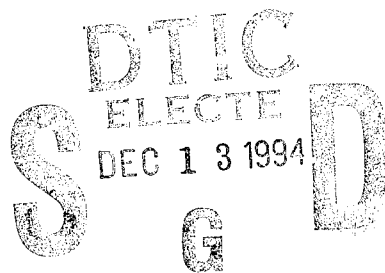


RAND

A Theory of Military Compensation and Personnel Policy

Beth J. Asch, John T. Warner



**National Defense
Research Institute**

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

19941207 039

The research described in this report was sponsored by the Office of the Secretary of Defense under RAND's National Defense Research Institute, a federally funded research and development center supported by the Office of the Secretary of Defense, the Joint Staff, and the defense agencies, Contract No. MDA903-90-C-0004.

Library of Congress Cataloging in Publication Data

Asch, Beth J.

A theory of military compensation and personnel policy / Beth J. Asch, John T. Warner.

p. cm.

"Prepared for the Office of the Secretary of Defense."

"MR-439-OSD."

Includes bibliographical references.

ISBN 0-8330-1544-3

1. United States—Armed Forces—Pay, allowances, etc. 2. United States—Armed Forces—Personnel management. I. Warner, John T. II. United States. Dept. of Defense. Office of the Secretary of Defense. III. RAND. IV. Title.

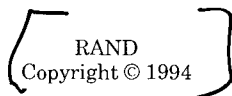
UC74.A834 1994

355.6'4—dc20

94-16956

CIP

RAND is a nonprofit institution that helps improve public policy through research and analysis. RAND's publications do not necessarily reflect the opinions or policies of its research sponsors.



Published 1994 by RAND

1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002; Fax: (310) 451-6915;
Internet: order@rand.org.

Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced <input type="checkbox"/>	
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

RAND

A Theory of Military Compensation and Personnel Policy

Beth J. Asch, John T. Warner

*Prepared for the
Office of the Secretary of Defense*

**National Defense
Research Institute**

PREFACE

A primary goal of military compensation is to enable the military to meet its manning objectives for force size, composition, and wartime capability. To attain these objectives, compensation must be appropriately structured to attract, retain, and motivate personnel at a reasonable cost, even when national security goals are changing. A key question facing military manpower and compensation managers is, How should military compensation be structured? The question of military compensation design has been actively debated over the years. But past studies have narrowly focused on the relationship between compensation and retention. Less attention has been paid to whether the military compensation system induces the best individuals to stay and seek advancements, and whether it motivates effective work. To address the issue of how military compensation should be designed in light of these considerations, a model is needed. The research presented in this report develops a model of compensation in a large, hierarchical organization such as the military. The model permits an analysis of the issues surrounding the design of military compensation.

The report is highly technical and is intended for analysts concerned with the structuring of compensation in large, hierarchical organizations such as the military, as well as for researchers concerned with military manpower issues.

This research was sponsored by the Assistant Secretary of Defense (Personnel and Readiness). The report was prepared within the Defense Manpower Research Center, part of RAND's National Defense Research Institute, a federally funded research and devel-

opment center sponsored by the Office of the Secretary of Defense, the Joint Staff, and the defense agencies. John Warner is a professor of economics at Clemson University and a RAND consultant. During the course of this research, he worked as a visiting scholar in the Office of Special Projects and Research, Office of the Assistant Secretary of Defense (Personnel and Readiness).

CONTENTS

Preface	iii
Figures	ix
Tables	xi
Summary	xiii
Acknowledgments	xix
List of Symbols	xxi
Acronyms and Abbreviations	xxv
Chapter One	
INTRODUCTION	1
Chapter Two	
THE MILITARY COMPENSATION AND PERSONNEL SYSTEMS: DESCRIPTION AND ISSUES	5
The Compensation System	5
Personnel Management: Officers	7
Personnel Management: Enlisted Personnel	12
Issues	16
Criticisms of the 20-Year Retirement System	19
Criticisms of the Active-Duty Pay System	23
Discussion of the Criticisms	25
Past Studies of the Military Retirement System	28
Chapter Three	
ORGANIZATIONAL STRUCTURE: DEFINITIONS AND ASSUMPTIONS	37

The Internal Labor Market and Personnel Flows in the Hierarchy	37
Production Relationships in the Hierarchy	42
Organizational Goals	43
The Budget Constraint	44
Chapter Four	
MODELING INDIVIDUAL DECISIONMAKING	47
Modeling Retention Decisions	47
Individual Retention Decisions	49
Modeling Effort Supply	53
Promotion Contests	53
Involuntary Separation Policies and Work Effort	57
Intragrade Pay Policies and Work Effort	59
Model Summary	61
Comparative Static Results	63
Determining the Individual Retention Decision	63
Variations in Effort and the Probability of Promotion	65
Individual Effort Decisions	67
Aggregation: Derivation of Cohort Retention Rates	70
Chapter Five	
THE EFFECT OF POLICY ON INDIVIDUAL DECISIONMAKING	73
The Effects of Active-Duty Policy	73
Increasing the Promotion Probability	73
Increasing Intergrade Pay Spreads	74
Up-or-Out Policy	77
Minimum-Performance Standards	78
Intragrade Contingent Pay	79
The Effect of the Structure of Retired Pay	80
Vesting Provisions	82
The Pension Formula	84
Early-Retirement Provisions	86
Chapter Six	
ORGANIZATIONAL DECISIONMAKING: SETTING ACTIVE PAY, RETIRED PAY, AND PERSONNEL POLICY	89
Managing Personnel Flows	89
Setting Active Pay	92
Intergrade Pay-Setting	92
Setting Intragrade and Intergrade Pay	97

Setting Personnel Policy	98
Up-or-Out Promotion Policy	98
The Roles of Retirement Pay	100
Structuring the Retirement System	103
Setting the Vesting Date	103
Setting the (Earliest) Retirement Date and (Early) Retirement Benefits	104
Chapter Seven	
A PRELIMINARY EVALUATION OF THE CURRENT MILITARY PAY AND RETIREMENT SYSTEM	107
The Active-Duty Pay Table	107
Pay Table Skewness	107
Service- and/or Occupation-Specific Pay Tables	110
Automatic Versus Contingent Within-Grade Increases	111
The Retirement System	112
Chapter Eight	
CONCLUSIONS	115
The Model: Individual Decisionmaking	115
Organizational Decisionmaking	116
Conclusions: Active Pay	116
Intergrade Pay Spreads	117
Intrgrade Pay	118
Up-or-Out Policies	118
Conclusions: Retired Pay	118
Initial Evaluation of the Military System	119
Pay Table	119
Retirement System	120
Appendix	
A. ILLUSTRATING THE DERIVATION OF THE ORGANIZATIONAL STRUCTURE	121
B. SETTING PAY TO PREVENT CLIMBING AND SLUMMING	123
C. SIMULATING CONTESTS	125
D. DERIVING SECOND ORDER CONDITIONS AND THE EFFECT OF POLICY CHANGES ON EFFORT SUPPLY	131
References	135

FIGURES

2.1. Continuation Rates for Unrestricted Line Officers, Fiscal Year 1990	11
2.2. Continuation Rates for Enlisted Personnel, Fiscal Year 1990	17
3.1. Personnel Structure of the Hypothetical Organization	40
4.1. Probability of Staying	64
4.2. Equilibrium Work Effort	70
6.1. Ability, Productivity, and Pay in the Hypothetical Organization	95
6.2. Hypothetical Pay and Productivity Profile over an Individual's Career	96
7.1. Annual Basic-Pay Increase (Average YOS at Promotion)	108

TABLES

2.1. Years of Service at Promotion and Selection Rates into Officer Pay Grades: DOPMA and Actual Service Statistics, Fiscal Year 1990	8
2.2. Grade by Years-of-Service Distribution: Officer Forces, Fiscal Year 1990	10
2.3. Mean Years of Service at Promotion and Current High-Year-of-Tenure Rules by Service: Enlisted Personnel, Fiscal Year 1990	12
2.4. Grade by Year-of-Service Distribution: Enlisted Forces, Fiscal Year 1990	15
2.5. Years-of-Service Distribution and Percentage Distribution by DoD Occupation Group: Enlisted Forces, Fiscal Year 1990	22
3.1. Probability of Each Stay-Leave Sequence	41
7.1. Mean AFQT of Army Enlisted Personnel by YOS and Grade, FY 1974 Cohort	110
A.1. Survival and Promotion Probabilities	121
A.2. Steady-State Force Based on Probabilities in Table 3.1	122
C.1. Simulation of a 10-Person Homogeneous Contest . . .	126
C.2. Simulations of Homogeneous Contests	127
C.3. Simulations of Heterogeneous Contests of Various Sizes	128

SUMMARY

How to efficiently and fairly compensate military personnel has been a topic of ongoing concern for policymakers. This concern has been punctuated over the past 20 years with the advent of the All-Volunteer Force, and by the dramatic technological improvements that have increased both the overall skill level required among military personnel and the diversity of manpower requirements. More recently, the force drawdown has highlighted the question of whether compensation and force-management policies that were suitable for maintaining a 2.1-million-person force will be equally suitable in the postdrawdown period. Despite these concerns, the current compensation system has changed little since the draft ended almost 20 years ago. In fact, a perusal of military compensation policy reveals little basic change since World War II. Some characteristics seem to reflect historical factors (e.g., past recruiting and retention problems, as well as how personnel were located around the world) rather than the needs and circumstances of the post-cold war national security environment.

A number of past study groups and commissions have analyzed the military compensation system. Some have recommended sweeping changes, particularly to the retirement system. Despite the well-intentioned analyses contained in their reports, it is apparent that they focus narrowly on the relationship between compensation and retention, and the resulting experience distribution of the military forces. Less attention has been paid to understanding other consequences of the system. In particular, (1) Does the system induce the most able personnel to stay and seek advancement to the highest

military ranks? and (2) Does the system motivate and encourage effective work?

The purpose of this study is to develop a unified model that will permit an analysis of all the various issues relating to the military compensation system. To do so, we marry recent advances in the modeling of military compensation and retention with the emerging economic literature on compensation and incentives in large, hierarchical organizations. The latter literature examines how large organizations use compensation and other personnel policies to motivate work effort and induce the proper *ability-sorting* (i.e., the motivation of high-quality personnel to stay and seek higher ranks) within the organization. Besides permitting us to address the traditional macroeconomic issue of the force-structure and -size implications of alternative military pay and personnel policies, our model permits us to examine the microeconomic issues of effort supply and ability-sorting that heretofore have been ignored.

The report has eight chapters, six of which define, develop, and apply the model. The first analytical chapter, Chapter Two, provides an overview of the military compensation and personnel systems and reviews past studies of the compensation system, focusing on recommendations regarding the military retirement system, a topic of constant debate.

The second, Chapter Three, discusses the key feature of the military personnel system: It is a hierarchical organization with little lateral entry, a feature that is a major constraint on the design of compensation and personnel policies.

The third, Chapter Four, begins the analysis by modeling the behavior of individual workers in a hierarchical organization. Individual decisions include whether to stay or leave (*retention*) and how hard to work (*effort supply*) when the organization cannot perfectly monitor individuals' efforts. Propositions about how effort and retention are affected by a number of pay and personnel policies are derived in Chapter Five, which also examines how the various policies affect the flows of personnel through the organization and the characteristics of those personnel (e.g., their abilities and tastes for the organization).

The fifth, Chapter Six, focuses on organizational decisionmaking. One of the most important goals of a hierarchy (the military) is to induce the best personnel to stay and seek advancement to the higher ranks. Conversely, it must discourage the less able from seeking the higher ranks (i.e., prevent "climbing") and it must identify and separate those unqualified to hold any rank. Because lateral entry is prohibited, the military can obtain its senior leaders only from its inventory of younger personnel. To provide promotion opportunities for the younger personnel, therefore, it must maintain personnel flows by separating senior personnel at some point. This chapter uses the model developed in Chapter Four to analyze efficient policies with respect to various policy levers at the organization's disposal: pay levels and pay spreads (intergrade and intragrade), up-or-out rules and involuntary separation, and retired pay.

This chapter reaches four conclusions. First, in a hierarchical system, pay spreads need to rise with rank to provide personnel with continuing incentives to work hard and seek promotion, and to induce the most able personnel to stay. Second, intragrade pay should to some extent be contingent upon performance and not be provided lockstep with seniority. Third, up-or-out rules are necessary to induce the separation of unpromotable personnel when pay is set administratively and one-on-one bargaining is costly. Fourth, although there is no unique theoretical role for retired pay, it may be offered for a number of reasons: to provide old-age insurance, encourage effort and retention of nonvested personnel, encourage the voluntary separation of senior personnel, and reduce *ex post* regret arising from earnings losses individuals may suffer when transitioning to other employment rather than full retirement.

Chapter Seven begins to evaluate the current military compensation system in light of the theoretical model. The current active-duty pay table is very "flat"; that is, the pay system is not skewed to any great extent: Compared with private-sector pay, entry pay is fairly high, but the rank differentials are not large and they do not increase much with rank, certainly not enough to offset declining probabilities of promotion. Intragrade pay rises with length of service (YOS, years of service), and there are a number of instances in which lower-ranking personnel with more longevity are paid more than higher-ranking personnel with less. Furthermore, the intragrade increases are automatic and not based on performance. All in all, the active-duty pay

system appears more aimed at attracting and retaining personnel than at providing them with effective incentives to work hard and seek advancement.

There is a theoretical rationale for the relatively flat pay structure. Because lateral entry is prohibited, personnel must be hired not just for what they can do today but for what they can potentially do in the future. To attract an entry pool of persons with enough ability to fill upper ranks in the future, pay must be set higher than would be the case if entrants were going to always fill low-level positions. As a result, entrants are initially paid more than their productivity. Eventually, as entrants progress through the ranks, their productivity will in fact exceed their pay in the higher levels. (But, because of the specificity of military skills, as workers age, they will be paid more than the wage they can command at older ages in the private sector.) Paradoxically, even though their productivity exceeds their pay once they reach the upper ranks, high-ranking, senior personnel must be induced to leave at some point to make room for younger personnel who also have high abilities and potential productivities. This is where retired pay comes in.

Military retirement benefits are available only to those who complete 20 or more years of service. These benefits induce voluntary separations. Furthermore, these benefits, which are sizable, effectively skew the pay system, because higher-ranking personnel earn larger benefits because they have more years of service and have been promoted to higher grades. They also provide retention and effort incentives for mid-career personnel, especially when stringent up-or-out rules make qualification for retirement benefits contingent upon performance, not automatic. (Because of the necessity of coping with the drawdown, the services have recently begun to apply tighter retention standards and more stringent up-or-out rules to enlisted personnel.)

Despite its virtues, the military retirement system poses four difficulties. First, it creates an implicit-contract problem: The services appear to "demand" large numbers of mid-career personnel because the personnel are there and will not quit, and separating them prior to YOS 20 would be viewed as unfairly breaking an implicit-contract. The services are therefore constrained to retain personnel who would not be retained were the terms of separation different.

Second, since the reward for an intermediate-length career is low, personnel must decide early on whether they want to be long-term careerists or leave. Some personnel who might have stayed longer under an alternative system leave very early. Third, the mid-career "bulge" slows down promotion opportunities for younger personnel and blunts the rewards to "fast-trackers," i.e., high-ability people who should move up more quickly. Finally, although the system effectively skews the pay system for younger personnel who are still trying to advance, it reduces skewness for those who are vested and may thereby diminish their effort and advancement incentives.

ACKNOWLEDGMENTS

This study was made possible by the foresight of Mr. Christopher Jehn, former Assistant Secretary of Defense for Force Management and Personnel, who recognized the need for a general theory of compensation and personnel policy. In the Office of the Assistant Secretary of Defense for Personnel and Readiness, Curtis Gilroy and Nicholas Timenes have provided continuous support and encouragement for this project. Thanks go also to our project officer, Saul Pleeter, and to John Enns.

Professor Sherwin Rosen of the University of Chicago and LTC Thomas Daula of the United States Military Academy provided painstaking reviews that materially improved the report. Professor Walter Oi of the University of Rochester commented on an early draft. At RAND, Glenn Gotz, Jim Hosek, and Bernie Rostker provided many helpful comments and useful discussions. We are also grateful to Peter Greenston at the United States Army Research Institute for providing data on Army enlisted personnel.

LIST OF SYMBOLS

A	Per capita accrual cost of retirement
B^*	The present value of compensation that the organization expects to pay to a new hire
$b_{i,t}$	Bonus in grade i at time t
$C_{i,t}$	Expected future civilian earnings stream
$c_{i,t,j}^*$	Expected earnings based on the individual's observable characteristics, in period j in years of service t in grade i
ci,t,j	Expected earnings in alternative employment in period j for someone leaving, in years of service t in grade i
D	Period of death
E_{it}^w	Performance threshold for grade i in period t
$E_{\min(it)}$	Minimum performance criterion for grade i in period t
E_{it}	Individual's score on evaluation administered in grade i at the end of period t
e_{it}	Work effort by individual in grade i and period t
$G_{i,t}$	The gain to staying for someone in grade i at the end of period t ($G_{i,t} = S_{i,t} - L_{i,t}$)
$G_{i,t}^*$	Gain to staying, given tastes and ability, but not conditional on $\varepsilon_{i,t}$ ($G_{i,t} = G_{i,t}^* + \varepsilon_{i,t}$)
I	Grade
k	Parameter of E_{it}
l	Loss

$L_{i,t}$	Return to leaving for someone in grade i at the end of period t
M^*	Expected present value at hire of future active-duty payments
$m_{i,t}^n$	Nonpecuniary factors, such as in-kind benefits, in grade i and period t
$m_{i,t}^p$	Wage paid in grade i and period t
$m_{i,t}$	Current pay (including in-kind benefits) the organization establishes for an individual in grade i at period t ($m_{i,t} = m_{i,t}^n + m_{i,t}^p$)
N_i	Organizational personnel requirements for grade i
n	Number of evaluations (or those vying for promotion) in grade i at time t
Pr	Probability
Q	Organizational output
$q(i, t, \alpha, e)$	Productivity of an individual in grade i with experience level t , ability level α , and effort e
R^*	Expected present value of the new hire's retirement liability
$R_{i,t}$	Future retirement stream for someone retiring in grade i at years of service t
r	Firm's discount rate
$r_{i,t,j}$	Retired pay received in period j by someone who retires in grade i and year of service t
r_t	Number of vacancies in the next-highest grade at time t
$S_{i,t}$	Return to staying for someone in grade i at the end of period t
$s_{i,t}$	The fraction of an initial cohort who survive to the end of period $t - 1$ who also survive until the end of period t (<i>conditional retention rate</i>)
sv_t	Survival rate into period t (i.e., the fraction of those surviving until the end of period $t - 1$ who also stay for period t)
T	Time or period in grade
$V_{i,t}^*$	Expected lifetime utility in grade i at the end of period t

w	Personnel
$Z(e_{it})$	Monetary value of the disutility individuals place on e_{it} units of work effort at period t [$Z(e_{it}) = \theta_1 (e_{it}^{\theta_2}) / (\theta_2 - 1)$]
α	Ability level
β_α	Proportionate effect of a unit change in ability on alternative earnings
β	The discount factor on future income [= $1/(1 + \delta)$]
δ	Individual's personal discount rate
$\varepsilon_{i,t}$	The random error in grade i and years of service t
$\Phi(G_{i,t}^* / \sigma_\varepsilon)$	Standard normal distribution function evaluated at the standardized gain to staying $G_{i,t}^* / \sigma_\varepsilon$ (same as $\Phi_{i,t}$)
$\Gamma_{i,t}$	Nonpecuniary benefit offered by alternative employers
ρ	Parameter of correlation between tastes and ability
$\pi_{i,t}$	The probability of promotion from grade $i - 1$ to grade i during period t
π_t^*	The aggregate promotion rate at t ($\pi_t^* = r/n$)
θ_i	Parameter i of the disutility of effort function $Z(e_{it})$
Σ	Covariance matrix
σ_ε	Standard deviation of $\varepsilon_{i,t}$
τ^c	Permanent taste factors associated with moving to an alternative employer
τ^m	Permanent taste factors associated with remaining with the current employer
τ	Net preference for the current employer ($\tau = \tau^m - \tau^c$)
ξ_{it}	A random factor determining the individual's evaluation score
$\psi_{i,t}$	Probability that the individual will surpass $E_{i,t}^w$

ACRONYMS AND ABBREVIATIONS

AFQT	Armed Forces Qualification Test
AVF	All-Volunteer Force
COLA	Cost-of-living adjustment
CPI	Consumer price index
DMC	Defense Manpower Commission
DoD	Department of Defense
DOPMA	Defense Officer Personnel Management Act
ERISA	Employee Retirement Income Security Act
ETS	Expiration of term of service
FY	Fiscal year
GAO	General Accounting Office (U.S.)
GPO	Government Printing Office (U.S.)
HYT	High year of tenure
IAC	Interagency Committee on Uniformed Services Retirement and Survivor Benefits
MDE	Marginal disutility of effort
MRE	Marginal return to a unit of work effort
MRRA	Military Retirement Reform Act
PCMC	President's Commission on Military Compensation
POF	Program objective forces
QMP	Qualitative Management Program (U.S. Army)
QRMC	Quadrennial Review of Military Compensation
REDUX	Name for MRRA
RMA	Retirement Modernization Act
USRBA	Uniformed Services Retirement Benefits Act
YOS	Years of service

The question of how to efficiently and fairly compensate military personnel is an ongoing concern for policymakers. This concern has been punctuated over the past 20 years by the end of the draft, the advent of the All-Volunteer Force (AVF), and the dramatic technological improvements that have increased both the overall skill level required among military personnel and the diversity of manpower requirements. More recently, the force drawdown has highlighted the question of whether the compensation and force-management policies that were suitable for maintaining a 2.1-million-person force will be equally suitable in the postdrawdown period.

Despite these concerns, the current compensation system has changed little since the draft ended almost 20 years ago. In fact, a perusal of the history of military compensation policy reveals little basic change since World War II.¹ Some characteristics seem to reflect historical factors (e.g., past recruiting and retention problems, as well as how personnel were located around the world) rather than the needs and circumstances of the post-cold war national security environment. Other characteristics, such as the use of the same pay table by all the services despite the differing manpower needs of each and a heavy reliance on retired pay,² appear to reflect equity considerations and a paternalistic philosophy in compensation design.

¹The Department of Defense (DoD) publication *Military Compensation Background Papers* (1992a) provides interesting discussions of the historical development of the various elements of military compensation.

²The retirement system offers an annuity to those who complete 20 years of service, but nothing to those who leave prior to completion of 20 years. The cost of funding for

Several past study groups and commissions have examined the military compensation system, and some have recommended sweeping changes, particularly to the retirement system.³ However, these past efforts generally limit their focus to the relationship between compensation and retention and, therefore, the experience structure of the armed forces. Except for recent exploratory work by Rosen (1992) and Asch (1993), little attention has been paid to understanding the relationship between compensation and motivation and, therefore, productivity. By *motivation*, we mean the incentive of personnel not only to train and learn skills, but also to work hard and effectively and to seek the grades and positions for which they are best suited from the service perspective. Since current models are incomplete as tools for evaluating compensation and personnel policies, whether current or alternative policies will best enable the Department of Defense (DoD) to meet its national security goals remains an open question.

The purpose of this report is to provide a unified framework of individual decisionmaking in the context of large, hierarchical organizations. The framework is intended to permit an analysis of the interrelationships between compensation and personnel policy, retention and force structure, and productivity incentives. In companion pieces (Asch and Warner, 1993, and forthcoming), we use this model to analyze the efficiency of the current military retirement system. To build this framework, we marry recent advances in the modeling of compensation and retention with the emerging literature on compensation and incentives in hierarchical organizations.⁴ The latter literature examines how organizations use compensation and other

the future retirement liabilities of current active-duty personnel is about one-third of their basic-pay costs.

³Some of the most notable commissions and study groups to have examined the retirement system include the First Quadrennial Review of Military Compensation (DoD, 1969) and the Fifth Quadrennial Review (DoD, 1983), the Defense Manpower Commission (1976), and the *Report of the President's Commission on Military Compensation* (1978). The latter two studies recommend fundamentally different retirement systems from the one that now prevails.

⁴Examples of the former work include Gotz and McCall (1984), Warner and Goldberg (1984), Black, Moffitt, and Warner (1990), and Gotz (1990). Pathbreaking studies in the latter literature include Lazear and Rosen (1981), O'Keeffe, Viscusi, and Zeckhauser (1984), and Nalebluff and Stiglitz (1983). McLaughlin (1988) provides a cogent review of this literature.

personnel policies to motivate work effort and to induce the proper *ability-sorting* (i.e., the motivation of high-quality personnel to stay and seek higher ranks) across jobs and organizational tiers. Besides enabling us to address the traditional macroeconomic issue of how military compensation and personnel policy affect the size and structure of the armed forces, our model permits us to examine microeconomic issues that have heretofore been ignored in both the military manpower literature and principal-agent literature. For example, our framework introduces not only the general role of compensation in encouraging effort supply, ability-sorting, and retention, but also the role of compensation structure in the pattern of and relationships between interlevel (or intergrade) pay, intragrade pay, and retired pay. In the process of incorporating ability-sorting and effort-supply decisions into a model of retention and compensation, we address a number of specific questions that have faced current and past study groups regarding the efficient design of military compensation, including: How should basic pay be structured across and within grades? and What should be the size and structure of retirement benefits relative to the size and structure of basic pay?

The report has eight chapters, six of which define, develop, and apply our new compensation model. Chapter Two gives an overview of the current military compensation system and personnel management. It also discusses the main criticisms of the military pay and retirement system and reviews the major studies aimed at reforming military compensation. Chapter Three formally defines the hierarchical structure of our hypothetical organization. In this chapter, we also delineate our assumptions regarding the production relationships within the hierarchy and define the organization's budget constraint. In Chapter Four, we model individual worker decisions regarding retention and effort supply and identify a number of personnel and pay and retirement policies that affect them. Chapter Five formally examines the effect of the policies identified in Chapter Four on individual retention and effort decisions. In Chapter Six, we derive implications for organizational decisionmaking and examine how pay and retirement policy should be set. In Chapter Seven, we begin to apply our general analysis of compensation and personnel policy in hierarchical organizations to the military setting. Although we do not draw specific conclusions about the efficiency of the military's

current system in this chapter, we discuss potential problems with it in light of our model. In the final chapter, Chapter Eight, we summarize our findings.

THE MILITARY COMPENSATION AND PERSONNEL SYSTEMS: DESCRIPTION AND ISSUES

Even after the drawdown is accomplished, the military will remain the largest single employer in the United States. As in the past, the military will continue to recruit and train personnel to perform a wide range of tasks and to require a complex system of compensation and personnel policies to manage its labor force. This chapter briefly describes current military compensation and personnel practices, presents basic statistics on system outcomes, then discusses the criticisms that various analysts and study groups have leveled against those practices. Later chapters evaluate those criticisms in light of the model developed in Chapters Three through Six.

THE COMPENSATION SYSTEM

A perusal of the Department of Defense publication *Military Compensation Background Papers: Compensation Elements and Related Manpower Cost Items* (1992a) reveals a military compensation system made up of a complex patchwork of active-duty pays and allowances, and retired pay. The large number of active-duty pays and allowances can be grouped into three categories. The main category is *basic pay*, which is the same for everyone with the same rank and years of service (YOS). In fiscal year (FY) 1991, the basic-pay costs of the active force were \$37.4 billion and made up 75 percent of active-duty cash compensation. The second-largest category, *allowances for housing and food*, was \$8 billion and accounted for 19 percent of FY 1991 cash compensation. The third category, *special and incentive pays*, contains the many elements of compensation that depend on individual circumstances, such as occupation (e.g.,

enlistment and reenlistment bonuses) and assignment (e.g., sea pay, flight pay, and submarine pay). Although the number of special and incentive pays is large, in FY 1991 these pays were \$2.9 billion, only about 6 percent of active-duty compensation.

After basic pay, *retired pay* is the major element of compensation. In FY 1991, DoD's retired-pay accrual charge was \$16 billion, 43 percent as large as outlays for basic pay.¹ The distinguishing feature of the military retirement system is that it requires 20 years of service before personnel are entitled to receive any benefits. Unless they join the reserves and accumulate enough reserve credits to begin receiving reserve retired pay at age 60, personnel who separate prior to 20 YOS receive no retirement benefits. In contrast, private-sector plans, which must conform with the 1986 modifications to the Employee Retirement Income Security Act (ERISA), must generally vest workers after five years or, if they use a graduated-vesting method, after seven years.

Three retirement systems are currently in effect. Pre-FY 1981 entrants receive retired pay according to the formula $(0.025 * YOS * \text{final basic pay})$, so that 20-year retirees receive 50 percent of final basic pay and 30-year retirees receive 75 percent. Importantly, retired pay for this group is fully inflation-protected. Retired pay for those who entered between FY 1981 and FY 1986 is calculated similarly, except that pay is based on the individual's high three years' *average* basic pay rather than *final* basic pay. It is also fully indexed for inflation.

The Military Retirement Reform Act (MRRA) of 1986, also known as REDUX, implemented several important changes. First, the annuity formula was changed to $\{[0.40 + 0.035 * (YOS - 20)] * \text{high-3 average basic pay}\}$ for the years between separation and age 62, at which time pay reverts to $(0.025 * YOS * \text{high-3 average basic pay})$. Consequently, retired pay during the transition between military service and full retirement ranges between 40 percent of high three years' average basic pay at YOS 20 and 75 percent of high three years' basic pay at YOS 30. Second, rather than indexing retired pay for inflation, the annual cost-of-living adjustment (COLA) between separation and

¹Since 1984, the retirement system has been funded on an accrual basis. Each year, the DoD Actuary determines how much DoD must set aside to fund the accumulating retirement liability of those on active duty.

age 62 is 1 percent less than the percentage growth in the consumer price index (CPI). At age 62, retired pay is then fully adjusted for the CPI growth since separation; thereafter, it again increases according to the CPI-minus-1-percent rule. The 1986 reforms thus changed the system by (1) reducing the amount received at YOS 20, (2) raising the growth in retired pay for each year served after YOS 20, and (3) reducing the real value of the stream of retired pay in an inflationary environment.

A final salient feature of the military compensation system is that much pay is received *in kind*. Health care benefits for service members and their families are a major component of this source of compensation, as are commissary privileges.

PERSONNEL MANAGEMENT: OFFICERS

The management of officer personnel is governed by the Defense Officer Personnel Management Act (DOPMA). DOPMA sets limits on the permissible officer-grade distribution and specifies promotion points or zones, promotion opportunities, high-year-of-tenure rules,² and procedures for the mandatory separation of officers. Officer ranks range from O-1 to O-10. Officers normally enter military service in grade O-1 (ensign in the Navy; second lieutenant in the other services).

Promotion from O-1 to O-2 (lieutenant junior grade in the Navy; lieutenant in the other services) and from O-2 to O-3 occurs early in the career (Table 2.1). If qualified, all O-1s are promoted to O-2 in the first or second year of service. Almost all O-2s are promoted to O-3 around the fourth year of service. Officers then spend between five and seven years before promotion to O-4 (lieutenant commander in the Navy; major in the other services). Beyond the grade of O-3, promotion becomes more difficult. DOPMA specifies a target promotion rate to O-4 of 80 percent, i.e., 80 percent of those in O-3 are permitted to be promoted to O-4 (Table 2.1). Actual rates vary from year to year. In FY 1990, the selection rate to O-4 ranged be-

²A *high-year-of-tenure rule*—also known as an *up-or-out rule*—stipulates that an individual who has reached a prespecified year of tenure within a pay grade without earning a promotion to the next higher grade must be separated.

Table 2.1

Years of Service (YOS) at Promotion and Selection Rates into Officer Pay Grades: DOPMA and Actual Service Statistics, Fiscal Year 1990

Source	Criterion	Officer Grade				
		O-3	O-4	O-5	O-6	O-7+
DOPMA	YOS at promotion	3-5	9-11	15-17	21-23	NA
	Selection rate (%)	100	80	70	50	NA
	High year of tenure	~13	20	28	30	
Army	YOS at promotion	4	11	17	23	27
	Selection rate (%)	95.8	78.6	73.1	46.3	NA
Navy	YOS at promotion	4	10	15	22	27
	Selection rate (%)	96.5	85.7	71.8	55.6	NA
Air Force	YOS at promotion	4	11	16	22	25
	Selection rate (%)	100	89.3	73.5	55.2	NA
Marine Corps	YOS at promotion	5	12	17	22	27
	Selection rate (%)	89.0	71.1	65.3	51.2	NA

DATA SOURCE: Officer and Enlisted Personnel Management, Office of the Assistant Secretary of Defense (Personnel and Readiness), The Pentagon, Washington, D.C.

NOTE: NA—Data not available.

tween 71.1 percent in the Marine Corps and 89.3 percent in the Air Force.

DOPMA specifies a target promotion year to O-5 (commander in the Navy; lieutenant colonel in the other services) and a target selection rate of 70 percent. In FY 1990, the promotion timing and selection rates to O-5 were close to these targets. Officers spend about five years in the grade of O-5, with promotion to O-6 (captain in the Navy; colonel in the other services) occurring around the twenty-first or twenty-second year. The DOPMA selection rate to O-6 is 50 percent. In FY 1990, the services were close to this target. Officers again spend about five years in grade O-6 before being promoted to the rank of general officer (O-7 and above). Although exact statistics on the selection rate to O-7 and above are not published, the selection rate to O-7 and above is less than 10 percent.

Table 2.1 also shows the high-year-of-tenure (HYT) points for officer personnel as specified by DOPMA. The HYT point for O-3s is approximate. Officers are promoted by year group, and O-3s are "in the zone" at about their eleventh year of service. An O-3 is mandatorily separated if he or she fails both promotion in the zone

and a second "above-the-zone" chance at promotion. Those O-4s who have not been promoted to O-5 are separated at the 20-year point, when they become eligible for retirement benefits. The notional HYT for O-5s is 28, but DOPMA permits the mandatory separation prior to YOS 28 of 30 percent of O-5s who have twice been passed over for promotion to O-6. Likewise, YOS 30 is the notional high year of tenure for O-6s, but DOPMA permits the separation of 30 percent of O-6s after they have served four years in grade. There is no formal high year of tenure for general officers.

