walmart workers tended to overuse emergency rooms and underuse prescriptions and doctor visits, perhaps from previous experience with Medicaid.
   (a) What can be done to combat these problems? Discuss cost and public relations implications.

Further Reading

Artz, B. (2010). "Fringe Benefits and Job Satisfaction." *International Journal of Manpower*, 31, 626–44.

Baughman, R., DiNardi, D., and Holtz-Eakin, D. (2003). "Productivity and Wage Effects of 'Family-Friendly' Fringe Benefits." *International Journal of Manpower*, 24, 247–59.

Carrington, W. J., McCue, K., and Pierce, B. (2002). "Nondiscrimination Rules and the Distribution of Fringe Benefits." *Journal of Labor Economics*, 20, S5–S33.

Cooper, R. (2018). "Recent Changes to Fringe Benefits." *TAXtalk*, 71, 52–55.

Dale-Olsen, H. (2006). "Wages, Fringe Benefits and Worker Turnover." *Labour Economics*, 13, 87–105.

Kolstad, J. and Kowalski, A. E. (2016). "Mandate-Based Health Reform and the Labor Market: Evidence from the Massachusetts Reform." *Journal of Health Economics*, 47, 81–106.

Royalty, A. B. and Hagens, J. (2005). "The Effect of Premiums on the Decision to Participate in Health Insurance and Other Fringe Benefits Offered by the Employer: Evidence from a Real-World Experiment." *Journal of Health Economics*, 24, 95–112.

Simon, K. I. and Kaestner, R. (2004). "Do Minimum Wages Affect Non-Wage Job Attributes? Evidence on Fringe Benefits." *Industrial and Labor Relations Review*, 58, 52–70.

Zawacki, A., Vistner, J. P., and Buchmueller, T. C. (2018). "Why Are Employer-Sponsored Health Insurance Premiums Higher in the Public Sector than in the Private Sector?" *Monthly Labor Review*, September, 1–19.

12 Turnover Management and Talent Retention

In the 2009–2010 academic year, I unexpectedly got a "raise" of -10%. Yes, you read that correctly. My "raise" was, indeed, a pay cut! I was not alone. All CSUEB faculty and staff suffered the same fate. The grim occasion was a one-year "furlough" for California State employees, in response to a severe state budget crisis. It doesn't feel good at all, let me tell you, to learn that for the next year you're only going to get 90 cents for every dollar of your salary, before taxes. It's enough to make you want to . . . well . . . *quit*.

And some employees did just that. Though not quite as many as you might guess. One reason why there weren't more quits is that the pay cut was known to be temporary, lasting "only" a year, so it wasn't really worth the hassle of searching for a new job that restored your original salary, because by the time you'd find and start such a job your original salary would be restored anyway. Another reason is that the same deep recession that triggered the pay cut made it a bad time to look for work; jobs were scarce. So, although a number of employees quit, most of us "sheltered in place" to wait out the storm. I remember hearing a lot of grumbling at the office, though, and talk of quitting. Even though the number of people who actually quit was small, many others were definitely cranky and were dusting off their résumés and perusing online job ads. If the pay cut had been deeper or more prolonged, others would have jumped ship, and I would have been among them.

Is such turnover bad or good for organizations? The answer isn't clear and depends on various factors, first and foremost being whether the employees who depart are your stars or your deadwood. Turnover can be very costly for organizations, especially when the workers who leave are your stars rather than your less-productive workers. You'll probably have to replace your departing stars, which means incurring fresh recruitment, screening, and training costs, not to mention the costs of being understaffed until you successfully hire a replacement. Probably it will take time for the new employee to get up to speed, so there will be an initial period of reduced productivity. Moreover, your workers who left had likely amassed firm-specific skills (see Chapter 8) during their time with your company; all of that is lost when they leave and takes time to rebuild in a new employee. Departures may also cause a drop in the productivity of your remaining employees if their work was closely interrelated with that of the worker who left, which is particularly likely in a team setting. There is also the risk that your former employees may behave in ways that hurt your organization. After all, once they leave, their loyalties no longer lie with your organization. They might take trade secrets to a competing firm. Or if they are disgruntled, they might harm your organization's reputation, which can impede your recruitment and talent retention . . . such risks are amplified in the age of the internet and social media (e.g., disgruntled former employees can post damaging online reviews on Glassdoor).

But turnover isn't always bad. Sometimes the departing workers are those you're happy to see leave. Even if you're sorry to see them leave, the replacement workers that you'll hire will bring fresh ideas and outside perspectives that can be valuable and inject dynamism into the workplace. The vacancies resulting from turnover also create promotion opportunities (see Chapter 13) that can be a powerful source of incentives for your workforce. Notice that this paragraph is shorter than the last one, and for good reason. Even though the benefits of turnover can be significant, more often than not the downsides of turnover loom even larger.

12.1 Turnover and the Level of Compensation

The example that opened this chapter illustrates the tight connection between compensation and turnover. Deep cuts in pay tend to trigger quits, and deeper cuts trigger even more quits. On the other side of the coin, generous raises encourage workers to stay. This implies that, as a manager, you can use the compensation system as a tool to manage turnover. But raising and lowering the compensation level isn't the only way to accomplish that, and simply raising your workers' pay is an expensive retention strategy. Changes in the design of an employee's compensation package that leave the overall level, or generosity, of compensation unchanged can also affect turnover. I'll elaborate in the next section, but first let's think about the consequences of a simple increase or decrease in the level of pay.

Which CSUEB employees quit in response to the 10% pay cut? Were they the stars of the organization? Or were they the deadwood? Unfortunately for CSUEB, they tended to be the stars. That's commonly the case in organizations, and it's an important point. If you cut pay across the board for everyone, the workers who quit aren't randomly drawn from your staff. Your better workers tend to be the ones who are most likely to quit, because they're typically the ones with the highest market values. But I need to be a bit more careful with this argument . . .

If your organization is about to undergo an across-the-board pay cut, and you're wondering which workers are most likely to quit as a consequence, it might not be the better workers who are at greatest risk. I said that *tends* to happen, but it doesn't always. Remember that the key factor determining whether a worker stays or quits is not whether they're one of your better workers . . . it's their valuation of what they could get paid at another company compared to their valuation of what you're currently paying them. It may happen that your best worker is very handsomely paid and that even after a big pay cut she'd be making more than most competing companies would be willing to pay her. So she's not a flight risk; even if she has a very high market value (perhaps higher than the market values of any of your other workers) she'd still be getting paid more by you, even after the pay cut, than she could hope to make elsewhere.

You've hopefully noticed a challenge that surrounds all this. It's often difficult to accurately assess what your workers could likely make at another company. Sometimes you have a rough idea, and other times you have the misfortune of finding out exactly (when one of your workers gets a competing offer and brings it to your attention, expecting you to match it and threatening to leave if you don't!) But often you just have to make an educated guess based on whatever clues and bits of information you've been able to cobble together. Unfortunately, the problem is even harder than I just described. I claimed that the key piece of information that you need to know is "what they could get paid at another company compared to what you're currently paying them". You might think that even if you're in the dark about half of this information (i.e., "what they could get paid at another company"), at least you know the other half exactly (i.e., "what you're currently paying them").

The sad truth, however, is that you're also partially in the dark about "what you're currently paying them". Of course, you know exactly what salary, annual bonus, and fringe benefits you're paying them. But you don't know exactly how much they *value* the entire compensation package. Remember the broad definition of compensation from section 1.1: *everything the person likes about a job*. That includes an array of features like the geographic location, commuting time, office size and characteristics, organizational culture, rapport with co-workers, etc. Maybe one of your key employees places such high value on these things that they would stay even if offered a considerably higher salary at another company.

In short, your challenge is to guess what value your employees place on their current compensation compared to their valuation of their best offer someplace else. That's very hard to do. But improving even a little on that difficult task will pay big dividends in terms of a successful talent management strategy. Coaxing such information out of your employees is difficult, because it's not necessarily in their interests to reveal it honestly. As a creative manager, however, you should be on the lookout for ways to squeeze bits of this information out of your workers. I discuss some strategies for doing this in Chapter 14, which covers negotiation and bargaining. That chapter and the present one are closely related. Devising a sound talent retention strategy (Chapter 12) and negotiating effectively over compensation packages (Chapter 14) both require collecting accurate information on how your workers value their current job relative to their next best option at another firm.

One way to elicit such information truthfully from workers was implemented by Merrill Lynch during the fall of 2001. Anticipating a big round of layoffs in late 2001, Merrill Lynch offered all of its workers a voluntary buyout package. Workers who volunteered to leave the company would receive a generous severance package. But not all workers who stepped forward to take the deal would necessarily be granted the deal. Workers who expressed interest in the deal were evaluated by their supervisors, who decided whether or not to grant them the deal. In structuring the voluntary buyout plan this way, Merrill Lynch was able to extract valuable information about how attached its workers were to their jobs. Any worker who voluntarily stepped forward to take a buyout deal with a severance package of $X obviously valued their job at Merrill Lynch at less than $X, and that information could potentially be useful to Merrill Lynch in designing future retention (or anti-retention) strategies for that worker or for other similar workers. Merrill Lynch's voluntary buyout offer is explored in a pair of related case discussions, the first of which concludes this chapter.

Pay cuts can certainly induce quits, and sometimes your goal as a manager is to induce quits. The most extreme form of this is when you fire a worker, which is effectively a pay cut of 100%. But tread carefully. The external and internal constraints that we discussed in Chapters 4 and 5 can complicate your efforts to reduce wages (e.g., consider minimum-wage laws, prevailing wages, living wages, and union contracts, all of which create wage floors) or to fire workers, particularly when they are members of demographic groups that enjoy legal protections (e.g., women, non-whites, older workers, and disabled workers). Sometimes, even if you face no legal hurdles to firing a particular worker, that decision could backfire. For example, your former employee might find a job at a competing firm and may use the information gleaned at your firm to enable the new firm to compete with you more effectively (e.g., by sharing trade secrets).

Or your former employee might take vindictive actions that tarnish your brand and hinder your future recruitment and retention efforts. An example would be the departure (by firing) of the *Assistant to the President and Director of Communications for the Office of Public Liaison* from the Trump White House in early 2018. The White House concluded that the costs of continuing to employ Ms. Omarosa Manigualt-Newman in the White House were exceeded by the costs of having her leave, which is why she was fired. But did the White House accurately measure those costs, and in particular the costs of her leaving? After her involuntary departure, Ms. Manigualt-Newman published a book (*Unhinged*) that harshly criticized President Trump, she did several interviews in which she vigorously attacked him, and she released several damaging audio recordings that she had secretly taped while in the White House. All of this eroded the credibility of the White House, and one might wonder whether it would have been easier for the Trump administration to keep her or to fire her under more generous terms (see sections 12.5 and 12.6).

12.1.1 Salary Ranges, Range Spreads, Compa-Ratios, and "Compe-Ratios"

When considering your employee's likelihood of leaving the company, you should think about how their compensation compares to what "the market" would offer them. By "the market" offer I mean the *best* offer that your worker could obtain from a competing firm. The higher the compensation you're offering your worker relative to the best that they could get from a competitor, the less likely they are to quit. The challenge here is that you typically won't know in advance what the best possible competing offer is that your worker could get. Therefore, a different and imperfect approach is often taken in the professional "comp" world based on readily available information.

. . .

Compensation professionals like to create a numerical score for each worker that aims to summarize the market competitiveness of the worker's compensation. That score is called the ***compa-ratio*** (pronounced *COMP*-uh), which is short for "comparative ratio". It is simply the ratio of the worker's current salary to the so-called "market salary". But to avoid confusion, please remember that I'm not using the term "market salary" the same way I used it in the preceding paragraph! In the preceding paragraph, it meant the *best* salary offer that your worker could obtain from a competing company. That's the theoretically ideal measure of what the market could offer, but again, it's generally impossible for you to observe it. In the present paragraph, the term "market salary" means the midpoint of the ***salary range*** for the job the worker holds.

The salary range for a given job consists of three numbers (namely the ***minimum salary***, ***maximum salary***, and ***salary midpoint***) that you and the rest of the management team choose. Typically, to define a salary range you assign the midpoint first. You might set the midpoint to the average salary in the industry for that position (which presumably you know or can estimate using market salary data). Or alternatively you might set it to the median salary in the industry for that position. Nothing requires you to choose the industry average (or median) for the salary range midpoint. You could choose a more generous midpoint or a less generous one. A more generous one is expensive but might improve your applicant pool and lower your turnover rate. A less generous one might yield a more modest applicant pool and increase your turnover rate, but it will also lower your salary costs.

Once you've chosen the midpoint of the salary range for the position in question, you set the ***range spread*** (i.e., the difference between the maximum salary and the minimum salary for that salary range) to your desired width, but choosing the maximum and minimum so that the salary midpoint is always the average of those two numbers. For example, if you set the salary midpoint at $60,000 and the maximum of the range at $72,000, then the minimum of the range must be $48,000. The midpoint lies exactly $12,000 below the maximum and $12,000 above the minimum, so that the average of the maximum and minimum is $60,000. The range spread is $24,000. The salary midpoint is what's used as the "market salary" in the compa-ratio's denominator.

To illustrate all of this with a concrete example, let's revisit the CSU chemistry and biochemistry professors from section 7.1. Before developing the example, I want to make clear that it's just for illustrative purposes and that I chose it simply because you're already familiar with the data. It brushes some important institutional features under the rug. In particular, in the CSU system there are internal constraints (see Chapter 5) in the form of a collective bargaining agreement that determines salaries. But the subsequent discussion assumes that you and the rest of the management team can set the salary range freely for each job, as opposed to the reality in which the relevant ranges are contractually set and the result of bargaining with the faculty union.

Recall that throughout section 7.1 you were a compensation consultant who was hired to analyze data for the CSU Chancellor's Office. Imagine that you are now reprising that role, and the Chancellor's Office wants you to compute compa-ratios for the 61 professors in your data from 3 CSUs. How should you proceed? You first need to define the salary range for each of the three jobs in your sample, namely assistant professor, associate professor, and "full" professor (or simply "professor"). Let's do this for the "assistant professor" job title. To define the salary range for assistant professors of biochemistry and chemistry you begin by defining the midpoint of the range. Suppose that the average salary for assistant professors of chemistry or biochemistry in the industry is about $91,000, which you assign as the CSU's salary midpoint for that job title. You then decide that a $40,000 range spread is appropriate for that job. This implies a maximum salary of $111,000 and a minimum salary of $71,000. These two numbers are $40,000 apart and have their average (or midpoint) at $91,000.

The range spread should be set so that the minimum and maximum salaries are sensible as the lower and upper bounds for salaries in that job. Assuming that the data sample from the three universities in section 7.1 is reasonably representative of the industry as a whole, a minimum salary of $71,000 – as implied by a $40,000 range spread – is realistic. In fact, the minimum salary of the 61 workers in your data is $72,083, as shown in Table 7.2. A maximum salary of $111,000, as described above, might seem unrealistically low when you consider that the highest-paid of the 61 workers in your data has a salary of $158,584 (again, see Table 7.2). But remember that the 61 workers on which Table 7.2 is based include all three job ranks. In contrast, our focus here is only on assistant professors. It's reasonable to expect (and indeed this is the case) that the lowest salary in the sample ($72,083) belongs to an assistant professor, whereas the highest salary in the sample

($158,584) belongs to a full professor. A maximum salary of $111,000 is pretty accurate if we focus only on assistant professors.

With the salary range defined (i.e., minimum = $71,000, maximum = $111,000, midpoint = $91,000), we're now ready to compute compa-ratios. Suppose that Linus Pauling is an assistant professor of biochemistry at CSU Long Beach and that his annual salary is $87,369. His compa-ratio is $87,369 / $91,000, which is about 0.96. The benchmark level for a compa-ratio is 1, which means that your worker is paid exactly the salary midpoint of the salary range. Your workers who are paid more generously than the midpoint have compa-ratios above 1, and those who are paid less generously than the midpoint have compa-ratios below 1. Higher compa-ratios indicate higher levels of pay relative to some market benchmark, which unfortunately is not the ideal (though typically unknowable) benchmark that I stated in this subsection's first paragraph. And the fact that it's not the ideal benchmark poses a significant limitation of the compa-ratio, at least from the standpoint of turnover management and talent retention. Let me harp on that point for a couple of paragraphs . . .

Suppose that Marie Curie is an assistant professor of biochemistry at CSU Long Beach, in the same department as Linus Pauling, whose compa-ratio we just found to be 0.96. Marie's salary is $96,510, so her compa-ratio is about 1.06. Suppose that Marie is recognized within the department as a higher overall performer than Linus, both in teaching and in research, and that is why her salary exceeds his. Compensation textbooks frequently state that compa-ratios above 1 (such as Marie's) mean that the employee's pay is highly competitive with the market. Such textbook discussions would lead you to believe that Linus's pay (with his compa-ratio of only 0.96) is less market competitive than Marie's. The problem is that both compa-ratios are computed using the same, crude midpoint of $91,000, when in reality the correct point of comparison with a worker's actual salary is the *best* salary offer they could land from a competing firm, as I mentioned in the first paragraph of this subsection. The best offer that Marie could obtain might well differ from the best one that Linus could obtain, and in fact there's good reason to believe hers would be higher. After all, if she's known to be the better teacher and researcher within their own department, then competing departments are also likely to recognize (and pay for) these talents.

Suppose, hypothetically, that the best competing offer that Marie could land is $110,671, whereas for Linus it is $85,721. Let's compute new ratios for Marie and Linus, in which we replace the midpoint (of $91,000) in the denominator of the compa-ratio with the best salaries they could each obtain from competing employers. Those new ratios are about 1.02 for Linus (i.e., $87,369 / $85,721) and 0.87 (i.e., $96,510 / $110,671) for Marie. Of course, we need a fancy name for these new ratios. Let's call them ***compe-ratios*** (pronounced *COMP*-ee), which is short for "competitiveness ratios". The compe-ratio is a measure of the market competitiveness of the worker's salary. Marie's compe-ratio is above 1, and Linus's is below 1, whereas the reverse was true for their compa-ratios. The compe-ratios suggest that, in fact, Linus's salary is more market competitive than Marie's, even though her salary and compa-ratio exceed his. So if you're going to worry about having a worker poached, you should focus your attention on Marie, because she can potentially land an outside offer that would far exceed her current salary, whereas Linus can't. The bottom line is that from the standpoint of turnover management and talent retention, compe-ratios are better than compa-ratios as measures of the market competitiveness of pay.

Okay, enough harping on the compa-ratio. Let's wrap up with a point on terminology. The term "compa-ratio" is well known in the world of professional compensation, and if you "google" it you'll find lots of material. In contrast, "compe-ratio" is a term that I apparently just invented! A "google" search on it appears to reveal nothing that relates to our compensation discussion. Before you get too excited about my ostensible new invention, however, remember that the big practical limitation of the compe-ratio is that you usually won't have advance knowledge of the *best* competing offer that your worker could obtain. Usually you'll never learn that number unless your worker shows up at your office one day with a competing job offer in hand, in which case you will likely be plunged into the bargaining waters that we'll navigate in Chapter 14.

12.2 Turnover and the Timing of Compensation

For the first decade of my employment at CSUEB, the commencement ceremony occurred in mid-June in an open stadium, under full sun exposure, high on a hill in Hayward, California, during the peak of the blistering afternoon heat. The event was an excruciating ordeal involving the national anthem, interminable speeches, musical breaks, and the conferral of honorary degrees. Only at the very end were the student degrees conferred, one by one, with each student walking individually across the stage to collect a diploma. I was required to sit on the central stage, sweltering, bundled in heavy academic regalia. The bleachers were packed with parents, grandparents, and friends, all of them there only to see *their student* receive a diploma. As more and more students collected their diplomas, the crowd in the bleachers thinned. As soon as families saw *their student* cross the stage, they quickly left the stadium. By the time the surnames beginning with "Z" were read, the bleachers were virtually empty. Every year, before reading the name of the first student, the university president implored the crowd to "please be respectful, and remain seated until the last student's name is read!" No one ever listened.

A basic, powerful principle – the most important one of this chapter – is apparent in that story . . .

> *people stick around when there's something to look forward to, and they leave when there's not.*

The principle pervades many aspects of life, including compensation. So far in this chapter I've described how turnover can be managed by changing the compensation *level*. But another way to encourage (or reduce) turnover is to leave the overall level of compensation unchanged but to shift the time at which it's granted. If it's back-loaded (or deferred, meaning that workers get a bit less today in exchange for a bit more tomorrow), then workers have something to look forward to and will stick around. If it's front-loaded, meaning that workers get a bit less tomorrow in exchange for a bit more today, then they're likely to leave after getting paid today, so that they can pursue more lucrative options elsewhere. Imagine how fast the CSUEB stadium would clear out on commencement day if the sequence of events on the program *started* with the conferral of student degrees, followed by the national anthem, a string of interminable speeches, and everything else!

This discussion relates to the wage-theft material from Chapter 2. There, timing was everything, and the key issue was turnover. If the employee and the employer simultaneously exchange work and pay, there's no problem. But back-loading, or deferring, the compensation, so that the work precedes the pay, exposes workers to the risk that employers will "take the labor and run". And front-loading it, so that the pay precedes the work, exposes employers to the risk that workers will "take the money and run". Having the pay precede the work is the most extreme form of compensation front-loading; it creates the possibility that a worker may quit without ever having done the work for which he was paid. This breaks a contract, even if the contract was an implicit one (i.e., a mutual understanding between the employer and employee) rather than a formal, signed one.

What we have in mind in this chapter is similar but not exactly the same. In the wage-theft discussion, we were focused on a relatively short interval of work time, such as a pay period or even a shift. In the current chapter, we have in mind a longer interval of work time, such as the worker's whole career at the firm. In the present context, by "front-loading" we don't mean extreme front-loading where the employee gets paid before any work is done. We simply mean that the employee's compensation in the short run is somewhat higher than it "should" be from a productivity standpoint, because of a mutual understanding that it will be decreased in the future somewhat below what it "should" be from a productivity standpoint. Your worker could quit at any point, and you wouldn't have a legal claim against them. However, if your worker ends up quitting "early" in their career then you would regret having front-loaded the compensation. Signing bonuses are examples of front-loaded compensation, and sometime are granted with "clawback provisions" attached that require your employee to pay back the signing bonus if he leaves your company within a specified time period after the hiring date. Clawback provisions create conditions similar to vesting; in either case your workers effectively have to leave money on the table if they leave your company prematurely.

As a manager, when you alter the timing of compensation it's usually to back-load it rather than to front-load it, because you're usually trying to reduce turnover. You can achieve this in various ways. One way is through vesting. For example, as discussed in Chapter 11, my job at CSUEB is covered by a defined-benefit pension plan run by CalPERS, which has a five-year "cliff" vesting schedule such as the one depicted in Figure 11.1. What that means is that new employees must remain with CSUEB for five full years before becoming eligible for pension payments after their retirement. If they leave CSUEB (voluntarily or involuntarily) before the five-year mark, they are ineligible to receive any pension benefits upon retirement, and they are given a lump sum when they leave the university that reimburses them for the contributions they made to the pension fund. That vesting schedule creates a strong incentive to stick around for at least five years. Even after the five years are up, the design of the pension formula is such that with each passing year it becomes harder for your workers to quit because they'd be forgoing a larger pension. So the pension schedule as a whole, but particularly the vesting component, is a good example of deferred compensation. A consequence is reduced turnover during the early-to-mid career stages.

Stock options are also sometimes included as part of compensation packages (particularly for executives, and in startups), as discussed in Chapter 10. Options typically come with a vesting schedule, meaning that they cannot be exercised until the vesting period has passed. If employees are given new options periodically, then if they ever decide to quit, they must do so leaving money on the table, because they will have to walk away from their unvested options.

Another way to defer compensation is simply to offer a low wage upfront with the promise of generous raises later on, if this can be done in a way that workers believe. This was the strategy taken by Subway in an online case discussion.

Deferred compensation's greatest strength is also its greatest weakness. By that I mean that it's a very successful way to reduce turnover ... sometimes too successful. Even if you tried to prevent one of your workers from quitting during part of their career, the time may come when you actually *want* them to quit. Late in their careers, if you've given them deferred compensation, they will probably be making a lot of money, and their productivity (which need not remain constant over time, and might even decrease) might no longer justify the generous payments you're giving them. You'd then like them to quit, but they have no reason to quit. So while deferred compensation is helpful for reducing turnover during a career, it can make it difficult to eject workers late in their careers when their productivity stagnates or sags.

Combining deferred compensation with human resource management policies that encourage quits later in the career is one approach for dealing with this problem. And in the case of the defined-benefit pension (CalPERS) mentioned earlier, such incentives are built right in. Early in the career the pension acts like deferred compensation and discourages quits, but as the worker's years of service accumulate and the pension benefit becomes more generous, retirement starts to look more appealing to the worker. Eventually, if the worker delays retirement for too long, the effective salary from working actually becomes negative, as we discussed in Chapter 11.

As a manager, most of the time when you actively alter the timing of your employees' compensation, it's to back-load it to reduce turnover. Usually you won't purposely front-load compensation without some form of protection from your workers quitting. Sometimes, however, it can't be avoided. Training offers a good example, as discussed in Chapter 8. Training typically happens early in the employment relationship, and often the training equips workers with skills that are marketable in other companies as well. The worst possible outcome is that you pay a lot of money to train your workers, and then they get hired away by another firm that benefits from the training that you financed. Because training increases your workers' market values (and therefore their future wages), it can be considered a form of compensation, i.e., it's something that workers like about their job, because it leads to higher future pay.

