

Focused Operations Management for Health Services Organizations

Based on the book by

Boaz Ronen

Joseph Pliskin

Shimeon Pass

BUDT 758O, Jan. 2009

Professor B. L. Golden

The Modern Health Care and Business Environment

- New managerial approaches are needed
- Why?
 - Massive increases in health care costs
 - Rapid changes in the business environment
 - Transition from sellers' market to buyers' market

What is a Sellers' Market?

- Supplier or service provider largely dictates terms
 - Customer is charged for full costs plus a reasonable profit
 - Response time: We are doing our best
 - Quality: We are doing our best
 - Performance: We know what our customers need

- Today's business market is more of a buyers' market

What is a Buyers' Market?

- A buyers' market is characterized by
 - Globalization of the world economy
 - Fierce competition
 - Global excess capacities (production, services, etc.)
 - New managerial approaches
 - Access to data and knowledge
 - Cheap and rapid communication
 - Timely availability of materials and services

A Buyers' Market -- continued

- A buyers' market is characterized by
 - Ease of global travel and transport
 - Advanced technologies for production
 - Extensive use of advanced IT & communications
 - Shortened life cycles of products and services
 - Customer empowerment

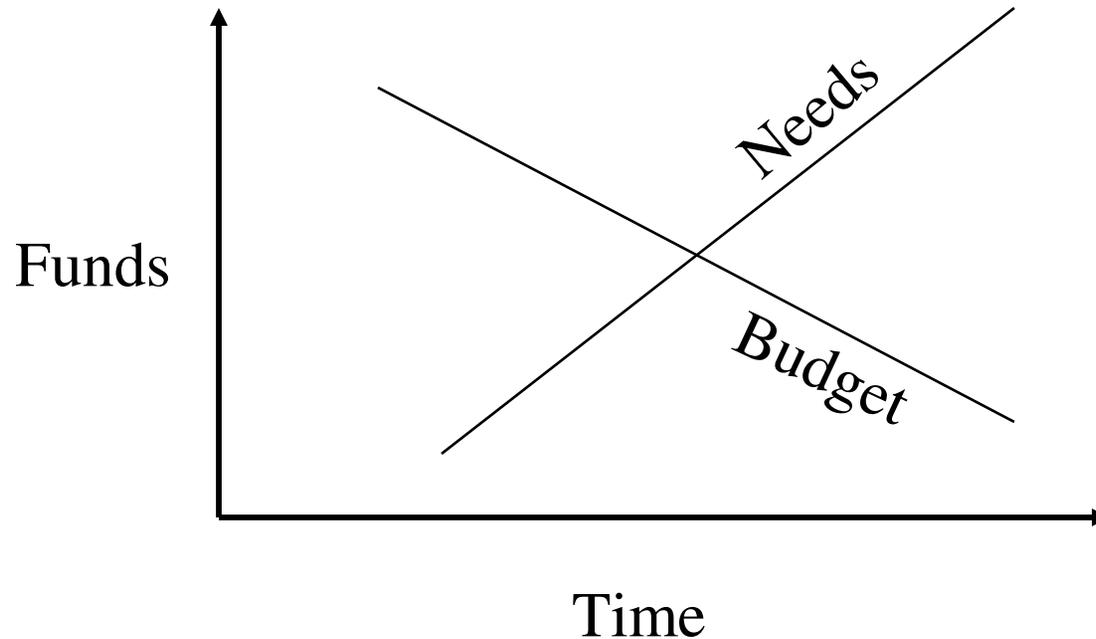
A Buyers' Market -- continued

- Price: Determined by the market
- Response time: Determined by response time of best in the market
- Quality: Determined by quality of best in the market
- Performance: Driven by customers

The Health Care Market

- In health care today, demand for services is up while budgets are being reduced
- Customers know more
- Customers demand more
- Technology is developing rapidly
- Life expectancy is increasing

Health Care Market: Recent History



- Scissors Diagram of Needs versus Budget

U.S. Health Care Expenditures, 1990 and 2003

Category	% in 1990	% in 2003	CAGR
Hospital care	36.5	30.7	5.6%
Physician and clinical services	22.6	22.0	6.8%
Other professional services	2.6	2.9	7.8%
Nursing homes and home health	9.4	9.0	6.6%
Prescription drugs	5.8	10.7	12.2%
Administration and cost of private health insurance	5.7	7.1	8.8%
Other	17.4	17.6	7.1%

U.S. Health Care Expenditures, 1990 and 2003

- Total expenditures in 1990 = \$696 billion
- Total expenditures in 2003 = \$1,679 billion
- CAGR = compound annual growth rate
- Overall CAGR from 1990 to 2003 = 7.0%
- This and the previous page are taken from Redefining Health Care by Porter and Teisberg (2006)
- One more point: the annual rate of increase in the consumer price index from 1990 to 2003 was approximately 3.5%

The Remedy: New Managerial Approaches

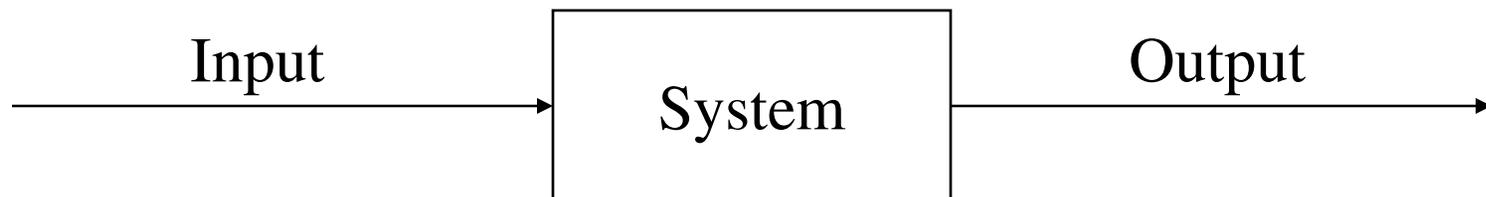
- Necessary conditions
 - Advanced technology
 - Powerful information systems
 - Professional personnel

- But, in addition, it is essential to manage differently

- Result: Increasing the value of an organization to its owners, workers, and the community

New Managerial Approaches: Characteristics

- They are based on common sense
- They evolved out of practice
- Later on, academics studied these approaches
- They contradict the myth of the input-output model