Table 2.2 shows the grade by YOS distribution of the officer forces in FY 1990. About two-thirds of the officers are in the junior ranks (O-1-O-3), and a similar proportion has less than 12 years of service. About one-fourth of the officers have between 12 and 20 years of service. Less than 10 percent of the officer forces have more than 20 years of service. The Air Force has a slightly richer grade and experience mix than the other services, i.e., a greater fraction of the force is in upper grades and YOS.

Figure 2.1 illustrates the retention behavior of military officers by plotting the FY 1990 continuation rates of unrestricted line officers by years of service.³ (The *continuation rate* is the fraction of officers in service at the start of the year who were also in service at the end of the previous year.) The figure indicates very high officer continuation rates during the first four years of service, during which time officers are serving their initial obligations and are therefore not eligible to leave. Officer continuation rates fall around the fourth or fifth year of service as initial obligations expire and some choose to leave. Because of their generally longer initial obligations, Air Force officers have higher continuation rates than officers in other services until about the seventh year of service. Continuation rates rise until about the eleventh year of service. At that point, continuation rates fall as those who are not selected for promotion to O-4 leave. Officers then have almost 100 percent continuation to YOS 20 and

³ *Unrestricted line officers* make up the majority of the services' commissioned officers. Excluded from this category are doctors, nurses, and other categories of restricted duty.

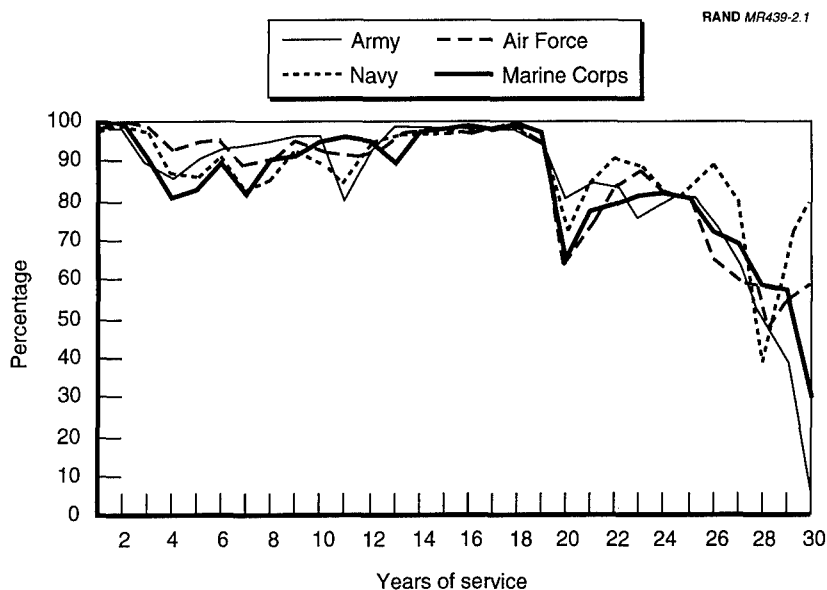
Table 2.2
Grade by Years-of-Service (YOS) Distribution:
Officer Forces, Fiscal Year 1990
(in percent)

Service	YOS	Officer Grade					TOTAL
		O-1-O-3	O-4	O-5	O-6	O-7+	
Army	1-5	35.6	0.0	0.0	0.0	0.0	35.6
	6-11	29.3	0.2	0.0	0.0	0.0	29.5
	12-20	3.0	16.9	5.2	0.0	0.0	25.1
	21-30	0.0	0.0	6.0	3.7	0.1	9.8
	TOTAL	67.9	17.1	11.2	3.7	0.1	
Navy	1-5	45.7	0.0	0.0	0.0	0.0	45.7
	6-11	23.8	2.9	0.0	0.0	0.0	26.7
	12-20	0.0	12.5	7.0	0.0	0.0	19.5
	21-30	0.0	0.1	2.7	4.3	0.2	7.3
	TOTAL	69.5	15.5	9.7	4.3	0.2	
Air Force	1-5	30.4	0.0	0.0	0.0	0.0	30.4
	6-11	33.4	1.0	0.0	0.0	0.0	34.4
	12-20	0.9	16.4	8.0	0.0	0.0	25.3
	21-30	0.0	0.2	4.1	4.7	0.2	9.2
	TOTAL	64.7	17.6	12.1	4.7	0.2	
Marine Corps	1-5	37.5	0.0	0.0	0.0	0.0	37.5
	6-11	26.3	0.2	0.0	0.0	0.0	26.5
	12-20	5.1	17.4	5.6	0.0	0.0	28.1
	21-30	0.0	0.0	3.6	3.7	0.2	7.5
	TOTAL	68.9	17.6	9.2	3.7	0.2	

DATA SOURCE: Defense Manpower Data Center, Monterey, Calif.

eligibility for retirement. After that point, retention declines sharply as a result of the voluntary incentive to leave provided by the retirement system and of the operation of the promotion system in conjunction with the application of HYT rules.

Based on the FY 1990 continuation rates in Figure 2.1, the cumulative portion of an entry cohort of unrestricted line officers that is expect-



SOURCE: Defense Manpower Data Center, Monterey, Calif.

Figure 2.1—Continuation Rates for Unrestricted Line Officers, Fiscal Year 1990

ed to stay for a 20-year career is about 40 percent in the Air Force, 35 percent in the Army, and 25 percent in the Navy and Marine Corps.⁴

⁴Data from the DoD Actuary indicate that the aggregate cumulative survival of all U.S. military officers to YOS 20 is 65 percent, which is much higher than we have calculated from the FY 1990 data, for two reasons. First, the actuary's continuation rates include gains from those in the enlisted force who become officers, usually in the first several years of service. Including these gains would raise by several percentage points the survival probability to YOS 20 calculated from FY 1990 data. Second, most of the difference is due to the fact that the actuary's continuation rates, which are based on a 5-year average of aggregate officer continuation rates from the mid-1980s, are much higher than the FY 1990 rates for unrestricted line officers, especially in the YOS 4-8 range.

PERSONNEL MANAGEMENT: ENLISTED PERSONNEL

There are nine enlisted pay grades. Enlisted personnel enter in grade E-1 and advance to the higher grades according to time-in-grade and time-in-service regulations, performance on advancement exams, and, importantly, vacancies in the higher grades created by turnover. Promotion to grade E-2 comes automatically after completion of initial skills training. Promotion to E-3 usually occurs around the completion of one year of service and, again, promotion is automatic. Table 2.3 shows the average years of service at promotion to grades E-4 through E-9. The effect of turnover on promotion timing shows up clearly here: Promotions in the Air Force, which, as we show below, has higher retention and lower turnover than the other services, generally lag behind promotions in the other services.

Current high-year-of-tenure points for enlisted personnel also are shown in Table 2.3. Having HYT points for senior enlisted personnel

Table 2.3

Mean Years of Service (YOS) at Promotion and Current High-Year-of-Tenure Rules by Service: Enlisted Personnel, Fiscal Year 1990

Service	Criterion	Enlisted Grade					
		E-4	E-5	E-6	E-7	E-8	E-9
Army	YOS at promotion	2.0	4.7	8.2	13.4	17.6	20.8
	High year of tenure	8.0	13-15	20.0	22.0	24.0	30.0
Navy	YOS at promotion	2.0	4.0	8.2	12.5	16.8	20.1
	High year of tenure	10.0	20.0	23.0	26.0	28.0	30.0
Air Force	YOS at promotion	2.8	6.8	12.1	15.4	18.7	21.7
	High year of tenure	10.0	20.0	20.0	24.0	26.0	30.0
Marine Corps	YOS at promotion	3.0	5.3	9.5	14.0	18.0	21.7
	High year of tenure	8.0	8-13	20.0	20-22	22-27	30.0

DATA SOURCES: Officer and Enlisted Personnel Management, Office of the Assistant Secretary of Defense (Personnel and Readiness), The Pentagon, Washington, D.C.; and Defense Manpower Data Center, Monterey, Calif.

who have reached retirement eligibility is not new, but the rules for E-4s and E-5s are.⁵ Prior to FY 1990, the services placed few constraints on the continuation of career enlisted personnel. But rising retention in the mid-1980s gave rise to concern that the enlisted experience mix was becoming too "rich" (i.e., too much of the force is in higher YOS) and the enlisted forces too costly. After vigorous debate within DoD, the services began imposing more stringent HYT rules on their enlisted forces around 1990. Now, enlisted personnel who have not advanced to E-5 by their eighth year must separate. The Army dropped the HYT for E-5s from 20 to the 13-15 range (effective October 1, 1993). Those E-5s in this YOS range who have not been promoted to E-6 by October 1, 1993, must separate. The Marine Corps' nominal HYT for E-5s is 13, but, in the future, Marine Corps separation policy will be based on failure to be promoted rather than YOS. Those Marines who have twice failed promotion to E-6 must separate, even if they have less than 13 YOS. Although the Army and Marine Corps have moved to more aggressively separate E-5s prior to YOS 20, the Navy and the Air Force HYT rules will continue to permit E-5s to remain until retirement eligibility.⁶

In addition to HYT policies, the services apply standards for the reenlistment of enlisted personnel. Some of the standards are explicit: Only those in grades E-4 and above are permitted to reenlist after the first enlistment. Standards relating to weight and physical condition are also imposed. At times, the services have also screened potential reenlistees on a number of implicit criteria. We were told, for example, that, beginning in the early 1980s, the Army began selectively denying reenlistment to personnel with low education levels and low Armed Forces Qualification Test (AFQT) scores, although it had no explicit written regulations for doing so. Until the drawdown in 1990, the authority to determine eligibility for reenlistment and to grant waivers to those who fail the formal criteria rested with indi-

⁵Not shown are HYT rules for those below grade E-4. Few enlisted personnel fail to reach the rank of E-4 prior to completion of an initial enlistment, but those who do fail to reach E-4 are ineligible to reenlist.

⁶We are told by sources in DoD that in adopting their HYT rules for E-5s, the Navy and the Air Force cited the need to keep in operations and maintenance billets experienced personnel who were not necessarily promotable to supervisory billets. On the other hand, a much higher percentage of Army and Marine Corps billets are in combat units, in which there is apparently little need for senior E-5s.

vidual unit commanders. As part of the drawdown, and to gain more control over retention, the services have recently implemented more centralized administration of reenlistment.

Unlike data for officers, promotion-opportunity data are not readily available for enlisted personnel. Officers are promoted in narrow YOS intervals based on their commissioning year, and it is easy to track the percentage that is promoted. But enlisted personnel are eligible for promotion over fairly wide time intervals, and statistics are not kept on the fraction entering a promotion zone that is eventually promoted. There have been some studies that estimate the cumulative promotion probability to various grades, and the Marine Corps publishes a target promotion percentage to each grade in its manpower plan. On the basis of these studies and the Marine Corps plan, it appears that the cumulative probability of promotion approaches unity up to grade E-5, then declines for promotion to the senior enlisted grades. For example, the Marine Corps plan currently calls for a cumulative promotion opportunity of 80 percent to E-6, 75 percent to E-7, 70 percent to E-8, and 65 percent to E-9.

The distribution of enlisted personnel by grade and YOS in FY 1990 is displayed in Table 2.4. The table reveals several facts about the services' requirements for enlisted personnel. The first is that most of the services' enlisted-grade requirements are in the mid-range E-4–E-6, the post-training journeyman operator and/or maintainer billets. Requirements for senior supervisory personnel (E-7–E-9) account for only about 10 percent of the total. The grade distribution is limited partly by Congress, which has said that no more than 1 percent can be in grade E-9, no more than 3 percent in grades E-8 and E-9, and no more than 50 percent in the top five grades (E-5–E-9).

Despite a big gain in the fraction of personnel in the services' career forces (those with more than four years of service) over the period of the AVF, the enlisted forces remain young. About 60 percent of the Marine Corps enlisted force and 50 percent of the Army and Navy enlisted forces are still in their first term of service (see Grissmer, Hosek, and Eisenman [1989] and Horowitz [1991] for analyses of the AVF trend in career manning). The Air Force is the only service with a very senior enlisted force, whereas personnel with more than 10 years of service account for less than one-quarter of the Army, Navy,

Table 2.4
Grade by Year-of-Service (YOS) Distribution: Enlisted Forces,
Fiscal Year 1990 (in percent)

Service	YOS	Enlisted Grade			
		E-1-E-3	E-4-E-6	E-7-E-9	TOTAL
Army	0-4	28.7	21.5	0.0	50.2
	5-10	0.7	25.6	0.0	26.3
	11-20	0.0	12.8	8.4	21.2
	21-30	0.0	0.0	2.3	2.3
	TOTAL	29.4	59.9	10.7	
Navy	0-4	32.7	17.2	0.0	49.9
	5-10	0.6	26.5	0.1	27.2
	11-20	0.0	13.2	7.1	20.3
	21-30	0.0	0.4	2.3	2.7
	TOTAL	33.3	57.3	9.5	
Air Force	0-4	23.5	12.2	0.0	35.7
	5-10	0.2	31.1	0.0	31.3
	11-20	0.0	21.0	6.8	27.8
	21-30	0.0	0.5	4.8	5.3
	TOTAL	23.7	64.8	11.6	
Marine Corps	0-4	47.4	11.9	0.0	59.3
	5-10	1.2	22.2	0.0	23.4
	11-20	0.0	8.9	6.4	15.3
	21-30	0.0	0.0	1.8	1.8
	TOTAL	48.6	43.0	8.2	

DATA SOURCE: Defense Manpower Data Center, Monterey, Calif.

and Marine Corps enlisted forces. Less than 5 percent of the enlisted forces have more than 20 years of service.

Figure 2.2 illustrates enlisted continuation behavior, again with data from FY 1990. The YOS profile of enlisted continuation rates is much like the officer profile shown in Figure 2.1. Enlisted continuation rates are high during the period of the initial enlistment (although not as high as officer rates). Rates drop in the YOS 3-6 zone as initial enlistment contracts expire and personnel leave.⁷ Rates then rise

⁷The continuation rate at a given YOS is composed of the retention of those who are at expiration of term of service (ETS) and are thus eligible to make a voluntary retention

uniformly to YOS 20, where they drop sharply from the influence of the retirement system. The Air Force generally has the highest continuation rates; the Marine Corps has the lowest. Based on the continuation rates in Figure 2.2, 10 percent of enlisted Army and Marine Corps entrants would be expected to stay for a 20-year career, and 13 percent of Navy entrants and 16 percent of Air Force entrants would be expected to do so. The DoD average is 12 percent. Enlisted continuation rates began declining in the late 1980s because of the above-described drawdown and service policies that have been aimed at reducing enlisted retention; therefore, these survival probabilities are lower than those calculated with data from earlier years.⁸

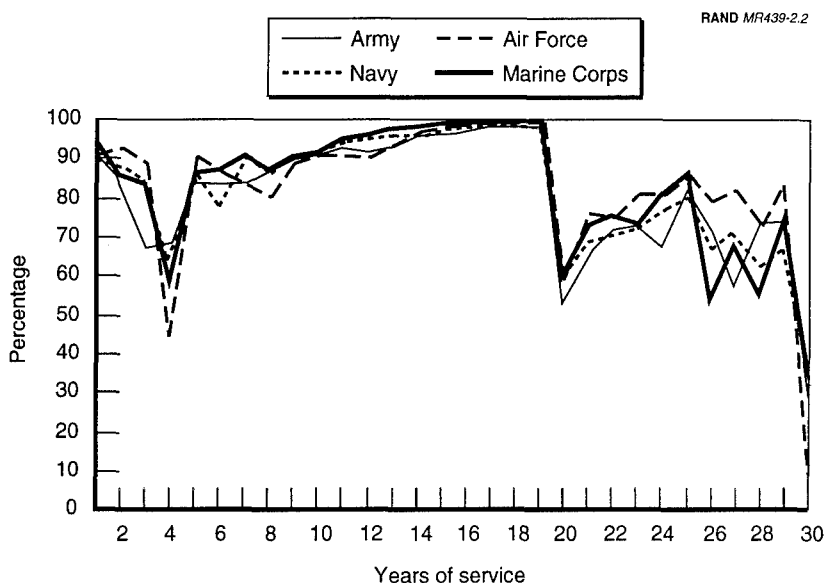
In addition to lower pre-YOS 20 survival, enlisted personnel have much lower post-YOS 20 retention than officers. In fact, we calculate that those enlisted personnel who reach YOS 20 serve only about one additional year on average before leaving. Officers who complete 20 years serve an average of about four additional years. These differences may be due partly to differences in the nature of the work and partly to differences in promotion and compensation policy: By the time they reach YOS 20, most enlisted personnel have been promoted to their terminal grades; proportionately more officers expect additional promotions beyond YOS 20. Furthermore, beyond YOS 20, longevity-pay increases for enlisted personnel are minimal. The structure of compensation seems to encourage the earlier departure of enlisted personnel.

ISSUES

Questions about both the level and the structure of pay in the military compensation system continually arise. The *level* of pay concerns how military personnel are paid relative to private-sector employees. The *structure* of pay concerns the setting of pay by grade, longevity, and occupation; the general allocation between active and

decision and those who have more than one year left on their enlistment contracts. The retention rate of those at ETS is much lower than the total continuation rate in a given YOS.

⁸Using a 5-year average of continuation rates from the mid-1980s, the DoD Actuary reports a DoD average enlisted survival rate to YOS 20 of 17 percent. It uses this 5-year average in its calculation of the accrual cost of the retirement system.



SOURCE: Defense Manpower Data Center, Monterey, Calif.

Figure 2.2—Continuation Rates for Enlisted Personnel, Fiscal Year 1990

retired pay; and the parameters of the retirement system. Regardless of the structure of pay, the pay level can always be adjusted sufficiently to attract and retain a force of a certain quality. This fact was demonstrated in the early 1980s, when two successive pay raises totaling 25 percent cured the recruiting and retention problems the services were experiencing at that time. Furthermore, it is clear that the level and structure of pay are intertwined: If pay is improperly structured, the level of pay may need to be set higher than would otherwise be required for DoD to accomplish its manpower goals.

The structure of military pay has been the focus of much past analysis. Most of the attention has been paid to the allocation between active and retired pay and the structure of the retirement system.⁹

⁹Studies of the compensation system mandated by the Executive Branch or the Congress begin with the 1947 "Joint Army-Navy Pay Board Study" (unpublished),

Analysis has focused on the following three traditional efficiency questions:

1. Does the system enable DoD to attract and retain personnel in sufficient numbers and of sufficient quality to accomplish its missions?
2. Does it promote an efficient distribution of personnel by grade and YOS?
3. Does it accomplish these objectives at least cost?

To these traditional macroeconomic questions we add two micro-economic questions that have received little attention: (1) Does the system motivate personnel to work hard and effectively? and (2) Does the system promote the best person/job/rank matches, i.e., does it sort people properly by rank and experience according to their individual skills and abilities?

Two factors should be kept in mind that, it is fair to say, have not been in most past analyses. The first is that the military compensation system must be designed to support a closed hierarchical personnel system with little lateral entry. Personnel enter the bottom ranks in large numbers, then progress through the ranks according to performance and other factors. Future senior leaders must be grown from these entry pools. In a triangular hierarchical system, promotion opportunities diminish as personnel advance through the system. Thus, the compensation system must be structured in a way that provides personnel with a continuing incentive to work hard and compete for promotion *and* reveal their true abilities and talents to personnel managers. As we detail in Chapters Four through Six, it is important for the military to select the most able personnel for promotion to the upper ranks. Consequently, the focus should not be just on retention per se, but on *who* is retained as well. Furthermore, given the triangular hierarchical personnel structures,

which was soon followed by the 1948 Advisory Commission on Service Pay (Hook Commission, *Career Compensation for the Uniformed Services*). More recent studies include those by the Seventh Quadrennial Review of Military Compensation conducted by DoD (1992b), the Defense Manpower Commission (1976), the President's Commission on Military Compensation (1978), and the Grace Commission (1985). In addition to officially mandated studies, the military compensation system has been the focus of numerous independent studies.

to get a return on their investments in training and experience, the services want personnel to stay long enough, but not too long: Turnover in the upper ranks must be enough to provide promotion opportunities for those in the lower ranks. Retention can be excessive, even if it is among very able personnel. Consequently, the compensation system must be structured not only to provide the proper retention and effort incentives, but also to provide the incentive for personnel to separate when it is in the services' best interest for them to do so. The system must be incentive-compatible *ex post* as well as *ex ante*. Many studies have stressed the *retention* aspect of compensation and ignored the *separation* aspect of compensation.

The second factor to keep in mind is that the different parts of the compensation system cannot be examined in isolation from one another. One must look at the system as a whole. Consequently, some previous analyses have recommended changes to the retirement system without considering how the active-duty pay system would need to be modified to compensate for changes in retired pay.

Having said all this, what are the issues? What criticisms have been leveled at the current military compensation system? In the following subsections, we first outline the criticisms that have been voiced against the military retirement system, then criticisms of the active-duty pay system. After that, we assess the criticisms and discuss factors that need to be considered in a more complete evaluation of the compensation system.

Criticisms of the 20-Year Retirement System

At one extreme, some critics have wondered why the military needs a retirement system at all under the AVF, and one that is more generous than the typical old-age annuity systems prevalent in the private sector. The argument critics offer runs as follows. Under the draft, the retirement system helped the military discriminate between one-termers and career personnel. But under the AVF, when more enlistees want to stay for a career, a large back-end (retirement) pay policy has outlived its usefulness because such discrimination is more difficult. Further, these critics charge that young people (i.e., military personnel) are known to have high personal discount rates, much higher than the government's, and therefore value a dollar of deferred (retirement) compensation less than it costs the government

to provide. Thus, the government could reduce retired pay and maintain retention incentives by increasing active pay by less than the (present value of) the savings to the government, and save money. Taken to its extreme, this argument says that there need be no retirement system at all: The most efficient compensation system is an active-pay-only system. Some advocates of this line of reasoning recognize, of course, that it would be politically infeasible to eliminate the system altogether and therefore recommend a less-generous system that conforms to ERISA guidelines for private-sector pension plans.

Critics further charge that the 20-year system constrains force planning and management in undesirable ways. Under existing procedures, the services determine how many personnel they need in each pay grade according to a variety of factors. They then obtain desired experience distributions of their forces, called program objective forces (POFs), by translating requirements by pay grade into a YOS distribution.¹⁰ But the POFs are largely based on the experience distribution that can be supported with the retention patterns produced by the current compensation system and not necessarily a system that is most efficient.¹¹

Not only does the 20-year system constrain force planning, but critics argue that it inhibits the management of the mid-career force. The services are well aware of the financial cost that is imposed on mid-careerists who are involuntarily separated. Consequently, beyond a certain grade or YOS, personnel are treated as if they implicitly have tenure and the services are reluctant to separate all but the poorest performers prior to YOS 20 for fear of what breaking the implicit tenure contract might do to morale. In the officer force, the implicit tenure point is promotion to O-4, around YOS 10. Prior to 1989, almost all enlisted personnel who reenlisted once were permit-

¹⁰DoD Directives 1304.20 and 1300.14 place restrictions on the pay-grade distribution and career content of the enlisted forces: (a) a maximum of 3 percent of personnel in grades E-8 and E-9, (b) the number in the top 5 and 6 grades and the number with more than 4 YOS shall not exceed the number in the DoD-approved POF, and (c) the ratio of the number in the top 5 grades shall not exceed the career content of the POF. A recent General Accounting Office (GAO) report (1991) finds that the Army and Air Force generally satisfy these restrictions, except the one on career content.

¹¹That the services' objectives are moving targets driven in part by actual retention patterns is documented in GAO (1991).

ted to stay for a 20-year career; very few screens were applied beyond the initial reenlistment. In fairness to the services, that practice has changed, and they are now applying more stringent high-year-of-tenure rules and tougher standards for reenlistment to lower- and mid-grade enlisted personnel. Furthermore, the drawdown has forced, and will continue to force, the services to involuntarily separate many mid-career personnel whom they would not have separated in normal times. Yet, despite these actions, the presence of the 20-year retirement system has arguably made management of the mid-career force generally more difficult and has accentuated the problems caused by the drawdown.¹²

Another related criticism is that the equality of treatment under a 20-year system constrains force structures in different occupations. Some occupations require youth and vigor, whereas others do not. The Army and Marine Corps need many young infantrymen and have few infantry-specific billets at more senior levels. Mid-level personnel in the combat skills who are not yet vested in the retirement system are assigned to other "infrastructure" jobs for the remainder of their military careers, similar to the management of Air Force pilots: A normal flying career for pilots may be ten or 12 years, after which the 20-year career is accommodated by moving those who stay to nonflying jobs. The creation of (possibly unnecessary) infrastructure billets to accommodate 20-year careers is another illustration of the implicit-contract problem.

But in some military occupations in which youth and vigor are not primary job requirements and in which there are high training costs and/or a big payoff to job experience (e.g., doctors and nurses), it might be more efficient to have careers span well beyond the 20-year mark. As Table 2.5 shows, the 20-year system produces similar retention patterns across the wide spectrum of occupations, thereby limiting the possibility of exploiting the different experience-

¹²A recent *Navy Times* editorial (November 18, 1991) stated that "the unfairness of this system generally escapes the notice of anyone other than service people who, after serving honorably for five, 10 or 15 years, leave the military with nothing but a handshake. The drawdown, however, is shedding more light on the inequity."

Table 2.5

**Years-of-Service (YOS) Distribution and Percentage Distribution by DoD
Occupation Group: Enlisted Forces, Fiscal Year 1990
(in percent)**

Service	1-Digit Occupation Group ^a	Years of Service				Percentage of Total
		0-4	5-10	11-20	21-30	
Army	0	49.2	23.1	25.7	2.1	29.1
	1	46.8	28.3	23.5	1.4	4.4
	2	50.5	29.3	19.2	1.0	13.0
	3	45.4	29.1	24.2	1.4	6.6
	4	39.7	30.3	28.6	1.5	2.7
	5	30.9	29.2	33.2	6.7	16.4
	6	47.2	33.1	18.6	1.1	14.8
	7	49.6	30.8	18.6	0.9	2.0
Navy	8	47.5	31.4	19.6	1.5	11.2
	0	58.5	22.0	17.3	2.3	10.6
	1	45.0	31.3	21.3	2.4	17.9
	2	41.7	32.6	23.4	2.5	11.4
	3	43.5	30.6	22.8	3.1	6.7
	4	32.7	29.3	32.8	5.3	1.0
	5	31.2	31.7	32.8	4.3	10.9
	6	39.1	33.2	25.1	2.6	29.4
Air Force	7	37.2	31.8	28.5	2.6	6.5
	8	41.6	26.3	26.3	5.8	5.7
	0	33.1	36.3	27.2	3.3	7.2
	1	24.5	33.4	35.2	6.9	12.8
	2	26.9	33.7	33.7	5.6	6.6
	3	38.6	34.1	23.4	3.8	6.7
	4	30.2	33.5	30.5	5.7	4.0
	5	22.3	33.4	36.2	8.1	23.3
Marine Corps	6	27.9	34.6	33.7	3.7	24.2
	7	31.8	29.5	32.6	6.0	5.6
	8	37.1	31.5	27.1	4.4	9.5
	0	68.5	21.1	9.5	0.8	28.9
	1	38.8	33.1	25.4	2.7	7.6
	2	52.6	26.5	18.7	2.2	8.5
	4 ^b	44.0	28.6	23.9	3.5	2.3
	5	40.5	28.5	25.5	5.6	17.1
	6	47.8	29.0	21.2	2.1	18.4
	7	54.1	27.2	17.6	1.1	3.0
	8	56.4	25.8	16.5	1.3	14.2

^aGroup designations are 0 = Infantry, Gun Crews, and Seamanship Specialists; 1 = Electronic Equipment Repairers; 2 = Communications and Intelligence Specialists; 3 = Health Care Specialists; 4 = Other Technical and Allied Specialists; 5 = Administration and Functional Support; 6 = Electrical/Mechanical Equipment Repairers; 7 = Craftsmen; 8 = Service and Supply Handlers.

^bThere is no group 3 in the Marine Corps.

productivity relationships that might prevail in different occupational areas.

Criticisms of the Active-Duty Pay System

Critics also see problems with the active-duty pay system. A perusal of the active-duty pay scales reveals apparently small intergrade differentials in basic pay. Furthermore, some pay increments resulting from longevity exceed increments resulting from promotion (see the report of the Seventh Quadrennial Review of Military Compensation [QRMC] [DoD, 1992b]). In several instances, lower-ranking personnel with more seniority receive more than higher-ranking personnel with less seniority. The concern is that the system rewards *retention* more than it rewards *performance*. In a system that rewards people almost as much for longevity as for promotion, individuals will not be motivated to work hard and strive for promotion, and the system may not retain the most able people.

The disincentives caused by the apparently narrow intergrade differentials were first recognized by the 1948 Hook Commission. After a surprisingly modern and lucid discussion of the issues, the Hook Commission recommended a substantial upward skewing of the active-duty pay table (i.e., greater interranks pay spreads at higher ranks).¹³ Their recommendations were implemented in the Career Compensation Act of 1949. But much of the skewness was dissipated by the move to the AVF in the early 1970s, when pay in the junior grades was increased substantially and pay in the senior grades was left unchanged. The problem for policymakers now is that, without also changing the retirement system, the skewness cannot be restored without substantially increasing retirement costs.

Disincentives present in the structure of the basic-pay table are compounded by the allowance system, the many in-kind benefits (e.g., free medical care and commissary privileges), and the tax advantages associated with receiving such benefits in kind rather than in the form of income. The difficulty here is that allowances and in-

¹³We say "surprisingly modern" because their arguments and policy recommendations anticipated the formal analysis of promotion contests begun by Lazear and Rosen (1981) and continued by others.

kind benefits depend heavily on marital status. As the first QRMC discussed at length, payment based on need (number of dependents) rather than on performance weakens motivation and effort. According to critics, the system subsidizes marriage and childbearing, which, in turn, increases the cost of the system.

In addition to these disincentive effects, critics claim that the system is inefficient because individuals would be better off receiving all their compensation up front—in the form of income, or cash, rather than partially in the form of in-kind benefits. They argue that by doing so the DoD could make military personnel better off for the same dollar outlays.

Critics also charge that applying the same pay table to all services is inefficient because doing so fails to recognize the unique functions and, thus, unique personnel requirements of each service. Some have argued that a separate pay table should also be used for different occupations or occupational groups.

Another criticism leveled against the active-duty pay system is that intragrade pay is misstructured. Some have argued for a system that rewards time in grade instead of time in service. Although more-productive individuals do tend to get promoted more often and thus get pay raises quicker under a time-in-service system, they lose their current pay advantage over less-productive individuals once the less-productive ones are promoted. According to the proponents of the time-in-grade system, such a system could maintain the pay advantage and thus the performance incentives of those who are promoted to a grade quicker, the “fast-trackers.”

Critics also argue that relying solely on promotion to reward performance motivates “climbing,” i.e., gives those who may be well matched in their current grade but ill suited for more responsible positions an incentive to seek higher positions instead of honing their within-grade skills. Further, because high retention reduces the promotion tempo of lower-ranked individuals, the services cannot reward performance in an *ontime way* (i.e., around the same time the individual demonstrates greater skill or better performance); to motivate personnel who get rewarded only every several years, the pay differentials between grades must be large. Some have argued that a system that makes within-grade pay increases contin-

gent on performance instead of automatic would address these problems.

Discussion of the Criticisms

Despite the criticisms leveled against the military compensation system, it is not apparent that the system is grossly inefficient. Critics, especially those who have advocated a strictly up-front pay system with little reliance on retirement benefits, ignore other considerations that may be relevant. First, like many civilian careers (see Topel [1991]), much of military training and experience is not easily transferred to other employment. Personnel who serve for long periods and then leave service in their forties or early fifties may have worse civilian opportunities than those who either left earlier or civilians of similar age. That is, personnel may have a second-career earnings loss following their transition to the civilian sector, for which retired pay is compensation. Recent studies of veteran earnings (Borjas and Welch [1986] and Goldberg and Warner [1987]) find that military retirees suffer second-career earnings losses of as much as 25 or 30 percent. However, losses are not uniform across the spectrum of military skills: They are smaller for those trained in such readily transferable skills as electronics technician (Goldberg and Warner [1987]) and are probably (proportionately) smaller for officers than for enlisted personnel (Borjas and Welch [1986]). Nevertheless, retired pay may be needed, in part, to compensate for the cost of transitioning to the civilian sector, especially when the transition occurs well before the age of full withdrawal from the labor force.

A second-career earnings loss means that, without an adequate separation incentive, personnel who stay for 20 or 25 years are likely to want to stay in service as long as they are permitted to do so. Among personnel in the military-specific occupations, the incentive for those with longer service to remain is magnified—military pay generally increases with years of military service at the same time civilian opportunities are declining. *But these are precisely the occupations in which the services do not want personnel to serve longer.* The result would be excess retention of senior personnel, an aging force, and reduced promotion opportunities for younger personnel. The 1948

Hook Commission stated the problem well (*Career Compensation*, p. 40):

[A] sound retirement system is essential to solving the superannuation problem. The services must be kept young, vigorous, and efficient; a sound retirement plan with a proper compulsory retirement age will permit youth and brains to rise to the top in time to be effective. The emphasis on youth is more important in the military services than in most other activities of Government and industry because of the physical demands of war conditions when the Regular establishment becomes the nucleus around which the civilian elements are organized. This vitalization purpose is not new; it was the fundamental premise of the present retirement system when it was established 80 years ago. Other concepts of fair treatment and the traditional concepts of retirement for those taking up the profession of arms are also important and have been given consideration but the Commission does not consider them to be controlling.

The Commission's emphasis on the retirement system's role as a force-management tool and its de-emphasis of other reasons for paying retired pay are unmistakable.¹⁴ The historical record indicates that the primary reason for implementing retired pay was to induce voluntary separations of senior personnel.