You have to be very careful when providing training, because this is front-loaded compensation that exposes you to the risk that your workers will quit after reaping the benefits from the training. One way that you can insure against that possibility is to lower your workers' monetary compensation during the early part of their careers, when training happens. Then, even if they quit, you haven't lost as much money by training them, because you were saving on their monetary compensation during the time that they worked for you. You might wonder why a worker would be willing to accept lower monetary compensation than they could get at another company. The answer is that the training you provide is part of the employee's compensation; since workers like it (because it raises their future earnings potential) they should be willing to accept a pay cut in the short-run to acquire it. So training gives rise to a compensating differential (Chapter 3).

An important form of deferred compensation that I haven't mentioned is the prospect of getting promoted in the future, because promotions tend to come with large pay increases, along with other perks that make workers look forward to getting promoted. This subject is important enough to cover separately (see Chapter 13).

12.3 Workers' Perceptions of Risk

A big issue surrounding deferred compensation is whether your workers believe that they'll actually get paid in the future. Back-loading the compensation exposes your workers to the risk that you'll "take the labor and run". How successful you are in assuaging your workers' fears determines the success of your deferred compensation plan. Sometimes formal rules or contracts do the job for you and eliminate the risk. An example would be the five-year vesting schedule for the CalPERS pension that I have at CSUEB. There's little risk that something like that would not be honored, and it would be impossible to renege on just one worker . . . reneging would have to happen for large groups of workers, which would cause a political firestorm. In other cases, when the deferred compensation arrangement that you make is specific to just one of your workers (so that it would be possible to renege on just that one worker) the risks workers face are higher.

The more fearful your workers are that you'll renege on a deferred compensation arrangement, the more compensation they'll demand in exchange for bearing this risk. This was also true with wage theft (Chapter 2). The compensating differential can be interpreted as a risk premium. A similar idea appears in Chapter 10. Executives' fears of income loss (due to stock price volatility rather than to wage theft) when paid with stock options causes them to demand a hefty risk premium.

12.4 Sorting and the Timing of Compensation

So far we have focused on how the level (and timing) of compensation affects the behavior of a company's *existing* workers. In particular, we have focused on how those workers' likelihoods of quitting are affected by changes in the level or timing of compensation. But such changes also affect the behavior of *potential* workers, i.e., those who are not already employed by your company but might consider working there.

Potential workers who plan to stick around for a long time may be willing to accept a deferred compensation plan; they are willing to tolerate being underpaid early on because they are confident that they'll be around to make it up later, although you might have to assuage their fears that you'll "take the labor and run". Workers who don't anticipate sticking around for a long time will avoid deferred compensation plans, because they'd be underpaid in the early days and would leave before having the chance to recoup their early losses. So if your goal is to attract a workforce that has every intention of sticking around for a long time, deferred compensation plans are one way to achieve that. The Subway online case discussion is an example.

12.5 Severance Packages

A severance package is a payment that you give your workers on their way out the door, when they leave the company. Often such payments are larger for those who have worked at the company for a long time. Severance payments could be either *mandatory* or *voluntary*. Let's start with mandatory ones. One context in which these arise is executive compensation contracts (Chapter 10), which typically have a "golden handshake" clause that entitles the executive to a severance package if they lose their job through a firing, restructuring, or, in some cases, even a retirement. Another context is the law, in particular "employment protection regulations" that are imposed by the government. Such regulations are common in many European countries but do not exist in the United States. In the United States there is unemployment insurance provided by the government, but that works a bit differently and is not all paid at once to workers in a lump sum upon their departure from the company.

Mandatory severance is a type of deferred compensation because, by definition, it only occurs at the (very) end of the employment relationship. It's also important to realize that in the context of mandatory severance, compensation is lower throughout the life of the employment relationship than it would have been in the absence of the mandatory severance pay. To understand why, recall the logic of compensating differentials (Chapter 3). Suppose that a talented segment of the workforce can choose whether to work for "Type-A" firms or "Type-B" firms.

The two types of firms are identical in all respects (recall the importance of that phrase, which came up repeatedly in Chapter 3), including pay, except for the fact that Type-A firms offer a severance package upon termination, whereas Type-B firms do not. Everyone would obviously prefer to work for a Type-A firm in that case. So to successfully compete in the labor market with these Type-A firms, the Type-B firms would have to sweeten the deal in some other aspect of the compensation package . . . for example, by offering a higher base salary, larger annual bonuses, or better fringe benefits. Because these higher pay levels would be received during the employment relationship (as opposed to the very end of it) Type-B firms would be front-loading their compensation packages, relative to Type-A firms.

Like any form of deferred compensation, mandatory severance induces sorting effects (section 1.8). Workers who prefer to have their compensation front-loaded would prefer to work in a Type-B firm, whereas those who are willing to have their compensation back-loaded so that they can have some insurance if they lose their job would prefer to work in a Type-A firm. The resulting sorting effects vary according to what "strings" are attached to the mandatory severance. For example, a "no strings attached" mandatory severance package might stipulate that every worker is entitled to a lump-sum payment of 1% of his final salary for each year of employment, regardless of the reason for the separation. So a worker who retired after 21 years would get a payment of 21% of final salary, whereas a worker who was fired after seven years would get a payment of 7% of final salary. There is no uncertainty associated with that contract, because at every point in time, from the first day of employment until the last, the worker knows exactly how much money he's guaranteed to receive if he separates.

Contrast that with a severance package that has strings attached. Perhaps the worker is only entitled to the payment if the worker is fired, but not if the worker retires or quits. In that case, workers who are particularly nervous about getting fired (and wanting to insure themselves against that outcome) will be attracted to such a contract, whereas very confident (perhaps overconfident!) workers who think that they have virtually no chance of getting fired will find such a contract unappealing; they would prefer to work for a firm that offers no such severance package but that front-loads the compensation instead through, for example, a higher base salary. In short, when the severance package has certain strings attached, it will be most appealing to those workers who believe that there is a particularly high likelihood that they will leave the firm *and* that they will be eligible for the package.

Now let's consider *voluntary* severance packages, which happen when you're not required to give your workers severance payments when they leave the company (i.e., there is no employment contract or law that requires you to pay them anything), but you choose to do so anyway. First, note that some of the mandatory packages that we just talked about can also be interpreted as voluntary depending on the time at which we view them. For example, if the CEO of your company has an employment contract with a golden handshake provision, that should be seen as mandatory severance once the contract is signed. But your company had a choice about whether to enter that contract in the first place . . . before the contract was signed, it was the company's voluntary decision to engage in a contract that involved a golden handshake. In contrast, government regulations like the employment protection laws that prevail in many European countries are mandatory from every vantage point.

Back to the subject at hand, i.e., voluntary severance packages. Suppose that you've decided to dismiss one of your workers, and there are no laws or employment contracts that require you to pay severance. You might choose to do so anyway. One reason is to buy some goodwill with a former employee who has the ability to affect your reputation. Disgruntled former employees can create negative publicity (particularly in the age of social media, as I mentioned in Chapter 2 when discussing employee retaliation after wage theft) which can harm your organization's recruitment and talent retention efforts. Another reason is that a well-known severance-pay policy can induce positive sorting effects, because it signals to job applicants that the company treats its employees well, even on their way out the door.

My discussion of voluntary severance was brief, because much of our discussion of mandatory severance also applies to voluntary severance. One key difference between the two, from your standpoint as a manager, is that your company is always worse off under mandatory severance than under voluntary severance. Mandatory severance is a constraint, so the discussion in section 4.3 on the “3 Cs” applies. If you are subject to mandatory severance, then you are forced to offer severance even if doing so reduces your profits. And even if you would have voluntarily chosen to offer severance anyway, you will still be hurt if the rules dictate that you pay a severance package that differs from the one you would have offered if left to your own devices.

This is yet another illustration of a recurring theme in this book . . . a constraint generally can’t make you better off, because you could always simply voluntarily behave in the same way that the constraint mandates. And the fact that you weren’t already doing that must mean that it would be bad for your bottom line. If you already find it desirable to voluntarily pay severance packages that exceed the generosity of any mandatory severance rules that apply, then the constraints are irrelevant to you, just like minimum-wage laws are irrelevant for you if you are already paying a higher wage than the legal minimum, because doing so maximizes your profit.

12.6 Buyouts

Sometimes you would like a worker to leave your company, but you’d prefer that they leave voluntarily rather than as a consequence of you firing them. There are, of course, various reasons why you might wish for the worker’s departure. Perhaps the worker is not very productive relative to the cost of compensating them, which often happens late in a deferred compensation contract. Or perhaps they are very productive but the needs of the organization have shifted and their job is no longer relevant. Or perhaps there’s a recession, and demand for your organization’s product or service drops, so that it’s necessary to undertake layoffs.

Regardless of the reason why you want the worker to leave, there are various reasons why you might prefer that the departure is voluntary rather than a firing. One reason is that a firing may be impossible, as in the case of tenured professors or workers who are protected by the layoff provisions of a collective bargaining agreement (see section 5.5). In a less extreme case, a firing may be possible but quite costly, for example, if the worker is from a protected group and might have grounds for a lawsuit. Yet another reason concerns the negative publicity that accompanies large-scale layoffs. This is why Merrill Lynch pursued a voluntary buyout program in the month preceding a large round of layoffs during the recession of late 2001, as explored in a pair of case discussions, the first of which appears at the end of the chapter. Yet another reason concerns the morale of your remaining workers. If you start firing a lot of workers, then your remaining workers might become fearful and distracted, and they might look for or accept job offers elsewhere, whereas if most separations are voluntary, then your workers who want to stay will be less fearful of being sacked.

The basic concept of a buyout is simple. *You pay your workers to leave.* That's potentially an expensive proposition, so you're hoping to pay as little as possible to induce them to leave. Your workers, of course, have exactly the opposite goal . . . they want the buyout (i.e., severance package) to be as large as possible. The information that both parties possess in this compensation tug-of-war will have a big effect on the outcome, and many of the bargaining strategies for both sides focus on creative ways to collect this information, as we'll discuss in Chapter 14.

Your challenge is figuring out what the lowest possible severance payment would be that would induce your worker to "take the money and run", which in this case is exactly what you want to happen. Identifying that minimum amount is probably impossible, because you can't read your workers' minds, and often it's not in their interests to reveal that information to you truthfully. Sometimes you can collect information that tells you something about how much a worker would need to be paid to quit, and a good example is the aforementioned pair of Merrill Lynch case discussions. When thinking about the severance amount, and the worker behavior that it will induce, you should keep some basic principles in mind. To develop these, let's introduce a little notation. There are five variables of interest:

Cost = expected future compensation costs that you'll incur if your worker stays at your firm

RevStay = expected future revenue that your worker will generate if they stay at your firm

BestAlt = expected future monetary value to your worker of his best alternative option

ValueStay = expected future monetary value to your worker of remaining at your firm

Severance = severance payment that your worker receives if they accept your buyout offer

I've visually coded these to help you remember and keep track of them. The non-italicized, boldface ones are primarily of interest to you, the manager, so let's call these the "employer" variables. The non-boldface, italicized ones are primarily of interest to your worker, so let's call these the "worker" variables. The severance pay, ***Severance***, is of primary interest to both you and your worker, and it's displayed in both boldface and italics. Notice that, other than ***Severance***, which is a one-time, lump-sum payment that occurs if your worker accepts your buyout offer, the other four variables begin with the words "expected future ..." Those four variables require you to add up a series of monetary values from the current time until some future date, namely the date at which you expect your worker to eventually leave, assuming that they stick around rather than taking the buyout offer immediately. For example, suppose that you expect your worker to remain at your firm this year (which we'll call Year 2020), next year (which we'll call Year 2021), and four more years after that, so six years in total. In the current year, total compensation costs for employing the worker will be $\mathbf{Cost}_{2020}$. Next year they will be $\mathbf{Cost}_{2021}$, and so on. Then we would have ...

$$\mathbf{Cost} = \mathbf{Cost}_{2020} + \frac{\boldsymbol{Cost}_{2021}}{(1+r)} + \frac{\boldsymbol{Cost}_{2022}}{(1+r)^2} + \frac{\boldsymbol{Cost}_{2023}}{(1+r)^3} + \frac{\boldsymbol{Cost}_{2024}}{(1+r)^4} + \frac{\boldsymbol{Cost}_{2025}}{(1+r)^5}$$

Similar calculations underlie the variables **RevStay**, *BestAlt*, and *ValueStay*, and you can refer to Chapter 8 for an explanation of the logic underlying such "present discounted value" calculations. Some of these variables are easier to calculate (i.e., estimate) than others. For all of them you face the challenge of guessing how long the employment relationship is likely to last if the worker stays; in the preceding example calculation I assumed that you knew it would be six years, but in practice this can be hard to gauge. I'll return to that issue at the end of the section.

Even if you have a good sense of the expected length of the employment relationship, some of the variables are harder than others to estimate. From your standpoint as a manager, it should be fairly easy to determine **Cost**, because you know what you're currently paying your worker, and you probably have a reasonable sense of what those compensation costs will look like over the next few years if your worker stays. But a "worker" variable like *ValueStay* is much more difficult for you, as a manager, to gauge, because it's hard to know how much your worker values the full medley of job characteristics that you offer. This difficulty comes up a lot in compensation (see Chapters 3, 11, and 14, among others). Remember that *ValueStay* (which represents how much your worker values the compensation package you offer) need not equal **Cost** (which is the cost to you of providing that compensation package).

Before proceeding to actually use these five variables to develop some principles about how buyouts work, let me say a few words about *BestAlt*, because the four others are more straightforward. *BestAlt* represents the monetary value (as perceived by your worker) of whatever that worker would choose to do after taking your buyout offer and leaving your firm. That might be a job at another firm, in which case *BestAlt* is the value to the worker of the compensation package that firm offers. Or it might be the monetary value, in the worker's eyes, of having leisure time, plus unemployment benefits from the government, or Social Security in the case of a retirement. It's simply whatever monetary value your worker places on their next endeavor after leaving your firm. And that next endeavor will certainly be their *best* alternative option, because . . . why would anyone ever choose anything other than their best option??

We're now ready to use the five variables for some analysis. A buyout can only happen if both you and your worker are simultaneously on board. Whether you're both on board or not hinges on the values of the five variables. First, let's look at the problem from your standpoint as a manager. The fact that you're even considering a buyout at all must mean that having your worker stick around is making you lose money. So you have to choose the lesser of two evils, i.e., lose money if they leave, or lose (potentially even more) money if they stay. You should offer a buyout if the cost of the buyout (i.e., the severance package, ***Severance***) is smaller than the cost of retaining your worker. That is, you should offer a buyout if:

$$\boldsymbol{Severance} < \mathbf{Cost} - \mathbf{RevStay}. \qquad [1]$$

Notice that the right-hand side of [1] is the *negative* of expected future profit (i.e., profit is defined as *revenue minus costs*, but the right-hand side is defined as *costs minus revenue*). Because profit itself is negative (i.e., you are losing money by retaining your worker), the *negative* of profit is a positive number (which represents a loss to your firm, just like ***Severance*** is also a positive number that represents a loss to your firm).

Now let's look at the problem from the perspective of your worker, who will only accept the buyout offer if the severance package is sufficiently generous. By accepting the deal, your worker gets two things: the severance package plus the value of whatever new activity the worker will engage in after leaving your firm. Alternatively, by rejecting the deal and remaining at your firm the worker would continue collecting a pretty high compensation level . . . one that's high enough to generate an expected future loss for you, which is why you're considering buying the worker out in the first place. In summary, your worker will accept your buyout offer if:

$$\boldsymbol{Severance} + BestAlt > ValueStay. \qquad [2]$$

Subtracting *BestAlt* from both sides of inequality [2] allows it to be rewritten in the following form that reveals the level that the severance package must exceed to induce your worker to quit.

$$\boldsymbol{Severance} > ValueStay - BestAlt. \qquad [3]$$

Remember, for the buyout to actually happen, both sides need to be on board. You, the manager, are on board if condition [1] is met. And your worker is on board if condition [3] (or, equivalently, condition [2], but let's use [3]) is met. Both [1] and [3] must be true at the same time. Putting [1] and [3] together into a single expression allows us to write:

$$ValueStay - BestAlt < \boldsymbol{Severance} < \mathbf{Cost} - \mathbf{RevStay}. \qquad [4]$$

This tells us the range in which the severance pay must fall if a deal is to be struck between you and your worker. If ***Severance*** falls too low then your worker won't be willing to take the buyout. And if it increases too high then you won't be willing to offer the buyout because it's too expensive. That's the first main take-away from expression [4].

To reveal the second main take-away from [4], let's focus only on the "worker" and "employer" variables in [4] and not on the severance pay, so we can drop ***Severance*** and rewrite [4] more simply as:

$$\textit{ValueStay} - \textit{BestAlt} < \mathbf{Cost} - \mathbf{RevStay}. \qquad [5]$$

This says that a buyout will happen whenever the loss that you as a manager incur from retaining your worker is larger than the benefit that your worker reaps by staying. Notice that the benefit that your worker reaps from staying is not just *ValueStay*. That benefit must be reduced by your worker's valuation of his best outside option, because the more valuable your worker's outside option is, the more your worker would be sacrificing by remaining with your firm. This cost to your worker, i.e., *BestAlt*, is called the *opportunity cost* (of remaining at your firm). Notice that if this opportunity cost becomes large enough, while the three other variables in [4] remain unchanged, a buyout will definitely happen. In other words, if your worker's opportunities outside your firm become sufficiently attractive then you will always be able to strike a deal with your worker to persuade them to quit, with a severance package, ***Severance***, that lies within the range given in [4].

But where, within the range given by [4], will the actual severance pay be? This is a bargaining problem (see Chapter 14). Your worker wants the severance payment to be as close as possible to the *maximum* amount you'd ever pay, namely **Cost – RevStay**, whereas you want it to be as close as possible to the *minimum* amount that your worker would ever accept, namely *ValueStay – BestAlt*.

12.6.1 Collecting Information for Bargaining Purposes

The outcome of the bargaining problem that I just described hinges to a large extent on what information both sides are able to obtain about the preceding variables. There are four variables to consider, two "worker" variables and two "employer" variables, because the fifth one, ***Severance***, isn't relevant to this discussion. So let's consider those four variables . . .

As I mentioned earlier, for you as a manager, the "employer" variables are easiest to measure. You know how much you are currently spending on your employee's compensation, i.e., **Cost**, and you can probably also form a reasonable guess about what the future trajectory of that variable will look like. The same is true for **RevStay**. Your worker has a harder time gauging the "employer" variables but will certainly be on the lookout for any hints and clues that you drop. This relates closely to Chapter 14, where you will be warned to be careful about what information you reveal (purposely or inadvertently) to your workers, even as early as the hiring stage. Information is the weaponry with which bargaining battles are fought, so you should be wary of arming your opponent. If you inadvertently reveal information that allows your worker to see that **Cost** is even larger than they had guessed, that might embolden them to be even more aggressive in negotiating for a higher ***Severance***.

You will have a much harder time collecting information about the "worker" variables, but you should remain ever alert to whatever information your worker might reveal or that you can obtain from other sources. Any information that you can gain can help. If you learn that *BestAlt*, for example, is larger than you realized, then you can negotiate more aggressively for a lower ***Severance***.

For your worker, the "worker" variables are easiest to measure and the "employer" ones are the most difficult, so the situation is the reverse of what we've just discussed. Your worker knows how much he values his current job (*ValueStay*) and probably has a good idea of how much he'd value his next best alternative (*BestAlt*). Your worker will be less well informed about the "employer" variables, and he will certainly have less information about the "employer" variables than you have. He will be on the lookout for clues from you about those variables so that he can use them against you in the bargaining process (again, see Chapter 14).

Collecting information about these four variables need not be a passive approach of waiting to see what hints and clues the other side reveals. Creative, proactive steps can be taken to induce the other side to reveal some information. A good example is offered in the pair of Merrill Lynch case studies, the first of which ends this chapter.

Finally, let's return to an important variable about which I've said little so far, namely the *expected remaining length of the employment relationship if your worker stays at your company*. Both you and your worker will need to form an estimate of this variable before you can engage in any of the preceding analysis. The definitions of the "employer" and "worker" variables all start with "expected future . . ." but exactly how long is the future? Shortly after those definitions I gave an example of how you would compute **Cost**, which assumed a six-year future horizon. But in practice, how would you know that it's six years?

You normally wouldn't, and to complicate matters further, it's normally a variable that both parties can influence based on information that the other party might not observe. So, *expected remaining duration of the employment relationship if your worker stays* is yet another variable that both parties must try to gather clues about so as to be better armed for negotiating over ***Severance***. In some cases, the picture is clearer. For example, for research assistants at the Federal Reserve Bank of Richmond, this variable is perfectly known to both parties (see the online case discussion), and the same is true in settings where mandatory retirement laws apply.

12.7 Raiding and Offer Matching

Employers sometimes poach, or raid, each other's workers, and at various times in your managerial career you're likely to find yourself on either side of that operation. Let's start with the case in which another employer is attempting to poach one of your workers. There are two ways that this can happen. One way is that your employee shows up for work one day and announces to you that he's quitting to join another firm, and that's that. If he was one of your weaker employees, you might be relieved to see him go. But more commonly he'll be one of your better employees, since those folks have the most attractive alternative opportunities in the marketplace. Because the worker is leaving, there's nothing that you can do beyond starting the recruitment process to fill the vacancy (if it makes sense to do so) and to consider whether, in the future, any steps should be taken to reduce the likelihood of this kind of costly turnover where you lose one of your valued workers.

And here's where you should pause and be a bit careful. Just because one of your best workers got raided by another firm doesn't necessarily mean that you're doing anything wrong! So you shouldn't necessarily rush to redesign the company's compensation system to try to prevent future similar situations. Just because you had a bad event doesn't mean that your pay policy is flawed. No compensation system, no matter how well it's designed, can completely eliminate turnover of top talent. So the question you have to ask is, was the separation just a case of a talented worker being picked off by another firm despite you offering a well-designed and market-competitive pay plan? Or is it part of a broader pattern that can be blamed on an inadequacy in your pay system?

I recently spoke with an HR Director at a Silicon Valley startup, and when I asked him what the main human resources management problem was at the startup, he said "turnover". The startup hires a lot of young people, and many quit after a short time to work at other (often larger) companies in the industry, even though the monetary compensation is about the same. I asked the HR Director why he thinks his startup loses workers to these other companies, and he said that the other companies provide a lot of non-monetary benefits, like free gourmet meals on site for breakfast, lunch, and dinner, like the banquets at Airbnb that I cited in section 11.4.4. I asked him if his company wanted to provide those benefits, and he said "no, they're too expensive". The HR Director should perhaps not be so quick to call the high turnover a problem that needs to be fixed, particularly if his company is unwilling to match the generous benefits offers of the competitors. It's possible that the company's strategy of saving on compensation costs at the expense of a higher turnover rate is sensible, and if that's so, then the high turnover should just be understood as one of the costs of doing business. It's not necessarily worth "fixing" the problem via the compensation system, because the cure might be worse than the disease.

I mentioned that there's a second way in which a firm tries to poach one of your workers. Your worker might announce to you that he has an offer in hand from another company but that he has not yet accepted the offer and is considering his options. Perhaps he is lobbying for more money and using his threat to quit (which is now quite credible because he has an alternative offer in hand) to bargain for better pay (see Chapter 14). This second situation is more interesting than the first, because now you can decide whether or not to adjust compensation to retain this particular worker. This situation is the mirror image of the buyout situation that we discussed in the last section. There, you were trying to change your compensation system to *induce* the separations of your weaker workers, whereas in the raiding context you're changing your compensation system so as to *prevent* the separations of your top talent.

You might decide that you don't want to match offers at all . . . that it's better to let your worker leave, even if the worker is good. One problem with matching outside offers is that all of your other workers are watching your every move. If one of your workers is offered a job at another firm but with a 15% raise, and you match that raise in order to keep that worker, then your other workers might all try to get outside offers just to get a 15% raise, even if they have no intention of actually leaving your firm.

In deciding whether to match the outside offer or not, you must do a calculation that is similar to the one in the preceding section for buyouts. The variables **Cost**, **RevStay**, *BestAlt*, and *ValueStay*, are defined as in the previous section, but note that **Cost** is now computed assuming that you make a counteroffer that is sufficient to retain your worker (i.e., you choose to raise your worker's pay to match that worker's competing offer). So if you refer to the example computation that I did for **Cost** in the preceding section, the numbers that you would plug in for $\mathbf{Cost_{2020}}$, $\mathbf{Cost_{2021}}$, $\mathbf{Cost_{2022}}$, $\mathbf{Cost_{2023}}$, $\mathbf{Cost_{2024}}$, and $\mathbf{Cost_{2025}}$ are the higher compensation costs that would arise after you've matched the outside offer and retained your worker.

Unlike the buyout discussion, there is no severance pay, ***Severance***, in the present context, because we are talking about a productive worker who you'd like to *retain* if it's not too expensive, as opposed to an unproductive worker who you want to pay to *leave*. There is now, however, a new variable, **CostOriginal**, which is the expected future compensation costs that you'll incur if your worker were to stay at your firm *at his current salary* (i.e., if you were to choose not to match the competing offer). Obviously, $\mathbf{CostOriginal} < \mathbf{Cost}$.

The fact that your worker is productive for you means that **RevStay – CostOriginal** is positive, i.e., profit is positive. In the preceding section on buyouts, the corresponding difference was negative, which is why you were considering offering your worker a buyout to try to convince them to leave.

In the present situation, the key question concerns how much you will need to increase **CostOriginal** in order to successfully retain your worker who has the outside offer. Clearly, you would never be willing to raise **CostOriginal** to a level that exceeds **RevStay**, because then you'd be making losses on the worker, and you'd be better off letting the poaching firm win. But suppose that you could successfully retain your worker by increasing **CostOriginal** somewhat, but by a low enough amount that **RevStay – Cost** remains positive (notice that I've used **Cost** here, and not **CostOriginal**, because once you make the counteroffer sufficient to retain your worker, it's the *higher* level of compensation costs that applies henceforth). Then you should do it! So you should make a counteroffer to retain your worker if **RevStay – Cost** would be positive even after the higher level of compensation costs that arises from your counteroffer. So, you, the employer, will match an outside offer if

$$\mathbf{RevStay} - \mathbf{Cost} > 0. \qquad [6]$$

where, again, **Cost** reflects the higher compensation level that applies after your counteroffer.