The Input-Output Model

- If we want to improve outputs, we must increase inputs
- If we want to increase patient volume in a clinic by 20%, we need more personnel, space, advertising, etc.
- If we want to improve (decrease) waiting time for patients, we need more personnel and equipment

A Preview of Things to Come

- We can improve output without increasing input
- The focus of the course text is
 - Better management and utilization of existing resources, or
 - Doing more with what you have

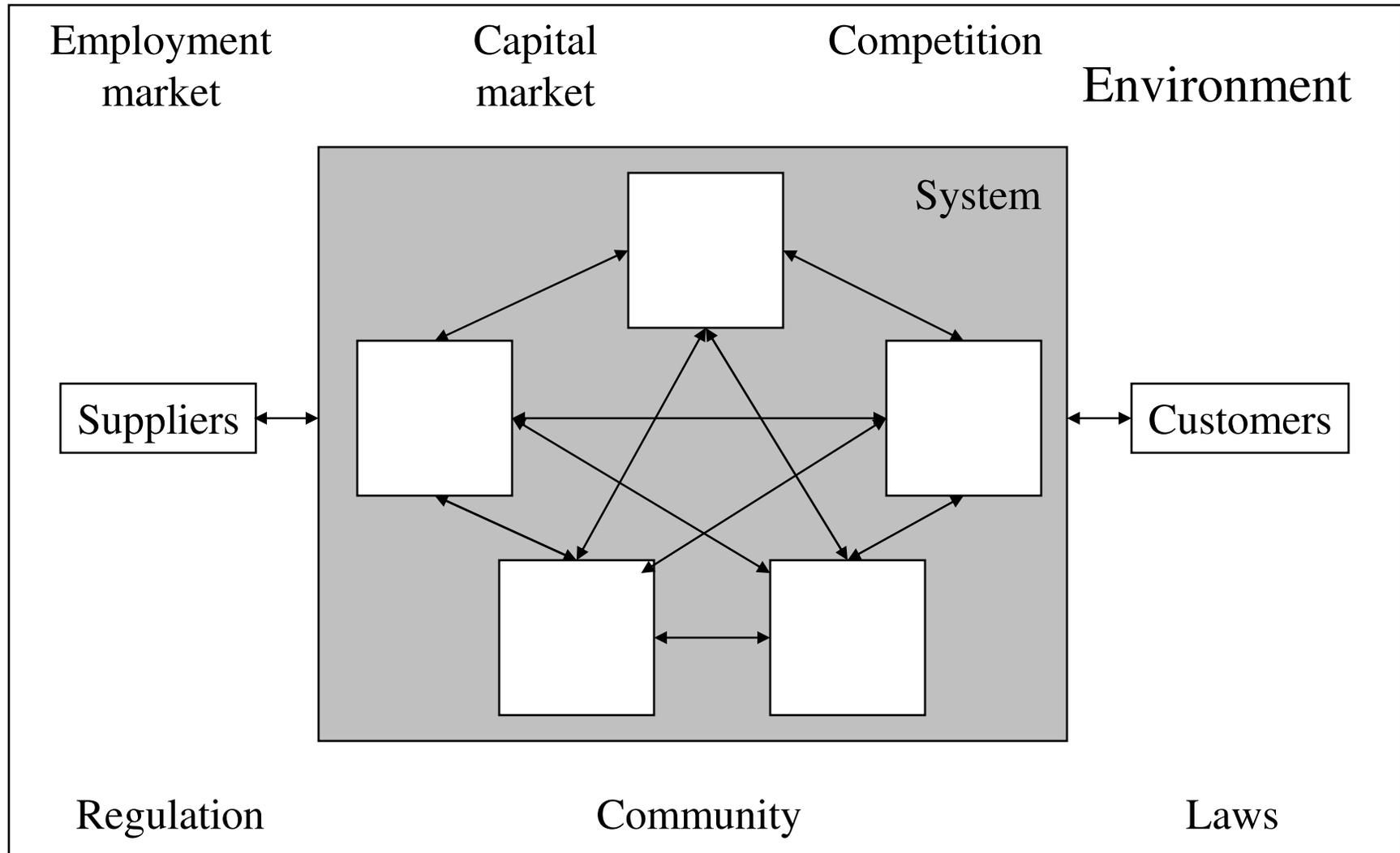
Principles of Management in a Dynamic Environment

- A system is a collection of interconnected components with a common goal
- There are quantitative objectives and performance measures
- There are subunits which interact within a hierarchical structure
- A process converts inputs into outputs and there may be feedback

A Traditional View of the Organization

- There is an internal system and an external environment
- Suppliers and customers are part of the external environment
- See page 17 for details