The discussion, of course, begs the question of why the military should need an inducement when it can simply involuntarily separate those senior personnel who it prefers would leave. The military could maintain the level of compensation by simply placing all of it up front in active pay instead of partially in retirement pay under this system. *More broadly, a justification for using voluntary inducements rather than involuntary methods (with greater active pay) is required.*

This discussion also begs the question of whether the benefits under the current retirement system are appropriate for the purposes of compensating for the second-career earnings loss and for inducing senior personnel to leave voluntarily. It needs to answer the

¹⁴Its de-emphasis of "fair treatment" may have been a reaction to the 1947 "Joint Army-Navy Military Pay Board Study," which recommended earlier vesting of military retirement benefits. The Hook Commission recommended not 20-year vesting but 30-year vesting.

question, Why must compensation for second-career earnings loss come in the form of retirement income and not as lump-sum severance pay or even higher pay while on active duty? The discussion overlooks the issue of whether separations are induced at the proper YOS point across the spectrum of military occupations. As noted above, the services would probably like to separate larger numbers of personnel in the combat skills prior to YOS 20. But they might like to keep personnel in the technical skills longer.

The discussion also needs to answer the question of whether a compensation system that is more reliant upon up-front active-duty pay would provide more or fewer work-effort incentives and more or less incentive for the more able personnel to stay and "percolate" to the top. It has been argued that deferring some compensation in the form of retirement benefits may provide substantial incentives for personnel to work hard and to undertake investments in military-specific skills, especially if individuals must pass through a number of promotion gates to reach the vesting point. According to some analysts, delayed vesting (coupled with sizable retired benefits) may have as much (if not more) motivational incentive as a more steeply skewed active-duty pay table.¹⁵

To evaluate the current military active-pay and retirement systems and any alternatives, many questions must be addressed. In the model we present in Chapters Three through Six, we provide a framework for addressing these issues and questions. Before that, we first overview past efforts to evaluate the military retirement system.

¹⁵A difficulty is that the compensation system must motivate both those who have strong preferences for military service (and are therefore likely to find a 20+ year career attractive) and those who do not (and are therefore unlikely to be attracted to a 20+ year career). At the same time that delayed vesting extracts extra work effort from those with strong preferences for service, it discourages the work effort of those with weak preferences. It does so because of the *budget constraint*: Higher retirement benefits for those who stay for 20 years are paid for by lower active-duty pay for everyone. It is not clear whether aggregate work effort would be higher under the current compensation system or under one that puts less in retirement and more in active-duty pay.

PAST STUDIES OF THE MILITARY RETIREMENT SYSTEM

A detailed description of the evolution of the military retirement system is provided in *Military Compensation Background Papers* (DoD, 1992a). Although a 20-year retirement system had been in existence for almost 80 years, at the end of World War II the Army and the Navy had different rules for retirement eligibility, and they treated officers differently than enlisted personnel. It was not until 1947 that inconsistent rules and regulations pertaining to retirement eligibility were (for the most part) eliminated and a common retirement system was adopted for officers and enlisted personnel.

Since 1947, a number of commissions or study groups appointed by the Executive Branch or the Congress have studied the structure of the military retirement system. All have grappled with a range of questions: (1) When should retirement benefits be vested? (2) When should benefits be paid? (3) How generous should the benefits be? (4) How should benefits be paid? (5) Should the system be defined benefit or defined contribution? (6) Should the system distinguish between officers and enlisted personnel? (7) Should the system distinguish between occupations requiring "youth and vigor" and other occupations? and (8) Should the system be contributory or noncontributory? Below, we identify these commissions or study groups and briefly review their proposals.

A common 20-year system was codified in 1946–1947 in several pieces of legislation. Almost as soon, it was severely criticized by the Joint Army-Navy Pay Board and the 1948 Advisory Commission on Service Pay (the Hook Commission). Both the Joint Pay Board and the Hook Commission thought that 20-year careers were too short and recommended immediate annuities for those who leave voluntarily only after 30 years of service (based on the $[0.025 * \text{YOS} * \text{final basic pay}]$ formula in existing legislation). However, expressing concern for the superannuation problem, both the Joint Pay Board and the Hook Commission proposal would have allowed immediate annuities to officers *over age 60* and enlisted personnel *over the age of 50* who retire with 20 or more YOS. Both study groups recognized that the services would need to involuntarily separate personnel prior to 30 years of service and would have given them tools to do so. Under the Hook Commission plan, for example, those involuntarily separated with 25 or more YOS would receive an immediate annuity

based on the above formula. Interestingly, its proposal would permit involuntary separatees with 20–24 YOS to choose among (1) a full annuity based on the standard formula to begin at age 60 in the case of officers, or age 50 in the case of enlisted personnel, (2) one-half the full annuity to begin immediately, or (3) a cash separation payment.¹⁶ The Hook Commission also recommended cash severance pays for those who are involuntarily separated prior to YOS 20.

A key difference between the Joint Pay Board and Hook Commission recommendations is that the Hook Commission plan maintained the 20-year threshold for receipt of retirement benefits. Remarkable for an internal DoD study group, the Joint Pay Board expressed concern with the apparent unfairness of such delayed vesting. Calling the 20-year system a “tontine,” something that benefits the few at the expense of the many, it recommended that all personnel who have 10 or more YOS be vested in an old-age annuity to start at age 62.¹⁷ Also remarkable was its recommendation that the system be *contributory*: Member contributions were to be invested in a fund that would pay interest, and those who left prior to completion of YOS 10 would receive their accumulated contributions and interest. Congress adopted many of the Hook Commission’s recommendations regarding the structure of active-duty pay in the Career Compensation Act of 1949, but it ignored the commission’s (or the Joint Pay Board’s) recommendations concerning retired pay.

Twenty-one years elapsed before the first QRMC provided the next formal review of the retirement system, in 1969.¹⁸ The first QRMC sided with the Joint Pay Board’s recommendation that the system be contributory. But it argued against vesting prior to YOS 20. What is notable about the first QRMC study is its attempt to define a rational basis for the *level* of retired pay. In its quest to do so, the first QRMC said a distinction should be made between the old-age period (i.e., the period after full withdrawal from the labor force) and the

¹⁶This plan would clearly reduce turnover at the 20-year point, but it provides weaker retention incentives for younger personnel.

¹⁷Named after the Italian Lorenzo Tonti, who devised a contest in which participants each venture a sum, with the winner being the one who lives the longest.

¹⁸The first QRMC was convened in 1967 but did not issue a final report on retirement until January 1969. Subsequent QRMCs have been convened at each 4-year interval since 1969.

"second-career" phase of retired pay. The first QRMC (DoD, 1969, p. S-13) said that "it was relatively easy to set the annuity in the old age period. . . . An annuity that pays military retirees a year of service percentage of salary equivalent to that paid by other employers should satisfy the management effectiveness and just treatment objectives." After recommending that the military move to a *salary* system, the first QRMC recommended that old-age annuities range from 33 percent of final military *salary* at YOS 20 to 75 percent at YOS 40. The old-age annuity equal to 33 percent of final salary at 20 YOS was not much different from the 50 percent of final basic pay being paid at that time.

But the first QRMC would have significantly reduced annuities during the second-career phase. It argued that benefits during the second-career period should be set at a level that just compensates for the second-career earnings loss. After analyzing data on retiree earnings, the first QRMC believed that the annuities in effect during the second-career phase in fact overcompensated retirees for their second-career earnings losses and recommended first-tier annuities (the annuity from separation to full retirement age) ranging from 24 percent of final salary for YOS 20 retirees to 51 percent for YOS 30 retirees, or about 34 percent of a 20-year retiree's basic pay and 73 percent of a 30-year retiree's basic pay.

Some writers had argued that a generous pension was just compensation for the rigors of a military career, with its exposure to danger, family separations, loss of freedom, and the like. The first QRMC rejected this view (DoD, 1969, p. S-13):

The argument that an additional increment should be included in the retirement annuity to compensate for the extra hazards, rigors, and inconveniences of military service was rejected on the grounds that inclusion of such an increment would be:

- inequitable, because only those military members who stayed until retirement would receive it,
- inefficient in attracting, retaining and motivating personnel because of the very high discount rate which personnel use to value retired pay, and

- inflexible and not readily adjustable to changing conditions of service or changing force management needs.

What is interesting about this argument is that it violates the first QRMC's own principle that retirement benefits during the second-career phase be set at a level that just compensates for the second-career earnings loss. The transferability of military training and experience varies considerably from one occupation to another, but the first QRMC's argument would seem to rule out varying the benefit by skill area. It would also seem to rule out varying the parameters of the retirement system to encourage different career patterns in different occupations.

After the first QRMC submitted its final report, its recommendations were reviewed by the Interagency Committee on Uniformed Services Retirement and Survivor Benefits (IAC), a study group appointed by President Nixon in 1971. IAC offered its own retirement proposal. Like the first QRMC plan, the IAC proposal would have also created a two-tier system of annuities. During the second-career phase of retirement, or first tier, YOS 20–24 separatees would have their annuities reduced by 2 percent for each year the member was under age 60 at separation; YOS 25+ separatees would have their annuities reduced by 2 percent for each year under the age of 55. Full annuities would be restored during the second tier, the retirement years, at ages 60 and 55, respectively. But rather than basing the full annuity on the linear formula ($0.025 * YOS * \text{final basic pay}$), IAC would base the annuity on the formula ($0.025 * YOS$) up to YOS 24 and $[0.6 + 0.03 * (YOS - 24)]$ up to YOS 30. In conjunction with the age- and YOS-based first-tier annuity reduction, the purpose of the full-annuity formula modification was to provide stronger retention incentives in the post-YOS 20 zone. IAC proposed basing the annuity on the individual's high-3 years' average basic pay rather than final basic pay.

The IAC proposal severely reduced pre-old-age annuities. The typical enlisted careerist who separates after YOS 20 at age 40, for example, would receive only 30 percent of high-3 years' average basic pay from separation to age 60. Many critics viewed this reduction as too severe. The IAC plan was examined by DoD and, after much modification, was submitted to Congress in 1974 as the Retirement Modernization Act (RMA). RMA departed much less dramatically from the prevailing (pre-1980) retirement system. With the same

second-tier (old-age) annuity formula as the one proposed by IAC, RMA proposed to reduce the first-tier annuity by a flat 15 percentage points, but only until such time as the individual would have completed 30 years of service. Further, it would base annuities on high-1 rather than high-3 years' basic pay. Thus, a 20-year separatee would get 35 percent of high-1 basic pay for 10 years, followed by the normal 50 percent provided by the then-prevailing system. A 30-year retiree would still receive 75 percent of (high-1) basic pay.

Finally, IAC and RMA continued a theme of the 1947 Joint Pay Board: 20-year vesting is unfair. Both proposals recommended vesting those who separate with between 10 and 19 years in an old-age annuity to begin at age 60. Interestingly, the IAC proposal would offer YOS 10-19 separatees a lump-sum payment in lieu of an old-age annuity.

Congress failed to act on RMA after its submission in 1974. RMA was reintroduced in 1975, and Congress again deferred passage. Instead, Congress called for DoD to establish a Defense Manpower Commission (DMC) to examine the full range of military personnel and compensation policy. A DMC was established, and it subsequently offered yet another retirement proposal in April 1976. The DMC also accepted the argument that 20-year vesting was inequitable, and, like the IAC and RMA proposals, it recommended vesting, after YOS 10, in an old-age annuity to begin at age 65 (with the option of a reduced annuity beginning at age 60).¹⁹ The other notable feature of the DMC analysis was its argument that, for the majority of military skills, a 20-year career was too short. It therefore recommended that eligibility for receipt of an immediate annuity be based on a point system. Those in noncombat skills would accumulate one retirement point per YOS, whereas those in combat skills would accumulate 1.5 points. Thirty points would be required for receipt of an immediate annuity. Consequently, personnel in noncombat skills would be required to serve 30 years to receive an immediate annuity, whereas those in combat skills would be required to serve only 20. (Those who complete 10 YOS but accumulate less than 30 retirement points receive only the old-age annuity.) The DMC pro-

¹⁹The DMC recommended an increase in the old-age-annuity multiplier from 0.025 to 0.0266.

posals is the only one that has explicitly distinguished jobs requiring "youth and vigor" from other jobs.

Not satisfied with past recommendations, then-newly elected President Jimmy Carter established the President's Commission on Military Compensation (PCMC) in July 1977 to again study the full range of compensation issues. This commission offered yet another proposal in April 1978. Like other studies, the PCMC agreed that 20-year vesting is unfair and also argued for 10-year vesting. It therefore recommended deferred annuities for all separatees with more than 10 YOS, with the age at receipt depending on YOS at separation.²⁰

Unlike other studies, the PCMC saw a need for vesting in something other than just an old-age annuity. Personnel bear costs in the transition from a military to a civilian career, both direct costs and the aforementioned indirect earnings losses arising from imperfect transferability of job skills. What is needed, the PCMC argued, is a system of transition benefits to go along with vested old-age annuities. The PCMC recommended that DoD establish a transition trust fund for each service member. DoD would make annual payments into the fund based on the member's basic pay and the contribution rate at the member's YOS. The PCMC recommended the following contribution rates: 0 for YOS 0-5, 20 percent for YOS 6-10, 25 percent for YOS 11-20, 15 percent for YOS 20-25, and 5 percent for YOS 26-30. DoD contributions would be invested in government securities and would accumulate interest. An individual at YOS 10 would have an accumulated amount equal to about one year's basic pay. At YOS 20, the fund would have over two years' basic pay; at YOS 30, it would have about four years' basic pay. At separation, individuals would be free to take the fund in lump-sum, receive it in monthly installments, or roll it over into another retirement account.

By providing YOS 10-19 separatees with significant cash benefits and by eliminating the large "cliff" at YOS 20, the PCMC argued that its proposal would allow DoD to trade off the 20-year rule for more flexible and thus more rational force management. The services

²⁰Old-age annuities for YOS 10-19 separatees were to start at age 62, whereas annuities for YOS 20-29 separatees were to start at age 60. Annuities for those who serve a full 30-year career would begin at age 55. Note the retention incentives created by the last discontinuity.

would no longer feel compelled to carry along unpromotable or unneeded members in this YOS range. The extra, immediate benefits provided at YOS 10, it argued, would also prove attractive to younger personnel and thereby stimulate pre-YOS 10 retention. The PCMC also foresaw that its proposal could cause retention problems in some (but not all) skill areas, but argued that the long-run savings its plan offered would increase DoD's flexibility to target more active-duty compensation at problem skills (i.e., those experiencing retention difficulties).

After scrutiny and modification of the PCMC proposal, DoD submitted the Uniformed Services Retirement Benefits Act (USRBA) to Congress in July 1979.²¹ USRBA modified the PCMC proposal three ways. First, like the PCMC proposal, all personnel with 10 or more years of service would be vested in an old-age annuity. But unlike the PCMC proposal, USRBA would begin the old-age annuity at age 60, regardless of YOS at separation.²² USRBA also changed the old-age multipliers.²³ Second, unlike the PCMC plan, USRBA would maintain annuities during the second-career phase of retirement. USRBA would pay pre-old-age annuities, which would range from 37.5 percent of high-2 years' basic pay for separation at YOS 20 to 62.5 percent at YOS 30. Third, instead of creating a transition trust fund for YOS 10+ separatees, USRBA would make cash payments in the YOS 10-19 range to both those who stay and those who leave. The scheme was to work as follows. At the completion of YOS 10, individuals could borrow against their prospective first-tier annuities. At YOS 10, the individuals would be permitted to borrow 10 months' basic pay regardless of whether they stay or leave. Upon completion of each of the next four years, individuals could borrow another two months' basic pay. It was hoped that these cash pay-

²¹There was no formal fourth QRMC. Although DoD is required by law to formally convene a QRMC every four years, DoD's internal review of the PCMC's recommendations served as an informal fourth QRMC.

²²See footnote 18 above. Analysts in DoD thought that the 5-year gap in the age at which YOS 30 separatees and other separatees would begin their annuities was too large and might inhibit management of the senior career force.

²³The old-age-annuity multipliers were 0.02 for YOS 1-5, 0.0225 for YOS 6-10, and 0.0275 for YOS 11-30. Also, the annuity was to be based on high-2 years' basic pay. A 10-year separatee would thus receive 21 percent of high-2 basic pay, whereas a 20-year separatee would receive 48.75 percent.

ments would prove attractive in retaining to the 10- or 15-year point those personnel who were not necessarily interested in pursuing a 20-year career. Personnel were not forced to leave to get the payments, because DoD did not want personal borrowing needs to be a reason for leaving and feared that individuals having financial problems might be tempted to borrow and then separate. Those who borrow against their future annuities and then stay for a 20-year career would have their first-tier annuities reduced by 0.45 percent for each month of basic pay withdrawn. A 20-year separatee who had withdrawn the maximum allowable 20 months of basic pay would get a first-tier annuity of 28.5 percent of high-2 years' basic pay rather than 37.5 percent.

In retrospect, it is clear that this rather complicated early-withdrawal scheme was based on a misunderstanding of the purpose of retired pay, which is, in fact, to encourage separation at the proper point and to assist individuals in their moves back to civilian life. If individuals are allowed to borrow now while remaining on active duty and thereby diminish their future annuities, they will have less incentive to leave in the future, and they might not have the wherewithal to make a smooth transition to civilian life.

USRBA was never enacted. Like the PCMC proposal, it met opposition on two fronts. One front, the services, opposed it (as well as the PCMC proposal) because of its departure from an annuity-based system of retirement benefits and because of uncertainty about how the plan would affect retention patterns and force structure. The second front was budgeteers in the Treasury Department and politicians in Congress, who thought the major drawback of both plans to be the significant changes in time pattern of governmental outlays that would result. While the move away from total reliance on annuities would eventually reduce outlays for military retirement, the payment into newly created transition funds or the payment of early-withdrawal benefits would add to near-term outlays.

Like the DMC and the PCMC, the fifth QRMC was charged with reexamining all previous proposals to modify the retirement system. It convened in 1983 and issued its final report in 1984. Going against the grain of the majority of past studies, the fifth QRMC sided with the first, arguing against earlier-than-20-year vesting in retirement benefits. But it sided with the previous study groups that found

benefits for 20-year retirees to be "excessive." Following the first QRMC, IAC, and RMA plans, it returned to the concept of a two-tier retirement system with reduced benefits during the second-career phase of separation from military service.

Rather than recommending one single plan, the fifth QRMC evaluated a range of alternatives that varied the reduction in 20-year benefits and the growth rate in benefits for service beyond YOS 20; it continued to examine plans with early-withdrawal options; and it considered alternatives to full COLAs for retirees. The fifth QRMC submitted its report to DoD in 1984. Following a mandate from Congress, DoD submitted legislation to modify the retirement system, in November 1985. Congress then implemented the Military Retirement Reform Act of 1986. Commonly known as REDUX, MRRA contains many features of past proposals for retirement reform. Notably, MRRA (1) distinguishes between the two phases of retirement and provides lower benefits during the second career, (2) increases the growth rate in the second-career annuity from 2.5 percent to 3 percent for each YOS beyond 20, and (3) provides less-than-full adjustment for increases in the cost of living. By diminishing the value of exactly a 20-year career, both absolutely and relative to longer careers, REDUX appears to implement many of the ideas of past study groups. But it still departs from the majority of past studies in its failure to provide any benefits for those who separate prior to YOS 20.

None of these discussions of retirement-system reform has addressed the issues of ability selection (or ability-sorting) and effort motivation (or effort supply) to any extent. They may have paid lip service to them, but just how these goals are achieved through the military's compensation system has not been discussed. These issues are addressed in our model, which we present in Chapters Three through Six.

ORGANIZATIONAL STRUCTURE: DEFINITIONS AND ASSUMPTIONS

The discussion in Chapter Two highlights the issues and questions surrounding the appropriate structure of military pay and retirement benefits. To evaluate the current and alternative systems, a framework is required that accounts for the ways that compensation and personnel policy affect individual decisionmaking. We set forth this framework in this chapter and in the following three chapters. In this chapter, we define the structure of the hypothetical organization that we analyze and delineate our main assumptions.

THE INTERNAL LABOR MARKET AND PERSONNEL FLOWS IN THE HIERARCHY

The organization that we are interested in is a *vertical hierarchy*, i.e., one in which the organizational structure is necessitated by technological and organizational considerations and the use of authority. Those who fill the higher grades (or levels or ranks) have the responsibility for, and authority over, the actions of those below them. The military is an example of such a hierarchy.¹

¹The distinction between vertical and horizontal hierarchies is attributed to Stiglitz (1975). In contrast to a vertical hierarchy, a *horizontal hierarchy* exists merely to confer differential pay and higher status on those workers with better skills and performance. Those with higher status or rank do not necessarily have responsibility for the actions of those with less status or rank. Academic institutions are an example of a horizontal hierarchy: academic ranks exist to distinguish individual accomplishments and performance. The rank or grade distribution of a horizontal hierarchy is not constrained by technological or authority considerations as is the grade distribution of a vertical hierarchy. For example, university faculty are promoted on the basis of indi-

We assume that the hypothetical hierarchical organization has a closed personnel system. Individuals enter at the lowest level, and positions at each upper level are filled by drawing from the pool of individuals in the level below. Thus, the firm does not fill the upper ranks with lateral entrants obtained from the open market, but with employees from within its internal labor pool. A body of literature suggests a number of reasons why hierarchical firms follow this practice. Employers are rarely well informed about workers' abilities and motivation before hiring and observing workers on the job. Even after expending substantial resources to collect information on job candidates—including interviewing the candidates, obtaining transcripts and letters of recommendation, and administering screening tests—some hires who looked good on paper turn out to be poor performers; others turn out to be pleasant surprises.² The substantial heterogeneity in workers' abilities and motivation, coupled with employers' inability to accurately measure those qualities beforehand, virtually dictates that large, hierarchical organizations hire workers at the lowest level, then promote them to fill higher levels, based on performance at lower levels.

Human capital investment considerations reinforce this practice. By *human capital* we mean the effect of training and experience with the current employer on the stream of potential earnings with the current and alternative employers. In the human capital literature, workers entering an organization receive either general or specific training. Workers receiving general training will tend to pay for their training costs in low wages during the training period because, after training, the market will bid away workers who are not paid for their post-training productivity (Oi, 1962). Workers receiving specific training are more problematic. If the firm bears all the training costs and the worker quits after training, the firm suffers a capital loss. Furthermore, if job skills are specific to the organization, the organization will be unable to fill the upper ranks by outside hiring. Workers receiving specific training must be offered advancement in the organization as an incentive to accept training that is not mar-

vidual merit, and the distribution of faculty by rank is not limited by technological or organizational factors.

²Recent work by Simon and Warner (1992) provides some evidence that information about job applicants collected through informal networks is superior to information collected via formal channels.

ketable elsewhere. Even if possible, lateral entry into upper levels would destroy such investment incentives.³

Another characteristic of large organizations is that they often set pay administratively rather than by one-on-one negotiations with the employee (*contracting*). That is, the firm sets a pay scale that may vary with such factors as grade, and experience with the firm or in a grade, and pays everyone in the same circumstances the same amount. The military and the federal government are examples of institutions that set pay administratively.⁴

Thus, our organization brings in untrained people at the bottom and then "grows" them to fill higher positions within the organization. In general, individuals may occupy I grades or levels and may remain with the firm for up to T time periods. For simplicity, we assume three grades and four time periods in the analysis below.⁵ All individuals enter at level 1 in period 1. Through promotions that occur in periods 2 and 3, some of the entrants rise to fill positions at levels 2 and 3. However, no promotions occur in period 4, and all employees are required to leave the organization after this terminal period. Individuals may fully retire after period 4, or they may work for an alternative employer during period 5.

³See Carmichael (1983). To guarantee that workers incur the full cost of quitting, including the organization's loss, Carmichael shows that the organization will, in fact, pay the worker more than his or her marginal product during the latter part of the post-training period. In Carmichael's model, overpayment comes through promotion to a job with higher status. Yet promotions occur only as a vehicle for paying the worker to accept specific training and not as a result of technological or organizational considerations, as in our analysis.

⁴Although we do not seek a full explanation for the practice of administrative pay-setting here, it is likely that the heavy transaction costs that would be incurred in one-on-one bargaining over pay drive large organizations to set pay administratively. See Milgrom (1988). Prespecifying a pay scale and punishing underperformers (by firing or not promoting them) may also be an efficient way for a large organization to motivate effort when the cost of monitoring effort is high. See Lazear (1991) and Carmichael (1989) for reviews of the literature on such incentive explanations.

⁵To maintain a connection between our abstract model and our application of it to the military, the reader may let the first period in the model correspond to the period of an initial military enlistment, the second and third periods correspond to the "early" and "middle" phases of a military career, and the fourth period correspond to the period after which military personnel are eligible for a pension.

Let $s_{i,t}$ denote the fraction of individuals who are in grade i at the end of period t who remain with the organization for at least one more period. Thus, $s_{i,t}$ is a conditional-retention probability. Furthermore, let $\pi_{i,t}$ denote the probability of promotion from grade $i-1$ to grade i during period t . We assume that no promotions occur in period 4. All promotions to level 3 thus occur during period 3. Promotions to level 2 can occur during period 2 or period 3. Therefore, there are two ways an individual can reach grade 2 but only one way to reach grade 3. A schematic of the process is shown in Figure 3.1.

Consider a new entrant's various possible stay-leave sequences. Some individuals will leave after period 1 in grade 1; this occurs with probability $1 - s_{1,1}$. Some individuals will leave after period 2 in grade 1; this occurs with probability $s_{1,1}(1 - \pi_{2,2})(1 - s_{1,2})$. The probabilities of the various possible stay-leave sequences are shown in Table 3.1.⁶

The steady-state grade structure for the organization (once the policy change has worked itself through) is derived as follows. Let $N_{1,1}$ denote the number of new entrants into grade 1 in period 1. Then the

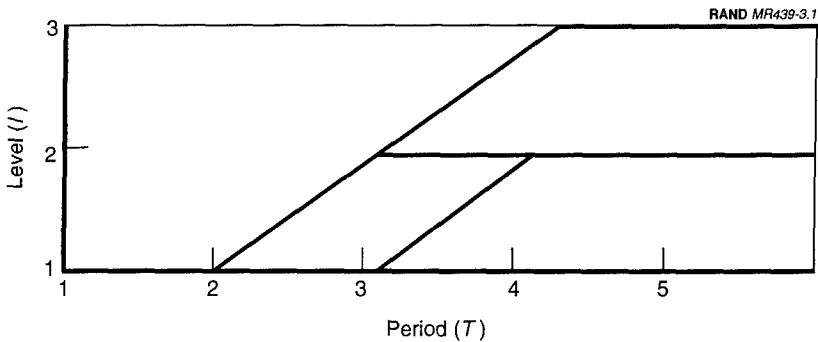


Figure 3.1—Personnel Structure of the Hypothetical Organization

⁶As our discussion in Chapter Four illustrates, the survival rates and the promotion probabilities are not independent. Greater survival rates reduce promotion rates because fewer positions need to be filled. The probabilities in the table abstract from this interrelationship to more simply illustrate the stay-leave sequence.

Table 3.1
Probability of Each Stay-Leave Sequence

Leave from Grade	In Period	Probability
1	1	$1 - s_{1,1}$
1	2	$s_{1,1}(1 - \pi_{2,2})(1 - s_{1,2})$
1	3	$s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{2,3})(1 - s_{1,3})$
1	4	$s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{1,2})s_{1,3}$
2	2	$s_{1,1}\pi_{2,2}(1 - s_{2,2})$
2	3	$s_{1,1}\pi_{2,2}s_{2,2}(1 - \pi_{2,3})(1 - s_{2,3}) + s_{1,1}(1 - \pi_{2,2})s_{1,2}\pi_{2,3}(1 - s_{2,3})$
2	4	$s_{1,1}\pi_{2,2}s_{2,2}(1 - \pi_{2,3})s_{2,3} + s_{1,1}(1 - \pi_{2,2})s_{1,2}\pi_{2,3}s_{2,3}$
3	3	$s_{1,1}\pi_{2,2}s_{2,2}\pi_{3,3}(1 - s_{3,3})$
3	4	$s_{1,1}\pi_{2,2}s_{2,2}\pi_{3,3}s_{3,3}$

total number in grade 1, N_1 , is $N_{1,1}$ plus the number flowing to period 2 in grade 1, $N_{1,1}s_{1,1}(1 - \pi_{2,2})$ plus the number flowing into period 3 in grade 1, $s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{2,3})$ plus the number in grade 1 at period 4, $s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{2,3})s_{1,3}$. Thus, the total numbers in grades 1-3 will be

$$\begin{aligned}
 N_1 &= N_{1,1} \left[1 + s_{1,1}(1 - \pi_{2,2}) + s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{2,3}) \right. \\
 &\quad \left. + s_{1,1}(1 - \pi_{2,2})s_{1,2}(1 - \pi_{2,3})s_{1,3} \right] \\
 N_2 &= N_{1,1} \left[s_{1,1}\pi_{2,2} + s_{1,1}\pi_{2,2}s_{2,2}(1 - \pi_{3,3}) \right. \\
 &\quad + s_{1,1}(1 - \pi_{2,2})s_{1,2}\pi_{2,3} \\
 &\quad + s_{1,1}\pi_{2,2}s_{2,2}(1 - \pi_{3,3})s_{2,3} \\
 &\quad \left. + s_{1,1}(1 - \pi_{2,2})s_{1,2}\pi_{2,3}s_{2,3} \right], \\
 N_3 &= N_{1,1} \left[s_{1,1}\pi_{2,2}s_{2,2}\pi_{3,3} + s_{1,1}\pi_{2,2}s_{2,2}\pi_{3,3}s_{3,3} \right]. \quad (3.1)
 \end{aligned}$$

Since the organization is a hierarchy, $N_1 > N_2 > N_3$. Appendix A presents a numerical illustration of the derivation of the hierarchical structure.

PRODUCTION RELATIONSHIPS IN THE HIERARCHY

We assume that the organization requires N_1 individuals at level 1, N_2 at level 2, and N_3 at level 3 and that these requirements are determined exogenously by technological and other considerations. There are N_2 units or divisions in the organization, each supervised by someone at grade 3. Each person in grade 3 supervises N_2 / N_3 personnel, who in turn supervise N_1 / N_2 personnel. Given this structure, the output, Q , that this hierarchy is capable of producing depends on the ability, α , and work effort, e , of its employees. We assume that the organization's output increases with the abilities and work effort of its employees. But ability and work effort affect Q differently at different levels. Rosen (1982) shows that, because of span-of-control considerations, an individual's productivity is proportional to his or her level in the hierarchy. In our model, level 3 jobs are the most important, and level 1 jobs are the least important. At the lowest level, variations in ability have minor effects on Q —more able and less able people perform low-level tasks about the same way. Because of span-of-control considerations, ability has a larger effect on Q at higher levels in the organization.⁷ It is important that the ability of the level 3 worker exceed the abilities of the level 2 workers, whose abilities in turn should exceed the abilities of the level 1 workers.

Work effort (i.e., how hard an individual works) also affects Q . When individuals put forth more effort, Q increases. Effort therefore can substitute for ability. A less able individual who puts forth much effort may be as productive as a more able slacker. However, the rate at which e can substitute for α diminishes with rank. Thus, a low-ability worker is unlikely to be very effective if assigned to a level 3 job, no matter how hard he or she tries. Finally, individual productivity may depend on the worker's experience with the firm, t , and aggregate output Q may therefore depend on the cumulative experience of its workforce. The size of the workforce N required to pro-

⁷We say this recognizing that there may be considerable variation in ability requirements at low levels in the organization. In the military case, some low-level jobs (e.g., electronics technicians) may require great mental aptitude, whereas others require less ability.

duce a given output Q is therefore not independent of the experience mix of the force.⁸

To put these arguments more formally, let $q(i, t, \alpha, e)$ represent the productivity of an individual in grade i with experience level t , ability level α , and effort e . Therefore, $\partial q / \partial \alpha$ is the effect of ability and $\partial q / \partial e$ is the effect of effort on productivity (i.e., the *marginal products*). Our arguments suggest that $\partial q / \partial \alpha$ is positive and that it increases with the individual's level in the organization. Furthermore, $\partial^2 q / \partial \alpha^2 > 0$, so that the increase in output that results from assigning a person to a higher level rises with the person's ability. We also assume that $\partial q / \partial e > 0$ and that it also increases with ability and level. However, our assumption that effort and ability do not substitute perfectly means that the marginal rate of substitution between effort and ability, $(\partial q / \partial e) / (\partial q / \partial \alpha)$, declines as i increases.

ORGANIZATIONAL GOALS

At the most basic level, the organization's objective is to maximize output given its constraints. Given the production relationships described above, this means that the hierarchy must solve both macroeconomic and microeconomic personnel problems. Its macro problem is to design compensation and personnel policies that will permit it to balance its actual personnel inventory with the inventory it requires, i.e., to fill the N_i slots. The military calls this balance matching "the faces with the spaces." Individuals must be attracted and retained to bring about the proper grade and experience distribution of the organization's labor force.

The necessary grade and experience distribution will vary from situation to situation. In some situations, technological and organizational factors dictate the need for an experienced labor force; in others, a relatively youthful labor force may be required. For example, the Marine Corps and the Army say they need relatively youthful labor forces capable of charging up hills and performing other physical tasks. Owing to the more technical nature of the jobs performed in

⁸During the AVF but prior to the drawdown, force size remained relatively constant but the average experience level of the force increased, suggesting that overall capability increased during most of the AVF period.

their services, the Navy and Air Force cite the need for more experienced personnel. In the context of our model, the effect of experience on productivity, $\partial q / \partial t$, is different across the services. The Navy and Air Force believe that $\partial q / \partial t$ rises with t ; the Army and Marine Corps implicitly believe it actually declines with t (after a certain point). Likewise, some situations may dictate the need for a "tall" grade structure, with many tiers and fewer people per tier, whereas others imply the need for a "flatter" one, with fewer tiers and more people per tier.

But given the role of effort and ability in affecting organizational output, the organization must also design policies that will encourage effort and induce the most able to aspire to the higher levels within the organization—the micro problem. This micro problem, which is discussed at length in Chapters Four through Six, has generally been ignored in previous models.