Your worker will accept your counteroffer if

$$ValueStay > BestAlt. \qquad [7]$$

Both conditions [6] and [7] must simultaneously be satisfied for your worker to be successfully retained. When you make a counteroffer that's more generous than what you're currently paying your worker, then *BestAlt* and **RevStay** remain unchanged, whereas *ValueStay* increases, and expected future compensation costs increase from **CostOriginal** to **Cost**. As those variables increase, [6] becomes *less* likely to hold, and [7] becomes *more* likely to hold. So the key determinant of whether you can retain your worker is whether you can make a counteroffer to increase your compensation costs (from **CostOriginal** to **Cost**) by enough to ensure that [7] holds while still being a small enough increase that [6] also still holds. If that is impossible, then you will lose your worker to the poacher!

But if it is possible, and you retain your worker, how generous must your counteroffer be? This is another bargaining problem (see Chapter 14) between you and your worker. Notice that **Cost** = **CostOriginal** + **Raise**, where **Raise** is the amount by which you must increase your worker's original compensation when making a counteroffer sufficient for retention. **Raise** increases your expected future compensation costs from **CostOriginal** to **Cost**.

Substituting **Cost** = **CostOriginal** + **Raise** into [6] and rearranging the inequality a bit, we can rewrite [6] as follows to obtain the *maximum* possible level for **Raise**:

$$\mathbf{Raise} < \mathbf{RevStay} - \mathbf{CostOriginal}. \qquad [8]$$

It is unsurprising that the *maximum* possible raise is determined by the "employer" variables of primary interest to you, the manager. It should also not surprise you that the *minimum* possible raise that would convince your worker to stay must be determined by the "worker" variables of primary interest to your worker, and that is exactly what [7] does. Remember that if you try to reduce **Raise**, then you also reduce *ValueStay*, and if **Raise** becomes too small then [7] fails to hold. So that determines the minimum possible value that you could choose for **Raise** . . . you can't choose a **Raise** that is so small that [7] fails to hold, since then you'd lose your worker.

The upshot here is that there's a range of possible values that **Raise** could take and that would allow you to retain your worker. You, as the manager, would like a value as close as possible to the *minimum* of that range, whereas your worker would like a value as close as possible to the *maximum* of that range. Where exactly **Raise** ends up is a bargaining problem (Chapter 14) between you and your worker.

All of the same issues concerning the challenges of measuring and collecting information on the "employer" and "worker" variables apply as in the buyout discussion from the preceding section. In fact, the amount of information that you, your worker, and the poacher have (and the price of obtaining better information) is central to determining whether raids will be attempted at all, and whether they will succeed once attempted.

You should only attempt to raid a worker from another firm if you believe that the worker you're targeting, if successfully poached, will produce future profits for your firm that are high enough to outweigh the costs to your firm of the raid. The costs of the raid include the costs of information acquisition and the reputational costs of alienating the other firm which you might need to interact with in future business dealings. The most important variables to try to measure, if you're considering poaching another firm's worker, are the expected future profits that the worker would generate in *your* firm and in his *current* firm. Those two numbers may well be different. If the latter exceeds the former, then you'll never succeed in raiding the worker, because their employer will always find it profitable to outbid you when the bidding war ensues. But if the former exceeds the latter, then you'll succeed in poaching the worker unless your costs of engaging in the raid are prohibitively high.

You're most likely to succeed in poaching another firm's worker when that worker has a unique skill set that is hard to find in a general applicant pool and when there's a lot of publicly available information about the worker's ability and output. The point about public information is important, because a worker's current employer virtually always has more (and better) information about that worker's productivity than other employers have. If you succeed in poaching another firm's worker, it's only because that firm allowed itself to be outbid by you. That should make you pretty nervous, because that firm likely has better information than you have about the worker's productivity. If the worker was so great, why didn't his employer make a counteroffer high enough to keep him?

This threat is worrisome enough to have its own name . . . the *winner's curse*. You might have heard about it in the context of auctions, and in fact the raiding situation is itself an auction . . . the thing being auctioned is the service of the worker who is the target of the raid, and the bidders in the auction are the worker's employer and any potential poachers. Your fear, as a potential raider of a target worker, is that the worker's current employer (who has better information about that worker's ability than you have) will use that information to its advantage during the auction-like bidding war for the worker's services. This should lead you to make very cautious bids, meaning that you'll bid conservatively (i.e., lower) to avoid suffering from the winner's curse.

12.8 Lessons for Managers

Increasing workers' pay will always reduce turnover, and sometimes that's the best approach. But it's an expensive retention strategy that may not always be worth it. Business strategies involving high turnover but low pay may be ideal for employers who face modest turnover costs. Changes to the *design* of a pay package (as opposed to its level of generosity) may also affect retention. When pay becomes less generous, the workers with the strongest outside options are the ones who are most likely to quit; these people typically are the firm's top performers, though that isn't always true. Targeted pay increases (or cuts) are preferable to across-the-board ones, though they are also harder to implement, and the risks of violating an internal or external constraint (e.g., those arising from anti-discrimination legislation) increase.

The compe-ratio is conceptually better than the compa-ratio for thinking about turnover management and talent retention, even though the latter ratio is a standard term and concept in the world of professional compensation whereas the former ratio is not. The compe-ratio's denominator is influenced by the *information* that competing employers have about your worker's skills. In the example involving Marie Curie and Linus Pauling, presumably the reason why the best competing offer available to Marie exceeded that available to Linus is that her greater talents in teaching and research were known at least partially by competitors. If competitors were unable to observe the talent difference, then both workers would get the same best outside offer despite their own employer knowing that Marie is actually better than Linus. The degree to which workers' skills are *general* or *specific* (see Chapter 8) also affects the compe-ratio's denominator. The more general (and therefore portable) are your workers' skills, the higher the bids that competing employers will be willing to offer your worker, thereby raising the compe-ratio's denominator. Teaching and research skills are pretty general in that they're easily transported to other universities, so it makes sense that Marie's best competing salary offer in the example exceeds Linus's.

If your worker gets a competing offer and allows you the opportunity to make a counteroffer, the worker's compa-ratio immediately becomes irrelevant. The reason is that the compa-ratio is based on the midpoint of the salary range, but the midpoint becomes irrelevant in the wake of a concrete competing offer that you can actually observe. In that situation, the preferred compe-ratio can be exactly computed and used because you've just learned its denominator. The central issue then is how your current salary offer to your worker (i.e., the numerator of the compe-ratio) compares to your worker's best competing offer (i.e., the denominator of the compe-ratio). This leads you into the world of bargaining (Chapter 14); you will need to fight a bidding war in an effort to retain your worker.

A limitation of compa-ratios and compe-ratios is that they tend to focus only on salaries, whereas total compensation (recalling the definition from section 1.1) is much broader. Compensation includes important non-monetary components (e.g., see Chapter 11) that workers consider in addition to their salaries when deciding where to work. Salaries are only part of the story.

The timing of compensation plays a key role in talent retention. Deferring some compensation has the consequence of rewarding your workers for sticking around and penalizing them for leaving early. This is true both for your stars (which is a good thing) and for your deadwood (which is a bad thing). Getting rid of your deadwood can be extremely difficult under deferred compensation. Even your stars might eventually become undesirably expensive if they remain for a long time on a deferred compensation plan; today's star may become tomorrow's deadwood, and that can happen surprisingly quickly. So if you're going to employ deferred compensation, it's good to pair that policy with other policies that make it easier to eject workers who eventually end up overpaid relative to their productivities.

Remember that if you renege on a deferred compensation promise, it will damage your reputation, thereby hurting your future recruitment and retention efforts, particularly in this age in which social media offers an easy forum for disgruntled workers to retaliate and to reach a large audience at a low cost. Also, to the extent that your workers expect that you might renege on a deferred compensation plan, they will demand a higher overall level of compensation upfront. That is, the compensation level that you pay will need to incorporate a risk premium, as is always true when there is a threat of wage theft. So finding creative ways to convincingly assure workers that your promise can be believed will translate directly into a savings on your compensation costs.

Severance pay, whether mandatory or voluntary, induces sorting effects that depend on its design (i.e., the conditions under which your worker is eligible to receive it). It also reduces the likelihood that disgruntled former employees will damage your firm's reputation after they leave.

Buyouts are strategies for inducing turnover (of unproductive workers) whereas counteroffers in the wake of raids are strategies for reducing turnover, but in either case changing the compensation affects turnover and retention. Keep in mind the conditions (both yours and your worker's) under which a buyout can happen, and try your best to form assessments of all the key variables, including the expected duration of the employment relationship if your worker stays. Once you have an idea of the range of values for severance pay that might occur in a mutually advantageous buyout deal, apply the forthcoming lessons in Chapter 14 to try to negotiate for lower values of severance within that range. Similar bargaining issues arise when deciding how generous a counteroffer to make (if you make one at all) when a competitor tries to poach one of your workers.

The information asymmetry that exists when a competing employer knows less about a worker's ability than does the worker's current employer creates the risk of a *winner's curse*, where the competing firm only succeeds in raiding the worker because it overpaid for the worker. As a manager, you have to understand your information and your assumptions and where you have an advantage or a disadvantage.

The biggest managerial challenge surrounding buyouts, raiding, and offer matching is collecting the relevant information (i.e., making the best possible guesses of the values of the "employer" and "worker" variables and other variables such as the expected duration of the employment relationship if your worker remains at your company). A sophisticated and data-driven approach for deriving educated guesses about the relevant variables might rely on compensation analytics (see Chapters 6 and 7). Broadly speaking, the most important managerial lesson here is to appreciate and remain aware of the great importance of information in affecting compensation outcomes, particularly from the standpoint of turnover and talent management.

Case Discussion 20: Merrill Lynch (Part A)[1]

In November 2001, during a recession, and after having eliminated 6100 jobs in 2001 (reducing total employment to 66,000 employees), Merrill Lynch undertook a new wave of layoffs involving thousands of workers. But in the month preceding those new layoffs, Merrill Lynch offered its employees the opportunity to voluntarily resign in exchange for a severance package that included a lump sum of cash based on the employee's years of service (up to a maximum of one year's pay) plus 40% of any year-end bonus received in the preceding year.

The interesting twist on this offer was that the volunteers who stepped forward to accept the package might not actually get it! Each volunteer's manager would decide on a case-by-case basis whether the volunteer could receive the package. When the offer was announced, all employees were told that the severance package would be exactly the same for those employees who were ultimately dismissed in

[1] Note: This case is based on an article in *The New York Times*, entitled "Merrill Lynch Will Offer Voluntary Severance Deals", published on October 20, 2001, by Patrick Mcgeehan.

November as for those who volunteered to take the buyout deal in October. The deal marked the first time that Merrill Lynch had ever had a formal voluntary buyout program.

Questions

1. Explain why a Merrill Lynch employee might volunteer for the buyout package and under what conditions you, as the employee's manager, would grant the buyout request.
2. If both conditions hold (i.e., the employee volunteers, and you accept) what are the implications?
3. Merrill Lynch presumably knows in advance which volunteers it will accept and which it will decline. So why bother offering the deal to all employees?
   (a) Why doesn't Merrill Lynch just offer the package to those employees whose requests it plans to accept?
4. For that matter, what's the point in offering a voluntary buyout deal at all? Simply laying off the appropriate number of employees in October would cost the same amount in total severance packages as the voluntary buyout plan, because the people who ultimately get dismissed in November get the same package as those who volunteered in October.
5. Why bother paying severance at all? Why not just dismiss the appropriate employees and save money on severance costs?
6. Summarize the advantages and disadvantages of Merrill Lynch's voluntary buyout plan and decide whether or not you think it is a good idea. Are there any changes you'd recommend? Explain your answer.
7. Suppose that you are doing compensation analytics for Merrill Lynch. Is there any data analysis that could be done prior to offering the voluntary buyout package that might improve the plan by suggesting changes to its design? Explain how you would do that analysis.
8. After the voluntary buyout plan is finished and the workers have left Merrill Lynch, can any useful (to Merrill Lynch) data analysis be done with the data that are generated from it? If so, explain how you would proceed in doing that analysis.

Further Reading

Barron, J., Berger, M., and Black, D. (2006). "Selective Counteroffers." *Journal of Labor Economics*, 24, 385–409.

Bhattacharyya, D. K. (2015). "Compensation and Benefits Program a Mediating Variable for Talent Retention: A Study of Two Century-Old Indian Organizations." *Compensation and Benefits Review*, 45, 75–80.

Fallick, B., Fleischman, C., and Rebitzer, J. (2006). "Job-hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster." *Review of Economics and Statistics*, 88, 472–81.

Fan, X. and DeVaro, J. (2020). "Job Hopping and Adverse Selection in the Labor Market," *Journal of Law, Economics, and Organization*, 36.

Lazear, E. P. (1986). "Raids and Offer Matching." *Research in Labor Economics*, 8, 141–65.

Messersmith, J. G., Guthrie, J. P., Ji, Yong-Yeon, and Lee, J.-Y. (2011). "Executive Turnover: The Influence of Dispersion and Other Pay System Characteristics." *Journal of Applied Psychology*, 96, 457–69.

Mitchell, T. R., Holtom, B. C., and Lee, T. W. (2001). "How to Keep Your Best Employees: Developing an Effective Retention Policy." *Academy of Management Executive*,15, 96–109.

Phillips, J. J. and Edwards, L. (2009). *Managing Talent Retention: An ROI Approach*, San Francisco: Pfeiffer.

Rau, P. R. and Xu, J. (2013). "How Do Ex Ante Severance Pay Contracts Fit Into Optimal Executive Incentives Schemes?" *Journal of Accounting Research*, 51, 631–71.

Stahl, G., Björkman, I., Farndale, E., Morris, S., Paauwe, J., Stiles, P., Trevor, J., and Wright, P. (2012). "Six Principles of Effective Global Talent Management." *Sloan Management Review*, 53, 25–42.

13 Promotions and Pay

Big increases in pay and prestige accompany promotions in the intelligence division of the New York City Police Department (NYPD). Not surprisingly, detectives want these promotions and are unhappy when they don't get them. And if they fail to get promoted for reasons that are not directly related to their job performance, they are particularly upset. This is what motivated three of the division's black detectives to sue the NYPD in the fall of 2017, alleging racial discrimination in promotions. A central issue in the federal lawsuit was that the intelligence division lacked a clearly defined promotion rule that laid out the performance criteria for getting promoted. Rather, the promotion decisions were made in opaque fashion based mainly on supervisor discretion, without much explanation or justification.

In addition to creating disgruntled workers, some of whom ended up suing their employer, a downside to the promotion practices in NYPD's intelligence division is that they erode detectives' incentives to perform well. It's demoralizing and demotivating to work hard while knowing that promotion decisions will ultimately be based on factors such as race that are not directly related to job performance. The situation in the NYPD is typical of a fundamental problem that many organizations face, which is how to strike the right balance between rules and supervisor discretion in promotion decisions. These challenges in the NYPD set the stage for the present chapter and are further explored in the case discussion that closes the chapter.

13.1 Promotion Prospects

In Chapter 12 we discussed how deferring some components of your workers' pay is a useful way to manage turnover and retain your top talent. Promotion prospects are another type of future reward that can help to retain your workers. Like the detectives in the NYPD's intelligence division, most workers are interested in career advancement and the higher pay and prestige that accompany promotions. When choosing employers, workers often pay attention to what the opportunities are for advancement within the company. Opportunities for future advancement are valued by workers and, therefore, should be thought of as a form of compensation.

As a manager, depending upon how high up in the organization you are, you might have some control over what those advancement opportunities look like for the workers in lower ranks. In other words, you might be able to design, or influence, promotion systems as well as compensation plans within your company.

My first job title after college graduation was "Research Assistant" at the Federal Reserve Bank of Richmond. After one year, I was promoted to "Assistant Economist". Often promotions come with changes in tasks, authority, and job description, but for me the only things that happened were a pay increase and a change in job title. An unusual feature of that job was that it was only a two-year position; I would not have been allowed to stay beyond the second year. All research assistants were promoted after one year to "Assistant Economist" and then left the Fed at the end of the second year. Each year about three to four new research assistants were hired, so at any given time there were about six to eight people doing essentially the same work.

Why, in a two-year position involving no changes in tasks, did the Fed promote the research assistants and change their job titles, as opposed to just giving them raises? I'll answer that question, and others, in due course. Why should the answers to such questions concern you? If you're a manager at virtually any but the smallest and flattest of organizations, you'll regularly have to think about promotions. Like the detectives in the NYPD, the people you manage will be striving for promotion, and you will have to evaluate some of them for promotion. Or you might be thinking about your own promotion prospects and those of the peers with whom you compete. Promotions and pay are tightly linked, so capably managing the compensation system in your organization requires an understanding of the issues surrounding promotions.

Promotions are job changes that elevate workers to higher "levels" in the organization and that often involve changes in tasks and job descriptions. The notion of "level" is sometimes ambiguous. Often the level change involves a real expansion in authority (e.g., if a high school's vice principal is promoted to principal, or a college's associate dean is promoted to dean). Other times, the level change is purely illusory and just amounts to changing the title to something that "sounds more prestigious", like my promotion at the Fed that involved no changes in tasks or reporting structure. Promotions typically come with raises, and the biggest raises usually coincide with promotions. Promotions can be *internal* or *external* according to whether the worker is promoted from within the organization or is hired from another firm. Employers use promotions to achieve two main goals in organizations: (1) *creating incentives for workers*, and (2) *matching workers to jobs in the ideal way*.

Promotion systems have four main parts: (1) a *job hierarchy*, (2) the *compensation* associated with each job level, (3) *vacancy creation*, and (4) a *process* for matching workers to jobs. All four of those parts fall under the control of those who run the company. Depending upon how high up you are in the company's management, you might have influence over all or some of the four parts. From the perspective of your workers, a promotion system is a deferred compensation plan (see Chapter 12) with some additional risk attached to it. Promotion "prospects" do not guarantee that a promotion will actually happen. Lots of factors enter into promotion decisions, and typically not every worker can get a promotion. Let's elaborate on each of the four parts of a promotion process.

A job hierarchy is an arrangement of jobs into different levels, where the CEO is the highest-level job, and entry-level workers are the lowest. What do we mean by "levels"? One consideration in defining job levels is pay. Higher-level jobs typically pay more than lower-level jobs. My guess is that your supervisor, in your current job or the last job you held, makes more than you do. Another consideration is authority. Workers in higher-level jobs typically supervise those in lower-level jobs and have more decision-making authority within the company.

For your workers to desire promotions, they need some idea of the value or desirability of a promotion. There may be a compensation range associated with each job level (see section 13.2 for further discussion), so that even if your workers don't know exactly what their compensation would be after a promotion, they have at least some idea. The "prize" that your workers win when getting promoted is the difference between their new compensation level (after promotion) and their old one (before promotion). The reason it's the *difference* that's relevant is that the workers already are getting paid their current compensation and will continue to do so if they remain in that job, so the reward from a promotion is whatever *additional* amount of compensation they would get after a promotion, above and beyond what they currently make and would continue to make if not promoted.

The larger the prize, the more your workers want to win it, the harder they will work for the promotion, and the more upset they will be if they don't get it. It might seem like a good idea to set an extremely large prize, which will really induce your employees to work incredibly hard. But large prizes are expensive. If you set an extremely high post-promotion compensation level, then you're stuck paying that to whomever you promote.

An alternative, and much cheaper, way to create a large prize is to reduce the pre-promotion compensation rather than raising the post-promotion compensation. But that approach comes with its own problems. Remember the furlough for California state employees that I mentioned at the start of Chapter 12; if you cut your workers' pay, don't expect them to stick around for long. Another concern with lowering the pre-promotion compensation is that you can only lower it so far before you start running into constraints (e.g., minimum wages, living wages, prevailing wages, or wage floors created by union contracts). Also remember that when choosing the size of the prize, you're not at liberty to do whatever you want. To a large extent, pressure from the external market will influence the prize structure that you can offer.

Promotion systems also require *vacancy creation*. If higher-level positions never open up, no one can get promoted into them. Vacancy creation can happen either because the company expands at the higher levels, or because there is turnover at those levels. At the very highest level, i.e., the level of CEO, expansion is not an option because a company typically only has one CEO. So the only possibility of vacancy creation at that level is turnover, i.e., the CEO leaves.

13.2 Pay Structures, Job Analysis, and Job Evaluation

My main goal in this book is to speak directly to managers about important general concepts and theories in compensation, without getting too bogged down in the details that make for tedious reading in standard compensation texts. Acquiring a certain amount of vocabulary is helpful, however, for those readers who will become compensation specialists. This section develops some of that vocabulary. Even if you're a current or future general manager rather than a compensation specialist, it's not a bad idea to acquire some basic fluency in the "language of comp". But if you're feeling impatient, you can skip to section 13.3 with little loss of continuity.

13.2.1 Job-Based and Person-Based Pay Structures

Our discussion of job hierarchies in section 13.1 relates to what compensation professionals refer to as ***pay structures***. Pay structures consist of three components: (1) the job levels in the organizational hierarchy, (2) the pay differentials between job levels, and (3) the criteria or bases that determine those differences in pay across job levels.

Pay structures are influenced by job characteristics, such as tasks, responsibilities, and the degree of decision-making and supervisory authority. These characteristics influence the pay that is attached to a job, regardless of which employees actually hold that job. Structures in which pay is largely determined by job characteristics are called ***job-based structures***.

But pay structures are also influenced by employee characteristics, such as intelligence, training and education, knowledge, work ethic, and experience. Structures in which pay is largely determined by employee characteristics are called ***person-based structures***. The third component of pay structures, i.e., the "criteria", includes frameworks for recognizing differences in employee contributions and credentials. These frameworks serve as the gateway for introducing person-based elements into the pay structure.

Pay structures that are purely job-based or purely person-based are unusual. Most structures are a blend of both job-based and person-based aspects, as reflected in my discussion in section 13.1. For example, some of my remarks concerning the compensation "prize" that accompanies promotions suggest that you and the rest of the management team are able to *choose* the pay that accompanies each job level. This is reflected in my use of the word "set", which I've underlined twice in the following quote:

> *It might seem like a good idea to* <u>*set*</u> *an extremely large prize, which will really induce your employees to work incredibly hard. But large prizes are expensive. If you* <u>*set*</u> *an extremely high post-promotion compensation level, then you're stuck paying that to whomever you promote.*

My language here suggests job-based structures in which the pay at each level is "set" (by you and the rest of the management team) regardless of which of your employees actually fill those positions. I concluded that discussion, however, with the following point:

> *Also remember that when choosing the size of the prize, you're not at liberty to do whatever you want. To a large extent, pressure from the external market will influence the prize structure that you can offer.*

My language here suggests person-based structures in which the outside market (i.e., the compensation offers of competing firms) determines your workers' pay. Recalling the terminology of section 12.1.1, suppose that you have set a salary range for senior software engineers that has a range spread of $20,000, a minimum of $102,000, a maximum of $122,000, and a midpoint of $112,000. Suppose that you pay the recently promoted senior software engineers salaries within this salary range, but one of them gets an outside offer of $135,000 shortly after being promoted into the job. You then have an immediate problem on your hands. You will probably either have to pay that employee more than $122,000 (which means redefining the salary range to have a higher maximum) or risk losing them to a competitor.

Even if the competing offers all lie within your salary range for a given position so that they don't require you to reconstruct the range, they still introduce person-based elements to pay. For example, suppose that you promote two of your employees into the job title "senior software engineer" and that you pay each of them the salary-range midpoint of $112,000. One of the employees might obtain an outside offer of $120,000 shortly after promotion, which you decide to match rather than risk losing that high performer. Even though the employee's new salary of $120,000 lies within the salary range of $102,000 to $122,000, that employee is still paid more than the other senior software engineer who did not receive an outside offer and is paid $112,000. The salary dispersion that exists among your senior software engineers in this example is driven by person-based factors that come into play via labor market competition.

13.2.2 Job Analysis and Job Evaluation

Hopefully you've noticed the tension surrounding the differences in pay across levels of a job hierarchy. On one hand, as a manager you'd like to have control over these differentials and to "set" them in a manner that best advances your company's bottom line. On the other hand, labor market competition heavily constrains your ability to do so. But to the extent that you have managerial control over these pay differentials, you need some way to define jobs, how they differ, and how those differences should affect pay. That's where job analysis and job evaluation come in . . .

Compensation professionals use the term ***job analysis*** to describe the systematic process of collecting and analyzing information about particular jobs and about how the work differs between different jobs. Given that this chapter is on promotions, we're mainly interested in comparing jobs *across* different hierarchical levels. But job analysis also can be used to compare different jobs *within* the same hierarchical level. A main goal of job analysis is to produce a ***job description***, which is the list of tasks and responsibilities that a particular job encompasses. Embedded within the job description is the ***job specification***, also known as an ***employee specification***, which is the list of personal characteristics that an employee must possess to perform the job. These characteristics include such things as educational degree requirements, specific knowledge and skills, experience, and ability.

Creating formal job descriptions is helpful for establishing (the job-based component of) a pay structure. But they can also facilitate compliance with the government regulations we discussed in Chapter 4. For example, section 4.5 covers Title VII of the Civil Rights Act of 1964. That law prohibits discrimination against various groups of protected workers, though exceptions are made if the defining characteristic of the worker group is a BFOQ that is reasonably necessary to the normal operation of the job. If you're faced with a discrimination lawsuit and hope to prevail based on the BFOQ exception, you will need a formal job description that makes clear which employee characteristics are essential to perform the job duties.