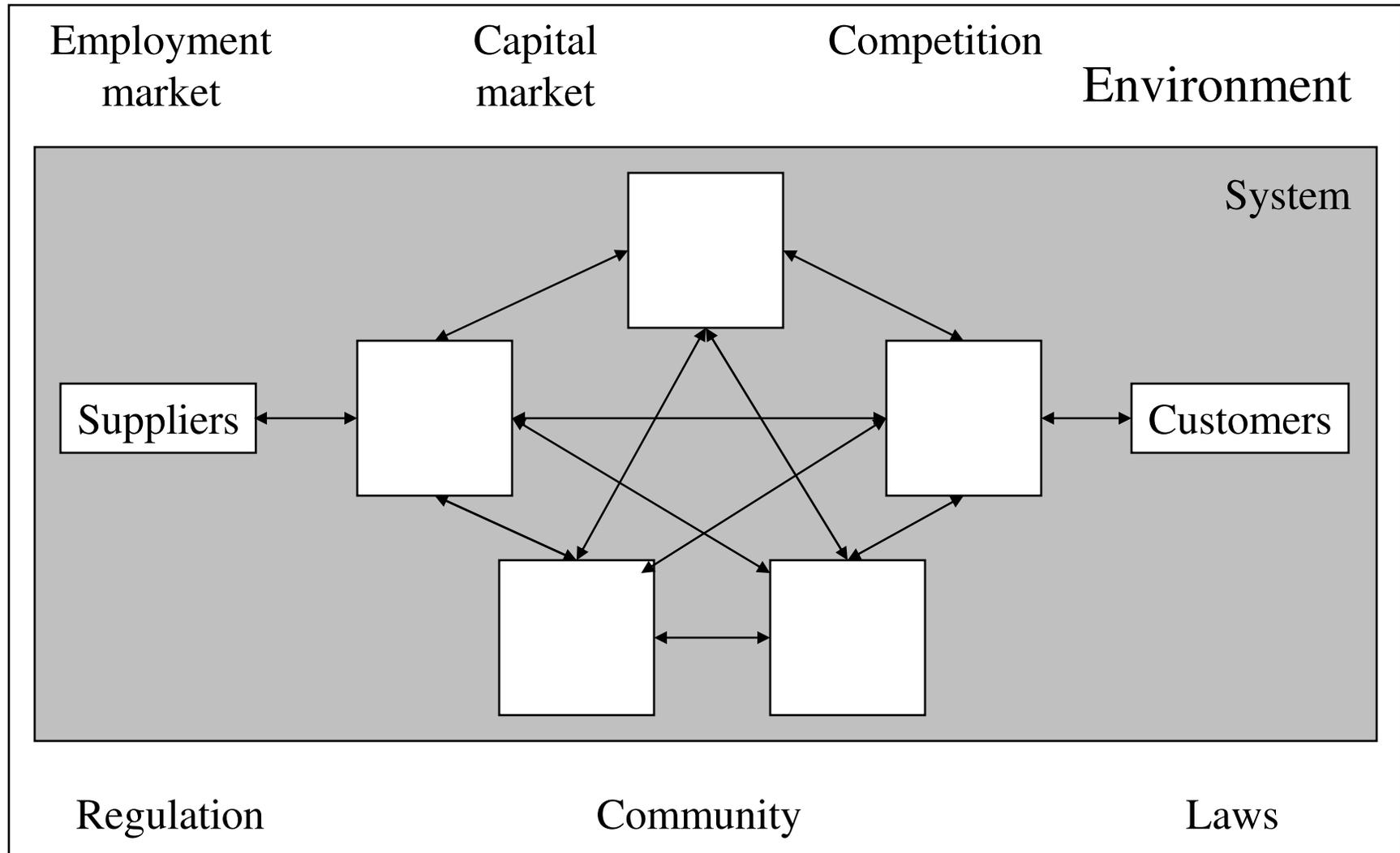
A Traditional Organizational System



Deming's (Modern) View of the Organization

- There is an internal system and an external environment
- Suppliers and customers are part of the internal system
- See page 19 for details

A Modern Organizational System



System Optimization and Suboptimization

- The performance of the whole system depends on a few factors - - the system constraints
- In a hospital, the operating rooms are often system constraints or bottlenecks
- If every subunit in an organization strives to function optimally, the entire organization may suffer
- This is called suboptimization or local optimization

Example of Suboptimization

- Purchasing Dept. in Hospital was judged based on purchasing costs
- It bought lower-cost, inferior-quality products
- Clinical and service failures resulted along with repeat hospitalizations

The Optimizer

- Optimizer: A decision maker who wants to make the best possible decision without consideration of time constraints
- To reach the optimal decision
 - One must generate all alternatives
 - Gather all the information
 - Build a model that will evaluate the alternatives
 - Choose the best one
- This requires time, effort, and money

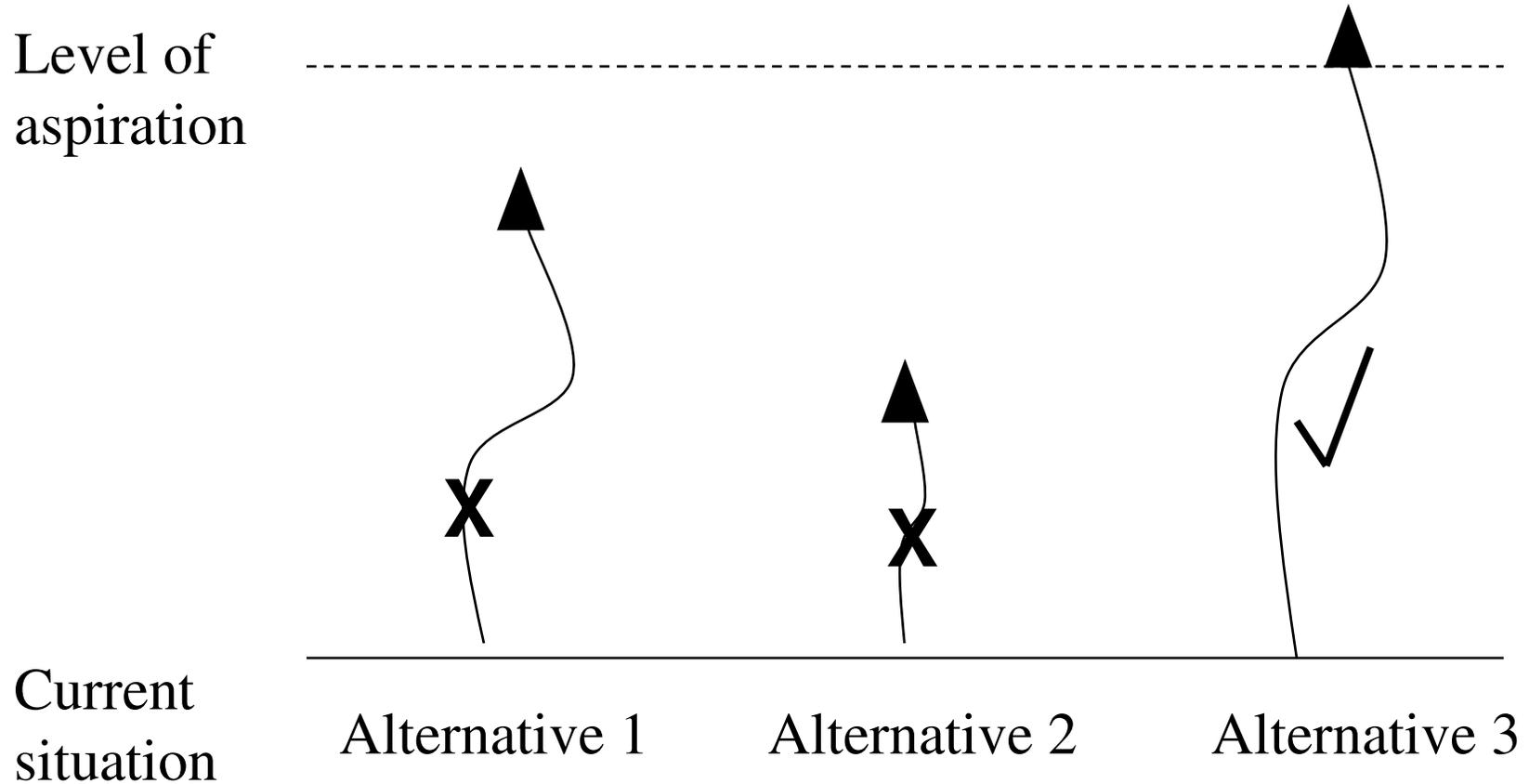
Challenges for the Optimizer

- Building the optimal model to evaluate alternatives is time and labor-intensive
- The optimizer may find the perfect solution, but it may come too late
- In a dynamic world, changes are frequent
- Timely decisions must be made
- This makes life challenging for the optimizer