THE BUDGET CONSTRAINT

Let $m_{i,t}$ denote the current pay (including in-kind benefits) the organization establishes for an individual in grade i at period t , and $w_{i,t}$ be the proportion of those personnel in period t who are in grade i so that $\sum_i w_{i,t} = 1$. Based on the grade distribution at each period of service, the mean pay in year t is $m_t = \sum_i w_{i,t} m_{i,t}$. We also assume that those who separate from the organization in grade i after period t receive the stream of retired pay, $r_{i,t,j}$, with $j = t+1, \dots, D$ (D denotes death period). We let $R_{i,t}$ denote the expected present value to the firm of retired pay of a loss from grade i at the end of period t , and $l_{i,t}$ denote the fraction of losses in period t who are in grade i such that $\sum_i l_{i,t} = 1$. Then $R_t = \sum_i l_{i,t} R_{i,t}$ is the mean expected present value of the retirement liabilities to the losses at period t . We also let sv_t be the *survival rate* into period t (i.e., the fraction of those surviving until the end of period $t-1$ who also stay for period t). If we define the firm's *per capita budget constraint*, B^* , as the present value of compensation that the organization expects to pay to a new hire, then

$$\begin{aligned}
B^* &= m_1 + sv_2 \frac{m_2}{(1+r)} + sv_3 \frac{m_3}{(1+r)^2} + sv_4 \frac{m_4}{(1+r)^3} + (1-sv_2) \frac{R_1}{(1+r)} \\
&\quad + (sv_2 - sv_3) \frac{R_2}{(1+r)^2} + (sv_3 - sv_4) \frac{R_3}{(1+r)^3} + sv_4 \frac{R_4}{(1+r)^4} \\
&= M^* + R^*,
\end{aligned} \tag{3.2}$$

where r is the firm's discount rate, M^* is the expected present value at hire of future "wage" payments (note that military basic pay is more than just an hourly wage), and R^* is the expected present value of the new hire's retirement liability.⁹ (Note that the firm may discount future dollars at a different rate than its employees discount those dollars.) The total pay budget is $N_1 B^*$.

We define the *per capita accrual cost of retirement* as the amount, A , that the firm must invest each year to fund the accumulating retirement liability of a new hire. That is, A is an amount whereby

$$B^* = \sum_{t=1}^4 sv_t \frac{(m_t + A)}{(1+r)^{t-1}} \text{ or } R^* = \sum_{t=1}^4 \frac{sv_t A}{(1+r)^{t-1}}. \tag{3.3}$$

Using these relations, we see that the per capita accrual cost is

$$A = \frac{\sum_{t=1}^4 (sv_t - sv_{t-1}) R_t / (1+r)^t}{\sum_{t=1}^4 sv_t / (1+r)^{t-1}} = \frac{R^*}{\sum_{t=1}^4 sv_t / (1+r)^{t-1}}, \tag{3.4}$$

where sv_t is the probability that a new entrant will remain until period t and $(sv_t - sv_{t-1})$ is the probability that the entrant will quit at exactly period t .

The accrual cost is an annualization of the expected present value of future retirement liabilities. The total annual accrual cost is $N_1 A$. We assume that the firm's output Q is somehow exogenously deter-

⁹For simplicity, we assume that active-service wage payments are made at the beginning of each period.

mined and that it therefore seeks to minimize the cost of producing Q or $N_1 B^*$. In addition to this *output constraint*, the firm must obey the constraints imposed by the hierarchical grade structure.¹⁰ Through its various compensation and personnel policies, the firm attempts to minimize costs by manipulating three variables that affect output: worker retention (or, conversely, the number of new hires each period), the effort exerted by its workforce, and the quality of the workers in each grade. These variables are modeled in Chapter Four for individual workers and the micro problem.

¹⁰Technological requirements dictate the proportions in which the firm must combine workers in levels 1 through 3. We assume that the number of workers can vary as long as those proportions are met.

MODELING INDIVIDUAL DECISIONMAKING

We now model individual worker's retention and effort decisions. Our analysis is patterned after that of Gotz and McCall (1984), who developed a dynamic programming model of Air Force officers' retention decisions. The model assumes that workers maximize their expected utility when deciding whether to remain with the organization. The model accounts for the dynamic nature of the retention decision by incorporating the fact that (rational) individuals take into account all possible future eventualities when making their decisions. We expand upon the model by incorporating work effort decisions. Effort enters into the analysis both positively and negatively. First, it represents a utility cost to individuals. On the other hand, effort can increase the utility gain to staying to the extent that future rewards (such as promotion) depend on the amount of effort supplied.

We organize the discussion as follows. First, we model basic retention decisions facing each individual. We then model the role of effort in the individual's expected utility gain to staying. Next, we derive the comparative statics results and show how retention and effort supply change when other factors in the model change. We then derive the cohort retention rates based on those for individuals. As shown in Chapter Three, the force structure depends on these cohort retention rates.

MODELING RETENTION DECISIONS

We begin our discussion of retention decisions with the following definitions. Let $c_{i,t,j}$ be the amount a searatee who leaves the firm

prior to full withdrawal from the labor force expects to receive from alternative employment (with $j = t + 1, \dots, L$ and $L =$ the labor force withdrawal period). We let $c_{i,t,j} = c_{i,t,j}^* \exp(\beta_\alpha \alpha)$, where $c_{i,t,j}^*$ is expected earnings based on the individual's observable characteristics (education, race, sex, etc.) and $\exp(\beta_\alpha \alpha)$ is the contribution of ability to civilian earnings. The parameter β_α shows the proportionate effect of a unit change in ability on alternative-earnings possibilities. We expect $\beta_\alpha > 0$. We let δ denote the individual's personal discount rate and $\beta = 1/(1 + \delta)$ be the discount factor on future income.

Nonpecuniary factors as well as pay with the current and alternative employers affect the decision to remain with the organization. Pecuniary factors are represented by pay within the organization, $m_{i,t}^p$, and the stream of retired pay, $r_{i,t,j}$. We let the nonpecuniary factors, such as in-kind benefits, be represented by $m_{i,t}^n$, so that the "pay" associated with each grade and year of service is given by $m_{i,t} = m_{i,t}^n + m_{i,t}^p$.

We model an individual's attitudes toward service as the sum of permanent (or time-invariant) and transitory (or random) components. Let τ^m and τ^c be the permanent taste factors associated with remaining with the current employer and with moving to an alternative employer, respectively. We assume that these factors are invariant to the individual's current period of service. Individuals with positive values of τ^m place a positive net value on the current employer's work environment and fringe benefits, for example. The individual's net preference for the current employer is thus $\tau = \tau^m - \tau^c$. Individuals who like (dislike) the current employer have positive (negative) values of τ .

Even if the individual does not like the current employer, random events can induce the individual to remain with the firm. Likewise, such events, or shocks, can induce those who like the current employer to leave. Such shocks might include a good or bad job assignment, winning a \$1 million lottery and deciding to quit, and an unexpected change in the economywide unemployment rate. We summarize these factors by the random error term, $\varepsilon_{i,t}$, which incorporates components that are totally idiosyncratic and components, such as economic conditions, that are potentially correlated across individuals.

We define the individual's supply of effort at time t in grade i as e_{it} and assume that individuals do not like to exert work effort. When they can, they shirk. Let $Z(e_{it})$ denote the monetary value of the disutility individuals place on e_{it} units of work effort at time t . We assume that $dZ(e_{it})/de_{it} > 0$, i.e., the marginal disutility of work effort is positive; and $d^2Z(e_{it})/de_{it}^2 > 0$, i.e., marginal disutility rises with the amount of work effort. For simplicity, we assume that everyone has a common $Z(e_{it})$.

Individual Retention Decisions

Consider the stay-leave decision faced by an individual who is at the end of the third period and deciding whether to remain for the last period. Since no one is promoted in period 4, the individual's return to leaving the organization at the end of period 3 is

$$\begin{aligned} L_{i,3} &= \sum_{j=4}^D \beta^{j-4} r_{i,3,j} + \sum_{j=4}^5 \beta^{j-4} c_{i,3,j} + \sum_{j=4}^5 \beta^{j-4} \tau^c \\ &= R_{i,3} + C_{i,3} + \Gamma_{i,3} . \end{aligned} \quad (4.1)$$

The return to leaving is the sum of the present values of the future retirement stream, $R_{i,3}$, the expected future civilian earnings stream, $C_{i,3}$, and the value the individual places on the nonpecuniary benefits offered by alternative employers, $\Gamma_{i,3}$. Similarly, the return to staying for the final period is

$$\begin{aligned} S_{i,3} &= m_{i,4} + \sum_{j=5}^D \beta^{j-4} r_{i,4,j} + \beta(c_{i,4,5} + \tau^c) + \tau^m - Z(e_{i4}) + \varepsilon_{i,3} \\ &= m_{i,4} + \beta R_{i,4} + \beta C_{i,4} + \beta \tau^c + \tau^m - Z(e_{i4}) + \varepsilon_{i,3} . \end{aligned} \quad (4.2)$$

Given our utility-maximization framework, the individual stays if the gain to staying is positive and leaves if it is negative. We define the gain to staying, $G_{i,3}$, as $S_{i,3} - L_{i,3}$, where $G_{i,3}$ is a utility gain and not just a financial gain. Combining Equations 4.1 and 4.2, we see that the individual stays if

$$\begin{aligned}
 G_{i,3} &= S_{i,3} - L_{i,3} > 0 \\
 &= \tau + m_{i,4} + (\beta R_{i,4} - R_{i,3}) - (C_{i,3} - \beta C_{i,4}) - Z(e_{i4}) + \varepsilon_{i,3} > 0. \quad (4.3)
 \end{aligned}$$

The gain to staying for the terminal period increases with the individual's net taste for the current employer, terminal-period pay, the growth in present value of retired pay, $\beta R_{i,4} - R_{i,3}$, and with the transitory disturbance (or shock), but diminishes with the present value of forgone earnings, $C_{i,3} - \beta C_{i,4}$, and the disutility of work effort. The individual chooses to stay if $G_{i,3}$ is positive for some optimal amount of effort, $e_{it} \geq 0$.¹

We specify the individual retention rate as the probability that an individual in grade i at the end of period 3 remains. Let $\varepsilon_{i,3}$ be normally distributed with zero mean and standard deviation of σ_ε . (Thus, the distribution of $\varepsilon_{i,t}$ is invariant to the grade and period.) Define the gain to staying given tastes and ability but not conditional on $\varepsilon_{i,3}$ as

$$G_{i,3}^* = \tau + m_{i,4} + (\beta R_{i,4} - R_{i,3}) - (C_{i,3} - \beta C_{i,4}) - Z(e_{i4}). \quad (4.4)$$

The individual stays if $G_{i,3}^* + \varepsilon_{i,3} > 0$. By the symmetry of the normal distribution, the probability of this event is

$$\begin{aligned}
 Pr(G_{i,3}^* + \varepsilon_{i,3} > 0) &= Pr(G_{i,3}^* > -\varepsilon_{i,3}) \\
 &= 1 - Pr(G_{i,3}^* < \varepsilon_{i,3}) \\
 &= Pr(\varepsilon_{i,3} < G_{i,3}^*) . \quad (4.5)
 \end{aligned}$$

The probability that an individual will stay in grade i for the terminal period, *conditional upon his or her tastes and ability*, is $\Phi(G_{i,3}^* / \sigma_\varepsilon)$, where Φ denotes the standard normal distribution function evalu-

¹ Note that $\varepsilon_{i,t}$ in Equations 4.2 and 4.3 refers to effort during period 4. But the individual is making his or her retention decision at the end of period 3. Therefore, we assume that the individual has all the information at the end of period 3 necessary to decide what effort he or she will supply during period 4 and, therefore, how that amount will affect his or her retention decision. Optimal period 4 effort depends on pay and personnel policies, as we show in Chapter Five.

ated at the standardized gain to staying, $G_{i,3}^* / \sigma_\varepsilon$. To reduce notation, we let $\Phi_{i,3} = \Phi(G_{i,3}^* / \sigma_\varepsilon)$.

Consider now the expected present value of the future utility of an individual at the end of period 3 who is yet to observe $\varepsilon_{i,3}$. This individual will leave before the terminal period and, hence, receive the present value of the earnings stream defined by Equation 4.1, with probability $1 - \Phi_{i,3}$. The expected present value of future utility conditional upon staying is

$$E(S_{i,3} \mid G_{i,3}^* > -\varepsilon_{i,3}) = \tau^m + \beta\tau^c + m_{i,4} + \beta R_{i,4} + \beta C_{i,4} + \sigma_\varepsilon \frac{\phi_{i,3}}{\Phi_{i,3}} - Z(e_{i,4}), \quad (4.6)$$

where $\phi_{i,3}$ is the standard normal density function evaluated at $G_{i,3}^* / \sigma_\varepsilon$.² Since the individual stays with probability $\Phi_{i,3}$ and leaves with probability $1 - \Phi_{i,3}$, his or her expected lifetime utility at the end of period 3 is

$$\begin{aligned} V_{i,3}^* &= \Phi_{i,3} E(S_{i,3} \mid G_{i,3}^* > -\varepsilon_{i,3}) + (1 - \Phi_{i,3}) L_{i,3} \\ &= L_{i,3} + \Phi_{i,3} G_{i,3}^* + \sigma_\varepsilon \phi_{i,3}. \end{aligned} \quad (4.7)$$

Expected utility is a weighted average of the return to staying and the return to leaving, with the weight being the individual's probability of staying or the return to leaving plus the expected gain to staying weighted by the probability of staying. The latter term can be thought of as the *option value of staying*. In Equation 4.7, the term $\sigma_\varepsilon \phi_{i,3}$ is always positive. Given Equation 4.7, the individual's expected utility at the beginning of period 3 is

$$m_{i,3} + \tau^m + \beta V_{i,3}^* - Z(e_{i,3}). \quad (4.8)$$

²The term $\phi_{i,3} / \Phi_{i,3}$ is called the inverse Mills Ratio (Maddala, 1983). The term $\phi_{i,3}$ is the ordinate of the standard normal distribution evaluated at $G_{i,3}^* / \sigma_\varepsilon$.

Individuals get the utility associated with their current pay and the value of net nonpecuniaries minus the utility cost of supplying effort during period 3 plus the discounted expected utility that they get as of the end of period 3.

Gains from staying and the probabilities of staying in earlier periods are derived recursively. Since promotions occur in period 3, promotion opportunities affect the gain from staying. Consider an individual who is in grade i at the end of period 2. His or her return to staying for period 3 is

$$S_{i,2} = \pi_{i+1,3} [\tau^m + m_{i+1,3} + \beta V_{i+1,3}^* - Z(e_{i3})] \\ + (1 - \pi_{i+1,3}) [\tau^m + m_{i,3} + \beta V_{i,3}^* - Z(e_{i3})] + \varepsilon_{i,2}, \quad (4.9)$$

where $\pi_{i+1,3}$ is the probability of promotion from grade i to grade $i + 1$ in period 3. As we discuss in the next section, the promotion probability in Equation 4.9 is the individual's assessment of his or her own promotion chances. Although each individual's calculations will be based on his or her own expected promotion chance, expectations should obey an *aggregation constraint*: Not everyone can have an expected promotion chance of 50 percent if the actual aggregate rate is only 30 percent. In the analysis in the next section, an individual's promotion probability is permitted to depend on personal work effort and ability, as well as on the aggregate promotion rate.

An individual who is in grade i at the end of period 2 can follow four possible future paths: (1) be promoted during period 3 but leave at the end of period 3, (2) be promoted during period 3 and stay to the end of period 4, (3) not be promoted and leave after period 3, and (4) not be promoted and stay to the end of period 4.³ In Equation 4.9, the returns to these different paths are weighted by their probabilities of occurrence. Since the individual stays for period 3 if there exists an $\varepsilon_{i,2} \geq 0$ such that $S_{i,2} - L_{i,2} > 0$, he or she stays with probabil-

³The exception is the possibility that the organization has an up-or-out rule that requires the individual to separate from the organization if he or she has not achieved a prespecified rank by a given period. The military and academic institutions are notable for their up-or-out rules. We discuss the role of up-or-out rules in more detail in Chapters Five and Six.

ity $\Phi_{i,2} = Pr(S_{i,2} - I_{i,2} > 0)$. Expected utility at the end of period 2 is derived analogously to Equation 4.7. Iterating again, we may derive the gains to staying, the probability of staying, and expected utility at the end of period 1. We present the fully specified model (iterated to period 1) below, in the "Model Summary" section of this chapter, after we incorporate the role of effort.

MODELING EFFORT SUPPLY

So far, we have allowed only effort to negatively affect individual utility via $Z(e_{it})$. However, if those who work harder get promoted more often or faster or achieve other gains within the organization, effort supply also has a positive role in determining the gain to staying. The role of effort and ability in affecting an individual's success within the organization depends crucially on the organization's compensation and personnel policy. Clearly, effort's role will be weak at best if the organization does not reward effort and ability or their covariates. Below, we incorporate into our model three types of compensation schemes and personnel policies that give effort a positive role in determining an individual's utility gain to staying: intergrade promotion contests, involuntary separation and/or minimum-performance standards, and intragrade contingent pay.

Promotion Contests

By assumption, information about ability and effort is asymmetric: Individuals know their own ability level and effort supply in each grade and period, but the organization cannot observe these or, alternatively, can do so only by incurring a positive cost. In cases where the firm cannot observe the worker's input or effort and/or ability, it might be able to observe the worker's output (as long as the firm produces a tangible output). But this output is a "noisy" indicator of the worker's effort/ability because output depends on both effort/ability and other random factors, none of which the firm can monitor directly. In situations where no tangible output is produced (e.g., the military), the situation is even worse. The potential for moral hazard and adverse selection abounds.

Lazear and Rosen (1981) suggest that in such circumstances the organization may use contests (or tournaments) to motivate work ef-

fort. Gifford and Kenney (1986) show that contests may be used to induce workers to reveal their true abilities. In contests, the firm ranks workers based on supervisors' evaluations, formal proficiency tests, etc., and rewards them based on their *relative* rather than their *absolute* performance. In such contests, smarter (higher- α) individuals are inherently more likely to obtain a higher rank-order. But, rank-order is not strictly proportional to ability; individuals can affect their place in the queue by exerting more effort, e_{it} . Less able people can surpass more able people by working harder.

We characterize *contests* as follows. At the end of a given period, the organization administers an evaluation to those in a given grade. Let the individual's score on the evaluation at time t , E_{it} , be a linear function of his or her ability, work effort during the period, and a random factor, ξ_{it} , so that $E_{it} = k_{i1}\alpha + k_{i2}e_{it} + \xi_{it}$.⁴ The random factor arises because other unobservable factors—how the individual felt on test day, gets along with his or her supervisor, and so forth—may influence the individual's performance on the evaluation. The test is therefore an error-ridden measure of the individual's true ability and work effort. We assume that ξ_{it} has zero mean, so that each individual's expected evaluation score is a linear function of his or her true ability and effort, i.e., $E(E_{it}) = k_{i1}\alpha + k_{i2}e_{it}$.

Suppose that the contest determines who is promoted at the beginning of t . The organization administers the evaluation, ranks the n evaluations from highest to lowest, and promotes those with the r_t highest scores, where r_t is the number of vacancies in the next-highest grade at time t . Thus, the aggregate promotion rate at t is $\pi_t^* = r_t / n$. Note that the number of vacancies in a grade depends on how many individuals survive into grade i at time t , $s_{i,t}$, as well as on the survival and retention rates in all higher grades and years of service. We denote these survival sequences as $\{s_{i,t}\}$. Therefore, the aggregate promotion rate, which depends on whole retention/survival sequences, is given by $\pi_t^* = r_t(\{s_{i,t}\}) / n$.

⁴This formulation of E_{it} , which assumes that effort and ability are perfect substitutes, is used for convenience. More generally, $E_{it} = (k_{i1}\alpha^{-\rho} + k_{i2}e_{it}^{-\rho})^{-1/\rho} + \xi_{it}$, where $-1 < \rho < \infty$; in this formulation, ability and effort would not be perfect substitutes.

Let us assume that the organization evaluates the contestants, ranks the evaluations from highest to lowest, then selects the top r for either a promotion or an intragrade bonus. If r of the participants are to be selected for promotion, then the contestant's evaluation must be among that of the top r participants in the contest. Thus, the contestant must beat out r of the n competitors. Consider any given contestant. The contestant beats his or her j^{th} competitor in rank i at time t if $E_{it} > E_{it}^j$ or

$$k_{11}\alpha + k_{12}e_{it} + \xi_{it} > k_{11}\alpha^j + k_{12}e_{it}^j + \xi_{it}^j, \quad (4.10)$$

or

$$k_{11}(\alpha - \alpha^j) + k_{12}(e_{it} - e_{it}^j) + \xi_{it} > \xi_{it}^j, \quad (4.11)$$

where α is the contestant's ability and α^j is that of the j^{th} competitor.

To derive the probability of various places in the contest, let us assume that the ξ are identically and independently distributed with mean 0 and standard deviation σ . Then the probability that the contestant beats the j^{th} competitor is

$$\begin{aligned} Pr(E_{it} > E_{it}^j) &= Pr\left[k_{11}(\alpha - \alpha^j) + k_{12}(e_{it} - e_{it}^j) + \xi_{it} > \xi_{it}^j\right] \\ &= F\left[k_{11}(\alpha - \alpha^j) + k_{12}(e_{it} - e_{it}^j)\right], \end{aligned} \quad (4.12a)$$

where $F(\bullet)$ denotes the cumulative density function of ξ_{it}^j evaluated at $\left[k_{11}(\alpha - \alpha^j) + k_{12}(e_{it} - e_{it}^j) + \xi_{it}\right]$. If we assume that ξ is normal, then

$$Pr(E_{it} > E_{it}^j) = \Phi\left[\frac{k_{11}(\alpha - \alpha^j) + k_{12}(e_{it} - e_{it}^j) + \xi_{it}}{\sigma_{\xi}}\right], \quad (4.12b)$$

where $\Phi(\bullet)$ denotes the standard normal cumulative density function.

To be promoted, the contestant must finish among the top r places in the contest. Let I_k denote a k^{th} place finish in the contest, such that I_1 denotes a first-place finish, I_2 denotes a second-place finish, etc. The probability of promotion is $\sum_{k=1}^r Pr(I_k)$.

An expression for the probability of promotion can be derived when the contestants are homogeneous (all have the same ability α). For convenience, assume that $k_{i2} = 1$ and that all the contestant's competitors supply effort e_{it}^* . Then the probability that the i^{th} contestant finishes in the k^{th} place is

$$\begin{aligned}
 Pr(I_k) &= \frac{(n-1)!}{(n-k)!(k-1)!} \int \left[\Phi \left(\frac{e_{it} - e_{it}^* + \xi_{it}}{\sigma_{\xi}} \right) \right]^{n-k} \\
 &\quad \times \left[1 - \Phi \left(\frac{e_{it} - e_{it}^* + \xi_{it}}{\sigma_{\xi}} \right) \right]^{k-1} \phi(z) dz \\
 &= {}_{n-1}C_{k-1} \int \left[\Phi \left(\frac{e_{it} - e_{it}^* + \xi_{it}}{\sigma_{\xi}} \right) \right]^{n-k} \\
 &\quad \times \left[1 - \Phi \left(\frac{e_{it} - e_{it}^* + \xi_{it}}{\sigma_{\xi}} \right) \right]^{k-1} \phi(z) dz,
 \end{aligned} \tag{4.13}$$

where $\phi(z)$ denotes the standard normal density function. The term ${}_{n-1}C_{k-1}$ is the number of ways that the i^{th} contestant can beat $n-k$ other contestants but lose to $k-1$ other contestants. It may be shown that if all contestants supply equal effort, then $Pr(I_k) = 1/n$. That is, the individual has an equal probability of any rank in the contest. The probability of promotion to rank $i+1$ at period $t+1$ is

$$\begin{aligned}
\pi_{i+1,t+1} &= \sum_{k=1}^r Pr(I_k) \\
&= \sum_{k=1}^r n^{-1} C_{k-1} \int \left[\Phi \left(\frac{e_i - e_{it}^* + \xi_{it}}{\sigma_\xi} \right) \right]^{n-k} \\
&\quad \times \left[1 - \Phi \left(\frac{e_{it} - e_{it}^* + \xi_{it}}{\sigma_\xi} \right) \right]^{k-1} \phi(z) dz.
\end{aligned} \tag{4.14}$$

Since the contestant has probability $1/n$ of any given rank in the contest, it is clear that when n contestants supply the same effort, each individual has a promotion probability $\pi = r/n$. Clearly, then, increasing the number to be promoted raises π , whereas increasing the number of contestants lowers it.

In general, an individual's promotion probability at time t depends on the aggregate promotion rate, $\pi_{i,t}^*$; his or her own ability and work effort; and the abilities and work efforts of all others against whom he or she is competing (denoted α^o and $e_{i-1,t-1}^o$, respectively). Therefore, $\pi_{i,t} = \pi_{i,t}(\alpha, e_{i-1,t-1}, \alpha^o, e_{i-1,t-1}^o, \pi_{i,t}^*)$; $\pi_{i,t}$ increases with α and $e_{i-1,t-1}$ but diminishes with α^o , $e_{i-1,t-1}^o$, and $\pi_{i,t}^*$ (i.e., the abilities and work efforts of his or her peers and the aggregate promotion rate). Note that the individual promotion probabilities, and thus the expected gain to staying at time t in grade i , depend on the cohort retention rates, because $\pi_{i,t}^*$ depends on $\{s_{i,t}\}$.

Involuntary Separation Policies and Work Effort

Involuntary separation brought about by either the failure to be promoted (i.e., *up-or-out*) or failure to meet certain minimum-performance standards may also provide effort incentives⁵ if it makes future gains (in the form of being permitted to remain in the organi-

⁵See Kanemoto and MacLeod (1991) for discussion of the differences between the pay and personnel practices of American and Japanese firms. Japanese firms stress the use of "carrots," that is, pay, to motivate performance. American firms are more likely to use "sticks," that is involuntary separation. The differences are explained by the relative ease of job mobility of American, as opposed to Japanese, workers.

zation and claim future benefits) contingent on performance. Of course, for involuntary separation to be an effort motivator, individuals must perceive that they will be worse off if separated or better off if retained.

To incorporate the role of up-or-out policies, suppose that the organization has a rule that says that anyone not promoted to grade 3 in period 3 must leave the firm at the end of period 3. Then in Equation 4.7, the individual's probability of staying, $\Phi_{2,3}$, is zero and expected lifetime utility, $V_{2,3}^*$, at the end of period 3 equals the value of leaving, $L_{2,3}$. If $G_{2,3}^*$ was positive without the separation rule, then $V_{2,3}^*$ is less than it would have been without the rule. Therefore, the up-or-out rule makes the individual's expected future success or utility contingent on retention and effort decisions in the current period, period 3. More generally, owing to the forward-looking nature of retention and effort decisionmaking, the policy affects decisions in earlier periods. We explicitly derive these effects in Chapter Five.

A *minimum-performance standard*, whereby the evaluation process identifies both poor performers and promotees, makes future employment contingent on current decisions. Of course, future employment is valuable only if the gains to staying are positive. We introduce minimum-performance standards as follows. Although promotion is based on relative performance, we also assume that the organization requires individuals to perform at time t in grade i to some absolute minimum, $E_{\min(i,t)}$, to be retained.⁶ Therefore, given our definition of the individual's evaluation score, E_{it} , the individual is terminated if

$$k_{i1}\alpha + k_{i2}e_{it} + \xi_{it} < E_{\min(i,t)} \text{ or } \xi_{it} < E_{\min(i,t)} - k_{i1}\alpha - k_{i2}e_{it}, \quad (4.15)$$

and involuntary separation occurs with probability $Pr(\xi_{it} < E_{\min(i,t)} - k_{i1}\alpha - k_{i2}e_{it})$. Obviously, the probability of involuntary separation

⁶This assumption is not without precedent; Malcomson (1984) assumes the same. Alternatively, the minimum-performance standard could be made a relative one by specifying that the lowest x percent on the evaluation be separated. The U.S. Army's Qualitative Management Program (QMP) operates in precisely such a fashion. A centralized promotion board meets annually both to select personnel for promotion and to decide upon involuntary separatees (other than those who have reached high-year-of-tenure points). Since 1971, the Army has annually selected about 2.5 percent of those in grades E-5-E-9 (and who are retirement-eligible) for involuntary separation.

risks with the minimum required score but declines as ability and effort increase.

We incorporate the risk of involuntary separation resulting from below-minimum performance by redefining the *individual retention probability* as the joint probability that an individual wants to stay and the organization wants to keep him or her:

$$\Phi_{i,t} = Pr\left(-\xi_{it} < -E_{\min(i,t)} + k_{i1}\alpha + k_{i2}e_{it}; G_{i,t}^* > -\varepsilon_{i,t}\right). \quad (4.16)$$

For simplicity, we assume that the random errors in the evaluation and retention decision equations, ξ_{it} and $\varepsilon_{i,t}$, respectively, are uncorrelated. Indeed, there is no reason to believe that there should be any relationship between them. Under this assumption,

$$\begin{aligned} \Phi_{i,t} &= Pr\left(-\xi_{it} < -E_{\min(i,t)} + k_{i1}\alpha + k_{i2}e_{it}\right)Pr\left(G_{i,t}^* > -\varepsilon_{i,t}\right) \\ &= \Phi_{i,t}^1 \Phi_{i,t}^2, \end{aligned} \quad (4.17)$$

where $\Phi_{i,t}^1$ is the probability that this individual will be permitted to stay and $\Phi_{i,t}^2$ is the probability that this individual will want to stay. As a result of this redefinition of $\Phi_{i,t}$, the inverse Mills Ratio term in Equation 4.6 is redefined to be $\Phi_{i,3}^2 / \Phi_{i,3}^1$. Thus, Equation 4.7 becomes $V_{i,3}^* = L_{i,3} + \Phi_{i,3} G_{i,3}^* + \sigma_\varepsilon \Phi_{i,t}^1 \Phi_{i,t}^2$.

Intragrade Pay Policies and Work Effort

The organization wants to motivate individuals to seek the grades for which they are best matched. However, once workers achieve their highest position, the organization also wants to ensure that they continue to supply effort. Further use of promotion contests as an incentive device is ruled out, because promoting these workers would reduce Q , the organization's overall output: By definition, these individuals are ill suited to perform the duties associated with higher grades. How can the organization motivate effort? It would seem that without the proper incentives, those who have reached their terminal grade will tend to shirk.

Consider an individual who is in grade 2 at the start of period 3 and who knows he has no further prospect of being promoted. What

work effort will he expend during the third period? Intuitively, he might not expend any effort, since he has no future promotion prospect. However, the organization can use several tools to motivate intragrade effort. Conceptually, the issue of how to structure within-grade pay for those with no promotion opportunities is basically the same as the issue of how to structure pay in a nonhierarchical setting. Generally speaking, the organization must make *current* pay or status contingent on the individual's *past* performance.

First, the organization can use involuntary separation as a threat; those who fail to meet within-grade minimum-performance standards will be fired. This threat is viable as long as individuals lose something by being separated. Thus, their future expected net utility within the organization must be positive. Those with a strong net taste for the organization (a large τ) will have a greater expected gain to staying. For others, the organization can ensure that staying is preferred by deferring some pay into the future. Following Lazear (1979), the firm might underpay workers in a grade relative to their true productivity profile initially but, contingent on satisfactory within-grade performance, overpay them relatively when they gain more within-grade experience. The contingent pay can take the form of an intragrade bonus or operate through the retirement system (as we discuss in Chapter Six).⁷ Of course, the intragrade pay spread must be set together with the intergrade spread; the organization wants to ensure that those whose best match lies in higher grades continue to be motivated to seek promotion while those who are better matched in their current grade are not motivated to move up. We discuss the problems of "slumming" and "climbing" in greater detail in Appendix B.⁸

To incorporate this role of involuntary separation, we set π_{it} equal to zero for those who have achieved their best match (i.e., terminal

⁷Underpaying workers relative to their productivity profile within the grade does not necessarily imply that workers are paid less than their outside alternatives, especially in the presence of match-specific capital or training.

⁸According to O'Keeffe, Viscusi, and Zeckhauser (1984), *slumming* occurs when high-ability individuals desire to compete in low-ability contests rather than in the high-ability ones, where they are more appropriately suited. *Climbing* occurs when low-ability individuals infiltrate contests designated for the more able (as was first considered by Lazear and Rosen [1981]).

grade) and no longer have promotion opportunities, and we let $\Phi_{i,t}$ be the joint probability that an individual wants to stay and the organization wants to keep him or her. Then the problem of intragrade incentives brought about by involuntary separation is the same as the problem described in the preceding subsection regarding the role of minimum-performance standards. Chapter Five examines how effort and retention decisions respond to this incentive method.

Second, of course, those who meet the minimum-performance criteria for the grade might still supply less than the efficient amount of effort. One way to continue to motivate those in their terminal grade is to offer (*contingent*) merit pay so that future within-grade pay is contingent on within-grade performance. For instance, the firm might evaluate those in grade i at the beginning of t and provide a bonus to those who meet a certain performance threshold $E_{i,t}^w$. Let $\psi_{i,t}$ denote the probability that the individual will surpass $E_{i,t}^w$, the within-grade threshold in grade i at time t . Implicitly, $\psi_{i,t}$ depends on work effort in the previous period (as with the case of promotion contests) and ability via $d_{i1}\alpha + d_{i2}e_{it-1}$, where d is a parameter of the intragrade performance evaluation. Thus, an individual who is not promoted can expect to receive $m_{i,t} + \psi_{i,t}b_{i,t}$ with probability $\psi_{i,t} = \psi_{i,t}(d_{i1}\alpha + d_{i2}e_{it-1} > E_{i,t}^w)$, where $b_{i,t}$ is the bonus in grade i at time t . Those newly promoted to a new grade do not receive the bonus until they have proven themselves in the new grade. In Chapter Five, we analyze how effort and retention decisions respond to such an intragrade pay scheme. First, we summarize our model specification and present the comparative static results.