Job evaluation picks up where job analysis leaves off. Whereas job analysis helps you to understand what a given job entails and how it differs from other jobs, job evaluation is the systematic process by which you assign the pay differentials between jobs. Job analysis is more of a descriptive exercise that aims at clearly defining what different jobs entail, whereas job evaluation requires some judgments about the value of that work to your company. Job evaluations are based on what are known as ***compensable factors***, which are job characteristics that companies use to establish relative pay rates within the organization. Some examples are employee effort, skill, responsibility, and working conditions, and these four are so widely used that they are known as ***universal compensable factors.***

There are different approaches to job evaluation. Some are internally focused, meaning that they focus on how valuable or important each job is to the company. An example is the ***point method***, which is a quantitative method job evaluation. The idea is to assign numerical values, i.e., "points" to the compensable factors in a job. Adding up these points within a job is then considered to be an overall measure of the "value" of the job to the organization, and relative comparisons can then be made across jobs that have differing numbers of points. Weights are also established that reflect the importance of each compensable factor to the job in question. For example, consider a lower-skilled manual job that is described by the aforementioned four universal compensable factors. In such a job that requires minimal skills, the most valuable factor is likely to be employee effort. Hypothetical weights might be 60 for employee effort, 10 for skill, and 15 each for responsibility and working conditions, so that the weights for all compensable factors in the job sum to 100. When the jobs are valued, pay rates are then assigned. These are reviewed for consistency with the external market by considering the pay rates attached to various external "comparison" jobs known as ***benchmark jobs***. The internal pay rates are then adjusted if necessary.

Other job evaluation methods are more externally focused than the point method, meaning that they place relatively greater emphasis on market-based considerations and less on internal comparisons and valuations of jobs. These approaches rely on data from compensation surveys, and they are often based on regression analysis (see Chapters 6 and 7).

13.2.3 The Connection to Promotions

The preceding methods concern how compensation professionals assign pay rates to jobs based on both internal and external considerations. The methods can be used to compare and value (in terms of pay rates) any two jobs in the organization, whether those jobs are in the same hierarchical level or in different levels. That's why I haven't said much about promotions yet in section 13.2. But promotions bring two additional important elements to the discussion.

First, promotions can be thought of as major "stress tests" for the pay structures that result from job design and job evaluation. This is because promoting one of your workers (as opposed to simply reassigning them laterally to a different job within the same hierarchical level) makes them a hot commodity on the market. This means that promotions have the potential to trigger bidding wars in which competing firms try to steal your top talent away. I'll elaborate in section 13.5. Market competition can blow a hole into all the meticulous and intricate work you've done on the company's pay structure, and you might be forced to reconsider the maximum of the pay range to prevent a star performer from leaving. When you're immersed in the many details and intricacies of job analysis and evaluation, particularly numerical approaches like the point method that seem very scientific, it can be easy to forget how strongly competing offers can rattle the entire internal compensation structure you've created.

Second, the pay structures that you create have important implications for the incentives that your employees face. I'll elaborate in section 13.3. Much of the preceding discussion of job analysis and evaluation concerns establishing the value of those jobs (and hence what levels of pay they should carry) to your organization. But an important part of that "value" concerns employee incentives. Keep that in mind when designing and managing pay structures.

13.3 Promotion-Based Incentives

As a manager, you can use the compensation system in various ways to motivate workers (e.g., see Chapter 9 on pay for performance). The prospect of a future promotion can also be a strong motivator. Workers compete among themselves, and potentially against external workers, to try to win the promotion "prize", which typically involves a big pay increase. To further that objective, workers take various actions to increase their promotion chances. Usually (but not always!) those actions benefit the organization. For example, workers may invest greater productive effort, logging longer hours with fewer breaks, taking some work home at night, etc. They might also be more inclined to invest in training and knowledge acquisition, because these investments improve their performance and, therefore, their promotion prospects.

Although I claimed that workers "compete among themselves" for promotions, that doesn't always happen. At the Fed, the research assistants were not competing with each other, because we all knew that everyone would get promoted after one year. So if the research assistant in the cubicle adjacent to mine performed extraordinarily well, that didn't hurt my own promotion chances. The Fed situation is more the exception than the rule. More often, there's a limited number of promotion slots (sometimes just one, as in the case of the CEO's job). Some workers will win, others will lose, and simply performing really well isn't enough to guarantee promotion . . . one must surpass one's competitors. The key point here is that promotion opportunities are usually *scarce*, and it's that scarcity that creates a competition.

My claim that workers at the Fed don't compete amongst themselves for promotions, and that there is no scarcity in promotion opportunities, is correct, *once they've been hired*. But let's take a step back and think about the pre-hiring stage, when lots of job applicants are trying to get hired at the Fed as research assistants. The fact that the Fed's policy is to promote people after one year to a more prestigious title (Assistant Economist) is a positive job attribute that applicants take into account when deciding to apply in the first place, or to accept a job there if offered one. So there is scarcity in the sense that there is a limited number of research assistant positions relative to the large number of applications received, and competition occurs among external *potential* workers who are all trying to get hired in the first place into a job that will guarantee promotion after one year.

The incentives that promotions create usually help the organization but can sometimes hurt it. The only things that are needed to create incentives are some sort of prize that accompanies the promotion . . . which is usually a big raise . . . and an understanding among the workers that promotion decisions will be based, at least in part, on job performance, as it is measured by the boss. (Note: It is that second part, involving "understanding", that hurts incentives in the NYPD's intelligence division, as explored in the case discussion that closes the chapter.) But whether promotions create bad incentives or not tends to hinge on whether promotion opportunities are scarce and on how accurately the performance metrics that are used for promotion decisions reflect true performance. If the situation is like it was at the Fed, there's little reason for workers to engage in bad behavior (e.g., undermining each other's performance) because no worker's performance affects another worker's promotion chances. But as soon as there's a competition, workers can improve their chances by sabotaging their competitors in an effort to lower their performances. Moreover, if flawed performance measures are used to determine promotion decisions, undesirable worker behavior can result even when promotions aren't competitive.

What's the difference between creating incentives via promotions or via the pay-for-performance methods discussed in Chapters 9 and 10? For one thing, promotions are more visible than straight compensation awards, and that visibility has implications for the behavior of other workers. Promotions also tend to be scarcer than direct compensation awards. Moreover, compensation-based bonuses are primarily designed to achieve one main purpose, namely creating incentives, whereas promotions are used for two purposes: (1) to create incentives; and (2) to match workers to jobs in an ideal way. This dual purpose has important implications that I'll discuss in section 13.4. Another difference between promotions and pay for performance is that with pay for performance you're focusing on one worker (or group of workers) at a time and thinking about how a change to the compensation system will affect that one worker or group. But with promotions you have to think about how one person's compensation relates to someone else's, in particular how the promoted employees' compensation compares to those who don't get promoted. That *difference in compensation* between winners and losers, i.e., the "prize" from promotion, is what drives incentives.

13.3.1 "Strategic Shirking" and Other Perverse Incentives

What are the bad incentives that promotions can cause? There are three big ones (sycophancy, sabotage, and strategic shirking). Sycophancy happens when your employees try to curry favor with you, to make you like them more than you like their co-workers. They are banking on the promotion decision being made not just on a cold calculation of merit but also on favoritism. Just like students who try to establish themselves as the "teacher's pet", employees do the same thing with bosses. At best, such sycophancy is an unproductive waste of time. To the extent that your workers invest time and mental energy in sycophantic activities rather than focusing on doing their jobs well, the bottom line suffers. This is a problem in the military, as illustrated in the "up-or-out" online case discussion. It is also likely a problem in the NYPD's intelligence division, where promotions are based heavily on supervisor discretion rather than on performance measures that are difficult to fudge.

A more pernicious form of bad incentives is sabotage. In their efforts to be perceived as the best performers, your employees might undermine their co-workers' performances to make those colleagues look weak by comparison. Often such sabotage behavior is difficult to detect, which makes it hard to punish and to combat. But it is obviously damaging to company productivity. As with sycophancy, with sabotage your workers spend part of the time that they should be spending on productive work engaging in activities that either don't help, or that actively hurt, the organization.

"Strategic shirking" is another example of bad worker behavior induced by promotions. Suppose that you run a division of a firm that has two types of workers (sales workers and sales managers). The sales workers are in the lower-level jobs and aspire to be promoted to sales managers. Both jobs have a range of tasks to perform, but to keep the discussion simple let's focus on only two tasks, namely "sales" and "leadership activities". Both tasks are needed in both jobs, but the sales task receives far greater emphasis in the sales-worker job, whereas leadership activities receive far greater emphasis in the sales-manager job. Maybe sales workers are expected to focus 90% of their time and effort on sales and only 10% on leadership activities, whereas for sales managers those proportions are reversed.

Now, suppose that a sales manager quits or retires, creating a vacancy that you must fill. Assuming that you want to promote internally rather than recruiting an external candidate, what should you do? Your first impulse is probably to promote the sales worker who has demonstrated the greatest promise in leadership activities, since those activities comprise 90% of the job of sales manager. But your workers are smart enough to see that coming . . . they know that's how you'll be inclined to make the promotion decision. So if they focus on doing their sales-worker job exactly as they should, focusing 90% of their time and effort on sales, they will be at a disadvantage compared to other workers who invest a higher fraction of time and energy demonstrating strong performance in the task that matters most for promotion, namely leadership activities.

In short, your workers will strategically shirk (i.e., purposely underperform) on sales while over-performing on leadership activities. A worker who appears to be weak in sales but exceptionally strong in leadership activities is a great prospect for promotion. The problem with this strategic shirking behavior is that it means that your sales workers' attention is diverted away from where it would most productively be focused in their current job, namely on sales. It's similar to the diversion of time and effort that occurs with either sycophancy or sabotage, only in the case of strategic shirking it involves focusing too much on certain tasks and not enough on more important ones.

Your workers might face powerful incentives to strategically shirk, and in fact, failing to do so might make them victims of "Putt's Law". Proposed by Archibald Putt in 1981, the "Law" – which is more of a management hypothesis than a law – states that the more competent workers in an organization often lose out to the less competent ones when it comes to promotion decisions. That in and of itself isn't such an interesting statement, because even those of us who don't work in the NYPD's intelligence division probably know of cases in which promotions happen for reasons other than merit. For example, they're sometimes based on department politics, and that's exactly the motivation behind the sycophancy discussed earlier. Putt's Law says something stronger and more disturbing. It says that the best workers are passed over for promotion not *in spite of* being the best, but precisely *because* they are the best! In other words, they're digging their own graves by performing too well. They get labeled as "too good to promote", because their strong performance in their current (lower-level) position makes them seem indispensable in that role.

Let's look at a concrete example of Putt's Law. In an academic department at a major university a few years ago, a staff member had the job of managing all matters related to graduate students. It was a big job with a lot of tasks and responsibilities, and from the perspective of the graduate students she did it exceptionally well and with an unusual level of energy and dedication. Eventually, her supervisor's position became vacant, and many of the graduate students were very surprised that she was passed over for promotion in favor of someone else who seemed less talented. Some of them expressed their surprise to a senior professor in the department, who said that one of the main reasons that the staff member didn't get the promotion was because the work she was doing in her current position was deemed too critical to the department's success. The lesson is that you can sometimes become so good at what you do in a critical job that people see you as indispensable and never want to see you leave that position, even at the expense of your career. If the worker in question had strategically shirked, she would probably have improved her promotion prospects.

But don't get the idea that shirking is a recipe for career success! The strategic shirking coin has two sides . . . one side involves shirking on the tasks that are emphasized on the current (lower-level) job but that don't matter much on the higher-level job, but the other side involves *over-performing* on the tasks that are emphasized in the higher-level job but that are less valued on the current job. Career success requires impressing the boss on at least some dimensions. If a job has only one task and workers decide to strategically shirk on it, that's more likely to get them fired than promoted. Strategic shirking is only a problem in *multiple-task jobs*, which are most jobs.

How much of a problem is strategic shirking likely to be in your organization? The external market plays a large role in answering that question. The key issue concerns how easily employers in the external market can learn about the workers' abilities. When a worker is hired, there's initially lots of uncertainty in everyone's minds about their ability. After some time on the job, their employer gets a much better reading on their ability. But competing firms also learn at least something about the worker's ability. One way is through social networks (i.e., "word of mouth") among potential employers. Or via sites like LinkedIn. So all employers collect some information about a worker's ability as the career progresses, and usually the worker's current employer has the most accurate and detailed information. Recall from Chapter 12 that employers' success in collecting this information has a significant bearing on whether raids will be attempted, whether they'll be successful, whether buyouts will occur, and whether a worker's competing job offers will be matched by his current employer.

Collecting information about workers' abilities may be the goal, but the notion of "ability" becomes complicated when there are multiple tasks. Some workers might be brilliant on one task and mediocre on others, whereas the reverse is true for other workers. Strategic shirking only happens when employers' learning is "asymmetric", i.e., the worker's current employer learns about ability at a faster rate than competing employers, at least on *some* tasks. Contrast that with "symmetric learning", where the competing employers learn exactly the same information about ability that the worker's current employer learns, on all tasks. So one way to assess how likely it is that strategic shirking will happen in your organization is to focus on the core job characteristics and ask how visible a worker's performance is on those to the external market.

How can you combat the strategic shirking problem in your organization? One way is to make promotion decisions more "balanced". To return to the earlier example involving sales workers, this would mean placing more weight on a worker's performance in the "sales" task when making the promotion decision, even knowing that that task will be relatively unimportant in the higher-level sales manager position. A second way is to design jobs (i.e., decide which tasks belong to which jobs) in such a way as to reduce strategic shirking, remembering that the "visibility" of performance on these tasks in the eyes of external firms is important. A third way, which may seem counterintuitive, is to publicly broadcast workers' performances on each task, to the outside market. This creates a "symmetric learning" environment in which strategic shirking disappears.

13.4 Matching Workers to Jobs Ideally

All of the following CEOs have something in common (other than being CEOs in 2018!) . . . Mary Barra (General Motors Company) Safra Catz (Oracle Corporation) Tim Cook (Apple) Steve Easterbrook (McDonald's) Mark Fields (Ford Motor Company) Marillyn Hewson (Lockheed Martin) Doug McMillon (Walmart) Oscar Munoz (United Airlines) Satya Nadella (Microsoft) Robert Niblock (Lowe's) Randall Stephenson (AT&T) Bernard Tyson (Kaiser Permanente) Devin Wenig (eBay).

They were all promoted to the job of CEO from within their organizations. That situation happens frequently. A vacancy opens, and it's filled internally. After the death of CEO Steve Jobs, Apple promoted Tim Cook from COO to CEO. Apple predicted that Cook could produce more value for the company as CEO than as COO. Matching workers to jobs in the best way possible is the second primary goal of promotions, in addition to creating incentives.

Sometimes those two goals conflict, and then the organization faces a difficult tradeoff. Consider an example. Virtually every dean in a university was once just a professor, before ascending to the dean's position. In the business school at CSUEB, as explained in the case discussion that closes Chapter 5, professors are engaged primarily in the following three tasks: Teaching (45%), Research (35%), Service (20%). The numbers in parentheses refer to the approximate amount of time devoted to each task. "Service" refers to lots of activities that help the university in various ways, other than through teaching and research (e.g., serving on committees to develop the curricula of new programs or to prepare reports for accreditation purposes). The dean's job involves the same three tasks, but in very different proportions: Teaching (3%), Research (3%), Service (94%). We saw the same type of idea in the preceding section, when sales workers and sales managers performed the same two tasks (i.e., sales, and leadership activities) but in different proportions.

Suppose that a new dean must be chosen to replace the current one who is retiring. Imagine that you're the provost, which is the dean's boss and the person who gets to make the promotion decision. Consider two extreme ways you might proceed. First, suppose that your only focus is getting the ideal person in the dean's office, i.e., promoting the professor who will perform best in the role of dean. You evaluate all of the professors, and the one you promote to dean is surely the one who has the most impressive performance record in "service", which is the dean's primary job task. That decision will certainly give you the best dean, but it may be harmful to incentives. The reason is that the person you just promoted might have mediocre performance records in both teaching and research, and the professors whose records are far superior in those areas will feel that they should have been promoted instead. After all, teaching and research are the primary activities on which professors' job performances are based, so the top performer in those areas feels entitled to a promotion as a reward. Once you pass that person over for promotion, all of the professors realize that striving to be the best in one's job will not lead to a promotion. That realization weakens professors' incentives to excel in teaching and research.

Now consider an alternative promotion decision. Suppose that instead of promoting the top performer in the area of service, you promote the top professor based on overall job performance. That person is a great teacher and researcher. Their record in service is mediocre, but since service was only 20% of their job description as a professor, it didn't stop them from being identified as the top professor overall. If you promote that person, you'll probably get a mediocre dean, because 94% of the person's tasks will be focused on an area in which they've had a so-so performance record. That might seem like a bad promotion decision, and perhaps it is. But unlike the first promotion decision, this one is great for incentives. The decision sends a clear message to the faculty that achieving the top performance as a professor will likely lead to a future promotion, and that gives professors strong incentives to excel in teaching and research.

Between the preceding two extreme cases lie many other intermediate possibilities, where the aim of getting someone good in the dean's office (which necessitates matching people and jobs ideally, meaning focusing heavily on service performance) is considered simultaneously with the aim of preserving strong incentives for professors to excel in their jobs (which necessitates basing promotion decisions on teaching and research performance, even though those tasks are downplayed in the dean's job). The intermediate cases won't achieve perfect incentives, nor will they achieve perfect worker–job matching, but they'll balance these two important objectives.

Incentives are an important byproduct of job matching. Somebody needs to fill the dean's shoes, and no matter who gets promoted there will be implications for incentives. The more heavily based the promotion decision is on service performance, the weaker the future incentives created for professors. This is inescapable. So promotions create incentives automatically and unavoidably, even if organizations are trying to use them only to achieve good worker–job matching. That's very different from other tools that you might use to create incentives. For example, if you introduce incentive pay, that's a conscious choice that you make, with the primary purpose of affecting incentives, whereas with promotions, the incentives occur automatically as a byproduct of your effort to match the right worker to the right job. The dual purpose of promotions makes them harder to fine-tune than things like performance-based pay . . . whenever you use a single tool for two different purposes, you're unlikely to get perfect results on both goals.

The CEOs at the start of this section have something else in common. None of them has any chance of an internal promotion, because they're already at the top of their organizations. So incentives created by (internal) promotion prospects disappear in the case of CEOs, and you need to find other ways to motivate and retain them. The obvious alternative is pay for performance, typically basing such pay on a broad measure of total organizational performance, like the firm's stock price. That's a reason why stock options are so heavily represented in CEOs' compensation, as we discussed in Chapter 10.

Let's now return to the question of why the Richmond Fed promotes its research assistants to the title of "Assistant Economist" after one year, even knowing that the job tasks won't change and that the assistant economists will be gone after just one more year. Obviously, the answer has nothing to do with ideal worker–job matching, because the tasks and responsibilities are identical under both job titles. The answer can't be based on incentives either, because the promotion is automatic . . . all research assistants get promoted after a year on the job, not just the highest-performing research assistants. So if the promotion doesn't create incentives and doesn't improve worker–job matches, what's the point of it?

You'd have to ask the Fed that! But here's my answer. Workers value the more prestigious job title of "Assistant Economist", and the promotion is permanently visible on their résumés. As I mentioned in section 1.1, anything that workers value should be understood as a component of compensation. This makes workers willing to accept lower monetary compensation than they'd receive in other jobs that were identical but that did not offer this automatic "résumé builder" after one year. That might allow the Fed to shave a bit off of compensation costs. The appeal of the "résumé builder" might also generate a stronger applicant pool via the sorting effect.

A downside to this strategy for most firms is that, as discussed throughout this section, promoting a worker broadcasts to the outside market that the worker is of high quality. That can trigger a bidding war, which necessitates either paying the worker a costly wage increase or allowing them to get poached away by a competitor. But the Fed has far less to fear than other firms about that possibility, because the position only lasts for two years anyway, and the promotion occurs just a year before the position terminates. By the time the worker would leverage the market value of the new job title and find and secure an attractive new position, the second year will probably already be nearly up anyway, and at that point the Fed would be delighted to see the worker move on to a prestigious next job with much higher pay. The more successful its "alumni" are in their careers, the easier it will be for the Fed to recruit new research assistants (and hire them at bargain prices, because part of their compensation will accrue in the form of enhanced future career prospects, which cost the Fed little or nothing to provide).

In summary, the Fed can possibly economize on compensation costs, strengthen its applicant pool, and also improve the future career prospects of its workers, essentially free of charge, by offering a prestigious-sounding promotion one year before the worker is guaranteed to leave anyway. All of this is possible because of the unusual, self-imposed two-year limit on the position.

13.5 Why Do Big Raises Accompany Promotions?

As is true for promotions into the detective positions in the intelligence division of the NYPD, most promotions come with big raises. Why? With certain exceptions like unionized settings, there's no "law" that requires you to give any raise at all when you promote someone. Moreover, awarding big raises is expensive for organizations. So suppose that, in an effort to save your organization money, you decide to be frugal and not give a raise to the worker you just promoted. What would happen?

You will have immediately created two problems. First, you will have broadcast to all of your employees that promotions come with no financial reward. Once you take the prize away, people have little incentive to care and to try, so future incentives will suffer. Second, the worker that you just promoted won't keep the promotion a secret. She'll update her résumé and LinkedIn profile, and she'll inform headhunters. She may apply for some new job opportunities, advertising her new credentials, or companies might even discover her with little or no effort on her part. These companies will infer that she's pretty good (after all, you promoted her!) and will make her a competitive compensation offer. In short, you'll be losing her. And probably sooner rather than later.

The preceding points illustrate that the two reasons why big raises typically accompany promotions are connected to the two reasons (i.e., incentives and ideal worker–job matching) that organizations use promotions in the first place. Focus for the moment on worker–job matching. A recurring theme of this book concerns the important role of the external market in shaping your organization's internal compensation policies. That role gets amplified even further in the case of promotions. Why? Because promotions are highly visible both inside and outside of your firm. Job titles appear on people's résumés, they get posted on their own (and company) websites, on LinkedIn, etc. There may even be press releases announcing promotions at high levels of large, well-known organizations. Job titles are easy for other firms in the market to see and to verify. Pay levels, in contrast, tend to be less publicly visible, at least outside the public sector. When you create a new worker–job match by promoting someone, you create a "feeding frenzy" as competing employers discover that you have a high-quality worker. A bidding war ensues, and you have to pay to stay in the game and retain your talent. The outcome of the bidding war determines the size of the raise. In short, to retain your high performer, you need to pay enough to at least match (if not slightly exceed) the total compensation offer made by the highest bidder.

That raise that the outside market effectively forces you to pay creates *automatic incentives* for your workers. Those incentives might be sufficiently appealing that you'd have chosen to offer a raise anyway, even if you weren't forced to by competing firms in the market. Suppose, hypothetically and unrealistically, that there's no outside market, so you don't have to worry about having a worker poached. Even in that setting, you'd want to combine a raise with the promotion, because of the incentive the prize would create for workers in the future who also aspire to be promoted.

But how large should you make the raise? The larger you make it the more intense the incentives will be. But those incentives can create bad behavior as well as good, as we discussed in the preceding section. Also, there are two ways to increase the size of the prize from promotion. One way is to increase the compensation that the winner gets, and another is to reduce the compensation that the loser gets. If you increase the winner's compensation, that becomes expensive, and you'll eventually reach a point where the stronger incentives that you'd gain from raising the compensation aren't worth the cost. Alternatively, if you decrease the loser's compensation, then even ignoring the fact that the compensation might dip below the market level (in which case you'll likely lose the workers who don't get promoted), you might start violating other "floors" on pay (e.g., minimum–wage laws, living or prevailing wages, and provisions of union contracts).

To summarize, the extent to which you have the discretion to strategically set the compensation prize to create ideal incentives in your organization hinges on how much "insulation" there is from outside market forces. Usually there's not much; the labor market is pretty competitive for most types of workers. The stronger the competitive forces, the more your hands are tied by bidding wars and the less control you'll have over compensation levels. But again, remember that incentives from promotions arise automatically as soon as there's a compensation prize to be won (and workers understand that promotions are based at least in part on job performance, as measured by the boss) . . . even if the size of that prize is dictated by outside market forces.

13.6 Internal versus External Hiring

Defining a promotion is more complicated for external promotions than for internal ones, because when a worker switches firms, the new firm might have a different organizational chart and collection of job titles. So when a worker switches from one firm to the next, it may be hard to discern whether the new job is of a higher or lower "level" compared to the original one. Even when the organizational charts and job titles are identical in the two firms, defining a promotion isn't always straightforward.

For example, in academia, there's a clear, three-level hierarchy of professors ("assistant professors" are at the lowest level, "professors" are at the highest level, and "associate professors" are in the middle). If a professor at a small liberal arts college moves to Harvard at the lower rank of associate professor, that job change cannot properly be understood as a demotion, because considerably more prestige and name recognition are attached to Harvard than to a small, liberal arts college. So whether employers in the market consider a job change a promotion or not, and how impressed or unimpressed they are by it, hinges on more than just the actual change in job titles. *A big fish who gets promoted in a small pond is seen as less impressive than the same fish who gets promoted in a big pond.*

Section 13.4 opened with a list of people who were internally promoted to the role of CEO. Although most CEOs enter their jobs via internal promotion, there are exceptions. Brian Cornell (Target Corporation) and Dara Khosrowshahi (Uber) were both externally recruited. There is an important difference between them and the other CEOs at the start of Section 13.4, other than the obvious one that they were the only external hires. All of the others were *promoted* into the role of CEO, whereas Cornell and Khosrowshahi entered their CEO jobs via lateral moves, i.e., by *moving horizontally* from a job level at one firm to the same job level at another firm. Cornell was the CEO at PepsiCo Americas Foods, a subsidiary of PepsiCo, from 2012 to 2014, before becoming Target's CEO. And Khosrowshahi was Expedia's CEO until moving to Uber in 2017. Their cases are typical. That is, even though external hires are relatively infrequent at high levels of the organization like CEO, when they do happen, they're more often lateral moves than promotions. Keep in mind, though, the cautions about the trickiness of comparing "levels" across different organizations.