The Satisficer vs. the Optimizer

- Herb Simon suggested that decision makers behave as satisficers
- They should seek to reach a satisfactory solution
- Satisficer: A decision maker who is satisfied with a reasonable solution that will clearly improve the system
- He does not seek an optimal solution

The Satisficer's Approach

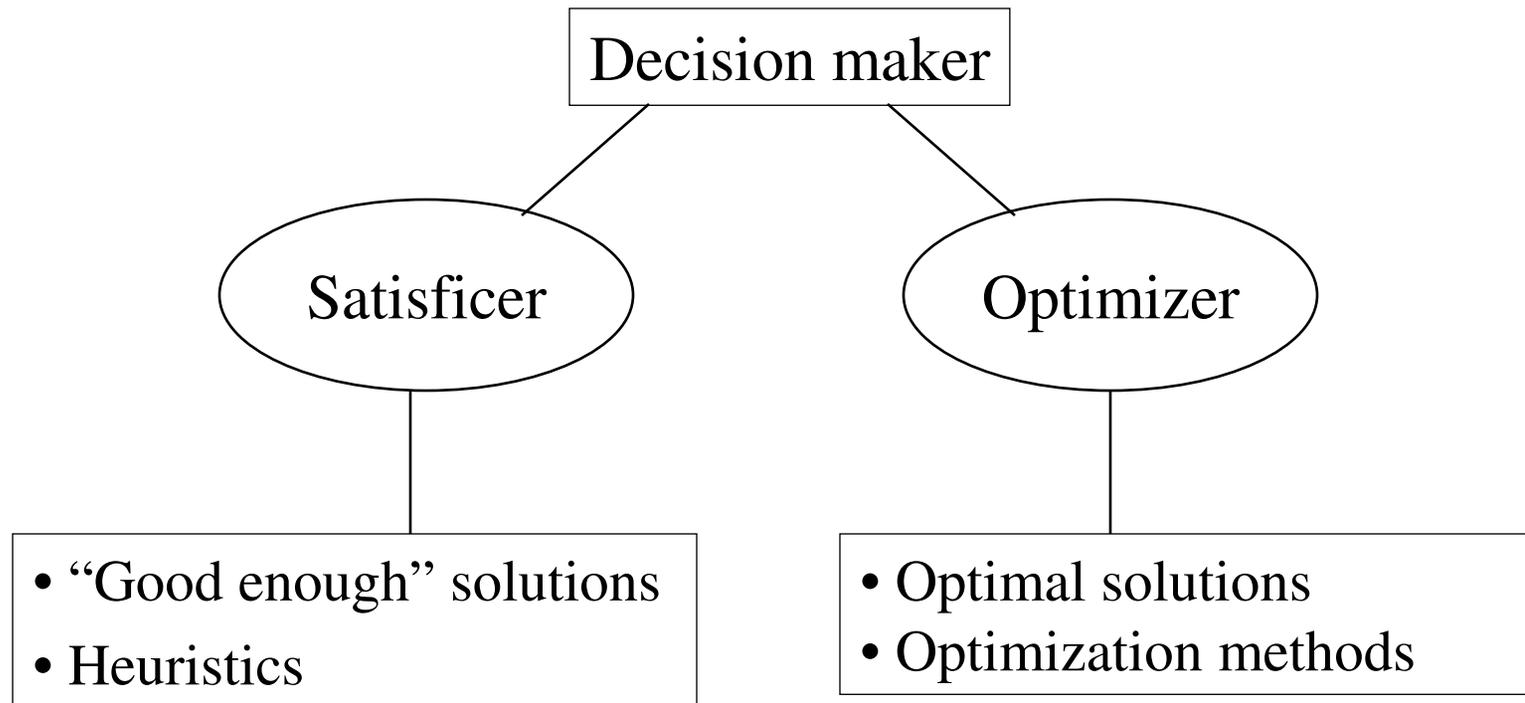


Driving Principles for the Satisficer

- A satisficer wins by complying with two principles
 - Set a high enough level of aspiration consistent with market conditions, competition, and investor expectations
 - Adopt an approach of continuous improvement
- A one-time improvement gives the firm a temporary edge
- Without continuous improvement, the temporary edge will be lost

Decision-Making Process: Optimizer vs. Satisficer

- The optimizer uses optimization techniques
- The satisficer uses heuristics
- The contrast



An Example

- Hospital A wanted to computerize patient records
- A consulting firm was hired
- It took six years to develop a system
- Two years later, the technology changed, making the system obsolete

An Example -- continued

- A competing hospital (Hospital B) adopted and adapted a computerized patient record system used in other hospitals
- Within one year, it worked reasonably well
- Hospital A sought an optimal solution
- Hospital B settled for a satisfactory solution

Focused Management

- The managerial approaches presented here are based on the satisficer approach and on heuristics
- We refer to them as focused management
 - Constraint management using the theory of constraints
 - Approaches to reduce response time
 - The value-focused management approach, etc.