MODEL SUMMARY

Before presenting the comparative static results and examining the effect of alternative pay policies on decisionmaking, we summarize the discussion so far by presenting the fully specified model, solved recursively back to period 1. Expected utility at the end of period 3 for an individual in grade i is

$$\begin{aligned} V_{i,3}^* &= \Phi_{i,3}E(S_{i,3} \mid G_{i,3}^* > -\varepsilon_{i,3}) + (1 - \Phi_{i,3})L_{i,3} \\ &= \Phi_{i,3}G_{i,3}^* + L_{i,3} + \sigma_\varepsilon \Phi_{i,3}^1 \phi_{i,3}^2, \end{aligned} \quad (4.18)$$

where

$$\begin{aligned}
 \Phi_{i,3} &= Pr\left(-\frac{\xi_{i3}}{\sigma_\xi} < \frac{-E_{\min(i,3)} + k_{i1}\alpha + k_{i2}e_{i3}}{\sigma_\xi}\right) Pr\left(\frac{G_{i,3}^*}{\sigma_\varepsilon} > \frac{-\varepsilon_{i,3}}{\sigma_\varepsilon}\right) \\
 &= \Phi_{i,3}^1 \Phi_{i,3}^2 \\
 G_{i,3}^* &= \tau + m_{i,4} + \psi_{i,4}b_{i,4} + (\beta R_{i,4} - R_{i,3}) - (C_{i,3} - \beta C_{i,4}) - Z(e_{i,4}) \\
 L_{i,3} &= C_{i,3} + R_{i,3} + \Gamma_{i,3} \\
 \psi_{i,4} &= \psi(d_{i1}\alpha + d_{i2}e_{i3} > E_{i,4}^w) .
 \end{aligned}$$

Note that, despite the lack of promotion opportunities in period 4, individuals can increase their chances of earning the intragrade bonus $b_{i,4}$ in period 4 by increasing period 3 effort. They also reduce the likelihood of being fired for insufficient performance. Expected utility at the end of period t (except for period 3) for workers in grade i is given by

$$V_{i,t}^* = \Phi_{i,t} G_{i,t}^* + L_{i,t} + \sigma_\varepsilon \Phi_{i,t}^1 \Phi_{i,t}^2 , \quad (4.19)$$

where

$$\begin{aligned}
 \Phi_{i,t} &= Pr\left(\frac{-\xi_{it}}{\sigma_\xi} < \frac{-E_{\min(i,t)}^* + k_{i1}\alpha + k_{i2}e_{it}}{\sigma_\xi}\right) Pr\left(\frac{G_{i,t}^*}{\sigma_\varepsilon} > \frac{-\varepsilon_{i,t}}{\sigma_\varepsilon}\right) \\
 &= \Phi_{i,t}^1 \Phi_{i,t}^2 \\
 G_{i,t}^* &= \pi_{i+1,t+1} \left[\tau^m + m_{i+1,t+1} + \beta V_{i+1,t+1}^* - Z(e_{it+1}) \right] \\
 &\quad + (1 - \pi_{i+1,t+1}) \left[\tau^m + m_{i,t+1} + \psi_{i,t+1}b_{i,t+1} \right. \\
 &\quad \left. + \beta V_{i,t+1}^* - Z(e_{it+1}) \right] - R_{i,t} - C_{i,t} - \Gamma_{i,t} \\
 L_{i,t} &= C_{i,t} + R_{i,t} + \Gamma_{i,t}
 \end{aligned}$$

$$\pi_{i+1,t+1} = \pi \left[\alpha, e_{it}, \alpha^o, e_{it}^o, \pi_{i+1,t+1}^* (\{s_{i+1,t}\}) \right]$$

$$\psi_{i,t+1} = \psi(d_{i1}\alpha + d_{i2}e_{it} > E_{i,t+1}^w) .$$

Promotion contests make future rewards contingent on current effort supply and retention. Those who are not newly promoted can increase their chances of earning the intragrade bonus $b_{i,t+1}$ by supplying more period t effort. An individual's expected utility over his or her career as of the beginning of period t is

$$\tau^m + m_{i,t} + \psi_{i,t} b_{i,t} + \beta V_{i,t}^* - Z(e_{it}) . \quad (4.20)$$

COMPARATIVE STATIC RESULTS

We next examine the individual's effort and retention decisions. We first derive the determinants of the individual's retention and effort supply decision. Then we look at the role of various policy alternatives in solving the organization's macroeconomic and microeconomic problems.

Determining the Individual Retention Decision

Exogenous changes within and outside the organization—*transitory disturbance*—will affect the individual's probability of staying, $\Phi_{i,t}$, in grade i at time t in Equation 4.19. Assuming that the transitory disturbance is normally distributed, Figure 4.1 presents a graphical illustration of the probability of staying. (We assume here that $\Phi_{i,t}^2 = 1$.) If $G_{i,t}^{*1}$ is the individual's gain to staying, then $\Phi_{i,t}$ is the area under the normal distribution to the left of $G_{i,t}^{*1}$. A rise in the gain to staying, from $G_{i,t}^{*1}$ to $G_{i,t}^{*2}$, increases the probability of staying. Factors that increase the return to staying, $S_{i,t}$, increase the gain to staying and hence the probability of staying. They include an increase in the individual's net preference for the current employer, a future increase in pay within the firm ($m_{i,t+1}$, $m_{i+1,t+1}$, $b_{i,t+1}$), a higher present value of future retired pay compared with that already vested, a lower disutility from work effort, a higher promotion rate,

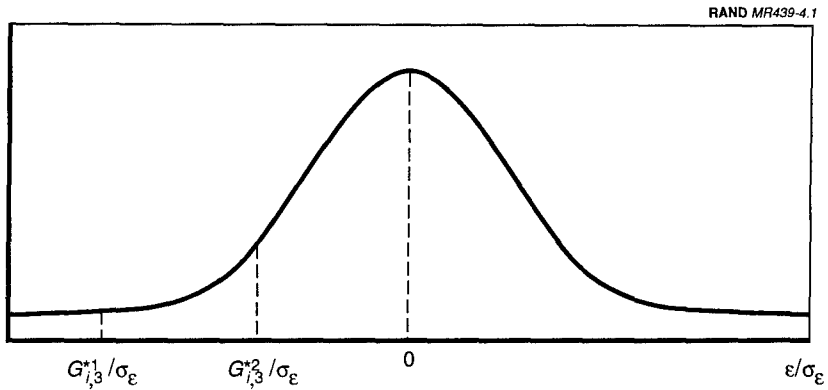


Figure 4.1—Probability of Staying

and a greater chance of meeting intragrade performance standards. Factors that increase the return to leaving, $L_{i,t}$, reduce the gain to staying and therefore the probability of staying. They include an increase in the expected present value of earnings with alternative employers, $C_{i,t}$.

The effect of ability on the individual retention decision is ambiguous. More able workers are not necessarily more likely to leave. Although they have better outside opportunities, they are also more likely to be promoted (given effort). If pay varies with grade and/or the more able are promoted faster, those with better promotion opportunities will have a higher return to staying than those with more limited opportunities. Therefore, the effect of ability on retention depends on how ability affects earnings possibilities outside the organization vis-à-vis inside possibilities. Much depends on the structure of pay and promotion.

Human capital considerations require some discussion. If training and experience are fully transferable (i.e., the human capital is *general*), the alternative stream $c_{i,t,j}$, with $j = t+1, \dots, L$, is independent of the leaving date t . But, without perfect transferability to other employers (the human capital is *specific*), the alternative po-

tential earnings stream declines with the length of service with the current organization.⁹

Human capital considerations have dynamic effects on an individual's probability of joining or remaining with the firm. Without differential pay incentives or other offsetting effects, an individual who is confronted with the prospect of a human capital loss because of nontransferability of skills will be less likely to join the organization initially (depending on how well informed he or she is) and will be more likely to leave early in his or her career. However, if the individual does stay, he or she will be more likely to remain later on.

Variations in Effort and the Probability of Promotion

Before deriving the optimal-effort condition, we first must address the question of how effort affects the probability of promotion, inasmuch as the relationship between effort and promotion probabilities is key to motivating effort supply. In this subsection, we first derive results for homogeneous contests, then consider how these conclusions are affected when the contestants are heterogeneous. Appendix C provides numerical examples of the propositions stated in this subsection.

Consider a cohort of n individuals in rank i at time t who are vying for promotion and r of them are to be promoted. If the candidates are *homogeneous*, then each individual has probability of promotion π_{it} , given by Equation 4.14. If a given contestant expands his or her effort while the other contestants' effort levels remain fixed, then differentiating Equation 4.14 with respect to e_{it} , the effect of the extra effort on the probability of promotion is

$$\frac{\partial \pi_{i+1,t+1}}{\partial e_{it}} = \frac{1}{\sigma_{\xi}} \sum_{k=1}^r n_{-1} C_{k-1} \int [n - k - (n - 1) \Phi(\bullet)] \times \Phi(\bullet)^{n-k} [1 - \Phi(\bullet)]^{k-1} [\phi(z)]^2 dz. \quad (4.21)$$

⁹Goldberg and Warner (1987) and Borjas and Welch (1986) find that longer periods of military service do detract from post-service earnings but that the extent of the earnings penalty varies from occupation to occupation.

It can be shown that Equation 4.21 is positive for all $r < n$. Thus, if the contestant expands effort and all other contestants' effort levels remain fixed, the contestant's probability of promotion increases. However, in the case where $r = n$ (everyone is promoted), increased effort has no effect on the probability of promotion.

The effects of changes in r , n , and σ on the marginal effect of effort on the promotion probability can be derived. First,

$$\frac{\partial \pi(r+1; n)}{\partial e_{it}} - \frac{\partial \pi(r; n)}{\partial e_{it}} = \frac{1}{\sigma_{\xi}} {}_nC_r \int [n-r+1-(n-1)\Phi(\bullet)] \times \Phi(\bullet)^{n-r} [1-\Phi(\bullet)]^{r-1} [\phi(z)]^2 dz \quad (4.22)$$

gives the effect on the probability of promotion when the number to be promoted increases by one. When this expression is integrated over the range of z , the term $[n-r+1-(n-1)\Phi(\bullet)]$ will have an expected value of $[n-r+1-(n-1)/2]$. This term, and therefore this expression, is positive when $r \leq n/2$ and negative when $r > (n/2) + 1$. Consequently, when the number to be promoted is less than half of the number of contestants, increasing the number to be promoted raises the marginal effect of effort on the probability of promotion. Conversely, when the number to be promoted exceeds half the number of contestants, increasing the number to be promoted reduces the effect of effort on π . Therefore, as the aggregate promotion rate r/n approaches either 1 or 0, the effect of effort on the probability of promotion diminishes. That is, when almost everyone, or almost no one, is to be promoted, effort has less effect on one's promotion chances.

From this result, it is clear that changes in the number to be promoted have an opposite effect. If the probability of promotion is less than 0.5 and n is increased, thereby reducing r/n farther below 0.5, the marginal effect of effort on the probability of promotion will decline. Conversely, when the probability of promotion exceeds 0.5 and n increases, so that r/n moves toward 0.5, the marginal effect of effort on the probability of promotion will increase.

It is apparent from Equation 4.21 that an increase in the standard deviation of the random factors in the evaluation, σ_{ξ} , causes the

marginal effect of effort on the probability of promotion to decline. That is, $\partial^2 \pi / \partial \sigma_\xi \partial e < 0$. *In other words, the more important are the random factors that influence promotion, the weaker is the link between effort and promotion.* In the limit, if an individual's evaluation were completely determined by factors outside of the individual's control, effort would have no effect on the probability of promotion.

Finally, an important result is that the effect of effort on the probability of promotion is not invariant to the scale of the contest. We can show from Equation 4.21, and the numerical examples in Appendix C confirm, that successive doublings of r and n (so that the ratio of promotions to competitors remains constant) *raises* $\partial \pi / \partial e$. This result suggests that, as the scale of the contest increases, each contestant has more opportunity to jump over his or her competitors by expanding effort. The consequences of this result for pay policy are discussed in the next chapter.

Mathematical expressions for the probability of promotion and the effect of effort on promotion are more complicated when the contestants are heterogeneous. But simulations of heterogeneous contests indicate that the results just derived for homogeneous contests continue to carry over to heterogeneous contests, i.e., the contest are qualitatively the same.

Individual Effort Decisions

The model suggests that individuals supply work effort only when there is a utility gain from doing so. Recall that $V_{i,t}^*$ in Equation 4.19 is the expected future utility of an individual in grade i at the end of period t . An important feature of this setup is that $V_{i,t}^*$ and the individual's expected utility as of the beginning of t (given in Equation 4.20) depends on current and future work effort. Individuals will choose current effort during t (e_{it}) that maximizes the expected utility given in Equation 4.20, or so that

$$\beta \frac{\partial V_{i,t}^*}{\partial e_{it}} - \frac{dZ(e_{it})}{de_{it}} = 0 . \quad (4.23a)$$

Using the definition of $V_{i,t}^*$ in Equation 4.19, the optimal effort supply e_{it} solves

$$\beta \left[\Phi_{i,t} \frac{\partial G_{i,t}^*}{\partial e_{it}} + G_{i,t}^* \frac{\partial \Phi_{i,t}}{\partial e_{it}} + \sigma_\epsilon \left(\phi_{i,t}^2 \frac{\partial \Phi_{i,t}^1}{\partial e_{it}} + \Phi_{i,t}^1 \frac{\partial \phi_{i,t}^2}{\partial e_{it}} \right) \right] - \frac{dZ(e_{it})}{de_{it}} = 0$$

or

$$\beta \Phi_{i,t} \frac{\partial G_{i,t}^*}{\partial e_{it}} + \beta \Phi_{i,t}^1 \left(G_{i,t}^* \Phi_{i,t}^2 + \sigma_\epsilon \phi_{i,t}^2 \right) \frac{k_{i2}}{\sigma_\xi} - \frac{dZ(e_{it})}{de_{it}} = 0 \quad (4.23b)$$

The first term in Equation 4.23a or 4.23b is the marginal return to a unit of work effort (MRE); the last term is the marginal disutility of effort (MDE). Individuals expend effort during t until MRE = MDE, i.e., the point at which the increase in future utility due to an extra unit of current work effort equals the disutility of an extra unit of effort.¹⁰

We assume the following effort disutility function:

$$Z(e_{it}) = \theta_1 \frac{e_{it}^{\theta_2}}{\theta_2}, \quad (4.23c)$$

which is defined for $\theta_1 > 0$ and $\theta_2 > 1$. The marginal disutility of effort is therefore $\theta_1 e_{it}^{\theta_2-1}$. When $\theta_2 = 2$, the marginal disutility of effort is the linear function $\theta_1 e_{it}$.

Without more structure (i.e., specific examples and/or functional specifications), this general condition tells us little about how the effort choice will vary across and within levels and about how retention, ability, and the effort decision interact. Therefore, in the rest of our discussion in this subsection, we take as an example the work effort decision of those who are in grade 2 in period 2 and who are

¹⁰If individuals plan to leave with certainty (as is the case at the end of period 4), so that $V_{i,t}^* = L_{i,t}$ in Equation 4.23a, then individuals expend effort only if $L_{i,t}$ depends on effort in Equation 4.1. Since, in our formulation, outside earnings do not depend on effort supplied within the organization, individuals supply zero effort in their last period.

working for promotion to grade 3 in period 3.¹¹ Individuals choose e_{22} to maximize expected utility at the beginning of period 2 (from Equation 4.19), or $\tau + m_{2,2} + \psi_{2,2}b_{2,2} + \beta V_{2,2}^* - Z(e_{22})$, where $V_{2,2}^*$ is expected utility at the end of period 2 in grade 2. This maximization requires choosing e_{22} so that

$$\beta \Phi_{2,2} \left[\left(\frac{\partial \pi_{3,3}}{\partial e_{22}} \right) \left(m_{3,3} - m_{2,3} - \psi_{2,3} b_{2,3} + \beta \left(V_{3,3}^* - V_{2,3}^* \right) \right) + (1 - \pi_{3,3}) b_{2,3} \frac{\partial \psi_{2,3}}{\partial e_{22}} \right] + \beta \Phi_{2,2}^1 \left[G_{2,2}^* \Phi_{2,2}^2 + \sigma_\epsilon \Phi_{2,2}^2 \right] \frac{k_{22}}{\sigma_\xi} - \theta_1 e_{22}^{\theta_2 - 1} = 0, \quad (4.24)$$

where $\partial \pi_{3,3} / \partial e_{22}$ is the rate at which the individual's chance of promotion in period 3 increases by working harder in period 2 and where the second term captures the effect of changes in e_{22} on the probability (and the rate of change in the probability) that the individual will meet the retention hurdle and be permitted to stay, given the individual's decision to stay. In the second term of Equation 4.24, $\Phi_{2,2}^1$ is the ordinate of the standard normal function evaluated at $[E_{\min(2,t)} - k_{21}\alpha - k_{22}e_{22}] / \sigma_\xi$. When no minimum-performance standards are in effect and everyone is permitted to stay, $\Phi_{2,2}^1$ is zero and the second term in Equation 4.24 disappears. Otherwise, the second term can be positive or negative, depending on the sign of $G_{2,2}^*$.

Figure 4.2 shows the optimality conditions. The MRE curve slopes downward and the MDE curve slopes upward, owing to our assumption of increasing marginal disutility of effort.¹²

¹¹The promotion probability is thus $\pi_{3,3} = \pi[\alpha, e_{22}, \alpha^0, e_{22}^0, \pi_{3,3}^*(\{s_{3,2}\})]$ and $\psi_{2,3} = \psi(d_{21}\alpha + d_{22}e_{22} > E_{2,3}^w)$ is the probability that an individual passes intra-grade retention hurdles.

¹²We assume that the second order condition for a maximum holds. This condition is presented in Appendix D.

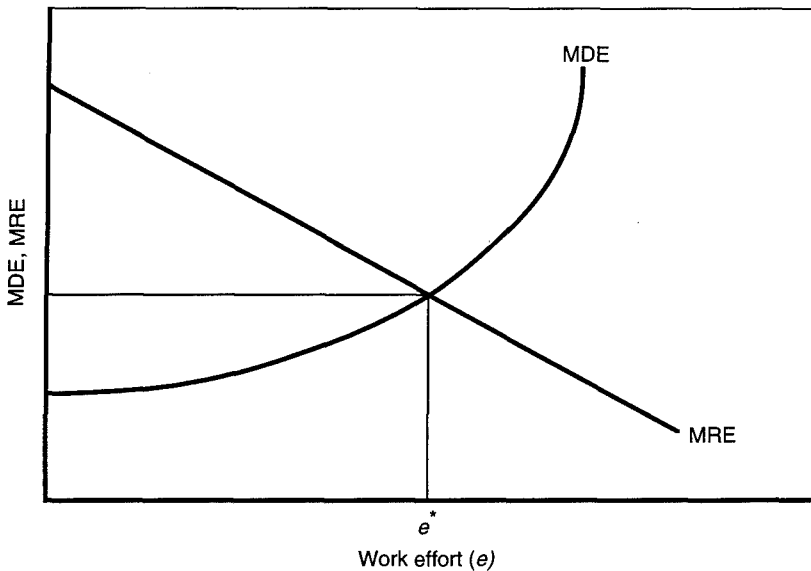


Figure 4.2—Equilibrium Work Effort

In Chapters Five and Six, we devote a considerable amount of our discussion to how effort supply responds to policy changes. Graphically, these changes in policy can be represented by shifts in the MRE schedule. Such shifts change the optimal effort supply. For example, a pay increase would shift the MRE schedule out to the right and the optimal effort supply would rise. We derive the effects of alternative policies on effort supply more formally in Appendix D.

AGGREGATION: DERIVATION OF COHORT RETENTION RATES

To this point, individual retention probabilities have been conditioned on the individual's tastes and ability. Here, we derive the fraction of an entry cohort that is expected to survive to each period of service, sv_t , and the entry cohort's conditional retention rate at each period, sv_t / sv_{t-1} . Consider first retention at the end of the initial period, when the accession cohort $N_{1,1}$ is still in grade 1. Each

individual in this cohort has a gain from staying $G_{1,1}^*$ that is conditional on his or her net taste, τ , and ability, α . Assume that τ and α follow a bivariate probability density $dF(\tau, \alpha)$ (e.g., normal) over this initial cohort with mean vector $\mu = [\mu_\tau, \mu_\alpha]'$ and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_\tau^2 & \rho\sigma_\tau\sigma_\alpha \\ \rho\sigma_\tau\sigma_\alpha & \sigma_\alpha^2 \end{bmatrix}. \quad (4.25)$$

The parameter ρ is the correlation between tastes and ability. The retention rate at the end of period 1, $s_{1,1}$, which also equals the survival rate into period 2, sv_2 , is

$$s_{1,1} = sv_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi\left(\frac{G_{1,1}^*}{\sigma_\varepsilon}\right) dF(\tau, \alpha). \quad (4.26)$$

Thus, the *cohort retention rate* at the end of period 1 is a weighted average of the individual retention probabilities.

To derive the three-period survival rate, sv_3 (the fraction of the $N_{1,1}$ hires who stay into period 3), we take the proportion who survive at the end of period 1 (defined by Equation 4.14) and determine the proportion of this group that will be promoted and stay at the end of period 2 and the proportion that will not be promoted but will still stay. Let $w_{1,2}$ be a discrete indicator variable equal to 1 if the individual occupies grade i at the end of period 2 and 0 otherwise. Then the fraction of entrants who survive to the start of period 3 is

$$sv_3 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[w_{1,2} \Phi\left(\frac{G_{1,2}^*}{\sigma_\varepsilon}\right) + w_{2,2} \Phi\left(\frac{G_{2,2}^*}{\sigma_\varepsilon}\right) \right] \Phi\left(\frac{G_{1,1}^*}{\sigma_\varepsilon}\right) dF(\tau, \alpha). \quad (4.27)$$

The aggregate conditional retention rate in the second period is sv_3 / sv_2 . Survival to later periods is defined analogously.

The model incorporates two noteworthy characteristics of cohort retention. One is the widely observed tendency for the conditional retention rate, sv_t / sv_{t-1} , to rise with job tenure, t , despite the fact that

there is a new drawing of the random shock, $\varepsilon_{i,t}$, in every period and grade. The reason for the rise is that those with low values of τ are more likely to leave early, leaving the group that remains composed of individuals who, on net, have stronger tastes for the current employer and are therefore more likely to stay in subsequent periods than the initial entry cohort would have been.¹³

The second characteristic is that conditional retention depends on past pay policy as well as on future policy. That is, the model is both backward-looking and forward-looking. For example, suppose the organization pays a one-period bonus at period $t - 1$. Retention will increase in period $t - 1$ and in earlier periods, primarily among those who are just on the margin of a stay-leave decision. Absent any future pay changes, some of these marginal individuals will leave at the end of the next period, thereby reducing that period's conditional retention rate.¹⁴ Conversely, a reduction in this period's pay will lower this period's retention rate but increase future retention rates.

¹³As a technical note, conditional retention rates do not rise with job tenure if $dF(\tau, \alpha)$ is degenerate (i.e., everyone has the same ability and tastes).

¹⁴Evidence from military reenlistment studies supports this conclusion. Goldberg and Warner (1982) find that the probability of a second reenlistment is negatively related to the size of the bonus for the first reenlistment.

THE EFFECT OF POLICY ON INDIVIDUAL DECISIONMAKING

In focusing on the decisions facing the organization, we noted in Chapter Two that the organization must solve both macroeconomic problems—designing policies that will enable it to balance its actual inventory (*supply*) with the inventory it requires (*demand*)—and microeconomic problems—designing policies that will motivate effort and induce the most able workers to seek the higher-level positions. Individual responses to policy form the basis of the micro problem, which is the subject of this chapter. We first examine the effect of various organizational compensation and personnel policies, including promotion contests, up-or-out policies, involuntary separation, and intragrade pay. Then we examine the effects of retirement policy on decisionmaking.

THE EFFECTS OF ACTIVE-DUTY POLICY

Increasing the Promotion Probability

Promotion has a direct effect on individual retention decisions. Individuals with a greater chance of promotion are more likely to stay because the expected gains to staying in Equation 4.19, and hence the individual retention rate, increase with the promotion probability, $\pi_{i+1,t+1}$.

Our model also reveals a more subtle, secondary effect of promotion contests that reduces the individual retention rate of employees in lower grades: As more individuals remain in grade i in response to the positive retention effects of a higher probability of promotion to the next grade, the grade-specific survival rate increases, leaving

fewer vacancies available for those in lower grades. Thus, the future sequence of aggregate promotion probabilities $\{\pi_{i,t}\}$ for those facing a retention decision in grade i is lower. Clearly, changes in the promotion rate of individuals in one grade have second-round effects on the retention rates of those in other grades.

Increasing Intergrade Pay Spreads

Increases in the intergrade pay spread also increase individual retention rates. From Equation 4.19, we find that $G_{i,t}^*$ increases with the intergrade pay spread weighted by the promotion probability, i.e., $\pi_{i+1,t+1}(m_{i+1,t+1} - m_{i,t+1})$. Promotion policy and intergrade pay spreads also affect current effort supply: Individuals vary effort in each period to satisfy Equation 4.24. We have seven results. First, like Lazear and Rosen (1981), Nalebluff and Stiglitz (1983), and O'Keefe, Viscusi, and Zeckhauser (1984), we find that larger pay differentials encourage effort; in Equation 4.24, effort rises with the period 3 intergrade pay differential $(m_{3,3} - m_{2,3} - \psi_{2,3}b_{2,3})$.¹

Second, in our multiperiod framework, current work effort also depends on all future payoffs to current work effort, not just on the next-period differential. For example, period 2 effort rises with $\beta(V_{3,3}^* - V_{2,3}^*)$, which depends on future pay differentials.

Third, a given pay differential (or present value of future pay differentials) induces less work effort in lower hierarchical levels than in higher ones because individuals do not work as hard when they have a lower probability of staying. The marginal return to effort increases with the individual's probability of staying ($\Phi_{2,2}$ in Equation 4.24). Individuals with stronger tastes for the organization have a higher probability of staying, because an increase in τ increases $\Phi_{2,2}$, and they will supply more work effort than those with a smaller value of τ .² The model predicts that aggregate work effort will be lower among those with less tenure because, *ceteris paribus*, workers with

¹In our model, the pay differential includes the intragrade bonus term weighted by the probability that the individual earns it, $\psi_{2,3}b_{2,3}$.

²Drafted armies are notorious for shirking; the result here makes clear why: Draftees probably perceive that the likelihood of remaining beyond the initial enlistment is low, in which case their work effort is likely to be low as well.

less tenure have a lower mean taste for the organization and therefore a lower mean probability of staying.

Fourth, the effect of a given pay spread on work effort depends on the rate at which promotion chances improve with effort, $\partial\pi_{3,3} / \partial e_{22}$.³ Individuals are likely to work harder when extra effort improves their promotion chances markedly than when it improves them only a little. Consequently, a given pay spread will have a greater effect when $\partial\pi_{3,3} / \partial e_{22}$ is larger. The size of $\partial\pi_{3,3} / \partial e_{22}$ depends, in turn, on the aggregate promotion rate. When the aggregate rate is either very high (promotion occurs almost with certainty) or very low (the individual has little hope of promotion), then $\partial\pi_{3,3} / \partial e_{22}$ will be small as well. The derivative $\partial\pi_{3,3} / \partial e_{22}$ is largest when $\pi_{3,3}$ equals 0.5. Therefore, as π departs from 0.5, the value of e_{22} that solves Equation 4.24 will be lower.

The size of $\partial\pi_{3,3} / \partial e_{22}$ also depends on the variance of ξ_{it} relative to the size of k_{i1} and k_{i2} . In other words, the rate at which effort improves the likelihood of promotion also depends on the relative importance of random factors ("noise" or "luck") in the promotion contest. A given pay spread will have more of an influence on effort when luck is a less-important factor in determining promotions (i.e., when k_{i2} / σ_{ξ} is larger). Note that because promotion in the lower ranks is based on explicit criteria or standards, luck has only a small influence on promotion outcomes in the lower ranks. Luck assumes a larger role as individuals progress through the upper ranks. Having the "right" assignment, working for the "right" mentor, etc., figure larger in the promotion outcomes at higher levels. Thus, a given intergrade pay spread will have less effect on effort in higher grades than in lower grades.

Fifth, $\partial\pi_{3,3} / \partial e_{22}$ will be bigger the more contestants there are. Intuitively, the more contestants there are, the more chance a given contestant has to surpass others by working harder. Effort has a larger payoff the larger the scale of the contest is. This scale effect provides another heretofore-unrecognized reason for a skewed inter-rank pay system in a triangular hierarchy.

³Although the argument in this paragraph refers to intergrade pay spreads, a similar argument can be made about intragrade merit-pay increases and the role of $\partial\psi_{2,3} / \partial e_{22}$ in motivating period 2 effort.

Finally, intergrade differences in rank-specific nonpecuniary factors will also affect effort. The privileges, power, and status associated with higher grades are a form of nonpecuniary compensation.⁴ We incorporated these factors in our model via the $m_{i,t}^n$ defined in Chapter Four. (Recall that *pay* is the sum of pecuniary and nonpecuniary benefits, i.e., $m_{i,t} = m_{i,t}^n + m_{i,t}^p$.) We have two results, the sixth and seventh. The sixth is that the more important tastes are for the $m_{i,t}^n$, the more retention varies with grade, irrespective of differences in monetary compensation. The seventh is that the larger the differences are in the $m_{i,t}^n$, the *smaller* are the differences in monetary compensation required to solicit a given amount of work effort. For example, with these factors explicitly incorporated into the model, Equation 4.24 becomes

$$\beta\Phi_{2,2}\left[\left(\frac{\partial\pi_{3,3}}{\partial e_{22}}\right)\left((m_{3,3}^n + m_{3,3}^p) - (m_{2,3}^n + m_{2,3}^p) - \psi_{2,3} b_{2,3} + \beta(V_{3,3}^* - V_{2,3}^*)\right) + (1 - \pi_{3,3}) b_{2,3} \frac{\partial\psi_{2,3}}{\partial e_{22}}\right] + \beta\Phi_{2,2}^1\left[G_{2,2}^*\Phi_{2,2}^2 + \sigma_\varepsilon\Phi_{2,2}^2\right]\frac{k_{12}}{\sigma_\xi} - \theta_1 e_{22}^{\theta_2-1} = 0 \quad (5.1)$$

The larger the nonpecuniary differential ($m_{3,3}^n - m_{2,3}^n$) is, the *smaller* the monetary pay differential between level 2 and level 3 needs to be to provide the same effort incentive.⁵

⁴Higher-ranked individuals may gain differential treatment by the organization. For example, individuals may get allowances for housing. In the military, higher-ranking personnel have access to better on-base housing, may receive faster and better treatment in military hospitals, get preferred starting times at the base golf course, and so forth.

⁵This result does not imply that pecuniary and nonpecuniary pay are entirely fungible, given that tastes for nonpecuniary benefits are unlikely to be uniform. Some individuals would prefer cash compensation because they have little use for some nonpecuniary benefits (e.g., young and healthy individuals are unlikely to use health benefits extensively). Similarly, some individuals prefer nonpecuniary benefits over cash because of the benefits' tax advantages. In either case, ignoring administrative costs, providing a uniform level of benefits is not efficient, because the organization

Up-or-Out Policy

Introducing up-or-out rules tends to reduce the gain to staying and, hence, the probability of staying. From Equation 4.19, an up-or-out rule that involuntarily separates those in grade i who have not been promoted to grade $i + 1$ at the beginning of t makes the probability of staying $\Phi_{i,t}$ at the end of t equal to zero and lifetime utility, $V_{i,t}^*$, equal to the value of leaving, $L_{i,t}$. If $L_{i,t}$ is less than what $V_{i,t}^*$ would have been in the absence of such a rule, then individuals in grade i at the end of $t - 1$ are less likely to stay. This condition is more likely to be satisfied when, in the absence of the rule, the organization defers a large part of a worker's compensation into the future because, in this case, workers who fail to be promoted (and thus would be subject to the rule if it is introduced) will tend to expect to get more by leaving than by staying: Current compensation is less if some is deferred into the future; thus, outside opportunities will seem relatively more attractive, so the workers leave.

Up-or-out rules are more likely to adversely affect the retention decisions of the least able individuals.⁶ As long as more able workers perceive a higher promotion probability and thus a lower chance of facing the up-or-out high-year-of-tenure point, this negative retention effect is strongest for those who view themselves as having the greatest chance of facing the up-or-out rule, i.e., the least able. From an organizational-policy standpoint, it is precisely these less able workers that the firm should prefer to release.

Up-or-out rules also affect effort supply. Given their effect on effort in Equation 4.24, up-or-out rules are effectively a method for increasing intergrade pay spreads. Reducing $V_{2,3}^*$ to the value of $L_{2,3}$ increases the spread $\beta(V_{3,3}^* - V_{2,3}^*)$. Thus, an individual at the margin must increase effort in period 2 if Equation 4.24 is to be satisfied.

could make personnel better off at the same cost by either paying entirely in cash or by varying the mix of cash and benefits according to tastes.

⁶By assumption, individuals know their own ability level. In a more general framework, in which information on ability is symmetric and individuals also learn their ability over time, up-or-out rules would affect those who lack confidence and believe that they are low ability (regardless of whether they are, in fact, less able).

Minimum-Performance Standards

Minimum-performance standards, whereby workers are fired if they fail to meet them, raise current effort if workers have a positive gain to staying because individuals must expend some work effort to meet the retention hurdle. When the probability that the individual stays depends on meeting the minimum standard, the expression in the second term in Equation 4.24 is nonzero, i.e., $\beta\phi_{2,2}^1(G_{2,2}^*\Phi_{2,2}^2 + \sigma_\varepsilon\phi_{2,2}^2)k_{i2} / \sigma_\xi \neq 0$. All terms other than $G_{2,2}^*$ are always positive. Therefore, if $G_{2,2}^*$ is positive, then this expression is positive. Thus, increasing effort increases the probability that the individual will be allowed to stay, $\Phi_{2,2}^1$, which increases the individual's expected utility at the beginning of period 2. The rise in expected utility means that the individual is willing to supply greater effort during period 2, at least to the point where MDE equals MRE again.

The effect of minimum-performance standards depends on whether the individual views staying as worthwhile, i.e., on whether $G_{2,2}^*$ is positive. One way for the organization to ensure that the gain to staying is positive is to defer some compensation into the future, either in the form of a bonus or retired pay, as we discuss below. This result is consistent with the general principal-agent literature, which finds that firing workers with unsatisfactory performance and making some compensation contingent on staying (and thus satisfactory performance) increases effort.