External hiring becomes a lot more common in jobs that are lower down in an organization's hierarchy. In fact, it becomes even more common than internal promotion. And for those jobs, as for CEOs, external hires are more often lateral moves than promotions. In some firms, a strong tendency to hire internally is part of the organizational culture. Such internal hiring policies can create strong incentives, because workers know that they only need to outshine their internal peers to achieve promotion, whereas when external hiring is a regular practice the effective pool of competitors is much larger and more threatening. An occasional external hire can keep internal workers who would otherwise get complacent on their toes and working hard, but extensive external hiring significantly diminishes the promotion chances of the insiders, thereby weakening their incentives.

13.7 Turnover and Promotions

The turnover patterns within an organization have important implications for promotion policies and incentives. Turnover creates vacancies, vacancies create promotion opportunities, and promotion opportunities create incentives. In organizations in which the senior management positions rarely experience turnover, middle managers face bleak promotion prospects unless the organization is expanding in the managerial levels. There's a tendency to think of turnover as "bad" and to focus only on the costs of turnover, which can be considerable. But turnover also has benefits, and creating incentives via fresh promotion opportunities is one of those benefits. At the Federal Reserve Bank where I worked, every year a cohort of assistant economists left the organization after having worked for two years there, which created room for the previous cohort of research assistants to be promoted to the rank of assistant economist, which in turn created room for a new batch of research assistants to be hired. When you're thinking of a change to the compensation system and its implications for turnover, you should consider the implications for future promotion opportunities within the organization.

Remember that it's not simply the overall turnover rate of the organization that matters. What's crucial is how that turnover is spread across the different levels of your firm. If virtually all of the turnover is in the lower, entry-level positions, that obviously doesn't create promotion opportunities. What matters for the incentives of your workers at a given level of the organization is the turnover rate of the workers in the *next level up*. Those are the positions that, if vacated, will create promotion opportunities for the workers in question.

Another consideration is that internal hiring creates a cascade of new vacancies that ripples all the way down the job ladder. For example, at Apple, when Tim Cook was promoted from COO to CEO, that left a vacancy at the level of COO that was ultimately filled in December 2015 via another internal promotion; Jeff Williams was promoted from Vice President of Operations to COO. Filling the COO vacancy via internal promotion of Jeff Williams created another vacancy one level down, and so on. This cascade of new vacancies means that eventually external hiring needs to happen at the lower ranks to replenish the firm's stock of workers. Similarly, at the Richmond Fed, the automatic promotion of a cohort of research assistants at the end of each year creates fresh vacancies at the "lower-level" job of research assistant, which in turn leads to a fresh batch of research assistants being hired each year.

13.8 Up-Or-Out Promotion Policies

The job that I held at the Fed might be called an "up-*and*-out" position . . . all research assistants get promoted at the end of the first year and then are forced out at the end of the second year. That's an unusual policy. Much more common than up-*and*-out are up-*or*-out promotion policies. In an up-or-out system, promotions are not automatic, and workers who fail to achieve them after a certain period of time are forced to leave the firm. Up-or-out policies are used in the military, law, academia, accounting, and consulting, among other industries.

Up-or-out policies generate intense incentives, because the prize that accompanies promotions is particularly large. Recall that there are two ways to increase the size of the promotion prize: increasing the compensation the winner gets, and reducing the compensation the loser gets. Up-or-out policies do this to an extreme degree, because the losers actually have to leave the firm, which is the greatest possible compensation reduction a firm can inflict on its workers. And the prize gets even bigger still, because typically, with an up-or-out policy, the reward that the winner gets is very significant, e.g., the status of partner at a law firm. Perhaps the most extreme version of up-or-out, with the largest prize, occurs in academia. Assistant professors begin on a "tenure clock" that runs for a certain number of years (usually six). When the clock runs out, either the person is promoted to the rank of associate professor (with tenure, which means a guaranteed job for life) or they are fired. In that case, the difference between the winner's prize and the loser's prize is a guaranteed job for life (with a pay increase) or getting fired!

One reason for using an up-or-out policy is the extra strong incentives it creates for your workers to perform at the highest level and to make investments in their future productivity. Another reason for up-or-out policies is that the resulting turnover generates a steady supply of new promotion opportunities within the firm, similar to how the up-*and*-out policy at the Fed created a new batch of promotion opportunities for research assistants every year. Both motivations for up-or-out can be quite powerful.

But there are also potential downsides to up-or-out policies. When the prize becomes big, the bad incentives intensify as well as the good ones. For example, in 2016, an untenured social psychologist at a German university was accused of manipulating the data in eight of his published research articles. A panel of statistical experts from the university at which he was previously employed did a thorough investigation of his published research articles and found strong evidence of data manipulation. The panel contacted the journals in question and recommended retractions of all eight articles. Some of the publications were subsequently retracted. The example highlights how the extreme incentives caused by up-or-out (which are further amplified by the prospect of tenure) can motivate workers to behave in ways that are unethical and damaging to the employer. Another downside of up-or-out is that the implied regular turnover imposes high costs of recruitment, screening, and training of new workers.

Perhaps the biggest downside of up-or-out is that the policy often requires firing workers who are highly productive and valuable to the organization. Let's return to the example of the "sales worker" and "sales manager" job hierarchy that appeared in our discussion of strategic shirking. In that example, sales workers strategically shirked on sales (i.e., the task they were supposed to be most focused on in their sales-worker job) while over-performing on leadership activities. An excellent sales worker with a stellar sales performance failed to get promoted in that example. In an up-or-out setting, the incentive to strategically shirk is further amplified because the size of the prize is so extreme. This hurts the firm. But on top of that problem, the policy requires that the firm fire the sales worker who fails to get promoted, even if that worker has an exceptionally high sales performance that would enhance the firm's bottom line. The firm voluntarily discards a highly productive, profitable worker.

You may be thinking, why not just keep the person in the sales job? Doing so would break the up-or-out promise that everyone expects to be enforced. Although that would have the short-term benefit of retaining a highly productive sales worker, it would hurt the incentives that your other workers would face in the future . . . they wouldn't be motivated to work really hard to get promoted and save their jobs, because they'd know that "up-or-out" was really an empty threat meaning, effectively, "up-or-stay". So an effective up-or-out policy forces you to make the painful move of firing a lot of highly competent, productive workers, in exchange for maintaining very strong incentives for your current and future workers.

Something else to consider if you're thinking of introducing up-or-out is that the composition of your workforce is likely to change. Recalling the sorting effect of section 1.8, up-or-out policies will attract a certain type of worker to your firm, and that can be a double-edged sword. The positive side is that workers who want to join an up-or-out firm are the ones who are pretty confident in their abilities . . . they expect to be among the "ups" rather than among the "outs". The negative side is that you might also attract the type of worker who will do almost anything to get ahead, even if that means engaging in bad behaviors like strategic shirking, sycophancy, or sabotaging co-workers' performances.

13.9 Lessons for Managers

Promotions are some of the biggest prizes that workers can achieve in organizations, and internal advancement opportunities are a major consideration for ambitious workers. Try to put yourself in the shoes of the workers you supervise, and think about how the organization looks from their perspectives. What do the career prospects look like in your firm for those workers? How frequently do promotion opportunities arise? When a vacancy opens, does it tend to be filled via internal promotion or external hiring? Who does each worker identify as the main competition for promotion? To what extent are promotions awarded based on demonstrated, strong past performance, as opposed to predicted future performance in the higher-level job? On what performance measures are promotion decisions based? Understanding the answers to these questions, and your role in shaping those answers, is part of being a good manager.

Think about how the compensation system (and other components of the human resource management system, such as those that might affect turnover rates) connects to promotion opportunities in the firm. Remember that, unlike performance-based pay which is introduced specifically to achieve a certain incentive-based objective, promotions create incentives automatically, as a natural byproduct of the inevitable process of worker–job matching that occurs in firms. Larger prizes intensify incentives, but those incentives could be for bad behavior as well as for good. Promotion incentives (both positive and negative) are automatic, whenever there's an organizational hierarchy and worker–job matching.

Negative incentives associated with sabotage, sycophancy, and strategic shirking can be countered with solutions like wage compression (i.e., shrinking the pay gap between the highly paid promoted workers and the non-promoted workers), intensive recruitment and screening, job design, and deciding who works with whom. As with all other aspects of compensation, the external market plays a key role when it comes to promotions, and indeed the ready visibility of promotions on LinkedIn and online job boards amplifies that role. In addition to determining the raises that accompany promotions, and the strength of the incentives from promotions, the external market determines whether strategic shirking happens. Unlike incentive pay, promotion incentives are more visible, scarcer, and harder to manage because of their dual role.

Case Discussion 23: New York City Police Department[1]

The intelligence division (hereafter "the division") of the New York City Police Department (NYPD) covers terrorism and other major crimes. In the fall of 2017, three of the division's black detectives filed a federal lawsuit alleging racial discrimination in promotions. Promotions in the division come with significant increases in pay and prestige.

A central issue in the lawsuit is the lack of a transparent and objective rules-based promotion policy that minimizes the discretion that can lead to discriminatory promotion decisions. A 2016 ruling by the federal Equal Employment Opportunity Commission (EEOC) referred to the division's promotion process as "wholly subjective and secret" and claimed that it promotes white detectives more rapidly than similarly qualified black detectives. Specifically, the EEOC used seven years of promotions data to show that white detectives spend a couple of years less time than blacks in the division's lowest level before being promoted and that experience or other worker characteristics cannot explain this racial difference.

Supervisor discretion plays a considerable role in detectives' promotions. Supervisors pass promotion recommendations up the chain of command, and then promotion decisions are announced with minimal explanation and justification.

[1] Note: This case is based on an editorial in *The New York Times*, entitled "Discrimination Inside the N.Y.P.D.", published on October 2, 2017.

This lack of a formal promotions policy for detectives is in contrast to the policies for officers seeking to become sergeants, lieutenants, or captains, all of which heavily depend on exam performances. Job assignments within the division's lower ranks also contribute to subsequent racial disparities in promotions. Black detectives, for example, tend to be concentrated in the "rap unit", which involves going to hip-hop concerts to provide security. Such low-profile assignments limit opportunities for advancement.

Questions

1. If the allegations against the NYPD are true, what are the implications (both for creating strong worker incentives and for matching workers and jobs in the best possible way)? Discuss these for white detectives and black detectives separately.
2. What are the advantages and disadvantages of going to an entirely rules-based promotion system that removes discretion from the hands of supervisors?
3. As a senior member of the NYPD, your job is to work with the legal team to develop a strategy for countering the claims in the lawsuit. What are the arguments you might make?
4. Suppose that, fearing future racial discrimination lawsuits, some supervisors start inflating the performance ratings of low-ranked black detectives. However, this creates a situation in which black detectives have stronger records (on paper) than white detectives who are (in reality) equally good or perhaps even better. If the black detectives are then passed over for promotion, their case for a lawsuit becomes even stronger. What ideas do you have to manage this problem?
5. Could changes to the compensation system solve or mitigate the NYPD's core problem? Explain.

Further Reading

Baker, G. P., Jensen, M. C., and Murphy, K. J. (1988). "Compensation and Incentives: Practice vs. Theory." *Journal of Finance*, 43, 593–616.

Connelly, B. L., Tihanyi, L., Crook, T. R., and Gangloff, K. A. (2014). "Tournament Theory: Thirty Years of Contests and Competitions." *Journal of Management*, 40, 16–47.

DeVaro, J. (2006). "Strategic Promotion Tournaments and Worker Performance." *Strategic Management Journal*, 27, 721–40.

DeVaro, J. and Gürtler, O. (2016). "Strategic Shirking in Promotion Tournaments." *Journal of Law, Economics, and Organization*, 32, 620–51.

Hoobler, J. M., Wayne, S. J., and Lemmon, G. (2009). "Bosses' Perceptions of Family-Work Conflict and Women's Promotability: Glass Ceiling Effects." *The Academy of Management Journal*, 52, 939–57.

Lazear, E. and Rosen, S. (1981). “Rank-Order Tournaments as Optimum Labor Contracts.” *Journal of Political Economy*, 89, 841–64.

Malhotra, N., Morris, T., and Smets, M. (2010). “New Career Models in UK Professional Service Firms: From Up-or-Out to Up-and-Going-Nowhere?” *International Journal of Human Resource Management*, 41, 1396–413.

Waldman, M. (1984). “Job Assignments, Signaling and Efficiency.” *RAND Journal of Economics*, 15, 255–70.

Waldman, M. (2003). “Ex Ante Versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion.” *Economic Inquiry*, 41, 27–41.

Waldman, M. (2013). “Classic Promotion Tournaments versus Market-Based Tournaments.” *International Journal of Industrial Organization*, 31, 198–210.

14 Negotiation and Bargaining

Some years ago, a friend of mine, who was an untenured professor, bought a house locally, told me about it, and said that the information was to remain top secret for the next year. The reason for total secrecy was that the professor was applying for tenure and promotion within the next year. In academia, at least in a unionized, public university like CSUEB, those events offer a rare opportunity to renegotiate one's base salary. Within the CSU system, tenure and promotion come with a minimum raise of 7.5% that's set by the collective bargaining agreement, but in certain rare cases it's possible to negotiate for a larger raise if one is believed to be particularly valuable to the institution (and potentially a flight risk).

The professor wanted to avoid having the dean learn about the house purchase, because that information might have given the dean peace of mind that the professor was firmly attached to the area and unlikely to move to another university simply for a modest pay increase, because of the big cost and hassle of selling a house. In the expensive San Francisco Bay Area, buying a house sends a signal that you plan to stay awhile. But during compensation negotiations you want to convey that you're highly mobile and can easily take a job elsewhere. The secret would give the professor a stronger bargaining position when negotiating with the dean!

. . .

If you're a manager, or aspire to be one in the future, you'll have to negotiate over compensation. At some point you'll likely have to hire a new employee to fill a job vacancy. Once you find someone you like, you'll make an offer and, if you're lucky, they'll accept it immediately. Often, however, they'll first try to negotiate to increase their compensation offer. Or perhaps one of your existing employees suddenly receives a higher-paying job offer from another firm, and their decision to stay with you or quit hinges on the outcome of compensation renegotiation in your firm (Chapter 12). Or perhaps your existing employee is getting promoted and would, therefore, have a strong outside market (Chapter 13). Alternatively, if you work in a unionized firm, you might have to represent the company's management in contract renegotiation with the union (Chapter 5). In all of these situations, successfully conducting negotiation is an important part of your job as a manager. It's also a challenging task involving both art and science. Some people find it fun, others distasteful. I'll share some techniques to equip you to be a better negotiator. But I can't make you a great negotiator. That requires more than just knowledge of the principles we're about to discuss. It requires a certain personality type and a lot of practice and experience.

There are two sides to any negotiation over compensation, and you're likely to experience both throughout your career. Sometimes you'll be trying to negotiate a higher compensation offer for yourself, when joining a new organization or engaging with your current one. Other times you'll be the one making the job offer and determining the compensation. Because this is a book for managers, I'll discuss mainly the latter situation in this chapter. Nonetheless, many of the principles I'll highlight can also be applied to the other side of the bargaining table.

Throughout the discussion, I often refer to the other party to the negotiation as your "opponent", to emphasize that your objective usually conflicts with theirs, though the term has an adversarial ring to it that need not describe all bargaining relationships. The generality of that term also captures the various identities that entity might take (e.g., a prospective employee, an existing employee requesting a raise or promotion, an employee who has received an offer from a competing firm, or a group of employees, perhaps represented by a union). For concreteness, I'll sometimes develop a point with reference to only one of those entities, but you'll easily see how the arguments would apply to other types of opponents.

The negotiating principles that we are about to discuss are general and also apply to areas other than compensation. To emphasize that generality, I'll occasionally compare the situation at hand to a non-employment bargaining situation you've almost certainly experienced, namely negotiating over the price of a car you want to purchase. Many aspects of that context mirror those from the compensation context, though there are also some important differences.

14.1 Define Your Objective

The starting point for negotiation is to *define your objective*. It's impossible to negotiate effectively without a clear objective. If you're willing to negotiate at all with a prospective new hire, you've already decided that hiring them would benefit the firm. But how much? Just how valuable are they? How much should your company be willing to pay for their services? In most cases, you can't *really* know how valuable an employee is until after they've actually started working for you, and even then it sometimes takes a long time to get an accurate picture. At the hiring stage, the best you can do is to make an educated guess based on all the information at your disposal.

You must figure out your *value* for the new hire, i.e., the *maximum* level of compensation that you (i.e., your organization) would be willing to pay to hire the person. You will obviously strive to hire them for less than that amount. But you should know the *maximum* amount that you'd be willing to pay if worst came to worst. Once you decide on that amount, it should only change if you receive new information. Recall that we have seen this concept of *value* elsewhere in the book. For example, in Chapter 11, we defined the value that a worker places on a fringe benefit as the maximum amount that the worker would be willing to pay for it.

This *maximum* "level" or "amount" of compensation is more complicated than just a salary number. It includes other components, like various fringe benefits. What you're negotiating over (whether or not your opponent understands that fact when sitting down at the bargaining table) is always *total* compensation. That complicates matters, though as I explain later, you can sometimes use that complication to your advantage. The bottom line is, when figuring out the maximum compensation you'd be willing to pay, don't think narrowly in terms of wages and salary only. Think in terms of the entire, total compensation package. What would it cost the firm to provide, and how much would the worker likely value that package?

Just as your starting point must be defining your objective, your opponent on the other side of the bargaining table must define their own objective at the start. In the rare event in which both objectives are aligned, there's nothing to negotiate, and a deal gets struck right away. But usually the two objectives conflict, and you must negotiate towards a compromise. For example, your objective might be to hire your opponent at the lowest possible compensation level that doesn't exceed a certain (*maximum*) amount, and your opponent's objective is to get hired at the highest possible compensation level that doesn't fall below a certain (*minimum*) amount. If a deal ends up happening and your opponent joins your firm, it will always be at some compensation level that lies somewhere between those two "certain amounts", or possibly exactly equal to one of them. An example would be when you negotiate with one of your employees over the amount of severance pay they should get in a buyout (see Chapter 12).

Typically, your opponent will want to negotiate the highest compensation level possible. As defined in section 1.1, compensation includes all aspects of a job that a worker likes, and your opponent may prioritize some of those aspects over others. Usually some components of the compensation system are fixed and ineligible for negotiation. For example, if your company offers a defined-benefit pension plan, the details of that plan are immutable from the standpoint of your opponent. So attention focuses on the aspects of compensation over which you, as the manager, have some discretion. The *level* of compensation is usually more negotiable than its *design*. But there are exceptions, and in some cases the design of compensation is also negotiable. For example, your prospective employee might want a higher commission rate and be willing to take a lower base salary in exchange for that; General Mills allows some of its managers to accept lower base pay in exchange for more stock options. Or your employee might want to opt out of certain fringe benefits in exchange for a higher salary, as allowed in some cafeteria plans (see section 11.5). Negotiation becomes more complicated when the design of compensation enters the discussion, since you are no longer just haggling over one number.

In the typical case, you'll be negotiating over the compensation *level*, as opposed to its design. But remember that compensation is a broad concept including "everything a worker likes about a job" ... so everything your opponent cares about becomes relevant to the negotiation, even if you have no control over those items and don't mention them during the negotiations. For example, your opponent might have a particular distaste for long, cold winters. In that case, if your company is located in Ithaca, NY, you're at a disadvantage with hiring and retention, and there's nothing you can do about that. But the information is still relevant to the negotiations because it means, recalling what we learned about compensating differentials in Chapter 3, that your compensation offer will likely have to be a bit higher to successfully attract this person and prevent them from accepting other job offers that are otherwise identical but in locations with better weather.

As a manager, you will generally wish to hire your opponent at the lowest possible compensation level. After all, the less compensation your employee gets, the higher your company's profit, and your job as a manager is to improve the bottom line. But we have to be a bit careful here. There are sound reasons why it might not be the best idea to hire your opponent at the absolute *lowest* compensation level that they'd accept. Think about what that would mean. If the final compensation level that emerges from negotiation turns out to be the absolute *lowest* level that your opponent would accept, that means that if it were to drop by even a dollar more, your opponent would want to quit and pursue another opportunity (or remain unemployed). If your opponent takes the job and then quits shortly thereafter, that creates a hassle for you because you must recruit to fill the position again and then start the training over from scratch.

You might think that this mess is easily preventable ... you'll just make sure that your opponent's compensation never drops by a dollar. But you can never make sure! When figuring out their *value* for the position you're offering them, i.e., the *minimum* compensation level they'd accept for the job, your opponent took account of *all* the positive and negative features of the job, based on all available information. Suppose that one of the positive features of the job was your opponent's expectation that the team he'd be joining would be filled with fun and nice people. Then, suppose that shortly after joining the team, your opponent discovers that the people on the team are actually boring and mean! That discovery amounts to a drop in compensation, just like a wage cut. Now your opponent's most attractive outside option becomes more appealing than the current job, and he will quit (see Chapter 12). If, instead, you had paid him several thousand dollars per year more than the *minimum* compensation level he'd accept, then even after discovering the disappointment with the co-workers your opponent may still find the job to be satisfactory. Leaving a bit of a "buffer" in the compensation level can save you future costs and hassles associated with frequent turnover, i.e., the need to recruit, screen, hire, and train new workers.

It might seem like I'm contradicting myself. After all, throughout the book, I've been preaching that your aim as a manager is to improve your company's bottom line, which should mean paying your opponent the *lowest* possible compensation level he'd accept. Offering anything more, i.e., including a "buffer" in the compensation offer, amounts to leaving money on the table, at the expense of the bottom line. There is no contradiction! The purpose of offering a buffer is not to give a gift to your opponent out of generosity and kindness. Rather, it's an insurance policy for your organization against future costs and hassles associated with turnover.

> *When we refer to the "bottom line" we don't mean just the immediate profit that would occur on today's balance sheet . . . we mean the long-run profitability of your organization, and that means accounting properly for things like the expected costs of future turnover, recruiting, screening, training, etc.*

So . . . throughout the book, when I say that you want to hire at the minimum compensation level possible, to maximize the bottom line, you can interpret this in most cases as incorporating some buffer. As a practical matter, you're almost certainly going to be paying a buffer anyway, whether you want to or not. The reason is that you typically won't exactly know your opponent's *minimum* acceptable compensation level, and they'll never willingly reveal it to you either. So even if your objective is to push the compensation level as low as possible, where it eventually settles at the end of negotiation will probably be higher than the lowest level your opponent would have accepted.

In contrast to the preceding discussion, when buying a car, it would never make sense to voluntarily pay any "buffer" beyond the minimum amount that the car dealership would accept for the car. Buying a car is a one-time transaction . . . an exchange for money and a physical good, after which the two parties will likely never see each other again. It's not the start of an ongoing relationship that, if it sours or severs, imposes future costs on the buyer or seller. So when buying a car, your objective is to extract every last penny you possibly can out of the seller. But when buying a worker's services, you might want to leave some buffer on the table as an investment in the future harmony and longevity of the employment relationship.

14.2 Collect Information about Your Opponent

Information is power when it comes to negotiation. The more information you have about your opponent, the more effectively you can negotiate. As a manager, one of your goals is to collect information that may be helpful in the final negotiation stage, in the event that the job candidate ever reaches that stage. The more information that you collect, the better, because some of it might eventually prove valuable. There are two main types of information that you want to collect about your opponent: *performance information* and *interest information.*

Performance information is anything that relates to the person's expected job performance, or productivity, if they were to join your firm. Interest information concerns how attractive a job at your firm would be to them, given their likely other options; in other words, how interested are they likely to be in the position, and how much do they value the position (see Chapters 3 and 12)? In the vignette that opened this chapter, the professor who secretly purchased a house was trying to prevent the dean from collecting interest information. You should begin collecting both types of information from the first moment when you encounter the job candidate in the recruitment process. From that moment forward, a continuous process of information collection and revelation occurs throughout the courtship stage of hiring. It can lead to only one of two outcomes . . . an offer that is made and accepted, or at least one of the two parties walking away and ending the employment relationship before it even starts.

Performance information is collected via recruitment and screening activities like reviewing résumés, interviewing candidates, checking references, giving performance tests, etc. Your challenge is that job candidates have an incentive to embellish and exaggerate their capabilities, so they are unreliable providers of performance information. Interest information is generally even harder to extract from your opponent than performance information. Often your opponent exaggerates interest in the early stages of negotiation (i.e., during the job interview, before an offer is made and both sides sit down formally at the bargaining table) and then sings a different tune once formal negotiations are actively underway. For example, if you try to ask your opponent directly during the job interview how interested they are in the job and whether they'd accept an offer, they're likely to seem very interested and enthusiastic. But once an offer is safely in hand, they'll often start "playing it cool" and saying lots of things that call into question whether they'd accept an offer. This is an effort to extract more compensation from you, once they know that they're getting an offer for sure.

Collecting performance information helps you to figure out how much you value your new prospective hire, i.e., the maximum amount you're willing to pay, whereas collecting interest information gives you a clearer picture of your opponent's objective. When collecting interest information, your goal is to learn as much about the candidate's preferences as possible. Subtlety is key. If you're based in a region with long, cold winters, like Ithaca, NY, and the candidate is from Florida's Atlantic coast, don't directly ask them, "How would you feel about moving to Ithaca?" It's a bad question that's too blunt and obvious. It is sure to trigger some version of the only acceptable answer, namely, "Great!", followed by a series of complimentary, positive opinions about Ithaca that may well be disingenuous.