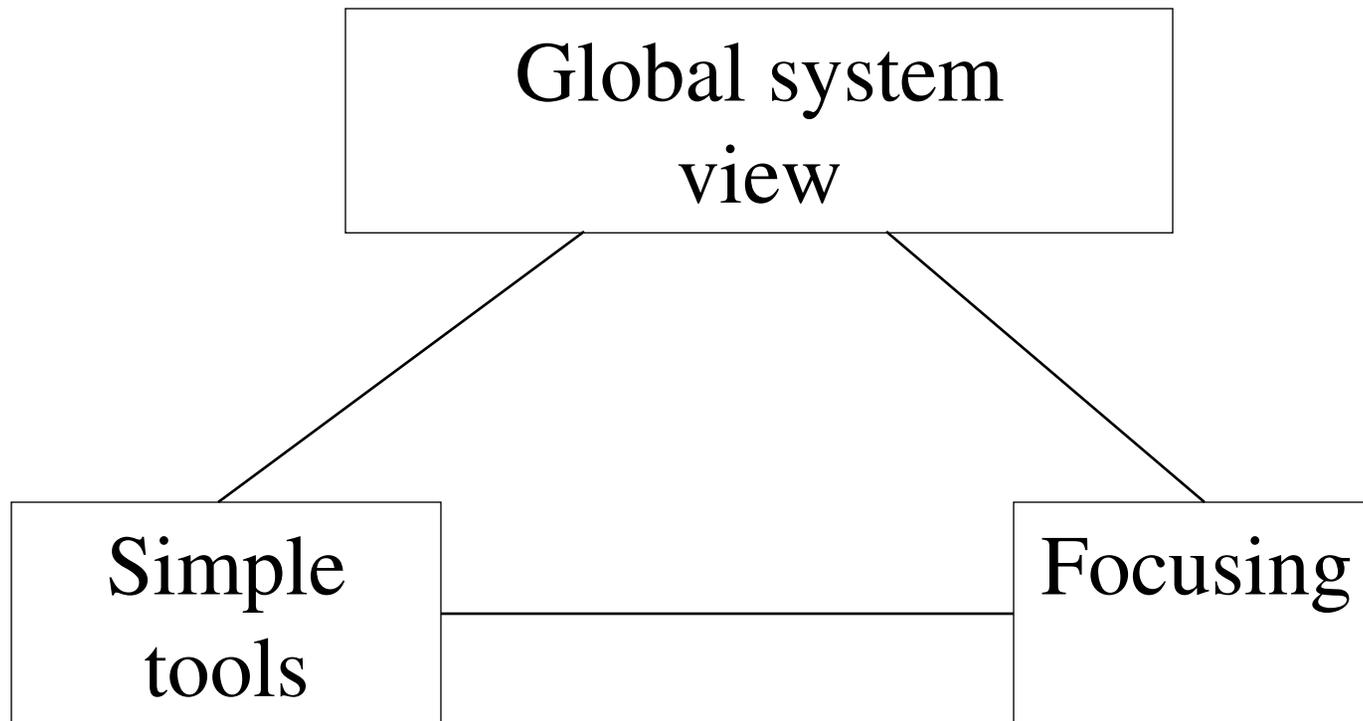
Definitions

- Focused management: Thrives on improving organizational performance and identifying the relevant value drivers and focusing on them
- Value drivers: Performance variables whose improvement will significantly increase the value of a business firm or the performance measures of a not-for-profit organization

Possible Value Drivers

- Increased contribution from sales
- Reduced time to market in developing products and services
- Increased throughput of operations and development activities in the organization
- Strategic focus
- Improved quality

Focused Management Triangle



The System Perspective

- In a private hospital, they tried to increase OR capacity
- More operating rooms and time slots were made available
- But throughput did not increase
- Why not? Because the recovery rooms did not have sufficient capacity to house additional patients

Expanding the Time Frame

- One needs to consider the total life cycle of a product
- A hospital was considering the purchase of a computer tomographic scanner
- There were several options available
- The initial price differences were hard to understand

Expanding the Time Frame

- After some research, it turned out that the machine with the lowest life-cycle cost had the most expensive initial price
- In subsequent purchases, suppliers were required to provide full life-cycle figures