A higher performance standard has an ambiguous effect on the individual's retention decision. If individuals know that they must expend additional effort in each future period to be retained, they will perceive a lower gain to staying because of the disutility of the extra effort required to surpass the minimum standard; therefore, they will be less likely to stay. But as with an up-or-out rule, minimum-performance standards have a greater effect on less able individuals or individuals who have a high disutility of effort, because they are more likely to be affected by the standard. By inducing separations of low-ability/-effort individuals, minimum standards reduce the probability of promotion and therefore reduce the retention of high-ability/-effort individuals because the competition becomes tougher as the quality of the promotion pool rises when low-ability/-effort individuals leave. But offsetting these declines in retention is the increase in

the number of available promotion opportunities, which increases the retention of higher-ability/-effort individuals.⁷

To capture the terminal-grade idea from Chapter Four, we let $\pi_{i+1,t+1}$ equal zero in Equation 4.19 and $\Phi_{i,t}$ be the joint probability that an individual wants to stay and the organization wants to keep him or her. We focus solely on the involuntary-separation policy for now and set $\psi_{i,t+1}$ equal to zero. Using Equation 4.24 as an example, individuals expend work effort in period 2 to the point where

$$\beta\phi_{2,2}^1 \left(G_{2,2}^* \Phi_{2,2}^2 + \sigma_\epsilon \phi_{2,2}^2 \right) \frac{k_2}{\sigma_\xi} - \theta_1 e_{22}^{\theta_2-1} = 0 . \quad (5.2)$$

When $G_{2,2}^*$ is negative, this equation will be satisfied at some positive level of effort as long as $\sigma_\epsilon \phi_{2,2}^2 > G_{2,2}^* \Phi_{2,2}^2$.

Intracgrade Contingent Pay

Another intracgrade policy option is to offer a bonus, $b_{i,t}$, to those who surpass intracgrade performance standards.⁸ Again, we let $\pi_{i,t}$ equal zero. Using Equation 4.24 as an example, an individual supplies effort in period 2 to the point where

$$\beta\Phi_{2,2} \left(b_{2,3} \frac{\partial \psi_{2,3}}{\partial e_{22}} \right) - \theta_1 e_{22}^{\theta_2-1} = 0 . \quad (5.3)$$

⁷The presence of an involuntary-separation policy provides yet another reason for retention to rise with tenure. The policy induces low-ability and/or low-effort individuals to leave early. As this sorting occurs, the incidence of separations for failure to meet minimum-performance standards is likely to diminish. Beyond some period of service, there may be no separation for failure to meet minimum-performance standards, even when the standards are quite effective. We thank Sherwin Rosen for this point.

⁸Of course, a more general policy would be to combine the carrot of a bonus with a stick of involuntary separation. Thus, the bonus policy could be combined with the minimum-performance standard described above. In the text, we focus on the carrot and assume that the organization does not use involuntary separation so that individuals in period 2 in grade 2 (in Equation 4.24) have $Pr(-\xi_2 < -E_{\min(2,2)}^* + k_{11}\alpha + k_{12}e_{12})$ equal to zero.

The first term is the (discounted) bonus weighted by the product of the effect of current effort on the probability of receiving it and the probability of staying to get it. Clearly, effort expands with the responsiveness of the probability of bonus receipt to work effort, $\partial \psi_{2,3} / \partial e_{22}$, the size of the bonus, $b_{i,t}$, and the individual's likelihood of staying.⁹

THE EFFECT OF THE STRUCTURE OF RETIRED PAY

Our framework enables us to derive the effect on retention and effort decisions of different aspects of defined benefit pension plans, such as the benefit formula, vesting provisions, and provisions for early retirement, the "normal" age of retirement, cost-of-living adjustments, and mandatory retirement.

For the retirement system to affect stay-leave decisions, it must affect the gain to staying. Consider the retirement decisions at the end of periods 2 and 3. From Equation 4.18, $[G_{i,3}^* = \tau + m_{i,4} + \psi_{i,4}b_{i,4} + (\beta R_{i,4} - R_{i,3}) - (C_{i,3} - \beta C_{i,4}) - Z(e_{i,4})]$ is the gain to staying at the end of period 3. The effect of the retirement system operates through the size and functional form of the term $(\beta R_{i,4} - R_{i,3})$ relative to the other terms in the equation, such as direct pay. For workers at the end of period 2, the expected gain to staying is from Equation 4.19:

$$\begin{aligned} G_{i,2}^* &= \pi_{i+1,3} [\tau^m + m_{i+1,3} + \beta V_{i+1,3}^* - Z(e_{i,3})] + (1 - \pi_{i+1,3}) \\ &\quad \times [\tau^m + m_{i,3} + \psi_{i,3}b_{i,3} + \beta V_{i,3}^* - Z(e_{i,3})] - R_{i,2} - C_{i,2} - \Gamma_{i,2} \\ &= \beta [\pi_{i+1,3} \Phi_{i+1,3} G_{i+1,3}^* + (1 - \pi_{i+1,3}) \Phi_{i,3} G_{i,3}^*] \\ &\quad + \beta E(R_3) - R_{i,2} - X - Z(e_{i,3}) - C_{i,2} - \Gamma_{i,2}, \end{aligned} \quad (5.4)$$

⁹These results are, of course, standard in the literature (see Lazear and Rosen [1981], O'Keeffe, Viscusi, and Zeckhauser [1984], and Nalebluff and Stiglitz [1983]). The bonus considered here is not to be confused with enlistment and reenlistment bonuses paid by the U.S. military, which are offered "before the fact" to encourage people to join or stay and which are not linked to performance.

where we use Equation 4.19 and where

$$X = \pi_{i+1,3} [m_{i+1,3} - (m_{i,3} + \psi_{i,3} b_{i,3}) + \beta(C_{i+1,3} - C_{i,3}) \\ + \beta(\Gamma_{i+1,3} - \Gamma_{i,3}) + \beta\sigma_\epsilon \Phi_{3,3}(\phi_{i+1,3}^2 - \phi_{i,3}^2)]$$

$$E(R_3) = \pi_{i+1,3} R_{i+1,3} + (1 - \pi_{i+1,3}) R_{i,3} .$$

The term $E(R_3)$ is the expected retired pay if an individual leaves at the end of period 3. The retirement system's effect on the stay-leave decision at the end of period 2 is captured by

$$\beta[\pi_{i+1,3} \Phi_{i+1,3} (\beta R_{i+1,4} - R_{i+1,3}) + (1 - \pi_{i+1,3}) \Phi_{i,3} (\beta R_{i,4} - R_{i,3})] \\ + \beta E(R_3) - R_{i,2} . \quad (5.5)$$

Equation 5.5 also depends on the probability of promotion and the probability that the individual stays at the end of period 3.

These equations show how retired pay enters into the retention decision; however, it does not necessarily follow that retired pay will alter individual stay-leave decisions, for two reasons. First, workers will be unaffected by the size or structure of retired pay at the end of period 3, when retired pay is actuarially fair (and workers are fully vested), because $\beta R_{i,t} - R_{i,t-1} = 0$ in this case.¹⁰ Second, even without actuarial fairness (whereby benefits have no effect on retention because they are the same regardless of the retirement date, so that $\beta R_{i,t} - R_{i,t-1} \neq 0$), the effect of the retirement system will be neutral if workers are fully aware of, and can take into account, the one-for-one trade-off between their pay in period 4, $m_{i,4}$, and the difference $\beta R_{i,4} - R_{i,3}$ that occurs through the organization's budget constraint: Any additional accumulation in retired pay earned by deferring retirement will be viewed as being offset by reductions in regular pay. Therefore, for the distinction between *retired pay* and *regular pay* to

¹⁰For retired pay to be completely neutral, the individual's rate of time preference, i.e., β , where $\beta R_{i,t} = R_{i,t-1}$, and the employer's discount rate must be the same.

be meaningful, workers must not be aware of (or are unable, because of capital market imperfections, to take account of) a direct trade-off between pay types.

However, if retired pay has a non-neutral influence, then both discounting and promotion influence its retention effect, as Equation 5.5 illustrates. A decline in the discount factor β (which is the same as an increase in the personal discount rate) reduces the effect of a given increment in retired pay, such as $\beta R_{i,4} - R_{i,3}$, on the retention decision. A rise in the worker's perceived promotion rate increases the influence of retired pay on retention. Once retired pay is allowed to influence decisionmaking, pension structure and provisions can have either a positive or a negative effect, or both. We examine the effect of some of the key aspects of retired pay in the following subsections.

If retired pay is not offset by pay policy and it is not actuarially neutral, it will also affect individual effort decisions through the expected gain to staying $G_{i,t}^*$. As is shown in Appendix D, under certain circumstances, the effect of $G_{i,t}^*$ on effort supply is positive. Thus, a change in retirement policy that increases $G_{i,t}^*$ will also increase effort in period t . When staying is viewed as better, individuals then place greater weight on increases in effort, which also increases their probability of staying. In what follows, we assume that $G_{i,t}^*$ and effort are positively related. Thus, by examining the effect of pension provisions on $G_{i,t}^*$ in the following subsections, we can infer the effects of such provisions on both retention and effort.

Vesting Provisions

Vesting provisions stipulate how and when workers earn the right to claim their retirement benefits. Under "cliff-vesting," workers have zero rights to their pension accrual before a certain age or year of service and have 100 percent rights after that time. Under *graduated vesting*, workers earn some portion of their rights at a certain age or year of service; this portion rises to 100 percent with time. Vesting stipulates only the *right* to claim the benefit. A 35-year-old worker who is 100 percent vested in his or her pension plan has the right to claim the benefit at age 65, for example, but *not* at age 35. Being

vested does not necessarily mean that the benefit can be claimed at any time.

The retention effect of vesting depends crucially on the relationship between the vesting age or year of service and the pension-taking age or year of service. In the military, these are the same, causing some confusion about what the effects of vesting really are. In the private sector, the difference between the two dates can be enormous. Under the 1986 Tax Act, private-sector pensions with cliff-vesting must, beginning in 1989, be fully vested at five years of service (and those with graduated vesting must fully vest at seven years). Even with an early-retirement age of 55, the difference between the pension-taking age and the vesting age is 30 years for a 25-year-old worker with five years of service.

The effect of cliff-vesting on the stay-leave decision is illustrated by considering the retention effect of the retirement system at the end of period 2 (which is given formally in Equation 5.5). Consider two cases: (1) when the pension-taking and the vesting dates are the same (as in the active-duty military system) and (2) when the pension-taking date exceeds the vesting date (as in the reserve retirement system and most civilian pension systems).¹¹ In the first case, cliff-vesting improves retention relative to full vesting if Equation 5.5 is larger under cliff-vesting.¹² However, even if cliff-vesting affects retention positively, the effect is smaller for younger workers because their tenure is farther from the cliff-vesting date. Young workers' future retirement benefits are discounted (1) by their rate of time preference β for more years and (2) by their lower probability of staying.¹³ In the latter case, cliff-vesting is not a sufficient condition for retired pay to have a large effect on the stay-leave decision, because the effect depends on the amount of time between the vesting

¹¹Military reservists who earn 20 years' worth of retirement credits are eligible for a pension beginning at age 60.

¹²Or, more formally, because under cliff-vesting, $R_{i,2}$, $R_{i+1,3}$, and $R_{i,3}$ are zero and $R_{i+1,4}$ and $R_{i,4}$ are positive in Equation 5.5, retention improves under cliff-vesting if $\beta[\pi_{i+1,3}\Phi_{i+1,3}(\beta R_{i+1,4} - R_{i+1,3}) + (1 - \pi_{i+1,3})\Phi_{i,3}(\beta R_{i,4} - R_{i,3})] + \beta E(R_3) - R_{i,2}$, which is the retention effect of the retirement system under full vesting, exceeds $\beta[\pi_{i+1,3}\Phi_{i+1,3}\beta R_{i+1,4} + (1 - \pi_{i+1,3})\Phi_{i,3}\beta R_{i,4}]$, which is the retention effect of the system under cliff-vesting.

¹³As shown below, the pension formula can exacerbate the smaller effect of cliff-vesting as age decreases.

and the pension-taking date. Cliff-vesting has a smaller effect on retention the earlier the vesting date is relative to the pension-claiming date, because more discounting occurs.¹⁴

These results explain why military vesting creates such a strong retention effect for those with between 10 and 19 years of service: The vesting date and the date when the pension stream may be claimed are the same. Although these soldiers are relatively young in age, they "act" like older, more senior workers in the private sector. The retention effect for these soldiers would be substantially less if they could begin claiming their pension benefit only at age 65 rather than at 40 (roughly). The strong retention effect associated with military vesting is generated by both its 20-year cliff and the pension-eligibility-date provision.

The Pension Formula

The pension formula may also alter the gain to staying. Under conventional formulas—such as the military's formula—benefits are proportional to both years of service and wages. Conventional formulas generally use the worker's average wages in the final years of service for two reasons. First, if wages rise with inflation, they index pension wealth to the inflation rate, as of the retirement date. The worker's pension wealth is increased, especially when inflation is high, and the administrative costs of making ad hoc cost-of-living adjustments are reduced. Second, if wages rise with productivity or level in the organization, such formulas index pension wealth to the worker's career performance.

To analyze how conventional plans affect retention, we abstract from the effects of vesting by assuming that all workers are fully vested. We also assume, without any loss of generality, that D , the date of death in Equation 4.1, equals 8. We also let the retirement annuity, $r_{i,t}$, be given by a conventional formula, $r_{i,t} = \psi t(m_{i,t} + \psi_{i,t}b_{i,t})$,

¹⁴More formally, if workers are cliff-vested at the end of period 3, then Equation 5.5 becomes $\beta[\pi_{i+1,3}\Phi_{i+1,3}(\beta R_{i+1,4} - R_{i+1,3}) + (1 - \pi_{i+1,3})\Phi_{i,3}(\beta R_{i,4} - R_{i,3})] + \beta E(R_3)$, since $R_{i,2} = 0$. But, if individuals are cliff-vested later, at the end of period 4, then Equation 5.5 becomes $\beta^2[\pi_{i+1,3}\Phi_{i+1,3}(\beta R_{i+1,4} - R_{i+1,3}) + (1 - \pi_{i+1,3})\Phi_{i,3}(\beta R_{i,4} - R_{i,3})] + \beta^2 E(R_3)$, which is clearly smaller.

where v is a constant multiplier, t is years of service, and $(m_{i,t} + \psi_{i,t}b_{i,t})$ is pay plus the expected intragrade bonus. Thus, given $D = 8$, the expected present value of retired pay at the end of period 2 is

$$R_{i,2} = r_{i,2}(1 + \beta + \beta^2 + \beta^3 + \beta^4 + \beta^5) = 2v(m_{i,2} + \psi_{i,2}b_{i,2}) \sum_{i=3}^8 \beta^{i-3}. \quad (5.6)$$

The retention effect of the retirement system at the end of period 2 is captured by the term $\beta E(R_3) - R_{i,2}$ in Equation 5.5. For $D = 8$, this term equals

$$\begin{aligned} -2v(m_{i,2} + \psi_{i,2}b_{i,2}) + v \left\{ 3 \left[\pi_{i+1,3}(m_{i+1,3} + \psi_{i+1,3}b_{i+1,3}) \right. \right. \\ \left. \left. + (1 - \pi_{i+1,3})(m_{i,3} + \psi_{i,3}b_{i,3}) \right] \right. \\ \left. - 2(m_{i,2} + \psi_{i,2}b_{i,2}) \right\} \sum_{i=4}^8 \beta^{i-3}. \quad (5.7) \end{aligned}$$

This term increases with increases in the intergrade pay spread, the probability of earning the intragrade contingent pay, the probability of promotion, and the intragrade bonus. Similarly, future random shocks, tastes, and organizational policy will also affect the retention decision. Note that a positive retention effect of a conventional plan also translates into a positive effect on effort supply, because effort rises as the gain to staying rises.

The retention effect of a conventional plan is particularly strong because each additional year of service increases pension wealth in two ways.¹⁵ First, holding wages constant, an additional year directly increases pension wealth via t in $r_{i,t}$. Second, each additional year increases pay, either through intragrade or intergrade increases. Workers who stay increase their pension wealth because they earn higher pay with each year of service and they accumulate more years of service, which multiply this higher pay. Leaving the organization

¹⁵Lazear (1985) shows that, in fact, individuals are *too* motivated to stay. As a result, conventional pension formulas are inefficient.

delinks regular and retired pay: The worker receives a lower total pension at retirement because final years' wages in the current and next job are multiplied by fewer years of service. By staying, final years' wages are multiplied by more years of service. This effect occurs regardless of worker seniority, but it is strongest for older workers who would be hurt the most by fewer years of service at retirement.

Conventional plans can motivate effort and retention among *fully vested* workers. Vesting is not essential for the retirement system to influence individual decisions. The retention and effort-supply effects of conventional plans are stronger the steeper the wage profile is, or the more wages rise with seniority and/or productivity; the steeper the wage profile is, the greater is the cost of leaving under a conventional plan, and thus the larger are retention and effort supply.

Early-Retirement Provisions

Early retirement has become more common in the private sector, and the age of early retirement has been decreasing. To partially offset the longer payout period, most early-retirement provisions reduce benefits in a graduated manner. Full benefits are given at the "normal" retirement age. The military's retirement system offers graduated early-retirement benefits for those who entered service as of August 1986; individuals receive a reduced benefit after 20 years of service but a full benefit after 30 years. Below, we analyze the effect of early-retirement benefits on the gain to staying. We abstract from the confounding effect of vesting by assuming that all individuals are fully vested.

Individuals who leave at the early-retirement age receive a reduced pension plus the value of their outside opportunities. If they stay, they get their current wages plus the option to receive higher wages and a more lucrative pension in the future. The retention effect of the pension depends on how benefits are reduced for early retirement. More specifically, when early-retirement benefits are not actuarially fair, the effect of early retirement provisions depends on whether the present value of benefits, $R_{i,t}$, is greater at earlier or later retirement ages, i.e., on the relationship between $R_{i,t}$ and $R_{i,t+1}$ dur-

ing the early-retirement-age window.¹⁶ For example, compared with the pre-1986 plan, the post-1986 military retirement plan accelerates the growth in the rate of retirement benefits for service beyond 20 years and, therefore, all else being equal, encourages greater retention in the post-20 YOS years.

Early-retirement provisions encourage retention among younger workers who have yet to reach the early-retirement age. Whereas, for a given retirement age, pension benefits are less valuable to younger workers than to older ones because of discounting, early retirement increases the present value of benefits to younger workers by allowing workers to draw benefits at earlier ages. As a result, early retirement increases the cost of leaving and the incentive to stay.

What happens when the early-retirement age is reduced? The answer depends on how benefits vary with retirement age. If the benefits are actuarially fair, only the relationship between wages and outside opportunities matters to those who have reached or surpassed the early-retirement age. Individuals will leave earlier if the civilian sector is more lucrative. If the present value of benefits is skewed toward retiring later, retention may be unchanged for those at or beyond the early-retirement age. For younger workers, the answer is clear cut: Reducing the early-retirement age increases the present value of benefits and encourages retention and effort supply.

Conversely, deferring the age when retirement benefits can be claimed reduces the gain to staying, even for fully vested workers. If the length of time between the vesting date and the pension-taking date is large, even when all workers are fully vested, the effect of an individual's earned pension benefit on the gain to staying is smaller at the vesting date. Offering the right to claim a pension benefit—*vesting*—by itself, does not necessarily improve retention or work effort.¹⁷

¹⁶Note that beyond the normal retirement age, the present value of retirement benefits typically falls drastically to motivate workers to leave.

¹⁷Allen, Clark, and McDermed (1991) find empirical support for this result.

ORGANIZATIONAL DECISIONMAKING: SETTING ACTIVE PAY, RETIRED PAY, AND PERSONNEL POLICY

We now turn to the question of how the organization should set personnel and pay policy given its macro and micro goals and given individual decisionmaking. Because of technical requirements for personnel at each hierarchical level and because of limited lateral entry, the organization must also consider the management of personnel flows in designing pay and personnel policy. We discuss first this aspect of personnel management, then examine alternative pay and personnel policies.

MANAGING PERSONNEL FLOWS

Without lateral entry, individuals in the organization are valuable for what they produce in their current grades and for what they are capable of producing in higher grades in the future. Thus, associated with each worker is the value of his or her current productivity and a shadow value equal to what he or she could produce in a higher grade in the future. On the other hand, some individuals may perform adequately in their current positions but may be unsuitable for higher positions in the organization. The clogging of promotion opportunities that these individuals create is a problem when the firm requires a large, upward flow of personnel to staff the higher ranks. Put differently, there is a shadow price associated with filling a position with a particular individual equal to the negative spillover effect (i.e., span-of-control consideration) on promotion opportunities of having the position filled plus the forgone value of filling that position with someone who is potentially more capable.

The organization must access and train a large enough pool of workers in the lower ranks to maintain the flow of workers necessary to staff the upper ranks. The individuals in this pool must also have sufficient ability to staff these future higher-ranked positions; otherwise, the firm faces the chance that an upper-level vacancy will not be filled with a qualified individual. The larger are the positive downward-productivity externalities that arise from filling the upper positions with more able people, the greater is the shadow value of a more able individual and the more emphasis this organization must place on recruiting, identifying, and promoting the most able workers.

The flow required to sustain the upper ranks depends on several factors. One is the shape of the pyramid. The taller the pyramid, the larger is the required flow. But, the more ability affects productivity, the more important it is for the organization to advance only those who are qualified for higher-level jobs. Consequently, the more heterogeneous is the talent pool from which the organization hires, the more it will have to hire and sample before identifying those with the talent to advance in the organization (recall our assumption that ability is difficult to observe before hiring and is measured only with error in promotion contests).

These considerations lead us to the notion of "overstocking": Compared with a flat organization or one that has lateral entry, a hierarchical firm that limits lateral entry must hire more entrants and maintain a larger low-level labor force if it is to fill its upper-level positions with persons of a certain ability. When initial screening is imperfect, when the true ability of an applicant can be learned only over time, or when applicants' tastes for the organization are an "experience good" that is learned over time, the hiring or promoting of some individuals will turn out to be a mistake (or individuals learn they do not like the organization or job after all). A sufficient supply of qualified individuals must be available to minimize these mistakes.

The extent of overstocking depends in part on the steepness (i.e., whether tall or flat) of the pyramid: A higher required rate of flow among grades requires a larger new-entrant pool. In addition, the

more ability affects productivity in the higher ranks, the greater will be the payoff to hiring a larger pool of new entrants from which to sample and select for the upper ranks. Overstocking will be accentuated by the variance in ability in the new-entrant pool, because (for a fixed mean ability) the organization will have to sample more new entrants to identify the portion capable of moving up. Overstocking will also be affected by the organization's ability and cost of screening for talent and of learning about talent over time. In the spirit of the human capital models, the *cost of overstocking* is the firm-specific investment a hierarchical organization must make to grow its upper-level employees. This cost is largely avoided by firms with flat organizational structures and by firms that are free to hire lateral entrants.

On the other hand, because of its pyramidal structure, the organization does not need, or indeed want, all entrants to stay and be promoted. After hiring and sampling employees who are part of the overstock, the organization will inevitably find some to be unsuited for continued employment. Paradoxically, however, the firm does not want even those individuals whom it retained and "percolated" to the levels to which they were best suited to stay too long. The longer they stay at the top levels, the more the firm must slow promotion opportunities for lower-ranking personnel, i.e., the greater the shadow price associated with their employment. As we saw in Chapters Four and Five, slower promotion opportunities reduce the work effort and retention of lower-grade individuals. Therefore, there is some optimal length of stay in, and flow out of, the higher-level positions. The organization must generate some turnover at the higher levels, even among qualified personnel, to maintain effort and retention incentives at the lower levels. Thus, simultaneous with the organization's incentive to overstock personnel will be its incentive to induce individuals to leave.

These macroeconomic considerations, together with the microeconomic aspects of individual decisionmaking, will affect the design of recruiting and retention policy, separation policy—i.e., up-or-out rules, minimum-performance standards, and the retirement system—and pay policy. We turn to policy design next.

SETTING ACTIVE PAY

Intergrade Pay-Setting

The analysis in Chapter Five shows why the organization should set compensation within the hierarchy so that intergrade pay spreads increase from one level to the next. That is, it suggests that the hierarchical pay scale should be convex, or skewed upward, for the following reasons. If, as we argued above, the marginal productivity of effort, $\partial q(i, t, \alpha, e_{it}) / \partial e_{it}$, rises with grade, then the pay differential must increase with grade to provide individuals in the higher grades with an incentive to supply more effort. Second, in a hierarchy, the aggregate promotion rate π^* declines as grade increases. Grade differentials must increase to just maintain a constant-effort incentive.¹ Furthermore, the rising relative importance of luck in promotion outcomes reinforces the need for a skewed pay profile to maintain effort incentives. Another factor accentuating the need for a skewed profile is that, as personnel progress through the ranks, the number of remaining promotions (and therefore promotion payoffs) that can be earned falls. Therefore, each subsequent payoff must be larger to maintain effort supply. Higher pay in upper grades also increases the likelihood of retaining the most able individuals. Finally, widening the intergrade differentials accentuates the competition and increases the likelihood that upper-level positions are filled by more able workers.

Other considerations may temper the intergrade differentials that are required to generate the desired amount of work effort and ability percolation. Rank-specific nonpecuniary factors are one such consideration. The privileges, power, and status associated with higher grades are nonpecuniary rewards.² The larger is the difference in the

¹Lazear and Rosen (1981) discuss why corporations might pay presidents, say, twice as much as vice presidents. The president earns twice as much not because his or her productivity rises by a factor of 2 upon promotion but because, with a large number of vice presidents, each of whom has a low probability of promotion to president, such a differential is necessary to induce work effort among the vice presidents.

²Higher-ranked individuals may gain differential treatment by the organization. For example, individuals may get allowances for housing. In the military, higher-ranking personnel have access to better on-base housing, may receive faster and better treatment in military hospitals, get preferred starting times at the base golf course, and so forth.

nonpecuniary benefits across grades, the more compressed the intergrade differentials can be to elicit the same levels of effort and retention.

The effect of pay differentials on cooperation and teamwork is another consideration. Rosen (1992, pp. 234–235) observes that “if rewards are skewed too much, a kind of cut-throat competition arises. Then competitors take steps to make others look bad rather than making themselves look good. . . . Some happy medium must be struck here.” Lazear (1989) develops the formal analytics of this argument. A prediction from his analysis is that pay differentials will be narrower the greater the team aspect of production is, since the greater the team aspect is, the more opportunity coworkers have to direct effort toward sabotaging others rather than contributing positively to output. These arguments assume that supervisors are unable to observe malfeasant behavior and discipline it. Clearly, the proper skewness of the pay system is an empirical matter.

The management of personnel flows provide reasons that tend to reduce intergrade pay differentials. Without lateral entry, recruits must be able to fill both entry-level positions and higher-level positions in the future. Recruits’ abilities must exceed the productivity requirements of their current entry-level positions and the abilities of entry-level workers in other organizations that do allow lateral entry. Consequently, the closed personnel system necessitates that the organization “cream” the pool of eligible applicants, i.e., offer pay that seems overly generous in light of the organization’s entry-level requirements. Pay must also be set higher because the supply of recruits must also be large enough to satisfy the firm’s need to overstock. Finally, to ensure that a large enough pool of applicants is attracted, pay must appear overly generous because overstocking reduces promotion opportunities and thus a recruit’s expected pay stream. Since youth have relatively high personal discount rates, initial pay will be relatively more valuable than pay offered far out in the future, and the necessary pay stream to attract sufficient youth will emphasize entry pay over deferred pay. Of course, the organization only appears to be setting pay “too” high. If, instead, we define *productivity* in terms of shadow values (i.e., internal productivity), *entry pay* equals the expected value of the individual’s marginal product.

Figure 6.1 shows the relationship between productivity, pay, and ability in the hypothetical organization.³ Ability are measured on the horizontal axis, and productivity and pay are measured on the vertical axis. The line AA represents grade 1, the entry level. Here, variations in individual ability do not alter pay or productivity much. The lines BB and CC represent grades 2 and 3, respectively. Because of the positive spillover effect of the productivity of an individual in grade 2 on the productivity of someone in grade 1, variations in ability in grade 2 matter much more; BB is steeper than AA. The segment CC is the steepest because ability matters the most in grade 3.

A firm that allows lateral entry would hire an α_1 worker at a pay and productivity q_1 in level 1, an α_2 worker at a pay and productivity level q_2 for level 2, and an α_3 worker at pay and productivity level q_3 in level 3. However, a firm that bars lateral entry must get a sufficient number of level 1 workers to fill level 2 and level 3 positions in the future. Thus, some grade 1 workers must have ability exceeding α_1 and entry pay exceeding q_1 . In the extreme case when all level 1 workers are destined to become level 3 workers, the firm hires only high-ability α_3 workers. Entry pay would be at least q'_1 , level 2 pay would be at least q'_2 , and level 3 pay would be at least q_3 . In a less-extreme case, entry pay exceeds q_1 but may be less than q_3 .

Note that a firm that allows lateral entry pays q_1 , q_2 , and q_3 . The intergrade pay spread increases with level to ensure the best ability- and position-sorting. Without lateral entry, the skewness of the pay profile is reduced. In the extreme case outlined above, the hierarchy pays q_1 , q_2 , and q_3 . The difference $q_3 - q'_2$ is smaller than $q_3 - q_2$. However, the effort considerations in the hierarchy discussed above add more skewness into the internal pay system.

Another consideration, related to personnel flows, that flattens the intergrade pay profile is firm-specific capital that is created by the lack of lateral entry. Since replacement workers must be grown from within, the firm does not want senior workers in upper ranks to leave too soon. Pay must exceed what these qualified individuals can earn

³Figure 6.1 is derived from Rosen (1982). We abstract from individual effort decisions temporarily. We discuss the role of effort later.

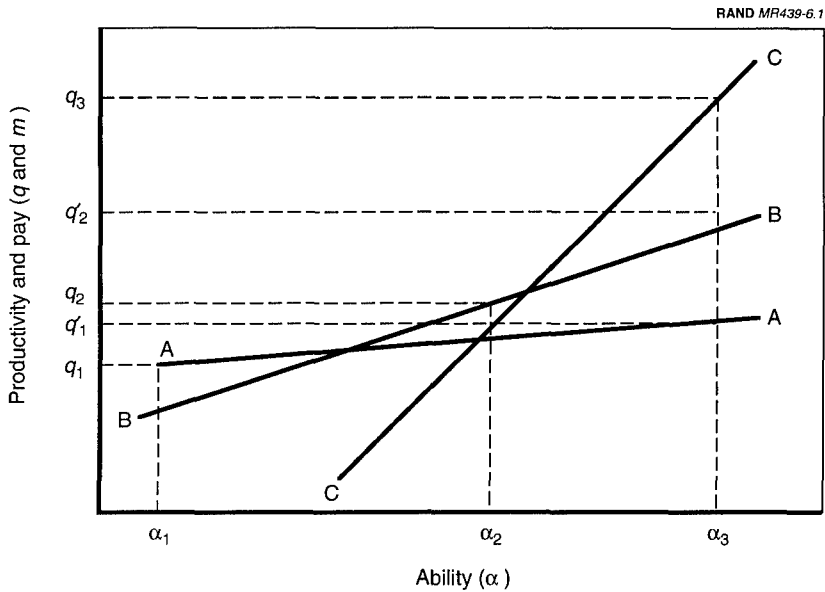


Figure 6.1—Ability, Productivity, and Pay in the Hypothetical Organization

elsewhere; otherwise, they will leave.⁴ But, to minimize cost, the firm must pay them less than their *internal productivity*, defined in terms of ability, which may exceed the productivity requirements of their current position (what the organization would pay if it permitted lateral entry) as in Figure 6.1. Abstracting from the role of effort, the pay structure will appear flat relative to the structure of internal productivity. The pay profile over an individual's career is shown in Figure 6.2 (where, for simplicity, we assume there are only two hierarchical levels).

To cream quality, the organization "overpays" junior workers, m_e , relative to their outside opportunities, q_e , but it pays them an amount

⁴An important constraint on the organization is its need to offer sufficient pay to attract and retain personnel. However, how it structures compensation within the organization relative to individual's productivity internally is another matter. The internal pay structure will be dictated by the organization's budget constraint and by its need to skew the pay system to provide effort and ability-sorting incentives.

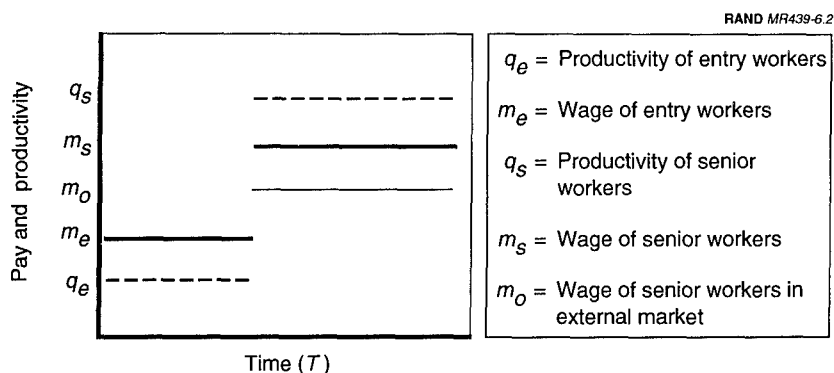


Figure 6.2—Hypothetical Pay and Productivity Profile over an Individual's Career

equal to their higher internal productivity. To ensure retention, it overpays senior workers, m_s , relative to their outside opportunities, m_o , but underpays them relative to their internal productivity, q_s , to minimize costs. Thus, to finance the creaming of the entry pool, the firm transfers the gains it gets from the senior workers—by paying less than internal productivity—to attract junior ones. It does not underpay recruits to minimize costs because such a policy would reduce the average quality of the entry pool. The specific capital created by the lack of lateral entry means that senior workers can be paid more than they can get elsewhere but less than their internal productivity.

The degree to which the pay profile is flattened depends on the retention and the size of promotion flows into the upper ranks, on the organization's overall personnel demand, and on the extent to which effort is reduced by a flattened pay system. If the intergrade pay profile is not properly skewed, effort incentives will be blunted. Further, a flattened profile reduces an individual's incentive to reveal his or her ability in the promotion contest. Therefore, the organization trades off the benefits of a skewed pay profile against the need to offer (and finance) sufficiently high pay to junior personnel.