An oblique approach is better. Ask casually and generally about interests outside of work . . . if they report that their favorite hobby is surfing, that reveals that Ithaca may be a tough geographic sell. Job candidates who let their guards down or who are not very savvy and experienced might reveal such information during casual conversation that makes them feel comfortable and at ease. But savvy ones are fully aware of your efforts to extract information and may interpret (correctly!) your question about hobbies as an effort to probe for information about geographic preferences . . . so their response might not be entirely truthful. Even more subtlety may be required.

How do these principles apply to negotiating over the price of a car? In that context the "buyer" is the person shopping for a car. But in the employment context the "buyer" is you, the manager, who is buying the labor services of an existing or prospective employee. A difference is that as a manager you're collecting both performance information *and* interest information, whereas when you're buying a car you're generally just collecting performance information, i.e., you're test driving the car, scrutinizing the exterior, evaluating how comfortable the seats are, learning about the safety features, etc. Interest information doesn't come into play, because there's nothing non-monetary that the seller cares about . . . the car dealership just wants to sell the car at the highest possible price and doesn't care at all about the buyer's characteristics. Whether the buyer is single or married, rich or poor, healthy or unhealthy, beautiful or homely, or residing nearby or far away, is irrelevant, whereas in an employment situation the buyer's (i.e., the employer's) characteristics are of immense interest to the seller (i.e., the job seeker).

Another important difference between the sale of a car and the sale of an employee's labor services is that cars are, for the most part, homogeneous products, whereas prospective employees are heterogeneous and differ in many ways that are difficult to observe. If you have your heart set on a certain type of car (e.g., a red 2020 Mazda Miata MX-5 with automatic transmission), then if you fail to reach a deal with one car dealership, you can simply go to another dealership and begin new negotiations over that exact same make, model, and color. Both you and the seller know that you have this option, which tends to simplify the negotiation problem. In contrast, in an employment setting each employee offers a unique package of attributes (e.g., ability, work ethic, personality, collegiality, experience, social skills, professional network) that is hard to replicate exactly with a different employee. So as a manager, if you fail to hire a particular employee, you might have difficulty finding another who is extremely similar.

14.3 Reveal Information Strategically

Often it's best to steer the conversation in a direction that will likely lead to useful information revelation without requiring you to ask a direct question. Job candidates are on high alert when asked direct questions during interviews, even ostensibly "innocent" questions. Sometimes you're more likely to successfully extract honest information by avoiding direct questions but rather tossing some random stuff out there and paying attention to how the candidate reacts. In a relaxed moment of the conversation, if you enthusiastically and truthfully describe, for example, the great time you had surfing off the Florida coast during a recent vacation down there, the job candidate's eyes might light up and – once he realizes he's in presence of another surfing aficionado – he's likely to let his guard down and blurt out that surfing is his favorite form of recreation and that he does it religiously four times a week. All of a sudden you've learned that Ithaca, NY, will probably be a tough sell.

The more you can get your opponent to talk, and react, the more likely it is that useful information will be inadvertently revealed. Generally, the more information that you reveal, the more your opponent will reveal. So revealing information can be a powerful tool for eliciting information from your opponent. But you must constantly gauge how likely it is, from various contextual clues, that the information being revealed by your opponent is truthful.

Moreover, every morsel of information that you reveal, inadvertently or purposefully, you should assume might be used against you in the future during direct negotiations. For example, if you reveal that three of your top five programmers unexpectedly quit within the last three weeks and that the software product that they were developing is due to be launched in less than a month, you have weakened your future bargaining position and given more leverage to the programmer you're interviewing for a job. Don't shoot yourself in the foot by revealing how desperate you are to hire this particular job candidate at this particular moment.

A related consideration is so-called "realistic job previews". These are approaches that organizations use to provide very honest descriptions of jobs to prospective candidates. Negative aspects of the job are freely shared, as well as positive aspects, and there is no attempt to hide the company's "dirty laundry" from the candidate. The approach has the advantage of securing better matches and reducing post-hire turnover, because the new hires are not subjected to any negative surprises upon starting the new job. A downside to giving realistic job previews, however, is that if you ultimately decide you want to hire the candidate, you have potentially weakened your position on the compensation negotiation front. The negative information that you reveal as part of the realistic job preview might diminish the candidate's interest in the job, thereby increasing the minimum compensation level that they'd be willing to accept.

14.4 Threats and Bluffs

Threats, from both sides, play an important role in negotiation. Your opponent's primary threat is simply walking away from the job or job offer. It's a threat that often goes unspoken, because both sides know that it's always there. Recall from section 1.6 that *mobility* is one of the three legs of the "compensation stool". Even after you successfully hire an employee, the threat never fully disappears, because at any moment the employee could quit and take a more attractive opportunity elsewhere. Although the threat is always there, it's not always clear to you how serious it is. Knowing in the abstract that your employee might quit for another opportunity is not the same as learning that your employee already has a concrete job offer in hand from one of your direct competitors, at a compensation level 20% higher than what you're currently paying. The latter situation is more alarming and demands your immediate attention (see Chapter 12).

As a manager, your primary threat is withdrawing the job offer or partially retreating from it by making the terms of compensation less attractive. The threat means that by choosing to persist with negotiations, rather than accepting the offer on the table, your opponent faces the risk of some kind of loss . . . either losing the offer altogether or losing some attractive feature of it. Withdrawing an offer that has been made, in writing or even verbally, is difficult and potentially illegal. At the very least, it seriously damages your reputation as a manager . . . word spreads, particularly in the age of social media, and prospective future job applicants learn that they cannot trust you and your organization to negotiate in good faith.

Even setting aside issues of legality, these reputational concerns alone are enough to prevent most managers and firms from withdrawing an offer. But less extreme threats are possible. For example, in CSUEB's business school, when prospective new assistant professors were negotiating total compensation with the dean, if the dean sensed that the deal was close to being closed, he would create a threat. Specifically, he would tell the job candidate in the morning that he could offer an additional signing bonus of several thousand dollars, but the offer would only be valid if the job candidate accepted the job offer by the close of that business day. If the candidate failed to accept the job offer that day, then the job offer still remained valid at the currently agreed upon terms, but the signing bonus was forgone. Thus, a real threat means that one side risks suffering a loss by delaying acceptance of the other side's most recent offer.

Another threat that your opponent faces by negotiating too aggressively is that you might become alienated and hold a grudge against your opponent even after a deal is struck. Job candidates worry about "pushing it too far" in negotiations with their future boss, because even if they "win" in terms of acquiring extra compensation, they might lose in terms of angering their future boss. Early in my career, I was chatting with a professor who had recently been poached by one university from another. I asked how the negotiations went and how far he was able to push his compensation. He replied that he did okay and believes that he could have successfully pushed for even more, but he chose not to. He explained that it's a good idea to "leave a little money on the table" so that your new employer still feels happy to hire you, whereas if you extract the absolute *maximum* amount that your new employer would be willing to pay you, then your employer is indifferent to whether you accept the offer or not (i.e., if the compensation were to increase by even one more dollar, your prospective employer would walk away). This same principle also applies on the other side of the bargaining table . . . as I mentioned, it often makes sense to pay a "buffer" that's a bit more than the *minimum* compensation that your employee would be willing to accept if necessary.

Threats matter only if they are *credible*. For the threat to be effective, the party being threatened, explicitly or implicitly, has to truly believe that the party making the threat will (or is very likely to) follow through with it. Otherwise the threat can be ignored. Both sides want to make their threats seem more potent and likely to be enacted than they actually are. The biggest and most credible threat that your opponent may present is an actual (attractive!) employment offer from another firm, as I mentioned in Chapter 12. This can happen before you've hired the person or at any point thereafter. Workers are mobile and are always free to move to another firm (section 1.6).

Typically, your opponent will present you with enough information about the outside offer to make it credible, so that you know it's actually a real offer. For example, they may present you with the written employment offer. You must then assess, based on all of the interest information you've managed to collect about your opponent, how appealing that external offer is likely to be to your opponent. Even if it pays a higher salary than you can afford to pay, it might have other undesirable features (e.g., the geographical location may be unlikely to appeal to your opponent, based on your information about your opponent's preferences). Remember that true threats rarely fully disappear. Even if your opponent chooses to take your offer rather than the competing one, your opponent can probably successfully obtain a similar competing offer (perhaps even renewing the exact same competing offer from before) in the future, if desired. Do not make the mistake of thinking you can relax once the employee takes your offer rather than a competitor's.

The nature and magnitude of your opponent's threat differs according to whether your opponent is already employed with your organization. Generally, the biggest and most worrisome threats are posed by your existing employees rather than by prospective ones. Your existing high-performing employees have already demonstrated themselves to be high performers, and presumably you've made investments in them (e.g., training, as discussed in Chapter 8). They have been in your organization long enough to learn its quirks, special features, office politics, etc., and they know how to get things done. They're not irreplaceable (no one is!), but they're costly to replace. So if they threaten to leave to accept a competing offer, that's a serious threat that imposes a real cost on you and the organization.

Obviously, however, if the insider is a relatively low performer, the threat of an outside offer is mild, because you wouldn't shed many tears if that employee were to leave. Also, in contrast to a high-performing insider, if a prospective new employee threatens to take a competing offer, the threat is milder and less worrisome. The reason is that your company hasn't yet invested significantly in a prospective employee, and it's not even clear how well that person would perform if they were to join your company. So the bottom line is, (high-performing) insiders' threats generally have more bite than the threats of outsiders (i.e., prospective new hires).

The preceding point is amplified significantly when you're dealing with *multiple* insiders' threats simultaneously, and that's exactly what happens during negotiations with a union. A significant source of the power that unions wield in negotiations is the threat of a strike. A strike involves a large segment of a company's workforce suddenly stopping work. That can create a massive, costly disruption that the company would be willing to pay a lot to prevent. For example, in December 2005, the transit workers who drive subways and buses in New York City went on strike for 60 hours, suddenly leaving vast numbers of commuters without transportation to and from work during some of the coldest days of the year. The threat of that kind of disruption far exceeds the damage that any one person could impose by stopping work or walking away from the bargaining table. Similarly, in the summer of 2013 in California's San Francisco Bay Area, the Bay Area Rapid Transit (BART) workers went on strike, imposing significant costs on commuters and attracting the attention of Governor Jerry Brown. To further amplify the threats, the timing of strikes is often chosen to inflict maximal damage. For example, the BART strike was scheduled during peak travel based on seasonal historical data. Similarly, as mentioned in an online case discussion, the Philadelphia Airport workers timed their strike to coincide with the 2016 Democratic National Convention, which was hosted in Philadelphia.

Strikes are costly for workers too, not just for management. Sometimes very costly. For example, on August 3, 1981, nearly 13,000 members of the Professional Air Traffic Controllers Organization (PATCO) engaged in a strike over wages and the length of the workweek, imposing significant costs on society that included the cancellation of around 7000 flights. That same day, President Ronald Reagan declared the strike illegal and ordered the strikers to report for work within 48 hours or they would be fired. On August 5, he fired 11,359 of the strikers and also banned all of them (for life!) from future employment with the Federal Aviation Administration (FAA). That's a memorable case in which striking didn't work out so well for the strikers. In less extreme cases, strikers "only" suffer the costs of foregone compensation during the strike.

The bottom line is that strikes are costly for both sides, and both sides have a compelling interest in reaching an agreement before a strike happens. That's why strikes are rare events. But they do happen sometimes, despite the costs that they impose on both sides. When they happen, it's usually because of a lack of information on at least one side of the bargaining table. For example, consider the situation on August 3, 1981. If the air traffic controllers knew with absolute certainty that President Reagan would fire them for striking (which was an unprecedented and extreme decision), then they almost certainly wouldn't have voted to strike in the first place. And once they did choose to strike and were given a 48-hour ultimatum, if they truly believed that President Reagan would fire them and ban them for life, surely many more of them would have caved in and returned to work. After all, they were fighting for higher wages, and instead what they "won" was a wage of $0 and a lifetime ban from the FAA. The strike happened because the president's threat was unprecedented and unanticipated, and even after he made it, the credibility of the threat was questioned . . . people didn't think he'd actually follow through with the firings.

The example illustrates the perils of situations in which threats are being made by parties at the bargaining table who lack complete information about what's happening on the other side of the table. As with one-on-one negotiation between an employer and a worker, the management's job in negotiating with a union is to collect as much information as possible. Such information is analogous to interest information, because it pertains to how eager the union is to reach an agreement and how tough they plan to be in negotiations. In the case of one-on-one negotiations, that often boils down to guessing what your prospective employee's best alternative option is if they walk away from the job that you're offering them. It's harder when dealing with a union, because union members might not all speak with one voice, and it can be hard to predict how they will vote (i.e., whether they will vote to strike or not). Section 5.5 explained the challenges that can arise when employees with very different preferences and agendas are represented by the same union. Clearly the air traffic controllers didn't speak with one voice. Nearly 13,000 of them went on strike, but only 11,359 ended up getting fired, which means that some of them decided to defect from the union and go back to work, in response to Reagan's ultimatum.

Bluffs occur when someone makes a threat with no intention of following through on it. They happen frequently during negotiation. Most likely, the air traffic controllers thought that President Reagan's threat was a bluff, because firing so many air traffic controllers simultaneously would immediately impose immense, and presumably intolerable, burdens on air travel. So the union thought that it "called the president's bluff", and the rest is history. The problem with bluffs is that once you start bluffing, the other side doesn't take your threats as seriously. What you ideally want is for the other side to always believe that you'll follow through on every threat.

Bluffing undermines that objective and is particularly damaging in a repeated bargaining relationship, such as between union and management. If the union threatens to strike, but the union members vote decisively against striking, then you don't have to be as worried as a manager when the union threatens to strike again in the future. Bluffing tends to be more effective in new bargaining relationships that are not expected to be repeated. If you don't have any history with the opponent on the other side of the bargaining table, it's hard to determine whether their threats are credible or bluffs, so they should probably be taken at least somewhat seriously. If the two parties are unlikely to have to negotiate again in the future, then neither party has to worry about its bluffs undermining the credibility of its future threats. A good example is buying a car, which is a one-time transaction between two parties with no history.

As noted earlier in this section, one of the deans at CSUEB, when negotiating compensation with new PhDs who had just been offered faculty positions, made the following threat to try to pressure candidates into a quick decision. An "exploding" signing bonus was offered in the morning, but it was only valid if the candidate accepted the job offer by the close of business that day, at 5pm. If the candidate didn't accept by that time, the job offer would remain open and all of the terms of the compensation offer would remain in place, but the signing bonus would be lost. Was that a credible threat, or was it a bluff?

I don't know the answer, because all of the candidates who received that offer accepted it! Suppose that they hadn't accepted, but a couple of days later they told the dean that they were ready to sign on the dotted line right then and there, as long as the signing bonus was put back on the table. Would the dean really have refused? If the dean was happy to take that exact same offer just two days before, he should be strongly tempted to take it two days later. The problem is that if he takes it, then he reveals to the job candidate that he was bluffing, which will undermine any future negotiations that he might have to undertake with that person (and with anyone else who catches wind of the story once that person joins the faculty). But if he doesn't take it, then he risks losing a good hire by rejecting a deal that he had been happy to accept just a couple of days prior, or he risks having someone join the faculty who is angry and disgruntled about the somewhat brutal negotiating tactic.

So what should you do if you are the dean? There's not necessarily a right answer. Each decision has plusses and minuses. Sometimes the circumstances suggest an answer. Suppose, for example, that you're in the last year of your term as dean and will then retire. In that case, you might want to put the signing bonus back on the table to close the deal and hire the assistant professor. The reason is that the costs of doing so won't be relevant until the future, and they won't be incurred by you because you'll be gone. They probably won't be borne by your successor either, because no one expects a new dean to replicate a predecessor's policies, so the university won't suffer. If you're not at the end of your term, then your job as a manager is to weigh the future costs of renewing the signing bonus against the immediate costs of failing to make a good hire. Then make whatever decision will most benefit the university, according to the university's objective (see section 1.4.).

14.5 Counteroffers

Recalling Chapter 12, as a manager you're likely to confront a situation in which one of your employees, or someone you're trying to hire, has an explicit offer from another firm that they're considering accepting. What should you do? Start by collecting as much information as you can, which will help you to assess how real the threat is that your employee will accept the outside offer. You already should have a lot of performance information and interest information in hand, because this is someone who has been working for you.

The interest information that you will have accumulated during your employee's tenure concerns their preferences, but another source of interest information concerns the details of their outside offer. Sometimes that information is easy to collect, because the person will provide it freely. They may even show you a copy of the offer letter, including all terms of employment. Other times, the person may be secretive and vague about it, and in that case you should be suspicious that they might be bluffing. Perhaps they didn't actually receive another offer, perhaps they received one but it had a short fuse that has already expired, or perhaps they have one but its terms are less appealing than those you're offering. You generally shouldn't be responsive to threats that might not even exist. But suppose they hand you the actual offer letter, and it's a serious offer. What should you do? Do you match offers or not? (See Chapter 12.)

One simple option that may work for *some* employers, either as a company-wide policy or as a solution you apply sometimes on a case-by-case basis, is to congratulate the employee, shake their hand, and say that you hope they stay with (or join) your firm but that you wish them the best either way. In other words, you don't budge on their compensation. That option has a few advantages. The biggest one is that it sends a message to all of your employees that attempting to get a raise, or to get hired at a higher compensation level, simply by presenting an outside offer, is unlikely to be successful. So it discourages your employees from hunting for outside offers they don't plan to take.

A further advantage is that if your employee is bluffing about the threat to go elsewhere, you will call them on their bluff, and you'll win without having to incur extra compensation costs. The main disadvantage is that you might lose some highly valued employees who would have been happy to stay if you'd raised the compensation offer a bit, and in that case you both would have been better off if you'd made a counteroffer. A disadvantage of making a counteroffer is that your other employees may become resentful if they see one of their co-workers get a huge raise or if someone new is hired at an exorbitant pay level. When you pay one of your employees more, you have to consider not only the direct cost of the higher pay but also the potential costs to morale of your other employees who aren't getting the higher pay. Recall from section 1.10 that problems can arise when workers know each other's pay levels, though some firms (e.g., see the Buffer case discussion that closes Chapter 1) feel that the benefits of pay disclosure outweigh its costs.

Keep in mind that you don't necessarily have to match the offer to retain the employee. Even if your employee would leave if you didn't budge on compensation, they might stay if you increased the compensation but still fell short of a full match on the monetary components. The reason is that the employee might value the non-monetary aspects of the job enough to stay. On the other hand, depending on the employee's preferences, the non-monetary aspects could mean that the person will leave even if you match (or maybe even exceed) the monetary outside offer.

Once you decide to respond to the outside offer, you are engaged in a bidding war with (at least) one competing employer. The way to think about a bidding war for an employee is the same way that you think about an auction, such as the ones that occur online at eBay. In this situation, you are at a competitive advantage relative to the other firm(s) that you're competing against. After all, as the employee's current employer, you have a lot more performance and interest information than the competing firm has. You might feel, after reviewing all of the performance and interest information, including the details of your employee's outside offer, that actually matching the outside monetary offer is unnecessary for retention. Perhaps you feel that it would be sufficient to offer a raise of 15%, which would partially bridge, but not fully close, the gap between your employee's current pay and the outside offer. The bidding war might end there, with you retaining your employee. Or the competing firm might sweeten the deal, and the bidding goes another round. And so on. In some situations you may feel that you must actually match (or perhaps even exceed) the outside compensation offer to retain the employee. In that case the bidding goes on, just like in an auction, until one party (either you or the firm that is attempting to raid your employee) gives up.

The last bid made by the party who gives up essentially determines the new compensation level, because the winning firm only needs to outbid that compensation level by a very tiny amount to prevail. You should be sure that you understand that, because it's a common point of confusion! The compensation value that determines where the bidding war ends isn't the *winner's value*; it's the *loser's value*, or more precisely the *highest-bidding loser's value* when the auction involves more than two bidders. By only having to pay the loser's value (which is lower than the winner's value) you, as the winning manager, get to reap an effective "bonus" for your organization, by hiring the employee at a lower compensation level than you would have been willing to pay if worst came to worst. (Side note: This "effective bonus" has a similar flavor to the one I describe in the last paragraph of section 3.6.1 in the context of compensating differentials. There, the marginal worker's preferences determine the magnitude of the market wage differential, and those workers whose preferences are the most distanced from those of the marginal worker reap the largest "effective bonuses".) This type of bidding war among competing firms often happens when the market learns new information about a worker's ability, such as when the worker achieves a prestigious promotion and advertises that on LinkedIn and other online forums (section 13.5).

In what I've just described, you're a manager who is responding to an attempted raid of one of your employees. Now, instead, let's put you in a managerial role at the other firm, i.e., the firm that is attempting the raid. Again, you're engaged in the same bidding war, but now you're on other side of it. What's different? Well, the main thing is that you're at an informational disadvantage. Because it's another firm's employee that you're trying to steal, you probably have way less performance and interest information than that person's current employer. Chances are that you have a good bit of information about the employee, because raids are most likely to be attempted when there's pretty good information about the target employee (see section 12.7), but even so, your information probably won't be nearly as accurate as that held by the person's current employer. This means that you're engaged in an auction against a competitor who knows more than you about the "product" (in this case, a person) being auctioned.

What's scary about that is, if you win the auction and successfully poach the employee, you have to ask yourself why the other firm (who has better information about the employee than you do) wasn't willing to offer a compensation package as generous as yours. Surely if they thought the person was a superstar, they would have been willing to put more money on the table to retain that star. So perhaps they know something that you don't know? Something bad about the person? It's a scary thought and might make you wonder whether "winning" was such a great outcome! This situation is aptly called *the winner's curse*, as I mentioned in section 12.7. It occurs when some parties to an auction are ill-informed about the value of what they're bidding on, and fear of the "curse" tends to have a chilling effect on bids. That is, you should bid more conservatively (i.e., not too generously) to increase the chance that, if you win the employee, it's a "real" win and not a "cursed" win.

As you've probably inferred by now, there's no easy solution to the question of what to do when another firm attempts to raid your employee by making them an offer. There are advantages and disadvantages of making counteroffers when presented with outside offers. If you decide to make a counteroffer, how high should it be? Again, there's no easy solution, though the tradeoff is clear. A more generous offer will make you more likely to win the bidding war in which you find yourself engaged with a competing firm. But it will also cut more deeply into your bottom line if the employee stays with you, and it may create larger morale problems (see section 1.10) if the employee ends up getting paid far more than peers in your firm. You must simply harness all of the performance and interest information that you've collected, and then make the decisions (about whether to make a counteroffer at all, and how generous it should be if you make one) that you believe will improve your company's expected profit the most.

14.6 Mix Things Up, or Simplify Them

Sometimes you can bamboozle your opponent by complicating the discussion. Generally, the more experienced negotiator benefits from complications. For example, suppose that your opponent is a prospective new hire and that he is singularly fixated on the starting salary. Let's say that your opponent demands a minimum salary of $100,000, and that you absolutely can't pay more than $85,000 due to hard constraints of the types discussed in Chapters 4 and 5. If you just keep repeating "85" and he keeps repeating "100", the negotiations will break down quickly, and a deal will never be struck. But what you might do is introduce another dimension to the negotiations that the candidate wasn't expecting but might find appealing. It can distract the opponent and potentially unravel their fixation on "100".

For example, suppose that at an earlier stage of the interviewing process, perhaps during a casual conversation over lunch, you managed to collect some critical interest information. In particular, you learned that the candidate has a spouse whose employer is located in a city 100 miles away from your location, so their plan (if the candidate were to join your company) would be to live equidistant from the two workplaces, each incurring a burdensome 50-mile commute. Armed with that information, if the candidate rigidly sticks to "$100,000", you might say that you can offer $85,000 but that in addition you will allow the candidate to telecommute three days per week. Telecommuting is something that the candidate probably wasn't expecting, and it makes things more complicated for the candidate, who now has to weigh the low salary against the convenience of being able to work from home three days per week. The concept here is a compensating differential (Chapter 3) attached to a particular non-monetary component of compensation (Chapter 11).

This example illustrates how the multi-dimensional nature of compensation allows one party to the negotiations to complicate matters by interjecting another component (in this case, a telecommuting fringe benefit) into the discussion. The more complicated the compensation package becomes, the harder it is to compare it to other offers. If the discussion were narrowly focused on salary alone, then if you were only offering the candidate $85,000 but a competing firm was offering them $92,000, those two offers can easily be compared, and the candidate would choose the other firm immediately if the two jobs were comparable in all other respects. But if you complicate matters by throwing in the telecommuting option, there are positives and negatives attached to both jobs, and the candidate has to think hard about whether he's willing to pay $7000 of pre-tax dollars per year for the right to work from home three days per week. The thought process here is the same that the workers had to go through in Chapter 3 when deciding how much monetary compensation they would be willing to sacrifice to live in San Francisco (versus Kansas).

Other times, you might benefit from simplifying the discussion. Suppose that you want to contract out janitorial services for your office building, and you're negotiating over the price of a year's worth of such services. If the service provider you're negotiating with is trying to sell you a bundle of services (e.g., landscaping, exterior maintenance, and plumbing, in addition to janitorial work) it becomes more difficult to compare that offer with competing bids from other vendors that offer different packages. If you simplify the discussion by insisting that it remain focused only on janitorial services, you will have an easier time comparing the offer to other offers and determining whether you're getting a good deal.