Focusing on Essentials

- Type A problems

- Few, but important

- Solving these will contribute greatly to the organization

- Type B problems

- Of medium importance

- Solving these will contribute to the organization

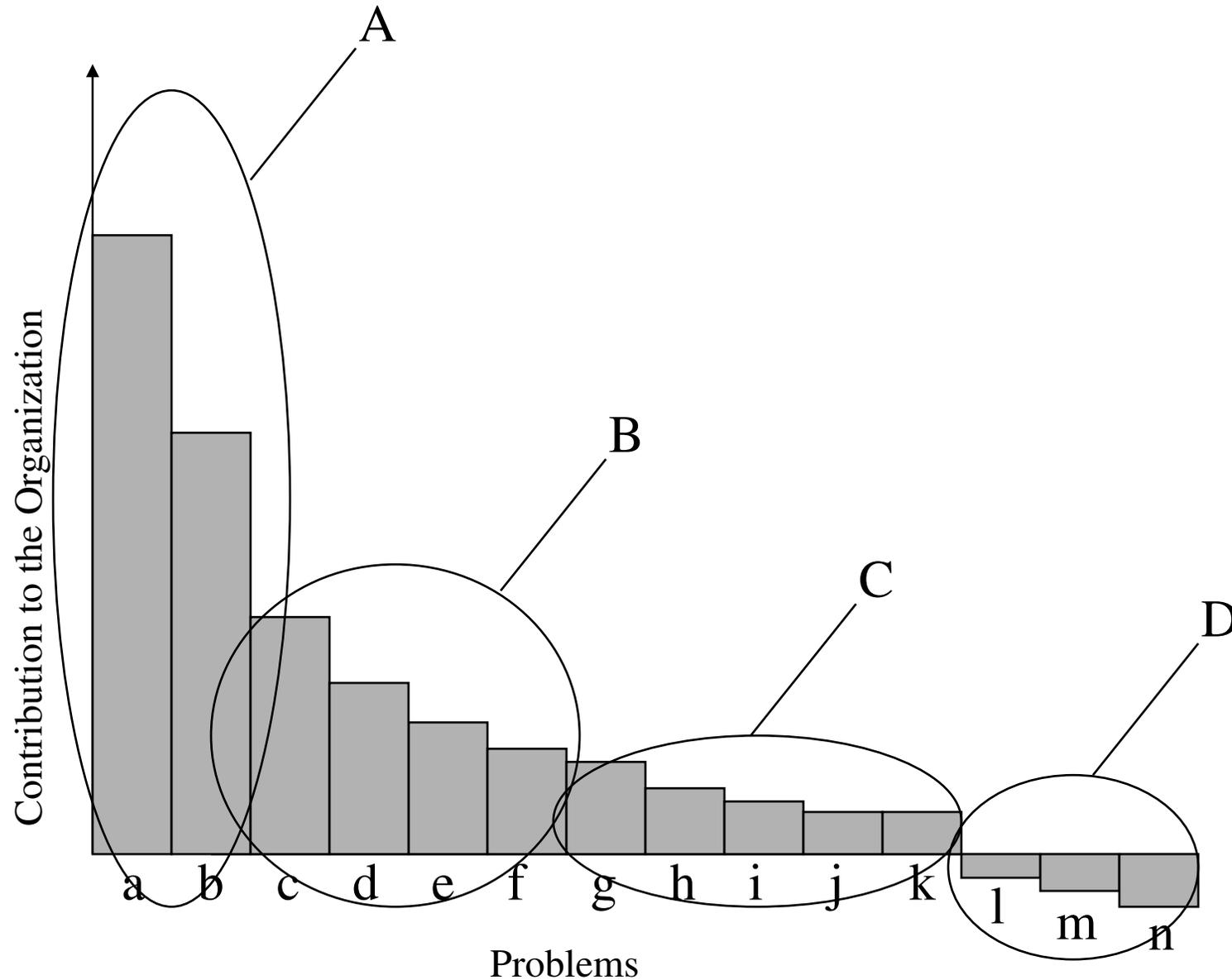
Focusing on Essentials

- Type C problems
 - Many routine problems
 - Solving these will contribute little
- Type D problems
 - Many easy-to-handle problems
 - Unimportant
 - Spending time on these has negative utility

Focusing on the Important Problems

- This classification scheme is presented on the next page
- It is tempting to deal with type C problems
- But, management must focus on type A problems
- Here is where managers can impact organizational performance in a major way
- Managers can delegate the small stuff

Classifying Organizational Problems



The Pareto Rule, Focusing Table and Focusing Matrix

- Pareto discovered that approx. 20% of the population has approx. 80% of world wealth
- This is called the “20-80 rule” and it describes many phenomena
- 20% of the patients in a hospital ward consume 80% of caregivers’ time
- 20% of patients consume 80% of medications

Other Examples of the Pareto Rule

- 20% of medications account for 80% of pharmaceutical costs
- 20% of laboratory tests account for 80% of laboratory costs
- 20% of suppliers provide about 80% of the value of products, materials, and components
- 20% of hospital inventory items constitute about 80% of the total inventory value

ABC Classification

- Group A: 20% of factors are responsible for 80% of outcomes
- Group B: 30% of factors are responsible for 10% of outcomes
- Group C: 50% of factors are responsible for 10% of outcomes
- An example follows

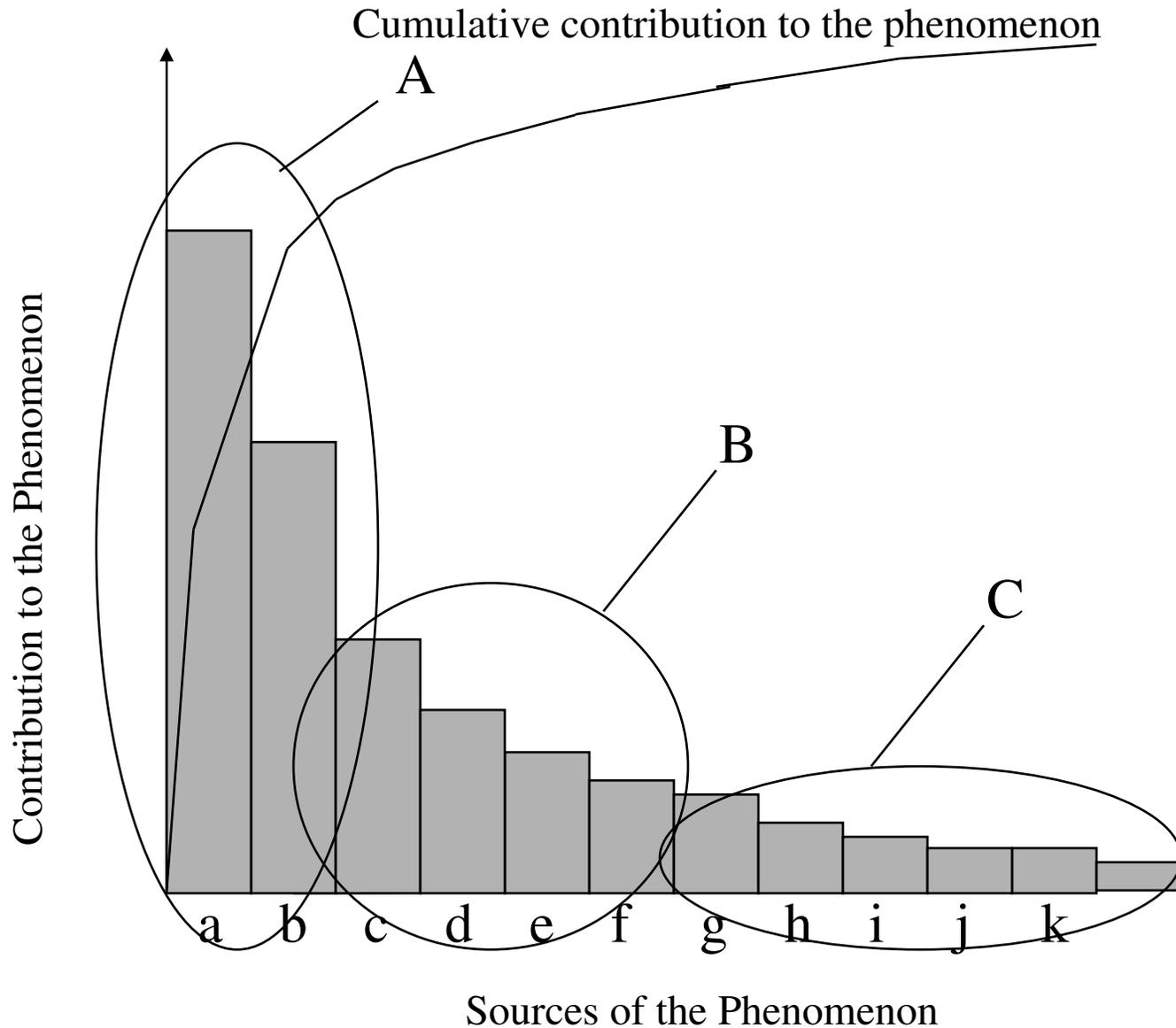
ABC Classification: An Example

- Group A: 20% of patients in a ward account for 80% of ward expenses
- Group B: 30% of patients in a ward account for 10% of ward expenses
- Group C: 50% of patients in a ward account for 10% of ward expenses