Setting Intragrade and Intergrade Pay

The preceding discussion, together with the analysis in Chapter Five, also shows a role for intragrade pay spreads. When personnel flows require a significant flattening of the intergrade pay profile, effort incentives can still be achieved by making intragrade pay increases contingent on performance. More generally, when promotion rates are small or when individuals have reached their terminal grade, intergrade pay spreads lose their effectiveness as an effort motivator. Some individuals may spend a long time in a given grade, often times learning skills that are specific to their current and (possibly) future grades. Supervisorial skill is an example. Contingent pay within a grade, such as step increases, can be an effective means of providing motivation.

In providing incentives through intergrade and intragrade pay policies, the organization must take care to structure such policies so that those best matched in upper ranks do not find it in their interest to slum. Similarly, it must structure them so that those better suited in lower grades do not seek to climb. More able workers must view their expected utility as higher if they are promoted, whereas less able workers must view their expected utility as lower if promoted.

There are several ways to structure intragrade pay, as we show formally in Appendix B. The firm can make contingent pay a larger fraction of total expected compensation in higher grades: The firm lowers pay in the upper levels and sets the standard for attaining the intragrade bonus high enough so that the less able have a small chance of earning it. Thus, low-quality workers have a smaller incentive to climb to the upper grades. To compensate high-quality workers for the stiffer intragrade standards in the upper grades, the bonuses in the upper grades must be larger. Although more of pay is made contingent on performance in the upper grades, high-quality workers have a higher chance of earning the bonus. If their expected pay in the upper grades (including their expected bonus) exceeds their expected pay in lower grades and the gain to staying is positive, their incentive to slum is reduced.

Alternatively, the firm can set the minimum-performance standard in the upper grades sufficiently high that low-quality workers fail it and high-quality workers pass it. Low-quality workers are worse off

by being promoted if the promotion results in their being involuntarily separated for substandard higher-grade performance. Similarly, high-quality workers are better off by the promotion because they are allowed to stay to enjoy the higher pay associated with the higher grade.

When pay in higher grades is such that a larger fraction of (intragrade) pay is contingent on performance, fast-trackers are underpaid relative to their internal productivity. Intragrade bonuses effectively tilt the intragrade pay-experience profile upward because, initially, the firm underpays workers within each grade and later overpays them, through the bonus, relative to their internal productivity. Fast-trackers move up the lower ranks relatively quickly. Thus, the most able workers will probably not be in the lower grade long enough to reach the intragrade overpayment periods. Although the organization skews intergrade pay and offers steepened age-experience profiles within grades, high-ability workers will experience the high pay only when they reach the level in the hierarchy where they are the best suited—the top. However, as discussed below, the firm does not allow them to stay too long and must eventually mandatorily separate them or induce them to leave with retired pay.

SETTING PERSONNEL POLICY

Up-or-Out Promotion Policy

Chapter Five illustrates how up-or-out policies affect retention and effort decisions, but we have provided no justification for those rules. As Lazear (1991) notes, justifying such rules is problematic because one must explain why a worker is considered productive one day but is fired the next. If the individual has lower productivity, why not just pay the individual less? Further, up-or-out rules negate the possibility of *comparative advantage*: An individual may not be well suited for higher-ranked positions, but he or she might be well matched in his or her current grade. Similarly, those who are best suited for the upper leadership grades may be poorly matched in lower grades. For example, a great mechanic might be a poor supervisor, whereas a good supervisor might make a poor mechanic. The organization might do well to retain those well matched in lower grades and promote as quickly as possible those well matched in upper grades. Up-

or-out rules force out the good mechanics who make poor supervisors (or motivate them to leave before they are forced out). Furthermore, a personnel system that particularly emphasizes promotion in the lower grades as a means of career advancement (as does one that relies heavily on up-or-out rules) is probably more likely to motivate climbing, i.e., to provide an incentive to move up (even if unqualified) before being forced out by an up-or-out rule.

Previous studies provide some reasons for up-or-out rules, but, those reasons do not directly relate to hierarchical organizations. For example, Kahn and Huberman (1988) model up-or-out rules as contracts that solve a two-sided moral-hazard dilemma: Employers that promise to pay workers who successfully complete organization-specific training can profit *ex post* by declaring the worker to be unsuccessful as long as workers are unaware of their true value to the firm. But an up-or-out rule forces the firm that declares the worker to be unsuccessful to fire him, which it would not want to do if the worker is in fact successful. Thus, an up-or-out rule keeps the employer honest and it thereby gives the worker more incentive to invest in specific training.

The problem with the Kahn-Huberman explanation for up-or-out rules is that, by virtue of the fact that a hierarchy has a prespecified set of positions it must fill that is determined by technological considerations, it has no incentive to incorrectly declare good workers to be poor ones. Regardless of the quality of its workforce, it must promote someone to fill the vacancies in higher positions. An employer-side moral-hazard problem is therefore not present here.

In a hierarchical system without lateral entry, a more cogent explanation may involve considerations about personnel flows. Up-or-out rules may possibly solve the problems that can arise when the organization observes ability and/or effort imperfectly and inadvertently hires or promotes workers whom, in hindsight, it should not have but is unable to induce to leave using voluntary means. The retention of these workers imposes a cost on the firm: It reduces the promotion opportunities of workers in the grades below, which, in turn, reduces their effort supply and/or the retention of high-ability individuals in those grades. Since the organization sets its pay scales administratively and cannot bargain pay on an individual basis, it cannot cut pay in an attempt to induce an individual to leave without cutting the

pay of others. Doing so would wreck the incentive structure for the other workers.

The organization could demote unsuitably matched workers and thereby effectively cut pay rather than use up-or-out policy. However, to use demotions, supervisors must have an incentive to detect and report unsuitable performers as candidates for demotion. In a hierarchical organization, supervisors may have blunted incentives to do so. Supervisors are not residual claimants the way that owners of firms are. Greater efficiency among workers is not likely to profit the supervisors if they are not rated on the basis of the performance of the group that they supervise. Singling out poorly matched workers for demotion may not produce any gain to supervisors; in fact, supervisors may be better off by ignoring poor performance, because disciplining workers might create poor worker morale among the other workers, as well as poor worker-supervisor relations (thereby lowering the supervisor's utility). Up-or-out policy might be a faceless bureaucratic way for the firm to weed out poor matches and mistakes when supervisors have little incentive to demote them.⁵

THE ROLES OF RETIREMENT PAY

The discussion so far has defined *pay* generically as any form of compensation. Although earlier we symbolically distinguished between regular pay, $m_{i,t}$, and retired pay, $R_{i,t}$, the distinction between pay types was unnecessary because of the one-for-one trade-off between types for a given budget.

What are the purposes of retired pay? Does retired pay have a unique role that cannot be accomplished with other forms of compensation or other policy tools? The purpose of retired pay in a hierarchical organization without lateral entry is much different from the purpose of retired pay in an organization with it. Because workers in the lat-

⁵Even when supervisors are rated on group performance, their ratings are unlikely to rely exclusively on the group if they have no control over who joins. If workers are assigned to tasks and supervisors must take who they get, the organization can make supervisors better off by not fully penalizing them for bad assignments. Thus, their incentives to reveal poor matches may still be somewhat blunted. Also, in the military, the frequent rotation of personnel reduces the cost to supervisors of failing to give poor evaluations to marginal performers.

ter firms can be hired directly into positions to perform similar tasks year after year, such organizations are not as concerned with generating turnover of older employees to create advancement opportunities for new hires. As a result, retired pay is less a tool for managing personnel flows and providing work effort incentives and more a vehicle for providing workers with tax-sheltered savings opportunities. In fact, to the extent that the retirement benefits provided by these employers are offset by lower wages, retired pay need not affect their hiring or retention decisions (Lazear, 1988, 1990).

The lateral-entry constraint places the hierarchical organization without lateral entry in a much different situation. Such an organization must access and train large numbers of entrants before identifying for advancement those who have the talent to perform the higher-level tasks in the organization. It therefore wants to provide incentives for the most talented to stay and seek advancement and for others to leave after they discover that they are unsuitable for the upper-level positions. That is, it must provide the proper incentives for personnel to *self-sort*. Salop and Salop (1976) were the first to recognize the use of two-part compensation schemes as a self-selection device. As discussed in Asch (1984), one such two-part scheme is a system of (1) active pay and (2) deferred, retirement benefits that are paid only to those who achieve a certain rank and longevity. Delayed vesting of retired pay induces self-sorting, because only those who think that they can achieve the requisite rank and longevity will stay early on and others will leave. Deferred retired pay can also motivate work effort, especially when combined with minimum-performance standards for retention and up-or-out rules that prevent low-ranking personnel from staying long enough to collect retirement benefits.

However, the question arises, Why should retirement benefits be part of the self-sorting mechanism? After all, why not just pay a bonus to all who reach the requisite grade and years of service? The answer has to do with retired pay's role as a *separation* incentive. At some point the organization wants everyone, including the best personnel, to separate, even when they may still be *individually* very productive (i.e., their own productivity exceeds their pay). The longer individuals remain in the top positions, the slower will be the promotion rates for younger (and potentially equally able) personnel. Unless offset by changes in the structure of pay, reduced pro-

motion opportunities in the junior grades will discourage work effort in those grades and will cause those junior personnel with the best external opportunities (i.e., the *more able*) to leave. Without the proper inducement, the senior personnel may not want to leave voluntarily if their pay exceeds their best outside alternatives. Such is especially likely to be the case for those trained in the firm-specific skills.

Retired pay can be used to induce voluntary separations of senior personnel. For example, once personnel become vested in the immediate annuities provided by the military retirement system, they have a much reduced gain from staying and are therefore more willing to depart voluntarily. The retirement system, therefore, induces the separations needed to control the age or experience structure of the force and to maintain promotion flows for younger personnel.

Thus, retired pay is not *retired pay* in the conventional sense of the term—a convenient vehicle for transferring consumption from the present to the future. Rather, retired pay may have a role in managing the distribution of experience and grade across personnel.

There is, of course, no reason why the separations required to maintain personnel flows could not be accomplished with other policy tools, such as up-or-out rules. However, excessive reliance on involuntary separation to control the inventory of senior personnel could be bad for morale, adversely affecting recruiting, retention, and work effort. The adverse effects might require the payment of a “regret premium” to compensate for the prospect of involuntary separation. In addition, personnel faced with the prospect of involuntary separation are likely to engage in activities aimed at getting the policy relaxed (e.g., complaining to the personnel managers, or, in the case of the military, writing to congressmen about the “unfairness” of the policy). Should their complaints prove successful, the organization would be compelled to modify the experience distribution of their personnel in unproductive ways. After Milgrom (1988), we call such extra financial costs and productivity effects the *organizational influence costs* of mandatory separation.⁶ Separation pay is the elixir

⁶The organizational influence costs of the military's drawdown are apparent today. There is at present much discontent in the mid-ranks over the likelihood of mandatory separation.

that eases termination from the organization, and it weakens potential criticisms about the capriciousness or arbitrariness of policy. Although clearly expensive, a system that provides voluntary-separation incentives may actually be cheaper because the adverse productivity effects of a much older force or the regret premium that might be required to maintain the current (younger) experience distribution, while difficult to calculate, could be substantial.

The other purposes of retired pay are, of course, not unique. Motivating effort, improving retention, and inducing personnel to properly self-sort within the organization could be accomplished through an appropriately structured active-duty pay table and through other personnel policies. So if there is a distinctive (if not unique) purpose for retired pay in the closed hierarchical organization, inducing *voluntary* separations at the appropriate points, and thereby minimizing the influence costs that accompany involuntary separation, must be it.

STRUCTURING THE RETIREMENT SYSTEM

Setting the Vesting Date

Lazear (1986) shows that anything other than full vesting of all workers is inefficient. Incomplete vesting causes "leavers" to leave earlier than would be efficient, to undersupply effort, and to underinvest in human capital. It also causes "stayers" to stay too long, to oversupply effort, and to overinvest in human capital relative to the efficient amount. The firm cuts pay of all workers, be they stayers or leavers, to finance the retirement system. Thus, those who leave prior to the vesting date get a pay cut but no (future) retirement income. Since they receive less total compensation than their alternative employment, they leave earlier than they would under a fully vested plan. The pay cut of the leavers cross-subsidizes the retirement income of the stayers (who do achieve vesting). This extra gain causes the gain to staying to increase for the stayers: They stay too long.

However, in the context of a closed hierarchy, where knowledge about which entry workers are the stayers and which are the leavers is uncertain, incomplete vesting makes more sense for two reasons. First, when the firm limits lateral entry, sorting the stayers from leavers early on is important, because an upper-level vacancy cre-

ated by the separation of a leaver is costly to the organization. The organization wants leavers to leave early and stayers to stay longer. But incomplete vesting induces precisely this self-sorting among workers. In fact, the more delayed the vesting date is (holding expected benefits constant) so that more compensation is deferred to the future, the more self-sorting occurs. Not only do leavers leave early, they never join.⁷

Second, the cutting of active pay necessitated to finance the extra costs associated with an earlier vesting date may be problematic in a closed hierarchy. If the firm cuts entry pay, a lower-quality pool of entrants is attracted. If it cuts the pay of more senior personnel, the organization may cause the pay of some personnel to fall below what they can earn elsewhere, and retention falls. In the presence of firm-specific investments (created by the lack of lateral entry), poor retention imposes a cost on the firm—lower-ranked individuals must be promoted faster than desired, and new entry-level workers must be recruited to replace those promoted to higher grades.

This discussion suggests that a vesting date should be later, i.e., more incomplete, when it is more important to induce self-selection among the leavers and/or stayers and/or less important to have high active pay to attract the requisite ability pool and to retain workers embodying firm-specific investments. Thus, organizations with tall hierarchical structures, in which the lack of lateral entry imposes large vacancy costs when a leaver leaves, should probably vest workers later than flat organizations that can hire replacements—even for senior workers—from the outside labor market.

Setting the (Earliest) Retirement Date and (Early) Retirement Benefits

When should individuals be allowed to begin receiving their benefits? Private-sector firms generally set the retirement date (for fully vested workers) when workers become old—around 60 to 65—although earlier retirement beginning at age 55 is becoming more

⁷Lazear (1986, p. 268) notes that when sorting is important, incomplete vesting may have a role. The role of pensions as a self-selection device is investigated in Asch (1984).

common. The military retirement system sets the earliest retirement date at the vesting date—20 years of service—which generally can occur when individuals are in their late thirties and early forties.

Benefits should begin when it is optimal for workers to leave, because retired pay can act as severance pay and induce those who would prefer to stay to leave instead. The role of active pay in providing effort incentives, sorting ability, and attracting and retaining a qualified pool can cause active pay to deviate from what workers can earn outside the firm. Therefore, pay policy can distort the individual's separation decision. An individual the organization may want to separate may not be inclined to do so without the inducement of retired pay. However, the optimal time to leave will not be the same for all individuals. For some occupations or functional areas, a long career makes sense; youth and vigor are relatively unimportant for physicians and teachers, for example. The nature of other occupations would suggest an earlier retirement date; youth and vigor are key components of success in combat occupations, for example.

A PRELIMINARY EVALUATION OF THE CURRENT MILITARY PAY AND RETIREMENT SYSTEM

The purpose of our theoretical model is to evaluate the current and alternative military compensation systems. In Chapter Two, we listed some of the main criticisms with the military's active-pay table and retirement system. In this chapter, we begin to evaluate the current military compensation system, using our theoretical model to determine whether these criticisms have merit.

THE ACTIVE-DUTY PAY TABLE

Some of the main features of the active-duty basic-pay table are (1) small intergrade pay spreads; (2) its use for widely differing services and occupational groups; and (3) automatic within-grade increases instead of contingent intragrade pay increases. We discuss these features in light of the implications of our theoretical framework.

Pay Table Skewness

Consider first the skewing in the basic-pay table. Figure 7.1 shows the annual increase in basic pay that personnel receive when promoted to the rank shown (calculated at the average time in service for promotion to that rank). Pay increments due to promotion are small and relatively constant through the first five enlisted grades, then increase in the last four enlisted grades. The officer grade differentials are not skewed: A smaller promotion increment is given to

O-4 than to grades O-2 and O-3. But the promotion increments do widen thereafter.¹

Some skewness is apparent in the basic-pay table. Still, Figure 7.1 paints a picture of a compensation system with a relatively flat intergrade pay structure. The policy question here is, Are these differentials (coupled with differentials provided by the retirement system) sufficient to motivate personnel to work hard and to sort properly? Our theory identified a number of factors—such as nonpecuniary benefits and the constraints that the lack of lateral entry impose on the pay system—that lead to a relatively flat intergrade pay structure, factors that at the present time are not quantifiable. However, a partial answer to whether the system is producing the desired effects is to examine the flows of personnel through the system, by aptitude level.

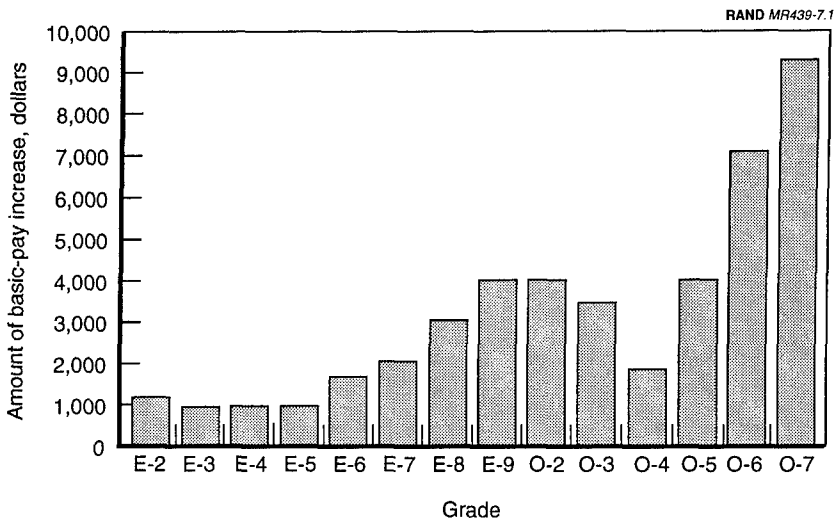


Figure 7.1—Annual Basic-Pay Increase (Average YOS at Promotion)

¹The Seventh QPMC (DoD, 1992b) recommended revisions to the basic-pay tables that will eliminate many of their anomalies, but their basic structure is not altered much. Importantly, the report recommended fundamental changes in the structure of allowances.

Table 7.1 provides rough evidence of how the system sorts personnel by their Armed Forces Qualification Test score, a measure of aptitude and trainability used by the military to screen recruits.² A one-quarter sample of all entrants into Army Infantry, Mechanical Maintenance, and Administration skills for FY 1974–1986 was tracked through FY 1990. We calculated the average AFQT score by rank and YOS as these cohorts flowed through the system. The resulting averages incorporate the effect of AFQT on retention and speed of promotion. Results for the FY 1974 cohort are displayed in Table 7.1.

Holding constant YOS, higher-AFQT personnel tend to occupy the higher pay grades. That is, high-AFQT personnel are, in fact, being promoted faster than low-AFQT personnel. However, because of lower retention of higher-AFQT personnel, the data reveal no tendency for average AFQT to rise with YOS. The average AFQT is about 51 in every YOS. Overall, the average AFQT increases slightly with rank. But the in-rank average AFQT tends to decline as YOS increases. The patterns for later-entering cohorts are similar: All display slight positive correlations between grade and AFQT but relatively flat average AFQT by YOS.³

This rough analysis indicates that in the enlisted ranks the system is at least not producing *adverse selection*, whereby the most able are leaving and the least able are staying and being promoted. But neither is it producing strongly positive selection effects. This analysis is, of course, limited, and we have no data whatsoever on how well officers are being sorted. Consequently, it would be hard to generalize too much.

²Although the AFQT score does not necessarily measure ability, it is thought to be a strong correlate of ability. Further, empirical evidence shows a strong correlation between AFQT and soldier performance on wartime outcome measures (see Orvis, Childress, and Polich [1992] and Winkler, Fernandez, and Polich [1992]).

³Perhaps most interesting is the tremendous variation in the quality of the entering cohorts. The entering average AFQT of the FY 1974 cohort was 51. The entering average began falling in the late 1970s, hitting 35 in 1980. As a result of improvements in pay for all service members and educational benefits targeted at the Combat Arms, the average increased steadily from 1981 to 1986. The FY 1986 cohort average AFQT was 58.

Table 7.1

Mean AFQT of Army Enlisted Personnel by YOS and Grade, FY 1974 Cohort

YOS	E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	TOTAL
<1	50	55	64	51	80				51
2	47	49	51	54	66	51	94		51
4	43	47	51	52	59	83	63		51
6	50	50	45	47	54	70	31		52
8	47	41	37	43	51	56	59		51
10	36		42	41	50	51	64		50
12	42		38	28	50	51	55		51
14			46	41	50	50	53	92	51
16	55			33	46	51	51	72	51
TOTAL	50	49	51	52	54	53	53	70	

Service- and/or Occupation-Specific Pay Tables

Our model gives several theoretical reasons why the level and structure of compensation might differ across occupations. The level of compensation might differ because of differences in the demand and supply of personnel in different occupations. The requisite skewness of the compensation distribution might also need to vary because occupations differ in the role of effort and ability in determining productivity at different grades and in the role of luck in determining the outcomes of promotion contests.

But whether these differences in the requisite skewness must be achieved using occupation-specific pay tables or by varying other forms of compensation by occupation, such as bonuses or job amenities, is an open question. In the past (and up until the draw-down), the military made wide use of bonuses to alter a recruit's occupational choices and to motivate individuals to reenlist. But the setting of these bonuses appears to reflect macro concerns (i.e., recruiting and retention) and not micro ones (effort supply and ability-sorting).⁴

⁴Several foreign militaries, such as the Canadian and British militaries, use compensation systems that explicitly vary by occupation.

Automatic Versus Contingent Within-Grade Increases

Since the services often have time-in-grade requirements before promotion is allowed, even the most able workers will spend some amount of time within a grade. The military pay system currently grants automatic longevity increases within a grade, so that only promotion is contingent on performance and within-grade pay increases are not. In contrast, our analysis suggests that contingent pay—or more generally deferred compensation contingent on worker behavior—will elicit more effort. This was true in the context of promotion and with a within-grade bonus system.

Proponents of the current system claim that the promotion system already offers effort motivation. But the question is, Does the promotion system motivate sufficient effort or is there an additional role for contingent within-grade compensation? Our analysis suggests that, indeed, contingent compensation within grades may have two important additional roles. First, the promotion system is relied on not only to reward effort and to induce the more able to seek the highest positions but also to handle pay *level* problems that relate to accession and retention issues (i.e., external market forces). Contingent within-grade compensation offers another policy tool to manage workers, especially the relatively less talented workers who are promoted less rapidly and who therefore spend most of their careers within a narrow range of grades.

Second, as was discussed in Chapter Six, a within-grade contingent pay system could help prevent climbing. Individuals who are insufficiently talented to successfully fill the top grades may be motivated to climb because of up-or-out rules that strongly penalize staying within a grade for too long, or because the pay table gives larger *intergrade* pay raises than *intragrade* pay raises. The organization cannot prevent climbing by simply reducing intergrade pay raises; doing so would wreck the incentive system for the more able, whose effort incentives and incentives to seek promotion would be dulled. But by holding back some of the within-grade pay and making intragrade pay increases contingent on performance, the organization could motivate those who should spend more time within a grade without negatively affecting the incentives of those who should move up. Such a system does not have to change the expected intergrade pay structure. As long as expected compensation within a grade is

maintained, the time profile of pay and how it increases with better performance can vary considerably. As discussed in Chapter Six, more within-grade pay should be held back in higher grades.

One caveat is needed. Minimum-performance standards could fill the same role. Such standards could prevent climbing if they are set so that upper grades have higher standards. On the other hand, there may be a limit to how far the military can rely on minimum-performance standards: By raising the standard, the military runs the risk of mistakenly dismissing those who had bad luck on test day.

Several aspects of a contingent intragrade pay system are noteworthy. First, such a system would reward time in grade as well as time in service. Second, up-or-out rules would take on a new meaning because those subject to such rules would be those who fail intragrade performance tests. Third, such a system would involve administrative costs that must be balanced against the system's benefits; therefore, the feasibility of introducing such a system needs to be explored.

THE RETIREMENT SYSTEM

The theoretical analysis also allows us to make some preliminary observations about the military's retirement system. These observations focus primarily on the pre-1980 retirement system—the plan upon which the subsequent 1981–1986 and post-1986 plans were based. These plans are described in Chapter Two.

Given the structure of the military's retirement system, the theoretical framework suggests that the system gives both an extremely strong effort incentive and retention incentive to mid-career personnel. The all-or-nothing aspect of cliff-vesting at 20 years of service, together with the sizable annuity that begins immediately at 20 YOS, makes the deferred compensation that retired pay represents highly valuable to those closest to the 20-year vesting date. Consequently, actions that increase the value of that compensation and the chances of receiving it are also highly valued. Insofar as greater effort supply translates into a higher probability of being allowed to stay and being able to accumulate years of service (i.e., avoiding involuntary separation), the vesting/immediate-annuity aspect of the retirement system encourages effort. The pension formula exaggerates this ef-

fect: Because the formula also depends on final pay, actions that increase final pay also increase retired pay. Thus, if greater effort means more promotions, then the formula rewards effort supply through basic pay.⁵

Several other aspects of the military's retirement system are noteworthy. First, the benefits induce voluntary separations. Thus, the system is incentive-compatible among service members *ex ante* and *ex post*. Second, the benefits, which are sizable, effectively skew the compensation system. Higher-ranking personnel earn larger benefits because they have more years of service and have been promoted to higher grades. Third, since the reward for an intermediate-length career is low, personnel must decide early on whether they want to be long-term careerists or leave. As discussed in Chapter Six, if filling an upper-level vacancy is costly because of the lack of lateral entry, then sorting the leavers from the stayers early on is important.

Despite these virtues, the theoretical analysis suggests that the military retirement system poses several difficulties. First, the strong retention incentive among mid-careers creates a mid-career bulge that slows down promotion opportunities for younger personnel and blunts the rewards to fast-trackers. Thus, the system stifles effort supply among the more junior personnel. Second, vested service members have a lower expected gain to staying than mid-careerists. By effectively skewing the compensation system for younger personnel who are still trying to advance, the system reduces the skewness for those who are vested and may thereby diminish their effort and advancement incentives relative to those of mid-careerists. Third, a virtue of the system is also a vice: Because leavers are induced to separate early on, some productive personnel who leave early might have stayed longer under an alternative system. Finally, the retirement system promotes management inflexibility, creating a kind of implicit-contract problem: The services "demand" large numbers of mid-career personnel because the personnel are there and will not quit; separating them prior to YOS 20 would be viewed as unfairly breaking an implicit contract. The services are therefore constrained

⁵As discussed in Chapter Five, Lazear shows that conventional pension formulas such as the military's give too much incentive to supply effort.

to retain personnel who would not be retained were the terms of separation different. The drawdown has highlighted this problem particularly well.⁶

⁶We analyze the military retirement systems in greater detail in companion pieces to this report (see Asch and Warner [1993 and forthcoming]).

This report develops a theoretical framework that allows an evaluation of the military pay and retirement system. Past attempts to evaluate the military compensation system have generally focused on the following traditional efficiency questions: (1) Does the system enable DoD to attract and to retain personnel in sufficient numbers and quality to accomplish its missions? (2) Does it promote an efficient distribution of personnel by grade and YOS? and (3) Does it accomplish these objectives at least cost? Our analysis aims at addressing these traditional macroeconomic questions and two additional microeconomic questions that have received little attention: (1) Does the system motivate personnel to work hard and effectively? and (2) Does the system promote the best person/job/rank matches, i.e., does it sort people properly by rank and experience according to their individual skills and abilities?

The organization that is the focus of our theoretical analysis is large, has a hierarchical structure, and does not permit lateral entry. Individuals enter at the lowest level, and positions at each upper level are filled by drawing from the pool of individuals in the level below.

THE MODEL: INDIVIDUAL DECISIONMAKING

In the context of this organization, we model individual worker's retention and effort decisions. This analysis builds upon Gotz and McCall (1984), who developed a dynamic programming model of Air Force officers' retention decisions. The model assumes that workers maximize their expected utility when deciding whether to remain

with the organization. The model accounts for the dynamic nature of the retention decision by incorporating the fact that (rational) individuals take into account all possible eventualities when making their decisions. We deviate from the Gotz-McCall model by incorporating work effort decisions. *Effort* in the model represents a utility cost to individuals, but it can also increase the utility gain to staying to the extent that future rewards (such as promotion) depend on the amount of effort supplied. Once we build this decisionmaking framework, we examine the effect on both the effort and retention decisions of various organizational active-compensation and -personnel policies, including promotion contests, up-or-out policies, involuntary separation, intragrade pay policy, and retirement policy.

ORGANIZATIONAL DECISIONMAKING

Given individual decisionmaking, we also consider organizational decisionmaking. A key factor in this process is the management of personnel flows into, within, and out of the organization. Without lateral entry, individuals in the organization are valuable for what they produce in their current grades and for what they are capable of producing in higher grades in the future. The organization must access, train, and retain a large enough pool of workers in the lower ranks to maintain the flow of workers necessary to staff the upper ranks. On the other hand, because of its pyramidal structure, the organization does not need, or indeed want, all entrants to stay and be promoted. The organization must generate some turnover at the higher levels, even among qualified personnel, to maintain effort and retention incentives at the lower levels. Thus, simultaneous with the organization's incentive to overstock personnel will be its incentive to induce individuals to leave.

CONCLUSIONS: ACTIVE PAY

Our analysis of individual and organizational decisionmaking leads us to several conclusions about pay.

Intergrade Pay Spreads

First, an important implication of the analysis is that the organization should skew its compensation structure. That is, intergrade pay spreads should increase with grade, for the following reasons:

- If the marginal productivity of effort rises with grade, then the pay differential must increase with grade to provide individuals in the higher grades with an incentive to supply more effort.
- In a hierarchy, the aggregate promotion rate declines as grade increases. Grade differentials must increase to just maintain a constant-effort incentive.
- Those in higher ranks have fewer promotion opportunities left to them—they are already near the top. The grade differentials need to be higher to induce individuals to supply the efficient amount of effort. Higher pay in upper grades increases the likelihood of retaining the most able individuals.

Second, other considerations will tend to reduce intergrade pay differentials:

- Rank-specific nonpecuniary factors will mean that intergrade pay spreads can be lower.
- Large pay differentials have a negative effect on cooperation and teamwork.
- As management of personnel flows suggests, there are other reasons for curtailing intergrade pay spreads:
 - To generate a large enough pool of qualified applicants, the organization must offer a sufficiently high pay stream. To cream on quality, the organization overpays junior workers relative to their outside opportunities but pays them an amount equal to their higher internal productivity.
 - On the other hand, to ensure retention, it overpays senior workers relative to their outside opportunities but underpays them relative to their internal productivity to minimize costs. The specific capital created by the lack of lateral entry means that senior workers can be paid more than they can get elsewhere but less than their internal productivity. Thus, the or-

ganization must trade off the benefits of a skewed pay profile against the need to flatten the profile.

Intracgrade Pay

Third, pay within a grade should depend, at least in part, on performance. Such a system rewards effort supply, especially among those who spend more time in a single grade or range of grades. Contingent intracgrade pay also offers the organization a means of preventing the climbing of those who are unqualified for higher-ranked positions but whom the promotion system motivates to move up and a means of preventing the slumming of those who are qualified to move up but who are unmotivated by the promotion system to work sufficiently hard to reach their best match.

Up-or-Out Policies

Fourth, up-or-out policies are necessary when the organization sets its pay table administratively and when worker ability must be learned about over time through observation. The organization is constrained in its ability to lower the pay of those who, in hindsight, it should not have hired. Lowering pay wrecks the effort-incentive effects of the pay system for other workers.

CONCLUSIONS: RETIRED PAY

Other conclusions relate to the roles of retired pay:

- As deferred compensation, retired pay can encourage both effort and retention of younger personnel.
- It provides individuals with an incentive to separate voluntarily.
- As a replacement for mandatory retirement, it is both *ex ante* and *ex post* incentive-compatible.

We also conclude that retired pay has no unique role in a hierarchical organization. All the roles that retired pay has can be accomplished by the appropriate pay and personnel policy. But if compensation is shifted away from retired pay and toward active-duty pay, the organization would need to use involuntary separations to control the

experience distribution of its workforce and maintain the promotion opportunities for its junior personnel. Such heavy reliance on involuntary separations would create *ex post* regret and subject the organization to what we call *organizational influence costs*, the costs of which equal the extra financial costs and negative-productivity effects that occur when the organization adopts policies that respond to employee complaints that the involuntary-separation policy is unfair. These costs could be substantial.

INITIAL EVALUATION OF THE MILITARY SYSTEM

In light of these conclusions, we make an initial evaluation of the military's pay table and retirement system.

Pay Table

The structure of the military's pay table is not skewed, as the theory suggests it should be. Intergrade pay spreads do not rise substantially with grade. However, it is difficult to conclude too much from this phenomenon because other considerations that are difficult to quantify at present (see "Intergrade Pay Spreads" above) lead to a relatively flat intergrade pay structure. To get a partial answer on whether the military pay table is sufficiently skewed, we investigated whether the system is producing desired effects by examining the flow of personnel through the system, by aptitude level. Our rough analysis suggests that, in the enlisted ranks, the system does not appear to be producing adverse selection, whereby the most able are leaving, but neither is it producing strong positive-selection effects.

Intraclass pay increases are automatic and are not contingent on performance. We conclude also that, whereas enlistment and reenlistment bonuses and special and/or incentive pays probably create the necessary pay-level differences across military occupations and across the branches of service, given their varying characteristics, it would only be a coincidence if the presence of these pays created the necessary differences in the structure of pay across occupations and services to reflect effort and sorting incentive considerations.

Retirement System

The following are general observations about how the military retirement system is likely to affect effort incentive, given our theoretical framework:

- The system gives a strong effort incentive, as well as retention incentive, to mid-career personnel.
- Retired pay helps skew the compensation system because those in higher ranks get larger retired pay.
- To the extent that mid-careers have a strong incentive to stay, promotion opportunities for more junior personnel are reduced; therefore, the system simultaneously blunts the effort incentives of more junior personnel.
- Because the system vests individuals with 20 or more years of service in an immediate annuity, the expected gain to staying is lower for vested personnel. As a result, the value of deferred compensation is lower for them and, thus, so is their effort incentive.