Similarly, when you're buying a new car, one of the first questions that the dealership asks you is whether you have a trade-in. If you do have a trade-in, the negotiation immediately becomes more complicated, because then you're then negotiating over two prices simultaneously. This gives even more of an upper hand to the dealerships, who are highly experienced negotiators who do this full time, every day, whereas you might do it only once every decade. If what you want to buy is a red 2020 Mazda Miata MX-5 with automatic transmission, then all you need is a phone and a half hour to call all of the local dealerships and find out whether they can beat the price that you've just been offered. But you can't quickly price shop in the more complicated case in which the deal also involves selling your trade-in. Pricing of used cars is more complicated and idiosyncratic than pricing of new cars, and no dealership will give you a price for your trade-in over the phone without first seeing and evaluating your used car.

The bottom line is that as a buyer in this situation, you're better off simplifying the discussion by limiting it only to the purchase of the Miata. Once that deal is closed, you can approach the problem of selling the trade-in. But the dealership has exactly the opposite incentive and will try to complicate the negotiation by combining the two deals, either implicitly or explicitly, perhaps offering you a "great" price on the Miata while simultaneously gouging you on your trade-in. And that's not even mentioning the warranties that they'll be trying to sell you to complicate things even further!

14.7 Lessons for Managers

This entire chapter provides practical lessons for managers, so I don't have much more to add. To recapitulate a key theme, always remember that negotiation involves collecting and revealing information, and for that reason the process begins far earlier than when both sides formally sit down at the bargaining table. Your challenge as a manager is to collect information that may be helpful later in the formal negotiation process, even before you know if the candidate will ever reach that stage, and to selectively and strategically reveal information that will similarly benefit you in that process. You should approach every aspect of communication with sensitivity to its implications for compensation negotiation at a later stage of the process. Remain open to the possibility that your objective might evolve as new information arrives. For example, if you acquire new, positive performance information, you might decide to increase the maximum compensation level that you'd be willing to offer.

Be prepared to switch hats quickly. Hiring involves multiple hats . . . a screening hat in which you try to figure out if you want the person, a wooing hat in which you try to entice the person to join your company, and a combat cap after an offer is made and you're negotiating pay. Know which hat to wear at which time, be able to switch quickly among them as needed, and when wearing one hat, behave with the knowledge that you'll probably be wearing other hats in the near future. If you find yourself in an auction-like bidding war with another firm over a particular worker, assemble all of the performance and interest information you can, and bid strategically. If you're the manager who is attempting the raid, keep in mind that you're probably at a disadvantage in terms of information, and be wary of the winner's curse when making your compensation offers. Fear of the winner's curse should induce caution, meaning that you bid more conservatively. Finally, behave ethically. Negotiate hard but always in good faith.

Case Discussion 24: Boston Symphony Orchestra[1]

As a manager at the Boston Symphony Orchestra (BSO), one of your responsibilities is to negotiate compensation with the musicians. The orchestra has a deep talent pool, its musicians are in high demand, and containing compensation costs is a tough job. Collective bargaining agreements guarantee a minimum base salary for all musicians, but some (particularly principal players, who tend to have the strongest external markets) can negotiate even higher pay levels. The principal flutist, a 44-year-old female, is trying to do just that. She is requesting a $70,000 raise, plus a $200,000 one-time payment to compensate her for past pay inequities.

[1] Note: This case is based on an article in *The New York Times* (July 6, 2018) by Jennifer Schuessler entitled "Star Flutist Sues Boston Symphony Over Pay Equity".

The flutist believes that she is underpaid relative to the other principal players, including those on oboe, French horn, trumpet, viola, and timpani, all of whom are males. She joined the orchestra in 2004, having previously held orchestral posts in Baltimore, Washington, and Indiana. As of 2018, she has been a featured soloist with the BSO 27 times, more than any of the BSO's other principal musicians. All of the other principal musicians in the BSO are men, except for the principal harpist. Both female principals are particularly visible members of BSO, e.g., they were the only members featured as soloists on a 2017 tour to Japan.

The flutist was hired at 154% of the minimum base pay rate set by the collective bargaining agreement. The flutist is particularly bothered that the pay of the principal oboe (a 62-year-old man who was paid $286,621 by BSO in 2015) exceeds hers by $70,000, i.e., she makes about 75% of what he makes. She believes that her current experience matches what the oboist had when he joined the BSO in 2001 at around her age. At that time the oboist was hired at 200% of the minimum base pay set by the collective bargaining agreement to match the pay he had received in his previous post at the Metropolitan Opera which he had held since 1986. The flutist claims that she is underpaid relative to the principal oboist even after adjusting for seniority.

Questions

1. What is your strategy for negotiating with the flutist? Carefully explain your arguments and how you would respond to her described claims and to other claims she'll possibly make.
2. Explain what the costs are of taking a hard line in negotiating with the flutist.
   (a) What are the costs of making significant concessions to her (beyond the obvious one of higher compensation costs)?
   (b) What steps can you take, if any, to mitigate both types of cost?
3. In comparing her pay to that of the (male) principal oboist, the flutist cites the Massachusetts Equal Pay Act, a law that went into effect earlier in 2018 requiring equal pay for "comparable work". You interpret her comment as an implicit threat to invoke that law in formal litigation against the BSO. Explain if and how this threat affects the strategies you outlined above.
4. The 2018 Massachusetts law says that salary history (e.g., the fact that the male principal oboist received high pay when he joined BSO in 2001, to match what his previous employer paid him) cannot be used to justify gender pay differences. Similar laws have been enacted recently in California, Delaware, NYC, and Philadelphia, prohibiting employers from asking job candidates about their current or previous salaries.
5. How do you anticipate that such laws will affect gender differences in pay? Explain.

Further Reading

Bazerman, M. H., Curhan, J. R., Moore, D. A., and Valley, K. L. (2000). "Negotiation." *Annual Review of Psychology*, 51, 279–314.

Bottom, W., Holloway, J., Miller, G. J., Mislin, A., and Whitford, A. (2006). "Building a Pathway to Cooperation: Negotiation and Social Exchange Between Principal and Agent." *Administrative Science Quarterly*, 51, 29–58.

Bowles, H. R. and Babcock, L. (2009). "Are Outside Offers an Answer to the Compensation Negotiation Dilemma for Women?" *Academy of Management Proceedings*, 1, 1–6.

Bowles, H. R., Babcock, L., and Lai, L. (2007). "Social Incentives for Gender Differences in the Propensity to Initiate Negotiations: Sometimes it Does Hurt to Ask." *Organizational Behavior and Human Decision Processes*, 103, 84–103.

Campagna, R. L., Mislin, A. A., Kong, D. T., and Bottom, W. P. (2016). "Strategic Consequences of Emotional Misrepresentation in Negotiation: The Blowback Effect." *Journal of Applied Psychology*, 101, 605–24.

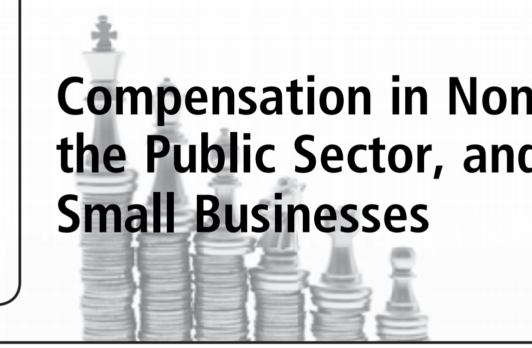
Cyert, R. M., Kang, S-H., and Kumar, P. (2002). "Corporate Governance, Takeovers, and Top-Management Compensation: Theory and Evidence." *Management Science*, 48.

de la Bruslerie, H. and Deffains-Crapsky, C. (2008). "Information Asymmetry, Contract Design and Process of Negotiation: The Stock Options Awarding Case." *Journal of Corporate Finance*, 14, 73–91.

Fisher, R., Ury, W. L., and Patton, B. (2011). *Getting to Yes: Negotiating Agreement Without Giving In*, 3rd ed., The Penguin Group.

Susskind, L., Mnookin, R., Rozdeiczer, L., and Fuller, B. (2005). "What We Have Learned About Teaching Multiparty Negotiation." *Negotiation Journal*, 21, 395–408.

Van Kleef, G. A. and de Dreu, C. K. W. (2010). "Longer-Term Consequences of Anger Expression in Negotiation: Retaliation or Spillover?" *Journal of Experimental Social Psychology*, 46, 753–60.



15 Compensation in Nonprofits, the Public Sector, and Small Businesses

In the future, or perhaps even now, you may be a manager in a nonprofit organization, the public sector, or a small business. How should you manage the compensation systems in those three settings? The answer, fortunately, is that most of the principles that we've developed in the preceding chapters are just as applicable in those settings as in conventional, medium-to-large, for-profit firms. Indeed, throughout the book I've drawn on specific examples from all three settings to illustrate key concepts. Nonetheless, there are some distinct features of these three environments that are important enough to highlight in a separate chapter.

I cover the three environments in the same chapter because, although they are distinct, they share some common features, so treating them together is efficient and facilitates interesting comparisons and contrasts among them. For example, government is a common element that connects all three. This is obvious in the case of the public sector. But if the public sector is government itself, then nonprofit organizations (henceforth "nonprofits") and small businesses are government *creations* in the sense that the very definitions of these types of organizations are determined by laws. The three environments are not mutually exclusive. For example, nonprofits can be small businesses. Even public-sector organizations can be small businesses, at least in the sense of being small in size. And organizations in the public and nonprofit sectors bear enough resemblances that they are often lumped together and compared to conventional, for-profit firms.

In much of this chapter we'll revisit familiar ground, but we will re-examine it through the lenses of the three organizational forms at hand. This repeated reference to lessons and themes from the preceding chapters offers a nice way to review and wrap up the book.

15.1 What Are Nonprofits, Public-Sector Organizations, and Small Businesses?

Public-sector organizations (henceforth "publics") are those that provide government services. Jobs at all levels of government (i.e., federal, state, county, and local) are included. But the sector is broader than government and includes more than just the people who make, enforce, and interpret laws. For example, CSUEB, the public-sector organization where I work, is a state university. A large fraction of CSUEB's budget comes from the California State Legislature, i.e., the funds come from tax revenue, as is the norm in publics. Moreover, as discussed in Chapter 11, CSUEB employees participate in a defined-benefit public pension plan called the California Public Employees' Retirement System (CalPERS), which covers California public employees.

Nonprofit organizations include a wide array of entities that are neither for-profits nor publics. Although nonprofits aren't government entities, again, they are government creations in the sense that there are laws that define the criteria for an organization to qualify as a nonprofit. There are different types of nonprofits, and the definitions vary across countries. In this chapter, I'll focus mostly on what are known in the United States as 501(c)(3)s, which comprise about 75% of US nonprofits. The definition comes from section 501 of the US Internal Revenue Service (IRS) tax code. According to the IRS, 501(c)(3)s enjoy tax-exempt status and must be organized and operated exclusively for certain designated purposes, mostly having to do with activities that improve the social good (e.g., education, animal welfare, and providing services to low-income populations). The law also requires that none of the earnings of the 501(c)(3) accrue to a private shareholder or individual. This rule, which is known as the *non-distribution constraint*, is an important defining characteristic of nonprofits.

Nonprofits and publics are not primarily focused on profit maximization but rather on advancing some aspect of the public interest. If the two types of organizations aren't trying to maximize profit, what exactly are they trying to do? The answer to that question isn't straightforward. But it's also not central to our purposes, for the following reason. Most of this book focuses on compensation in for-profit firms, where the organization's objective is *profit maximization*. But the crucial word in that phrase isn't "profit"; it's "maximization".

The essential point is that organizations need to define a clear *objective* that their managers understand (see section 1.4). In for-profit firms, that's profit. In other organizations, it might be something else. But once the managers understand that objective, they can take actions to advance it. So most of the principles developed earlier in this book can be applied in publics and nonprofits simply by replacing "profit" with whatever is the organization's objective. For example, consider public universities. At the end of section 14.4, I discussed how you should handle salary negotiation if you're a dean trying to hire an assistant professor. I concluded by urging you to "make whatever decision will most benefit the university", a phrase that purposely leaves open the nature of the university's objective.

As for small businesses, there is no universally accepted and singular definition. The definition of a small business varies by country, industry, and context.

Sometimes tax laws and other regulations differ between small businesses and other businesses, and in such cases the government offers precise criteria for defining what is a small business and what is not. Other times, outside of the regulatory context, the term "small business" is used informally in the business world, without a precise definition in mind, though it generally means a business with a "relatively small" number of employees and/or a "relatively small" flow of revenue.

What exactly do we mean by "relatively small"? Let's not worry about precise cutoffs to define "small", because for most of our discussion they won't matter. Let's just agree that in this chapter a small business is one that employs a fairly small number of workers (maybe a handful, maybe a good many more, but probably not hundreds), with the exception that if we discuss particular laws and regulations we may need to resort to the precise definitions provided by the government. Again, small businesses can be nonprofits, for-profits, or (at least in the sense of small size, as opposed to the sense of tax law and other regulations) publics. In contrast to nonprofits and publics, small businesses have the conventional "profit maximizing" objective that has been our primary focus throughout the book.

15.2 Organizational Mission and Workers' Intrinsic Motivation

All organizations have missions. Employees like some organizations' missions more than others. If one of your employees values your organization's mission, there are two implications. First, the employee's affinity for the mission should be thought of as a component of that person's compensation, which creates the basis for a compensating differential (Chapter 3). This means that the employee should be willing to accept lower monetary compensation to work for your organization than for an identical organization with a less appealing mission. Second, once hired, your employee's affinity for the mission may provide a source of *intrinsic motivation* that energizes the person to work hard to advance the mission. Intrinsic motivation is similar to performance-based pay in that it's a form of compensation that induces the employee to invest more effort on a regular basis, but in the case of a mission it's "free" in the sense that it's not a monetary cost that appears on the organization's balance sheet.

Here I'm talking about your organization's "mission", whereas throughout the book I've talked mostly about your organization's "objective". What's the difference? There's essentially no difference. They are two alternative terms to capture the notion of what the organization is striving toward. But we have to be a bit careful here, because what the organization is truly striving toward often doesn't exactly match what the organization publicly *claims* to be striving toward. For example, consider the following mission statement:

To organize the world's information and make it universally accessible and useful.

That sounds very much like a nonprofit's objective to advance the social good, particularly the part about "universally accessible and useful". But it's Google's original mission statement, and Google is clearly a for-profit company that's largely out to maximize the bottom line. The company makes money by selling online space for ads, and the ads that users see are the ones for which Google's clients pay top dollar. Google controls the price of the ads, who sees them, and how prominently they're displayed, and all of those decisions are made, first and foremost, to maximize profit rather than to "make information universally accessible and useful". Moreover, it's fair to say that the workers who join Google know that. But "*To maximize profits*" would sound a bit too crass and mercenary as a mission statement for Google, or for any other for-profit for that matter, so loftier goals are put forth instead for public consumption. Really the essential point here is that the mission or objective that gives rise to compensating differentials (Chapter 3), and that creates intrinsic motivation, is the one that employees actually *believe* directs the organization's behavior, and that's not necessarily the mission statement that's displayed on the organization's website.

Although all organizations have missions, those outside of the for-profit sector (i.e., in nonprofits and publics) are more likely to have missions that resonate with workers to the extent that they'd be willing to accept lower monetary compensation to work there. That is why public-interest lawyers are usually paid less than lawyers in other areas of law. So if you're a manager in a public or nonprofit, you should be aware that your employees' degree of affinity for the mission affects how much you must pay them in monetary compensation. Those who are the most passionate about the mission will be willing to sacrifice the most in monetary compensation to work for you, and they'll have strong incentives to work hard on the job to further advance the mission. Generally, the employees whose jobs are more directly related to the (desirable) social mission are those who experience the strongest intrinsic motivation. Those tend to be the employees in the higher ranks of the organization . . . the work of secretaries, for example, and of other workers who are employed at lower ranks of the hierarchy, isn't vastly different from what it would be in a for-profit, so intrinsic motivation isn't particularly strong.

15.3 Compensating Differentials

A friend of mine worked as a Senior Technical Writer at Salesforce.com, Inc., which is a successful American cloud computing company headquartered in San Francisco. His cash compensation in 2014 was an annual base salary of $120,000, plus an annual bonus of 15% of base salary. He had also accumulated restricted stock units (RSUs), which were granted on a 4-year vest cycle. Another significant component of his compensation was the employee stock purchase plan (ESPP), which essentially guaranteed a minimum 15% return on the investment (stocks always being purchased at 15% below a set strike price), twice per year. The return was often more than 15%, and typically added $15,000 to $30,000 to his annual pay, taxed as ordinary income, because he always cashed out immediately.

In 2015, he voluntarily switched jobs, moving to the Salesforce Foundation (Salesforce.org), which is a tax-exempt, 501(c)(3) nonprofit organization that provides technology grants to any qualifying nonprofit or higher education institution. Employees moving to the nonprofit branch of Salesforce typically receive a steep pay cut. Although my friend's base salary remained at $120,000, his annual bonus dropped from 15% to 10%. More significantly, the RSUs and ESPP plan disappeared. Any unvested RSUs (which in my friend's case amounted to about $12,000) were forfeited.

Why did he request to transfer from the for-profit to the nonprofit branch of the company, knowing that he'd be paid considerably less? The answer is that he found the Foundation's work more fulfilling. The first sentence of the Foundation's mission statement is:

> *Salesforce.org is based on a simple idea: leverage Salesforce's technology, people, and resources to help improve communities around the world.*

The opportunity to work towards the goal of "help[ing] improve communities around the world" excited my friend enough that he was willing to incur a big cut in monetary pay, given that such mission statements are more credible in nonprofits than in for-profits. Recalling the definition of compensation from section 1.1, the appeal of the mission at Salesforce.org served as a non-monetary component of compensation, for which my friend was willing to sacrifice some monetary compensation. The difference in monetary pay between the dot com and dot org branches of Salesforce is a good example of a compensating differential (Chapter 3). The differences in pay between the two Salesforce branches are further explored in a case discussion that closes the chapter.

Mission-based compensating differentials may arise for nonprofits like Salesforce.org, and for publics, but what about for small businesses? Small businesses are just conventional firms (albeit smaller ones that experience some difference in regulatory treatment), so their missions generally aren't especially inviting to workers. Some employees may simply like (or dislike) working in small businesses, in which case compensating differentials arise for reasons explained in Chapter 3. The reasons people may have preferences one way or the other derive from some unique features of the small business working environment, and I'll highlight some of these shortly.

Small firms generally pay less than larger ones. This is a well-known fact that has been recognized by compensation researchers for decades. Why do larger firms tend to pay more than smaller ones? Multiple potential reasons have been suggested, and delineating all of them would take us too far afield. But the pay difference between small and large firms could be interpreted, at least in part, as a compensating differential, if small businesses are seen as particularly desirable places to work. Small business employers could then still attract workers even when paying less monetary compensation than larger firms.

Why might small businesses be seen by workers as desirable places to work? One reason could be the same reason why many students prefer the intimacy and closer relationships in small classes than in large ones. The smaller number of employees in a small business might also imply greater variety in task assignments and stronger demand for multiskilling, which appeal to some employees. Or the smaller number of internal competing co-workers might make it easier to advance to leadership positions. There are also usually fewer "turf wars" in small businesses, which lessens one dimension of distasteful office politics.

Of course, there are arguments in the other direction, for why some employees may find small businesses to be relatively undesirable. Employees may perceive that large firms offer better career prospects, because having worked for a large firm may have more kudos on a résumé than having worked for a small business. Such perceptions may put small businesses at a disadvantage when competing with larger businesses on the basis of pay, even at the same levels of monetary compensation and benefits.

Another source of competitive advantage for larger firms is that there are certain efficiencies that go hand-in-hand with larger size. For example, consider firms that are negotiating with insurers to purchase health plans for their workers. Larger firms can negotiate better prices for health insurance, for reasons we discussed in Chapter 11. This means that to offer its employees the same amount of health insurance that a large firm offers, a small business must spend more. This is another reason to expect lower pay levels in small businesses.

15.4 External and Internal Constraints on Pay

External and internal constraints on pay (Chapters 4 and 5) apply in publics, nonprofits, and small businesses just as they do in other organizations. As we discussed in Chapter 4, external constraints refer mostly to laws and regulations. Nonprofits usually enjoy tax-exempt status, at least to a certain degree. In the United States they are exempt from federal taxes. However, if nonprofits engage in activities that are unrelated to their fundamental social mission, then they are taxed on the income generated from those activities. Nonprofits are also exempt from paying sales tax and property taxes. Nonprofits are not exempt from payroll taxes like Social Security and Medicare; they pay taxes on those items just like a for-profit company.

The tax advantages that are enjoyed by nonprofits don't come for free. Having the nonprofit designation subjects an organization to a legal constraint that doesn't apply to for-profits. In particular, the law allows nonprofits to earn profits (yes, despite their name, nonprofits can and do make profits!) but restricts what can be done with those profits. As I mentioned earlier, the key regulation is often called the "*non-distribution constraint*", and it prohibits nonprofits from disbursing profits to those who run the organization. Rather, the profit must be reinvested in activities that enhance the organization's social mission.

The non-distribution constraint is just one example of the external constraints that we discussed in Chapter 4. Do you remember what your managerial mantra should be concerning constraints? The 3 Cs! *Comprehend*, *circumvent*, *comply*. The "comprehend" part is easy, because all nonprofit managers are acutely aware of the non-distribution constraint and, unlike other labor laws that change over time, the non-distribution constraint is fairly static. So, focusing on the second "C", your primary challenge as a manager is to devise ways to loosen or circumvent the non-distribution constraint, bearing in mind the warnings of section 4.3 concerning ethics and the law.

For example, although the non-distribution constraint prohibits profit from being disbursed back to those who run the organization, it is possible to inflate the compensation of those employees, thereby reducing profits (which, again, are defined as total revenues minus total costs). By increasing the compensation costs the nonprofit organization can effectively "burn" profit (by spending it on higher compensation) before the profit even materializes, whereas distribution of that profit *after* it materializes, in the form of performance pay (see Chapters 9 and 10) would be in violation of the non-distribution constraint.

Small businesses are also subject to external legal constraints. As a small-business manager, you should be aware that some government regulations (including the tax treatment of businesses) vary by firm size. Moreover, the cutoffs for "small" size may vary from one law to the next. For example, the California state hourly minimum wage (as of January 1, 2019) is \$11 for employers with 25 or fewer employees, and \$12 for employers with 26 or more employees. This means that if you have been operating a small business with 23 employees and then expand to 28 to meet growing demand for your product or service, your size classification with respect to the state minimum wage has changed, and you must adjust your compensation plan accordingly.

But the federal Family and Medical Leave Act (FMLA) uses a different size cutoff. FMLA applies to employers with 50 or more employees (who live within a 75-mile radius of the business), though most employers with fewer than 50 employees are exempt. Small businesses whose employment headcounts hover around this critical threshold of 50 are subject to the law if they had at least 50 employees (who live within a 75-mile radius of the business) employed for 20 workweeks in the previous or current calendar year. So if you run a small business in California with 35 workers who live locally, then you're "large" with respect to the state minimum-wage law (i.e., greater than 25) but "small" with respect to the federal FMLA (i.e., less than 50). Generally, when laws differ by firm size, they tend to be less stringent for smaller employers, as the preceding two examples illustrate. The bottom line is that as a manager in a small business you must pay particular attention to how your business is classified – by size – with respect to each of the labor regulations (federal, state, and local) that apply to your business.

As we discussed in Chapter 5, unions are a significant source of internal constraints on pay. Of the three settings covered in this chapter, publics are definitely the places where unions are the most prevalent and relevant. Indeed, throughout the book I've drawn on a number of examples from CSUEB, which is the unionized, public-sector organization where I am currently employed. While unions have declined in strength and membership in the United States for several decades, they have remained relatively strong in the public sector, in part because of laws that govern the rights of public-sector unions to tax employees through payroll deductions. A unanimous 1977 ruling by the US Supreme Court (in *Abood* v. *Detroit Board of Education*) held that public-sector unions could legally force non-union members (who the unions were required to represent in addition to their own members) to pay dues or fees to cover the costs of collective bargaining. These fees that non-members (who, again, are represented by the union) were compelled to pay are sometimes referred to as the non-members' "fair share".

I obtained firsthand knowledge about fair shares in 2008, when I joined CSUEB. On my first paycheck I noticed a mysterious 1% payroll deduction that seemed to be an involuntary contribution to the California Faculty Association (CFA), which, as we discussed in Chapter 5, is the public sector union representing faculty within the CSU system. After discovering the unwanted deduction, I informed the payroll office that I wasn't a union member, and their response (consistent with the 1977 Supreme Court ruling) is that even as a non-member I was required to relinquish 1% of every paycheck to the union.

CFA members actually had to pay slightly more than the 1% "fair share" tax that applies to everyone (members and non-members alike). The reason is that the 1977 ruling allowed for non-members to be exempt from financing the union's ideological or political activities with which the non-member disagrees. Rather, employees could only be compelled to finance the costs of collective bargaining, contract administration, and grievance adjustments, which are the union activities from which non-members actually derive clear benefits. Thus, in the CSU context, professors who wished to opt out of the charges that are allocated to finance political activities could do so by writing a formal letter to the CFA, opting out of those fees. But the 1% "fair share" tax was unavoidable for everyone. The amount that was eligible to "opt out" of because of its connection to political expenditures was extremely small relative to the "fair share". This isn't surprising, because the union has a lot of discretion over choosing this number, and making that number very small (relative to the fair share) meant that non-members were forced to pay the union almost the same amount as members. Faced with such a negligible price difference between joining and not joining the union, many faculty members would join.