The Pareto Diagram

- A Pareto diagram visually displays the Pareto rule
- How does one construct a Pareto diagram?
 - List the sources of the phenomenon along with their contribution
 - Sort the sources by descending order of contribution
 - Draw a histogram as on the next page

A Pareto Diagram



Pareto Analysis of Drug Use in a Hospital

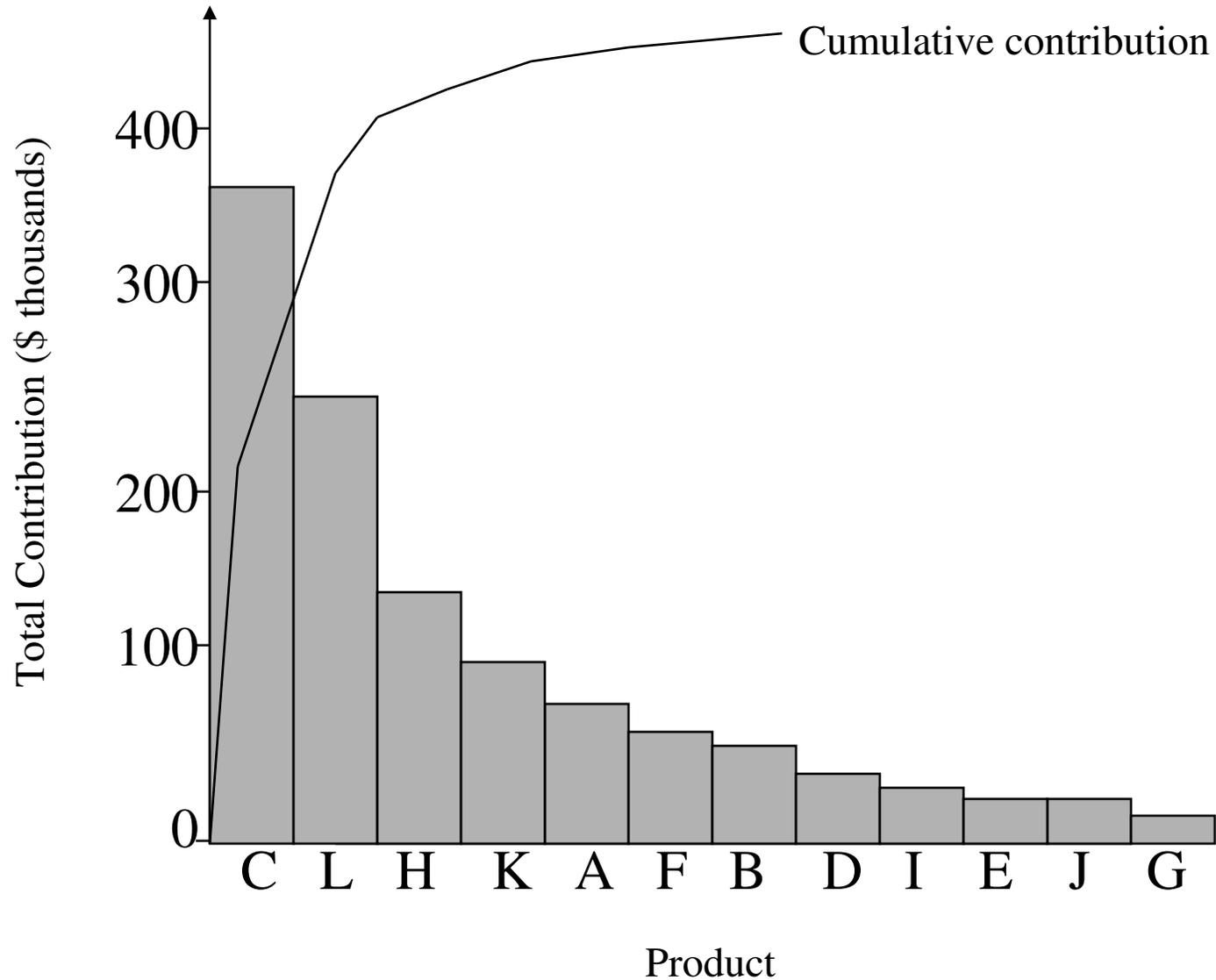
Drug Use Volume and Costs in a Hospital Ward			
Drug	Cost per Unit (\$)	Units Consumed per Month	Total Cost (\$ thousands)
A	180	361	65
B	250	160	40
C	950	347	330
D	90	389	35
E	75	267	20
F	560	89	50
G	1,350	11	15
H	650	169	110
I	220	114	25
J	15	1,333	20
K	56	1,518	85
L	150	1,367	205

Pareto Analysis of Drug Use in a Hospital

- Sort drug costs in descending order

Drug	Cost (\$ thousands)
C	330
L	205
H	110
K	85
A	65
F	50
B	40
D	35
I	25
E	20
J	20
G	15

Pareto Diagram of Analysis of Drug Costs



Building a Pareto Diagram

- List all sources of a phenomenon
 - Indicate the contribution of each source
- Rank all sources from largest to smallest contribution
- Draw a histogram of the sources (in rank order)
 - The y-axis reflects the size of the contribution
- The Pareto rule and Pareto diagram are especially useful in the presence of resource constraints or bottlenecks

Pareto-Based Focusing Method

- Classification: Classify the sources of the phenomenon
- Differentiation: Apply a differential policy
- Resource allocation: Assign resources appropriately
- An application in a large HMO is presented next
 - The purchasing department is a system bottleneck
 - It does not have the time to negotiate carefully with all suppliers

Classification of Suppliers by Purchasers

- Group A suppliers: The big suppliers are 20% of all suppliers and account for 80% of the dollar value of all purchases
- Group B suppliers: The 30% medium-size suppliers account for 10% of the total value of purchases
- Group C suppliers: The small suppliers constitute 50% of all suppliers but only 10% of purchase value

A Differential Policy for Each Supplier Group

- Group A suppliers: Comprehensive negotiations at the beginning of the year, detailed negotiations on the largest purchasing orders throughout the year
- Group B suppliers: A group of selected suppliers will be chosen, comparative price follow-up performed periodically
- Group C suppliers: Price discounts will be negotiated annually