A companion piece to this document (Asch and Warner, forthcoming) develops an empirical simulation of the theoretical model that incorporates effort and ability-sorting considerations. The simulations enable us to examine the force structure and cost (i.e., macro) implications, as well as the productivity (i.e., micro) implications of alternative compensation systems.

ILLUSTRATING THE DERIVATION OF THE ORGANIZATIONAL STRUCTURE

The following numerical example clarifies the model presented in Chapter Three. Suppose that the organization's survival rates, the $s_{i,t}$, and promotion rates, the $\pi_{i,t}$, are as in Table A.1. Half of the organization's accessions are retained at the end of period 1. Of that half, half are promoted to grade 2 during period 2 and half remain at level 1. Seventy percent of those who are promoted to level 2 in period 2 choose to stay for period 3. Of this number, 30 percent are promoted to level 3 in period 3. Thus, the probability that a new hire will reach level 3 by period 3 is $0.5 \times 0.5 \times 0.7 \times 0.3 = 0.053$.

We may compute the probabilities of other paths and build the organization's steady-state workforce. Based on 1,000 new hires

Table A.1
Survival ($s_{i,t}$) and Promotion ($\pi_{i,t}$) Probabilities

Grade	Period		
	1	2	3
$s_{i,t}$			
1	0.5	0.6	0.7
2	0.0	0.7	0.8
3	0.0	0.8	0.9
$\pi_{i,t}$			
1	0.0	0.0	0.0
2	0.0	0.5	0.3
3	0.0	0.0	0.3

each period, its steady-state force is provided in Table A.2. The unconditional survival rate, sv_t , in the table is the fraction of new hires that remains from period 1 to each future period. The conditional survival rate is the fraction that remains in period t given that they remained until $t-1$; it equals sv_t / sv_{t-1} . The fraction of the entry cohort that quits in period t is $sv_t - sv_{t-1}$. Finally, expected man-years are $\sum_t t(sv_t - sv_{t-1})$.

Retention or survival rates, promotion rates, and the number of new hires each period are not independent of one another. Note in our example that there are 5.52 level 2 workers for each level 3 worker and 2.59 level 1 workers for each level 2 worker. If this is the organization's desired grade structure, it will have to adjust promotion rates and timing in response to changes in retention. To fill its upper-level positions should retention fall, it will have to increase promotions among those who remain. An increase in retention will force the organization to reduce promotion opportunities. Higher retention also increases the expected man-years per new hire and reduces the number of new hires that are necessary to maintain a given workforce.

Table A.2
Steady-State Force Based on Probabilities in Table 3.1

Variable	Period				TOTAL
	1	2	3	4	
Grade					
1	1,000	250	105	74	1,429
2		250	168	134	552
3			53	47	100
TOTAL	1,000	500	326	255	2,081
Unconditional survival rate	1	0.5	0.326	0.255	
Conditional survival rate to start of period	1	0.5	0.652	0.782	
Fraction quitting at end of period	0.5	0.174	0.071	0.255	
Expected man-years per new hire					2.081

SETTING PAY TO PREVENT CLIMBING AND SLUMMING

For illustrative purposes, suppose that there are two types of workers, high quality, α_1 , and low quality, α_2 , with $\alpha_1 > \alpha_2$. To prevent slumming and climbing, the organization must ensure that the following conditions are met:

$$\frac{\partial V_{i,t}^*}{\partial \pi_{i+1,t+1}} > 0 \quad \text{if} \quad \alpha = \alpha_1 ; \quad (\text{B.1})$$

$$\frac{\partial V_{i,t}^*}{\partial \pi_{i+1,t+1}} < 0 \quad \text{if} \quad \alpha = \alpha_2 . \quad (\text{B.2})$$

A promotion must increase expected utility at the end of t for those with α_1 to deter the slumming of high-quality workers in grade i and must decrease it for those with α_2 to deter the climbing of low-quality workers to grade $i + 1$. Note that for $i = 2$ and $t = 2$,

$$\frac{\partial V_{2,2}^*}{\partial \pi_{3,3}} = \Phi_{2,2} \left[m_{3,3} - m_{2,3} - \psi_{2,3} b_{2,3} + \beta (V_{3,3}^* - V_{2,3}^*) \right] . \quad (\text{B.3})$$

Given that $G_{2,2}^*$ is positive for both worker types and the intergrade spread $m_{3,3} - m_{2,3} - \psi_{2,3} b_{2,3}$ is positive, deterring slumming, climbing rests on making the term $\beta (V_{3,3}^* - V_{2,3}^*)$ positive for high-quality workers and negative and sufficiently large to offset

$m_{3,3} - m_{2,3} - \psi_{2,3}b_{2,3}$ for low-quality ones. There are two ways to do this.

The term $\beta(V_{3,3}^* - V_{2,3}^*)$ can be made negative and sufficiently large for low-quality workers by setting the grade 3 minimum-performance standard sufficiently high that low-quality workers fail and high-quality workers succeed in meeting it. To see how this would work, note that $V_{3,3}^* - V_{2,3}^* = (\Phi_{3,3}G_{3,3}^* - \Phi_{2,3}G_{2,3}^*) + (L_{3,3} - L_{2,3}) + \sigma_\epsilon(\phi_{3,3} - \phi_{2,3})$. The term $\beta(V_{3,3}^* - V_{2,3}^*)$ can be made negative for low-quality workers by making $\Phi_{3,3}$ and $\phi_{3,3}$ zero and $\Phi_{2,3}$ and $\phi_{2,3}$ positive for α_2 , i.e., by ensuring that they fail the standard in grade 3 but pass that in grade 2. Low-quality workers are worse off by being promoted if the promotion results in their being involuntarily separated for substandard higher-grade performance. Similarly, high-quality workers are better off by the promotion because they are allowed to stay to enjoy the higher pay associated with the higher grade.

Alternatively, the term can also be made negative and sufficiently large for $\alpha = \alpha_2$ and positive for high-quality workers by ensuring that low-quality workers do not qualify for intragrade bonuses in higher grades and high-quality workers do, and by appropriately structuring the intragrade contingent bonuses and standards. Specifically, one way to make the term $\Phi_{3,3}G_{3,3}^* - \Phi_{2,3}G_{2,3}^*$ negative and large and thus the term $\partial V_{2,2}^* / \partial \pi_{3,3} < 0$ for the α_2 is to make $G_{3,3}^*$ small and $G_{2,3}^*$ large for the less able, which can be done by lowering the base amount of pay earned in grade 3, $m_{3,4}$, and by making the standard for attaining the intragrade bonus for grade 3 high enough so that $\psi_{3,4} = 0$ if $\alpha = \alpha_2$ in Equation 4.18. If the grade 2 bonus and/or the chance that a low-quality worker in grade 2 will earn it is sufficiently high, then low-quality workers are less inclined to climb to grade 3. To compensate high-quality workers for the lower base pay in grade 3 and the stiffer intragrade standard, the bonus, $b_{3,4}$, must be increased. Although more of grade 3 pay is made contingent on performance, high-quality workers have a higher chance of earning the bonus. They are not inclined to slum as long as their expected pay in grade 3 (including their expected bonus) exceeds their expected pay in grade 2 and the gain to staying is positive.

SIMULATING CONTESTS

This appendix simulates a homogeneous contest for various parameter values, then provides a simulation of a five-person contest comprising individuals with varying abilities. Simulations of the homogeneous contests are provided in Tables C.1 and C.2. Table C.1 begins by simulating a ten-person contest. For these simulations, it is assumed that the normal random factor in the contest, ξ , has a mean of 0 and standard deviation of 1. We first use Equation 4.19 of the main text to calculate each contestant's probability of promotion. The results, provided in the second column of Table C.1, show that when homogeneous contestants all supply the same effort, each contestant has a .1 probability of each possible rank in the evaluation. Therefore, if there are r promotions, each contestant's probability of promotion is r/n (column 5).

We used Equation 4.19 to calculate how a contestant's rank probabilities would change if the contestant supplies .1 more effort than all other contestants.¹ The results are shown in the third and fourth columns of Table C.1. Extra effort raises the probability of finishing in the top five places and reduces the probability of finishing in the bottom five places. Note the descending order of effect: Extra effort raises the probability of a first-place finish the most and reduces the probability of a last-place finish the most.

¹An alternative would be to calculate the instantaneous effect of effort on the probability of promotion using Equation 4.24 of the main text. The directional effects are equivalent to those shown in Table C.1.

The last two columns of Table C.1 show the effect of effort on the probability of promotion. Note that extra effort raises the probability of promotion the most when half the contestants are to be promoted. Conversely, when only a few of the contestants (or most of them) are to be promoted, effort has less effect on the probability of promotion.

Table C.2 provides some additional simulations of homogeneous contests using different values of the contest parameters. A key result alluded to in the text is that the effect of effort rises with the scale of the contest. For example, when half of the contestants are to be promoted, increasing the scale of the contest from two to four contestants increases the marginal effect of effort from .028 to .033. Other simulations (not shown) reveal that the effect of effort rises with the scale, but at a decreasing rate, and that, after a certain contest size, further increases in the number of contestants cease to matter. Table C.2 also reveals the inverse relationship between the noise in the contest, measured by the standard deviation of ξ , and the effect of effort.

Finally, we simulated heterogeneous contests of different sizes. To do this we assumed that the contests had three different ability types. The low-ability types have ability $\alpha = -1$, the average-ability types

Table C.1
Simulation of a 10-Person Homogeneous Contest

Evaluation Rank (and number to be promoted)	Base		Probability		Promotion	
	Probability of Rank in Contest	of Rank with More Effort	Change in Probability of Rank	Cumulative Promotion Probability	Probability with More Effort	Change in Promotion Probability
1	.1	.116	.016	.1	.116	.016
2	.1	.110	.010	.2	.226	.026
3	.1	.106	.006	.3	.332	.032
4	.1	.103	.003	.4	.436	.036
5	.1	.101	.001	.5	.537	.037
6	.1	.098	-.002	.6	.635	.035
7	.1	.096	-.004	.7	.731	.031
8	.1	.093	-.007	.8	.779	.024
9	.1	.090	-.010	.9	.914	.015
10	.1	.085	-.015	1.0	1.0	0.0

Table C.2
Simulations of Homogeneous Contests

Number of Contestants (<i>n</i>)	Number to Be Promoted (<i>r</i>)	Base Promotion Probability	Probability with More Effort ($\sigma_{\xi} = 1$)	Change in Probability ($\sigma_{\xi} = 1$)	Probability with More Effort ($\sigma_{\xi} = 2$)	Change in Probability ($\sigma_{\xi} = 2$)
2	1	.5	.528	.028	.514	.014
3	1	.333	.362	.029	.347	.014
3	2	.666	.694	.028	.680	.014
4	1	.25	.257	.026	.263	.013
4	2	.5	.533	.033	.516	.017
4	3	.75	.774	.025	.762	.013
5	1	.2	.224	.024	.212	.012
5	2	.4	.433	.033	.416	.017
5	3	.6	.632	.033	.616	.016
5	4	.8	.822	.022	.811	.011

have ability $\alpha = 0$, and the high-ability types have ability $\alpha = +1$. The ability types are represented in equal proportions in each contest. Individuals' ranks in the contest are determined by the evaluation $E = \alpha + e + \xi$. We assume that ξ has a standard deviation of 1. Thus, the low- and high-ability types are thus one standard deviation below, and one standard deviation above, average, respectively. Assuming first that each individual supplies the same effort, we generated a random error for each individual, each individual's evaluation E , and each individual's rank in the contest. This process was repeated 10,000 times. The frequency distribution of observed finishes for each individual approximates that individual's probability distribution of finishes. We then increased each contestant's effort by 0.1 while holding constant all other contestants' effort levels and observed the frequency distribution of that contestant's (and all other contestants') finishes in 10,000 replications of the contest.

Results are displayed in Table C.3 for contests of 3, 6, 9, and 12 individuals. The table shows the probability that a contestant of a given ability type will finish in the top one-third and the top two-thirds in the contest.

Table C.3
Simulations of Heterogeneous Contests of Various Sizes

Ability Type	Contest Size	Probability of a Finish in the					
		Top One-Third			Top Two-Thirds		
		Equal Effort	Extra Effort	Effect of Effort	Equal Effort	Extra Effort	Effect of Effort
Low	3	.046	.051	.005	.275	.301	.026
	6	.058	.068	.010	.313	.344	.031
	9	.057	.067	.010	.323	.355	.032
	12	.053	.060	.007	.321	.358	.037
Average	3	.228	.249	.021	.771	.793	.032
	6	.256	.286	.030	.738	.764	.026
	9	.271	.301	.030	.738	.768	.030
	12	.274	.306	.032	.717	.750	.033
High	3	.726	.750	.024	.954	.960	.006
	6	.677	.707	.030	.946	.955	.009
	9	.668	.700	.032	.946	.954	.008
	12	.669	.699	.030	.943	.954	.011

Thus, it is apparent that when all contestants supply equal effort, a low-ability individual has a very small chance of finishing in the top one-third, whereas a high-ability individual has over a .66 probability of a top-one-third finish. The low-ability individual has about a .3 probability of a finish in the top two-thirds, whereas the high-ability individual is virtually assured of a finish in the top two-thirds.

Extra effort raises the probability of a finish in the top one-third or the top two-thirds. But as in the case of homogeneous contests, the effect of effort depends on the base probability of reaching a given threshold. When the probability of reaching a given threshold is either very large or very small, extra effort has a small effect. As Table C.3 indicates, extra effort has a miniscule effect on the probability that a low-ability individual will place in the top one-third or that a high-ability individual will place in the top two-thirds. Effort has a larger effect on the outcomes of the individuals with average ability.

The results in Table C.3 preserve the finding that effort has a larger effect the larger the scale of the contest is. Not shown is that increasing the noise in the contest (the standard deviation of ξ) reduces the effect of effort.

It should be stressed that the base probabilities of various finishes and the marginal effect of effort on them shown in Table C.3 are not the final probabilities and effects that will prevail as a result of optimizing behavior. In theory, each contestant will solve a first order condition, similar to Equation 4.24 of the text, in light of the optimizing behavior of the other contestants. If all contestants have the same effort disutility function and the marginal reward for effort is the same, individuals will all vary effort until its marginal effect on the probability of obtaining the reward is the same. Suppose, for example, that one-third of the contestants are to receive some payoff and all contestants have the same disutility of effort. Then, beginning from the situation in Table C.3 in which all contestants supply the same effort, the high-ability individuals would reduce their effort because its marginal effect is smaller than the marginal effect of effort for individuals of average ability. As the high-ability individuals reduce their effort, their likelihood of a finish in the top one-third declines but the marginal effect of effort increases. Adjustments occur until the marginal effect of effort is identical for all individuals and the contestants' mean probability of finishing in the top one-third is .333.

**DERIVING SECOND ORDER CONDITIONS AND THE
EFFECT OF POLICY CHANGES ON EFFORT SUPPLY**

In maximizing expected utility with respect to effort supply in Equation 4.24, the second order condition for a maximum is

$$2\beta \frac{\partial \Phi_{i,t}}{\partial e_{it}} \frac{\partial G_{i,t}^*}{\partial e_{it}} + \beta \Phi_{i,t} \frac{\partial^2 G_{i,t}^*}{\partial e_{it}^2} + \beta G_{i,t}^* \frac{\partial^2 \Phi_{i,t}}{\partial e_{it}^2} + \beta \sigma_\epsilon \frac{\partial^2 \phi_{i,t}}{\partial e_{it}^2} - \frac{\partial^2 Z(e_{it})}{\partial e_{it}^2} < 0, \quad (D.1)$$

where we assume, for simplicity, that $\Phi_{i,t}^1 = 1$ (i.e., the probability of being allowed to stay is 1), so that $\Phi_{i,t} = \Phi_{i,t}^2$ (the probability of staying equals the probability that the individual wants to stay). Whether Equation D.1 holds depends importantly on the sign of $\partial G_{i,t}^* / \partial e_{it}$. This term equals

$$\left(\frac{\partial \pi_{i+1,t+1}}{\partial e_{it}} \right) \left[m_{i+1,t+1} - m_{i,t+1} - \psi_{i,t+1} b_{i,t+1} + \beta (V_{i+1,t+1}^* - V_{i,t+1}^*) \right] + (1 - \pi_{i+1,t+1}) b_{i,t+1} \frac{\partial \psi_{i,t+1}}{\partial e_{it}}, \quad (D.2)$$

which is positive if the organization sets policy to reward individual effort. Since such a reward system seems in reasonable accord with reality or with stated policy, we assume that $\partial G_{i,t}^* / \partial e_{it} > 0$. Thus, the first term in Equation D.1 is positive. The second term is negative, given that $\partial^2 \pi_{i+1,t+1} / \partial e_{it}^2$ and $\partial^2 \psi_{i,t+1} / \partial e_{it}^2$ are negative.

The term $-\partial^2 Z(e_{it}) / \partial e_{it}^2$ is also negative, given the increasing marginal disutility of effort. The signs of $\partial^2 \Phi_{i,t} / \partial e_{it}^2$ and $\partial^2 \phi_{i,t} / \partial e_{it}^2$ depend on the shape of the probability distribution function, $\phi_{i,t}$ (that is, its symmetry and variance, σ_ε), whether the expected gain to staying, $G_{i,t}^*$, is positive or negative, and whether the probability of staying, $\Phi_{i,t}$, is greater or less than 0.5. These terms need not be negative. However, if $\phi_{i,t}$ is normal and σ_ε is not too small, then even if these terms are positive, their magnitude can be small enough to be offset by the negative sign of $\partial^2 G_{i,t}^* / \partial e_{it}^2$. Below, we assume that the second order condition in Equation D.1 holds.

We can use Equations 4.24 and D.1 to see how various policies, such as increases in pay, retired pay, promotion rates, involuntary separation, and intragrade bonuses, affect effort supply. The mechanism by which these policies affect individual effort decisions is the expected gain to staying in the organization. For example, if the organization raises pay in grade 3 in period 3 ($m_{3,3}$), then staying provides more utility because $G_{i,2}^*$ is higher. When staying is viewed as better, individuals weigh actions that increase their ability to stay or that increase the value of staying, such as effort, as being more worthwhile (assuming for a moment that effort increases the gain to staying). More generally, if effort rises when the gain to staying rises, then policies that increase the gain to staying will also increase effort. The key link, then, between pay policy and effort is the expected gain to staying. How are effort and the gain to staying related? As we show below, whether effort rises when the expected gain rises depends importantly on organizational policy.

To illustrate the effect of some policy on the optimal effort supplied, e_{it}^* , we consider the effect of a change in some parameter or policy embedded in $G_{i,t}^*$, to be denoted μ , on e_{it}^* . The parameter μ can be thought of as some aspect of pay or personnel policy, such as $m_{3,3}$. Thus, to derive the effect of an increase in pay (or some other parameter) on the optimal effort supply, we derive

$$\frac{de_{it}^*}{d\mu} = \frac{de_{it}^*}{dG_{i,t}^*} \frac{\partial G_{i,t}^*}{\partial \mu}. \quad (\text{D.3})$$

We need to determine the sign of $de_{it}^* / dG_{i,t}^*$. Then, by knowing how the pay or personnel policy affects the expected gain to staying,

$\partial G_{i,t}^* / \partial \mu$, we can ascertain how the system or the structure of retired pay alters the optimal-effort decision. To derive $de_{it}^* / dG_{i,t}^*$, recall that e_{it}^* is the solution to the first order condition given in Equation 4.23b. The effect of the expected gain to staying on the optimal effort is derived by invoking the envelope theorem. This theorem allows us to totally differentiate Equation 4.23b with respect to e_{it}^* and $G_{i,t}^*$ and to solve for $de_{it}^* / dG_{i,t}^*$. We get

$$\frac{de_{it}^*}{dG_{i,t}^*} = \frac{-\beta \left(\frac{\partial G_{i,t}^*}{\partial e_{it}} \frac{\partial \Phi_{i,t}}{\partial G_{i,t}^*} + \frac{\partial \Phi_{i,t}}{\partial e_{it}} + G_{i,t}^* \frac{\partial^2 \Phi_{i,t}}{\partial e_{it} \partial G_{i,t}^*} + \sigma_e \frac{\partial^2 \phi_{i,t}}{\partial e_{it} \partial G_{i,t}^*} \right)}{\left(\beta G_{i,t}^* \frac{\partial^2 \Phi_{i,t}}{\partial e_{it}^2} + 2\beta \frac{\partial \Phi_{i,t}}{\partial e_{it}} \frac{\partial G_{i,t}^*}{\partial e_{it}} + \beta \Phi_{i,t} \frac{\partial^2 G_{i,t}^*}{\partial e_{it}^2} + \beta \sigma_e \frac{\partial^2 \phi_{i,t}}{\partial e_{it}^2} - \frac{\partial^2 Z(e_{it})}{\partial e_{it}^2} \right)} \quad (\text{D.4})$$

By the second order conditions for a maximum, the denominator of Equation D.4 is negative. Therefore, the sign of $\partial e_{it}^* / \partial G_{i,t}^*$ is the same as the sign of the enclosed term in the numerator, which consists of four terms. If effort increases the gain to staying, $\partial G_{i,t}^* / \partial e_{it} > 0$, and the individual performance evaluation, $\partial E_{it} / \partial e_{it} > 0$, and if the function $\phi_{i,t}$ is well-behaved, then the numerator is positive and Equation D.4 is positive.¹ Thus, the effect of a policy on effort supply depends on how the policy affects the expected gain to staying, $\partial G_{i,t}^* / \partial \mu$.

¹More specifically, the first two terms, $(\partial G_{i,t}^* / \partial e_{it}) * (\partial \Phi_{i,t} / \partial G_{i,t}^*)$ and $\partial \Phi_{i,t} / \partial e_{it}$, are positive given our assumption that $\partial G_{i,t}^* / \partial e_{it} > 0$ and $\partial E_{it} / \partial e_{it} > 0$. The sign and magnitude of the last two terms depend on the shape of $\phi_{i,t}$ and on whether $\Phi_{i,t}$ is less than or greater than 0.5. As with Equation 4.24, we assume that $\phi_{i,t}$ is normal and exhibits sufficient variance that these terms, if negative, are not large enough to make the numerator negative.

REFERENCES

- Abowd, John M., and Orley Ashenfelter, "Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Differentials," in Sherwin Rosen, ed., *Studies in Labor Markets*, Chicago: University of Chicago Press, 1981, pp. 141-170.
- Advisory Commission on Service Pay, *Career Compensation for the Uniformed Forces*, A Report and Recommendation for the Secretary of Defense, Washington, D.C., December 1948.
- Allen, Steven G., Robert L. Clark, and Ann McDermed, *Pensions, Bonding, and Lifetime Jobs*, Cambridge, Mass.: National Bureau of Economic Research, NBER Working Paper No. 3688, April 1991.
- Argüden, R. Yilmaz, *Personnel Management in the Military: Effects of Retirement Policies on the Retention of Personnel*, Santa Monica, Calif.: RAND, R-3342-AF, 1986.
- Asch, Beth, *Pensions, Wages, and Sorting: A Theoretical and Empirical Analysis*, Ph.D. Dissertation, University of Chicago, 1984.
- , *Designing Military Pay: Contributions and Implications of the Economics Literature*, Santa Monica, Calif.: RAND, MR-161-FMP, 1993.
- Asch, Beth, and John Warner, "Should the Military Retirement System Be Reformed?" paper presented at the conference "A Military of Volunteers: Yesterday, Today, and Tomorrow" at the U.S. Naval Academy, Annapolis, Md., September 1993.

- , *A Policy Analysis of Alternative Military Retirement Systems*, Santa Monica, Calif.: RAND, MR-465-OSD, forthcoming.
- Black, Matthew, *Personal Discount Rates: Estimates for the Military Population*, Arlington, Va.: SRA Corporation, 1983.
- Black, Matthew, Robert Moffitt, and John T. Warner, "The Dynamics of Job Separation: The Case of Federal Employees," *Journal of Applied Econometrics*, Vol. 5, July/September 1990, pp. 245–262.
- Borjas, George, and Finis Welch, "Post-Service Earnings of Military Veterans," in Curtis L. Gilroy, ed., *Army Manpower Economics*, Boulder, Colo.: Westview Press, 1986.
- Carmichael, H. L., "Firm-Specific Human Capital and Promotion Ladders," *Bell Journal of Economics*, Vol. 14, Spring 1983, pp. 251–258.
- , "Self-Enforcing Contracts, Shirking, and Life Cycle Incentives," *The Journal of Economic Perspectives*, Vol. 3, No. 4, Fall 1989, pp. 65–83.
- Daula, Thomas, and Robert Moffitt, "Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment," West Point, N.Y.: West Point Military Academy, draft paper, July 1992.
- Defense Manpower Commission, *Defense Manpower: The Keystone to National Security*, Report to the President and the Congress, Washington, D.C.: U.S. Government Printing Office (GPO), April 1976.
- Department of Defense (DoD), *Modernizing Military Pay*, Volume I, *Active Duty Pay*, Report of the First Quadrennial Review of Military Compensation, Washington, D.C.: U.S. GPO, November 1967.
- , *Modernizing Military Pay*, Volume IV, *The Military Estate Program*, Report of the First Quadrennial Review of Military Compensation, Washington, D.C.: U.S. GPO, January 1969.
- , *Report of the Fifth Quadrennial Review of Military Compensation*, Washington, D.C.: U.S. GPO, November 1983.

- , *Military Compensation Background Papers: Compensation Elements and Related Manpower Cost Items*, Washington, D.C.: U.S. GPO, 1992a.
- , *Report of the Seventh Quadrennial Review of Military Compensation*, Washington, D.C.: U.S. GPO, August 1992b.
- General Accounting Office, *Enlisted Force Management: Past Practices and Future Challenges*, Washington, D.C.: U.S. GPO, January 1991.
- Gifford, Adam, and Roy W. Kenney, "The Production of Information Through Labor Market Contests," *Journal of Law, Economics, and Organization*, Vol. 2, Fall 1986, pp. 305–313.
- Gilman, Harry J., *Determinants of Personal Discount Rates: An Empirical Examination of the Pattern of Voluntary Contributions of Employees in Four Firms*, Alexandria, Va.: The Center for Naval Analyses, Memorandum 76-0984-30, 1976.
- Goldberg, Matthew S., and John T. Warner, *Determinants of Navy Reenlistment and Extension Rates*, Alexandria, Va.: The Center for Naval Analyses, CRC 476, 1982.
- , "Military Experience, Civilian Experience, and the Earnings of Veterans," *Journal of Human Resources*, Vol. 22, Winter 1987, pp. 62–81.
- Gotz, Glenn A., "Comment on the Dynamics of Job Separation," *Journal of Applied Econometrics*, Vol. 5, July/September 1990, pp. 263–268.
- Gotz, Glenn, and John J. McCall, *A Dynamic Retention Model for Air Force Officers: Theory and Estimates*, Santa Monica, Calif.: RAND, R-3028-AF, 1984.
- Grace Commission, *Final Report of the President's Private-Sector Commission on Government Management*, Washington, D.C.: U.S. GPO, March 1985.
- Green, Jerry R., and Nancy L. Stokey, "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, Vol. 91, No. 3, 1983, pp. 349–364.

Grissmer, David W., James R. Hosek, and Richard L. Eisenman, "Trimming the Senior Enlisted Force: Estimating Cost Savings and Structuring Institutional Incentives," Santa Monica, Calif.: RAND, unpublished draft.

Guasch, J. Luis, and Andrew Weiss, "Wages as Sorting Mechanisms in Competitive Markets with Asymmetric Information: A Theory of Testing," *Review of Economic Studies*, Vol. 47, 1980, pp. 653-664.

———, "Self-Selection in the Labor Market," *American Economic Review*, Vol. 73, No. 3, 1981, pp. 275-284.

Hook Commission, *Career Compensation for the Uniformed Services*, Washington, D.C.: U.S. GPO, December 1948.

Horowitz, Stanley, "The Experience Mix of Enlisted Personnel in the Military," Alexandria, Va.: Institute for Defense Analysis, IPA Paper P-2568, November 1991.

Ippolito, Richard A., "Encouraging Long-Term Tenure: Wage Tilt or Pensions?" *Industrial and Labor Relations Review*, Vol. 44, No. 3, 1991, pp. 520-535.

"Joint Army-Navy Pay Board Study," 1947, unpublished manuscript.

Kahn, Charles, and Gur Huberman, "Two-Sided Uncertainty and 'Up-or-Out' Contracts," *Journal of Labor Economics*, Vol. 6, No. 4, 1988, pp. 423-444.

Kanemoto, Yoshitsugu, and W. Bentley MacLeod, "The Theory of Contracts and Labor Practices in Japan and the United States," *Managerial and Decision Economics*, Vol. 12, 1991, pp. 159-170.

Lawrence, Emily C., "Poverty and Rates of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, Vol. 99, No. 1, 1991, pp. 54-78.

Lazear, Edward P., "Why Is There Mandatory Retirement?" *Journal of Political Economy*, Vol. 87, No. 6, 1979, pp. 1261-1284.

———, "Incentive Effects of Pensions," in David A. Wise, ed., *Pensions, Labor, and Individual Choice*, Chicago: University of Chicago Press, 1985, pp. 253-282.

- , "Retirement from the Labor Force," in Orley C. Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, New York: North Holland, 1986, pp. 303–355.
- , "Pensions as Severance Pay," in Zvi Bodie and John Shoven, eds., *Financial Aspects of the U.S. Pension System*, Chicago: University of Chicago Press, 1988, pp. 57–89.
- , "Pay Equality and Industrial Politics," *Journal of Political Economy*, Vol. 87, No. 6, 1989, pp. 1261–1284.
- , "Pensions and Deferred Benefits as Strategic Compensation," *Industrial Relations*, Vol. 29, No. 2, 1990, pp. 263–281.
- , "Labor Economics and the Psychology of Organizations," *Journal of Economic Perspectives*, Vol. 5, No. 2, 1991, pp. 89–110.
- Lazear, Edward P., and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, Vol. 89, No. 51, 1981, pp. 841–864.
- MacLeod, W. Bentley, and James Malcomson, "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, Vol. 96, No. 4, 1988, pp. 832–854.
- Maddala, G. S., *Limited Dependent Variables in Econometrics*, Cambridge, England: Cambridge University Press, 1983.
- Malcomson, James, "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, Vol. 92, May 1984, pp. 486–507.
- McLaughlin, Kenneth J., "Aspects of Tournament Models: A Survey," *Research in Labor Economics*, Vol. 9, 1988, pp. 225–256.
- Milgrom, Paul R., "Employment Contracts, Influence Activities, and Efficient Organization Design," *Journal of Political Economy*, Vol. 96, No. 1, 1988, pp. 42–60.
- Nalebuff, Barry J., and Joseph E. Stiglitz, "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *Bell Journal of Economics*, Vol. 13, No. 1, 1983, pp. 21–43.

- Oi, Walter, "Labor as a Quasi-Fixed Factor of Production," *Journal of Political Economy*, Vol. 70, No. 6, 1962, pp. 538-555.
- , "The Fixed Employment Costs of Specialized Labor," in Jack E. Triplett, ed., *The Measurement of Labor Cost*, Chicago: University of Chicago Press, 1983.
- O'Keeffe, Mary, W. Kip Viscusi, and Richard J. Zeckhauser, "Economic Contests: Comparative Reward Schemes," *Journal of Labor Economics*, Vol. 2, No. 1, 1984, pp. 27-56.
- Orvis, Bruce R., Michael T. Childress, and J. Michael Polich, *Effect of Personnel Quality on the Performance of Patriot Air Defense System Operators*, Santa Monica, Calif.: RAND, R-3901-A, 1992.
- President's Commission on Military Compensation, *Report of the President's Commission on Military Compensation*, Washington, D.C.: U.S. GPO, April 1978.
- Rosen, Sherwin, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, Vol. 13, No. 2, 1982, pp. 311-323.
- , "Prizes and Incentives in Elimination Tournaments," *American Economic Review*, Vol. 76, No. 4, 1986, pp. 701-715.
- , "Some Economics of Teaching," *Journal of Labor Economics*, Vol. 5, No. 4, Part 1, 1987, pp. 561-575.
- , "Transaction Costs and Internal Labor Markets," *Journal of Law, Economics, and Organization*, Vol. 4, No. 1, 1988, pp. 49-64.
- , "The Military as an Internal Labor Market: Some Allocation, Productivity, and Incentive Problems," *Social Science Quarterly*, Vol. 73, No. 2, 1992, pp. 227-237.
- Salop, J. K., and S. C. Salop, "Self-Selection and Turnover in the Labor Market," *Quarterly Journal of Economics*, Vol. 91, 1976, pp. 619-628.
- Simon, Curtis, and John T. Warner, "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure," *Journal of Labor Economics*, Vol. 10, No. 3, 1992, pp. 306-329.

- Smith, D. Alton, Stephen D. Sylwester, and Christine M. Villa, "Army Reenlistment Models," in David K. Horne, Curtis L. Gilroy, and D. Alton Smith, eds., *Military Compensation and Personnel Retention: Models and Evidence*, Alexandria, Va.: Army Research Institute for the Behavioral and Social Sciences, 1991.
- Stiglitz, Joseph, "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," *Bell Journal of Economics*, Vol. 6, No. 2, 1975, pp. 552-579.
- Topel, Robert, "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy*, Vol. 99, No. 1, 1991, pp. 145-176.
- Warner, John T., *Military Compensation and Retention: An Analysis of Alternative Models and a Simulation of a New Model*, Alexandria, Va.: The Center for Naval Analyses, Research Contribution 436, 1981.
- Warner, John T., and Matthew S. Goldberg, "The Influence of Non-Pecuniary Factors on Labor Supply: The Case of Navy Enlisted Personnel," *Review of Economics and Statistics*, Vol. 66, 1984, pp. 26-35.
- Winkler, John D., Judith C. Fernandez, and J. Michael Polich, *Effect of Aptitude on the Performance of Army Communications Operators*, Santa Monica, Calif.: RAND, R-4143-A, 1992.