All of this changed abruptly in June 2018 with a 5–4 ruling by the US Supreme Court (in *Janus* v. *AFSCME*). The 2018 ruling overturned the 41-year-old precedent set by "Abood" and prohibited public-sector unions from compelling non-members to pay the "fair share". Thus, as of July 2018, I no longer pay a 1% "fair share" tax to the CFA. The 2018 ruling dealt a painful blow to public-sector unions in the United States. It will cost the unions a significant amount of revenue now that they can no longer compel non-members to contribute anything at all.

Moreover, it is likely that the number of employees who join the union will shrink even if the dues remain unchanged. The reason is that even if the union dues remain unchanged from their "pre-Janus" level, the cost of joining the union sharply increased after the 2018 ruling. Before June 2018, everyone (members and non-members alike) had to pay 1% of their salary to the CFA. Then the cost of joining the union is only a nominal fee that's far less than 1% of your salary (again, the union understandably made this small to encourage more workers to join the union). But after the 2018 ruling, the cost of joining the union is that nominal fee *plus* 1% of one's salary. Put another way, those faculty who joined the union prior to 2018 could have 1% of all future paychecks restored simply by quitting the CFA.

Apart from unions, another source of internal constraints is bureaucracy, i.e., excessively complicated administrative procedures that impede the smooth and efficient operation of business. Bureaucracy is often fueled by unions, because most collective bargaining agreements are awash with internal rules and regulations that you must follow as a manager. But bureaucracy certainly exists in non-unionized settings, and it tends to grow with the size of an organization. It tends to be less of an issue in small businesses. But the public sector is famous for it. As I mentioned at the outset of this chapter, if you're a manager in one of the three settings covered in the chapter, most of the lessons from the preceding chapters are applicable. But not always with equal weight. If you're a manager in a public-sector organization, then the material in Chapter 5 is likely to be far more relevant to you than if you're a manager in a small business.

15.5 Recruitment and Training

I've already mentioned that the organizational mission tends to be important in nonprofits and in publics, for attracting the right kinds of employees, for keeping those employees intrinsically motivated, and for serving as a potential source of compensating differentials (Chapter 3) that lead to lower wages than those found in comparable for-profit firms. What this means is that "fit" (between the employees and the job) is particularly important in nonprofits and publics, i.e., the organization wants to recruit those people who identify strongly with the mission. It might be expected that such organizations recruit particularly intensively, relative to for-profits, and there is indeed some empirical evidence of this in academic research.

But there's another reason, apart from the desire to secure a "good fit" who identifies strongly with the organizational mission, why nonprofits might recruit more intensively than for-profits. That reason is, again, the "non-distribution constraint", which prohibits nonprofits from disbursing the organization's income (i.e., profit) to those who run the organization. If those funds can't be distributed, then what can be done with them? The government's underlying motivation for imposing the non-distribution constraint was to require nonprofits to reinvest profit in furthering the social mission. An indirect way of doing this is investing resources in finding the perfect fit for each position ... basically, since the profit cannot legally be disbursed back to those who run the organization, it has to be "burned" on something, so it might as well be burned on something useful like intensive recruitment. This line of reasoning may lead nonprofits to over-invest in activities like recruitment, meaning they invest in them more heavily than they would in the absence of the non-distribution constraint.

Exactly the same argument applies to training. That is, because of the non-distribution constraint, it may be expected that nonprofits are more likely than for-profits to provide training. And indeed there is some empirical evidence of just that in academic research.

15.6 Performance-Based Pay

Performance-based pay is used more extensively in the for-profit sector than in either the public or nonprofit sectors. There are several reasons for that. One reason is a legal constraint, namely the aforementioned non-distribution constraint that prohibits nonprofits from disbursing profits to those who run the organization. The non-distribution constraint is why you do not see, for example, profit-sharing compensation plans in nonprofits. As mentioned in Chapter 9, profit sharing is a form of performance-based pay in which the performance measure is based on organizational performance (namely profit) as opposed to the performance of individual employees. Identifying a performance measure that reflects organizational performance in a public or nonprofit is more complicated than simply measuring profit, because the organization's objective might not lend itself to easy quantification and measurement.

Even performance-based pay that is based on individual employee performance (as opposed to organizational performance) is less common in nonprofits and publics than in the for-profit sector. One reason concerns measurement difficulties. If it's hard to quantify and measure the objective for the entire organization, it's likely to be hard to quantify and measure an individual employee's contribution to that objective. Another reason concerns the intrinsic motivation generated by the organizational mission. Those incentives are generally stronger in publics and nonprofits than in for-profits and can substitute, to some degree, for monetary motivators like performance-based pay. Another reason (applying to publics) is that, as I've mentioned, publics are considerably more likely to be unionized than non-publics, and unions generally dislike performance-based pay because of the dispersion in monetary compensation that it tends to create (see Chapter 5 and section 1.10).

Don't get the impression that nonprofits never use performance-based pay. Rather, that form of pay is considerably less prevalent in nonprofits than in for-profits. But there are instances in the nonprofit sector of compensation that is strongly performance based. Consider, for example, Amnesty International, a London-based nonprofit that focuses on advancing human rights. Canvassers employed by Amnesty International have the job of approaching people on public streets in an effort to raise money. There are strict quotas that canvassers must meet in terms of new donations to avoid losing their jobs. A quota system is a form of performance-based-pay that can create powerful incentives, as we saw in section 9.1.

In addition to lessening the need for the explicit pay-based incentives of Chapter 9, the presence of intrinsic motivation in nonprofits (and publics) also lessens the need for the promotion-based incentives of Chapter 13. Recall from Chapter 13 that promotions serve two main objectives in organizations . . . they create incentives for employees, and they match employees to jobs in an ideal way. The challenge is that it's hard to use one instrument (i.e., promotions) to perfectly achieve two very different objectives, and improving on one objective often means compromising on the other. The fact that nonprofits (and publics) can rely on intrinsic motivation lessens the need to use promotions to create employee incentives; instead, promotions can be used primarily to match employees to jobs in an ideal manner for the organization.

15.7 Turnover

We've discussed the tight connection between compensation and turnover (Chapter 12) and the fact that there are both positive and negative aspects of turnover. Turnover is often relatively low in the public sector, giving rise to the stereotypical "lifetime government job". Some of the reasons for low turnover in publics are institutional. As we've discussed, in the United States, unions have a stronger presence inside the public sector than outside of it, and unions tend to enhance job security by making it difficult for employers to fire workers.

Another reason is that publics, such as CSUEB, are more likely to have defined-benefit pension plans (see Chapter 11), which discourage turnover. Another factor that can contribute to lower turnover in publics (and also nonprofits) is the sorting effect (see section 1.8). The employees who end up in nonprofits and publics tend to be those who identify most strongly with the organization's social mission, whereas those who end up in for-profits are (at least in a relative sense) more interested in money. The employee–job matches may, therefore, be tighter in nonprofit and publics, and less easily sundered by offers of higher pay from competing firms. Employees in for-profits, in contrast, may be easier to tempt away to another firm that pays a little more.

Organizations should tread carefully when it comes to changing the organizational mission, either literally or via actions that may seem in conflict with the mission. Google, for example, once had the motto "Don't be evil", yet the company began working on defense contracts that leveraged artificial intelligence to physically harm people more efficiently. Google was working with the US Department of Defense on the Maven program, which uses artificial intelligence to interpret video images. The technology could be used to improve the targeting of drone strikes. There was an outcry among Google's employees, who worried that the contract would pave the way for incorporating the technology into advanced weapons.

Roughly four thousand Google employees signed a petition demanding a clear policy statement that the company wouldn't build warfare technology for killing people. The petition served as a strong warning to Google that the sorting effect (see section 1.8) was going seriously awry, and indeed, some of Google's employees actually resigned in protest. In June 2018, in response to its employees' strong concerns, Google made the decision not to renew a contract with the Pentagon for artificial intelligence work when the contract expired in 2019. The point here is that organizational missions, which drive the sorting effect, are only effective when the employees buy into them, and that can't happen if the organization's management takes actions that visibly conflict with the mission.

In small businesses, the costs of turnover can loom particularly large. The costs imposed when a Walmart cashier quits are pretty small for that Walmart store and are negligible for the company as a whole, whereas when a long-tenured employee in a "Mom-and-Pop" grocery or restaurant quits, the costs tend to be much higher and more disruptive. Small businesses tend to have fewer resources to devote to recruitment and training, and the institutional knowledge that is lost when employees separate can be more damaging.

To highlight the issues, let's compare a large and a small business in the restaurant industry. Chili's Grill & Bar is a well-known American chain restaurant serving Tex-Mex cuisine. As of 2017, Chili's operated over 1600 locations on five continents. There were actually locations on six continents, but in 2008 the company announced it would close all of its Australian restaurants after a fine was imposed on it by the New South Wales Office of Industrial Relations for wage theft and underpaying its employees (see Chapter 2).

Chili's has a set menu that has evolved over time, so if you enter a Chili's restaurant (as most American readers probably have at some point) you pretty much know what to expect in terms of menu items and quality. As is the case with most large chain restaurants, consistency of the product is ensured not only by setting the menu and recipes but by centralizing some of the production, so that each Chili's restaurant isn't producing every dish entirely from scratch. Such standardization also lessens the expectations on cooks in terms of their required skill levels. A given Chili's restaurant employs a number of different people working as cooks on different shifts, none of whom is indispensable, so that if one of them quits the costs imposed on that restaurant are modest and short-run. In fact, if you were to eat at Chili's several times in a given week, there's a good chance that your meal would be prepared by a different person, or team, each time.

In contrast, consider Kamdesh Afghan Kabab House, a small restaurant operating at a single location in Oakland, California, right around the corner from where I teach my MBA compensation class. According to one of my former students who is part of that family-owned small business, the biggest challenge that Kamdesh faces is turnover. Employees are often poached by other restauranteurs who are interested in learning Kamdesh's secret recipes, food preparation system, and kitchen layout, so that they can incorporate these features into their own restaurants. Competing restaurants offer as low as 50 cents more per hour to successfully recruit Kamdesh's best cooks. In addition to the concerns about having Kamdesh's recipes (i.e., trade secrets) shared when its employees are poached, the turnover implies high costs of hiring and training a replacement cook while maintaining the restaurant's continuous operation and the consistency of its food.

Unlike Chili's, if you eat at Kamdesh several times in a week, it is very likely that all of your meals are prepared by the same person and if that person is particularly skilled and poached by a competing restaurant then you may well notice a drop in the quality of the food. If that cook were to be poached by a nearby restaurant (and, indeed, in 2019 there is another Afghan restaurant on the same block as Kamdesh), along with Kamdesh's recipes and trade secrets, Kamdesh could lose market share to the competitor. Even if Kamdesh's cook isn't poached but rather calls in sick on a busy night, that could impose costs on the restaurant that are considerably larger than those a Chili's establishment would incur in the same circumstance.

Because of their smaller size, demands on employees in small businesses tend to differ from the demands on their counterparts in larger firms, and this also has implications for turnover. Fewer employees sometimes means that there's a greater need to multi-task, so being able to do a lot of things, solve problems, work independently, handle unexpected situations, etc., becomes important. This "task variety" and expansive responsibility may be appealing to some workers but distasteful to others. Like any other non-wage job attribute, task variety can give rise to compensating differentials (Chapter 3). As a small-business manager focused on using the compensation system to manage turnover, particularly when it comes to retaining your top talent, you should be aware of such task variety (and other unique features of the small-business working environment) that may be valued more by some of your employees than by others.

The costliness of turnover in small businesses means that "fit" (of employees with the job) becomes particularly important when hiring. While a large firm like Google might want to hire the smartest programmer they can find, a small business might place greater emphasis on stability and hire someone who is less capable but more likely to remain longer in the position.

Another aspect of "fit" is how well employees get along with co-workers. Employee attitudes are often even more important in small businesses than in larger ones. It's easier to tolerate (and retain) a high-performing employee with a bad attitude in a larger organization than in a smaller one. One reason is that such "difficult" employees tend to generate higher turnover (particularly of those employees who report directly to them), and those turnover costs are easier to absorb for a large business than for a small one.

15.8 "Distance" between Managers and Owners

There is often a shorter "distance" (in the sense of roles, rather than geography) between owners and managers in small businesses than in larger ones, and this has implications for compensation. In a large corporation, many decisions about compensation are made by people who are very far removed from the firm's ownership. This sometimes means that the person granting higher compensation to an employee often doesn't have to finance that pay increase directly. Therefore, compensation decisions are made that would look quite a bit different than if the firm's owner(s) were making them. A manager who is about to retire or switch firms, for example, might grant a large raise to a subordinate (and friend) as a parting gift, at the expense of the firm's owners.

The situation I just described is far less likely to happen in a small business like Kamdesh, where the owner is also managing the restaurant and making the hiring and compensation decisions. Because Kamdesh's profits are retained by the owner, an extra dollar of compensation that is given to employees beyond what is necessary to comfortably retain them is one less dollar of profit that goes to Kamdesh's owner. This shorter distance between ownership and management in small businesses may also contribute to the lower compensation levels that tend to be seen in small firms. The shorter distance between the owner and the managers (and other employees) can also create other tensions beyond those that I just mentioned concerning pay levels.

15.9 Lessons for Managers

Knowing your organization's objective is particularly important in nonprofits and publics, because that objective isn't as simple as "maximizing profit", and you need to understand the objective before you can manage the organization in a manner that advances that objective.

If you manage in one of the three settings covered in this chapter, know the rules and regulations, because some of them are specific to your setting. The rules also change over time, and if you manage a small business in which the number of employees fluctuates, then even if the regulations remain fixed you may face periodic fluctuations in which regulations apply to you. If you're a nonprofit manager, pay particular attention to the tax laws that apply to you and to the non-distribution constraint and its implications.

If you're in the public sector, be aware that you're subject to labor laws affecting compensation that wouldn't be relevant in the for-profit sector. For example, living wage ordinances act like minimum-wage laws but apply only to city (or state) employees and to the contractors and subcontractors that do business with the city. Also, in the public sector you're considerably more likely than in a for-profit to be in a unionized setting, which implies an additional set of compensation regulations beyond those required by law.

Remember the 3 Cs, and remain alert for creative (though ethical and legal) ways to circumvent or relax the constraints that you face. For example, when managing a nonprofit, increasing base salaries or fringe benefits is a way to effectively transfer profit to employees "before the profit materializes", simply by voluntarily lowering profit via higher labor costs. Similarly, you can invest heavily in training, which employees value as part of their compensation because it increases their future productivity and earnings potential.

Keep in mind that even though your employees have chosen the nonprofit (or public) sector, there are for-profit opportunities that they could have chosen and that they can move to in the future. So your competitors are not just other nonprofit (or public) organizations but also for-profit firms. If you're a small-business manager, be aware that larger firms tend to offer more generous compensation packages. This puts you at a competitive disadvantage, and you need to find other ways to compete. So the employee–job "fit" or quality of the match becomes very important, just as it does in nonprofits and publics due to the mission. Identify the advantageous (from the employee's standpoint) features of the working environment in your small business, and invest in recruiting and screening to identify those employees who value those features most highly.

Understand the extent to which the organization's mission generates intrinsic motivation in nonprofits and publics. If intrinsic motivation is strong for some employee groups, this may allow you to rely less heavily on incentives based on pay or promotions. Remember that intrinsic motivation hinges on finding (and retaining) employees who are excited about the mission, and that involves careful recruiting and screening. Those employees whose affinities for the organizational mission are weakest (which tend to be employees in the organization's lower ranks) have the weakest intrinsic motivation and are likely to be the easiest ones for competing firms in the for-profit sector to poach with higher compensation offers. Capitalize on the mission by investing carefully in recruiting and screening employees to find those who are most excited by the mission. That will save on compensation costs and lead to higher employee productivity.

Case Discussion 25: Salesforce.com versus Salesforce.org

Salesforce.com, Inc. is a successful, for-profit, American cloud computing company based in San Francisco, California. In early 2016, Salesforce.com had a market capitalization of $61 billion, making it one of the highest-valued American cloud computing companies. Salesforce.org, called the Salesforce Foundation, is a tax-exempt, nonprofit, 501(c)(3) organization that provides different kinds of grants, including grants for disaster relief assistance.

Salesforce.org has the following mission statement:

> *Salesforce.org is based on a simple idea: leverage Salesforce's technology, people, and resources to help improve communities around the world. We call this integrated philanthropic approach the 1-1-1 model because it started with a commitment to leverage 1% of Salesforce's technology, people, and resources to improve communities around the world. By encouraging and enabling companies to adopt the 1-1-1 model, Salesforce.org is helping to spark a worldwide corporate giving revolution.*

Compensation is generally lower at Salesforce.org than at Salesforce.com, consistent with the willingness of Foundation employees to sacrifice some monetary compensation in exchange for the opportunity to do work that advances the organization's mission.

Questions

1. Salesforce.org can hire workers either from Saleforce.com or from outside of Salesforce. What are the advantages and disadvantages of both approaches?
   (a) Similarly, Salesforce.com can hire workers either from Salesforce.org or from outside of Salesforce. What are the advantages and disadvantages of both approaches?
2. Should Salesforce encourage or discourage internal migration of its employees from one branch to the other? In which direction is internal migration more desirable, or less damaging?
3. How should Salesforce determine the size of the pay gap between Salesforce.com and Salesforce.org?
   (a) To what extent can Salesforce choose this pay gap, given the influences of the external market?
4. Suppose that Salesforce implemented the following job rotation policy. After serving five years at Salesforce.com, an employee would be required to spend at least one year at Salesforce.org, with a voluntary option to stay a second year at Salesforce.org, before returning to Salesforce.com.
   (a) What are the advantages and disadvantages of such a policy?
   (b) If the preceding policy is implemented, explain how it can be expected to affect compensation at Salesforce.org and Salesforce.com?
   (c) What are the implications for external (i.e., non-Salesforce) recruiting into either branch?

Further Reading

Ben-Ner, A., Ren, T., and Paulson, D. F. (2010). "A Sectoral Comparison of Wage Levels and Wage Inequality in Human Services Industries." *Nonprofit and Voluntary Sector Quarterly*, 40, 608–33.

Brunner, E. J. and Ju, A. (2018). "State Collective Bargaining Laws and Public-Sector Pay." *Industrial and Labor Relations Review*, 72:2, 480–508.

Davis, J., Batchelor, J. H., and Kreiser, P. (2018). "The Influence of Organizational Task Environment and Firm Size on Top-Executive Compensation Contracts." *Journal of Small Business and Entrepreneurship*, 31:1, 21–42.

DeVaro, J. (2010). "A Theoretical Analysis of Relational Job Design and Compensation." *Journal of Organizational Behavior*, 31, 279–301.

DeVaro, J. and Brookshire, D. (2007). "Promotions and Incentives in Nonprofit and For-Profit Organization." *Industrial and Labor Relations Review*, 60, 311–39.

Hurst, E. and Pugsley, B. W. (2011). "What Do Small Businesses Do?" *Brookings Papers on Economic Activity*, 73–142.

Kalmenovitz, J. (2018). "Pay Inequality and Public Sector Performance: Evidence from the SEC's Enforcement Activity." Available at SSRN: https://ssrn.com/abstract=3072695 or http://dx.doi.org/10.2139/ssrn.3072695

Kickul, J. (2001). "Promises Made, Promises Broken: An Exploration of Employee Attraction and Retention Practices in Small Business." *Journal of Small Business Management*, 39, 320–35.

Saitgalina, M., Bennett, A. A., Reina, C. S., and Coombs, J. E. (2018). "Serving Others at the Expense of Self: The Relationship Between Nonprofit CEO Compensation and Performance in Trade and Professional Associations." *Journal of Public and Nonprofit Affairs*, 4, 156–80.

Seyam, A. and Banerjee, S. K. (2018). "An Examination of Issues Surrounding Nonprofit Executive Salaries: 2010–2017." *Applied Finance and Accounting*, 4, 67–78.

Last Remark

My goal was to write a book that would help managers, and aspiring managers, to think about compensation in a more sophisticated and nuanced way, so that they can make better business decisions. Did I succeed? The answer remains to be seen and depends upon how we define success. My definition of success is not whether you learned a lot (although I hope you did!) . . . it is the extent to which you can translate what you have learned into better business decisions in the future. Suppose, for example, that the lessons that you have internalized from reading this book allow you to adjust your organization's compensation system to solve or mitigate a specific problem, to strengthen incentives for your existing employees, to recruit a more productive workforce, to retain your top talent, etc. Each such good decision, large and small, creates tangible value for your organization, and the cumulative value of those decisions (across all readers) is the best measure of this book's success. So . . . the *really* hard work – of applying the lessons from the preceding pages to solve business problems – begins where this book ends.

Further Resources

A good way to stay up to date on US labor law at the federal, state, and local levels is to monitor the websites at all relevant levels of government.

To stay current on federal labor law in the United States, consult the following website of the US Department of Labor: www.dol.gov/elaws

For labor law at the state and local levels, visit the relevant government websites. State-level information can also be found at www.dol.gov/whd/contacts/state_of.htm.

Data Sources

The Current Population Survey (CPS), which is sponsored jointly by the US Bureau of the Census and the Bureau of Labor Statistics (BLS), is the primary source of labor force statistics for the US population, and it contains a lot of information related to compensation. CPS data can be found here:

www.census.gov/programs-surveys/cps.html

The Survey of Income and Program Participation (SIPP) is a good data source for information on income and participation in government programs for both workers and non-workers. It tracks people over time, which allows investigation of changes in key variables such as economic well-being, family dynamics, education, assets, health insurance, childcare, and food security. SIPP data are sponsored at the US Census Bureau and can be found here:

www.census.gov/sipp/

Data from the National Compensation Survey (NCS) conducted by the BLS can be found here:

www.bls.gov/ncs/

The data are used to produce the Employment Cost Index (ECI), which measures changes in labor costs, and the Employer Costs for Employee Compensation (ECEC), which measures average hourly costs for employee compensation.

ExecuComp, which is the most widely used data set for executive compensation, can be obtained with an account with Wharton Research Data Services (WRDS):

https://wrds-web.wharton.upenn.edu/wrds/ds/execcomp/exec.cfm

The Panel Study of Income Dynamics (PSID) is collected at University of Michigan. It tracks the same US workers and households over a long period of time (significantly longer than the SIPP). The survey covers employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics.

https://psidonline.isr.umich.edu/

The National Longitudinal Surveys (NLS), sponsored by the BLS, track workers over time and have a large amount of information on wages, benefits, employment, work hours, occupation and industry, and many other variables.

www.nlsinfo.org/

The Health and Retirement Study (HRS) covers older workers and has a large amount of information related to work and compensation, including information on retirement and pension plans, health and life insurance, and health status.

Some Useful Websites

Cornell University's Institute for Compensation Studies (ICS) is an interdisciplinary center that provides a lot of compensation-related resources with the aim of bridging the gap between academic researchers and compensation practitioners. Its mission is to improve the teaching, research, practice, and public discourse around compensation and rewards to work:

www.ilr.cornell.edu/institute-for-compensation-studies/about-ics

World at Work
www.worldatwork.org
Society of Human Resources Management
www.shrm.org
European Industrial Relations Observatory
www.eurofound.europa.eu/observatories/eurwork
International Labour Organization
www.ilo.org
Employee Benefits Research Institute (EBRI)
www.ebri.org

Compensation Café is an interesting blog devoted to compensation issues:

www.compensationcafe.com/

For guidance concerning FMLA compliance (and dealing with employee abuses of FMLA) see:

www.fmlainsights.com/

For information on employee ownership, visit the following two websites:

https://rady.ucsd.edu/centers/beyster/ (The Beyster Institute)
fed.org (The Foundation for Enterprise Development)

Lots of HR and compensation-related information can be found at:

www.hr-guide.com

For information on pensions and retirement income see:

www.ssa.gov
www.medicare.gov
www.pbgc.gov

For tax law and information on nonprofits see the US Internal Revenue Service (IRS) website:

www.irs.gov

Journals that Publish Articles on Compensation

There are two broad categories of journals that publish articles related to compensation, and they are differentiated by their target audiences. The first is "practitioner journals", for which the target readers are mostly HR professionals, managers, and other readers outside of academia. The second is "academic journals", for which the target readers are academic researchers. Academic journal articles are harder to read, longer, and generally less interesting for business managers. They are heavy on academic theoretical jargon and on quantitative and statistical methods that go considerably beyond the basic tables and graphs that are the focus of most practitioner journals.

The following six journals focus in whole or in part on compensation and are oriented for practitioners rather than for academic researchers:

California Management Review (CMR)

http://cmr.berkeley.edu/

Compensation & Benefits Review (CBR)

http://journals.sagepub.com/home/cbr

Harvard Business Review (HBR)

https://hbr.org/magazine

IZA World of Labor

https://wol.iza.org/articles

WorldatWork Journal

www.worldatwork.org/journal/

Workspan

www.worldatwork.org/workspan/

A selected list of academic journals containing articles about compensation (and other topics):

Academy of Management Executive
Academy of Management Journal
Academy of Management Perspectives
Academy of Management Proceedings
Academy of Management Review
Administrative Science Quarterly
American Economic Review
British Journal of Industrial Relations
Economic Journal
European Economic Review
Human Resource Management
Human Resource Management Review
Industrial and Labor Relations Review
Industrial Relations: A Journal of Economy and Society
International Journal of Human Resource Management
Journal of Accounting and Economics
Journal of Applied Psychology
Journal of Empirical Finance
Journal of the European Economic Association
Journal of Finance
Journal of Financial Economics
Journal of Human Resources
Journal of International Business Studies
Journal of Labor Economics
Journal of Law, Economics, and Organization
Journal of Management
Journal of Occupational and Organizational Psychology
Journal of Organizational Behavior

Journal of Participation and Employee Ownership
Journal of Political Economy
Labour Economics
Management Science
Monthly Labor Review
Personnel Psychology
Quarterly Journal of Economics
Review of Economic Studies
Review of Financial Studies
Strategic Management Journal

Index