Resource Allocation

- Most resources should be devoted to negotiations with group A suppliers
- Few resources should be invested in dealing with group B suppliers
- Group C suppliers will be evaluated occasionally
- It may not always make sense to focus on monetary contribution - - one alternative is item criticality

Monitoring Drug Consumption

- A large HMO wants to control the drug consumption of its patients
- Classification: Patients were classified according to the monetary value of the drugs they consumed
- Group A patients are the 15% of patients who were responsible for 75% of the dollar cost of drug consumption

Classifying Drug Consumption

- Group B patients are the 25% of patients with moderate drug consumption, which accounts for 15% of total drug costs
- Group C patients are the remaining 60% of patients who consume only 10% of the drugs

Drug Consumption: Differentiation

- Group A patients will be evaluated by the medical director of the HMO and the chief pharmacist
 - Every prescription must be approved by the medical director
- A (random) 10% of prescriptions for group B patients will be screened to verify reasonable and cost-effective practice
- 5% of patients in group C will be randomly evaluated

Drug Consumption: Resource Allocation

- Most resources for managing drug consumption will be devoted to group A patients
- Limited resources will be targeted to patients from groups B and C
- Next, we introduce the focusing table and the focusing matrix

Building the Focusing Table

- The emergency department (ED) in a large hospital wants to improve its performance
- Numerous meetings take place
- Every suggestion is evaluated with respect to
 - Importance
 - Ease of implementation
- The table on the next page emerges

Emergency Department (ED) Focusing Table

Item Number	Suggestion	Importance ^a	Ease of Implementing ^b
1	Separate ED into surgical and internal wards	4	2
2	Change strategy regarding amount of testing	5	2
3	Open additional imaging room using same personnel	4	5
4	Increase frequency of visits by specialists	4	4
5	Increase frequency of lab workup	5	5
6	Measure average waiting times	4	4
7	Shorten discharge procedure	5	4
8	Redesign admission process	3	3

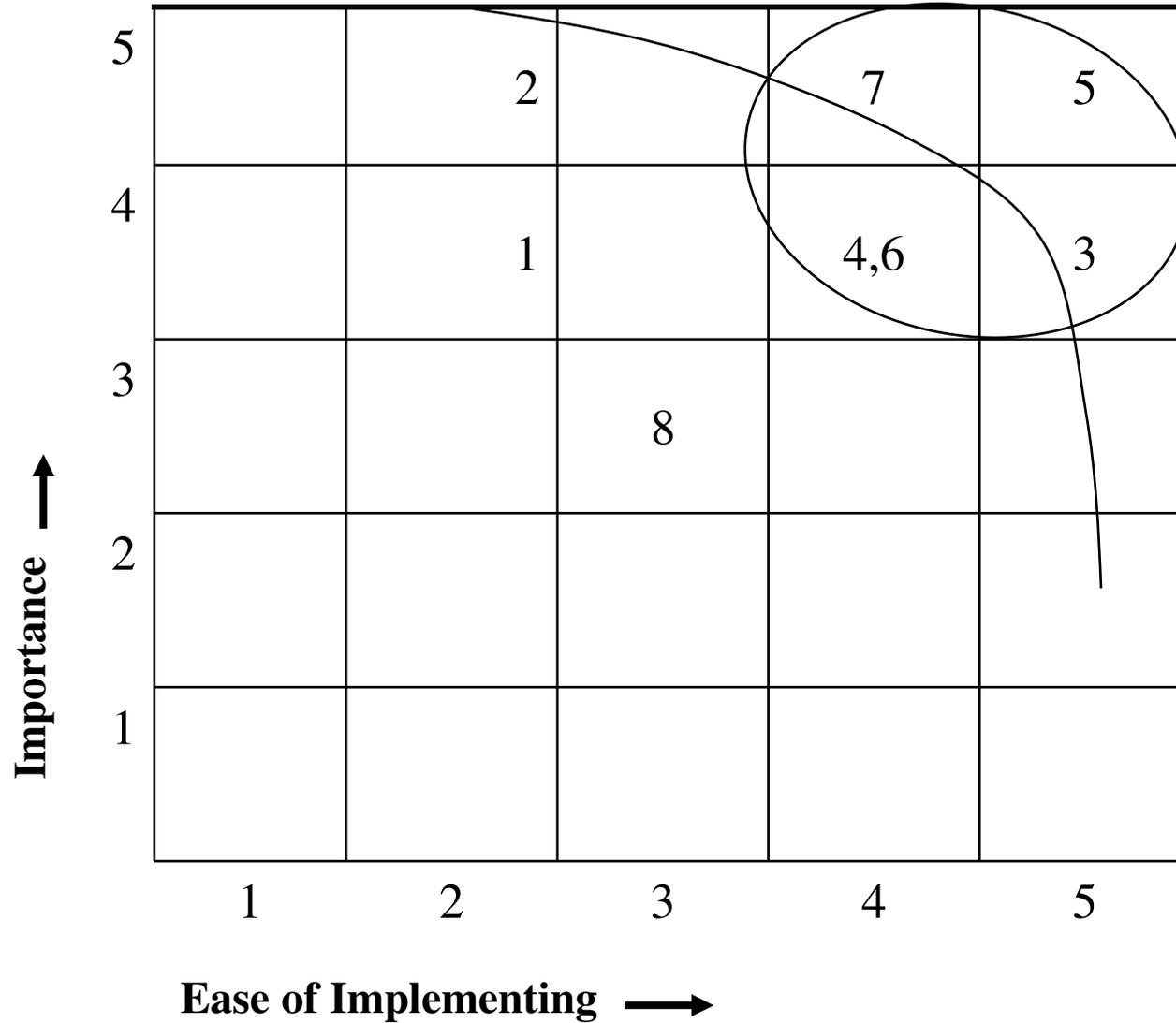
^a Scale is 1 (unimportant) to 5 (important).

^b Scale is 1 (very difficult) to 5 (very easy).

Generating the Focusing Matrix

- The focusing matrix is an extension of the focusing table
- The preferred suggestions are those near the upper right corner of the matrix (see the next page)
- Suggestion 5 dominates the others
- Suggestions 3 and 7 also look good, no dominance

Focusing Matrix for the ED Example



Applications of the Focusing Matrix

- Choosing among patient case studies to be discussed in morning rounds
- Choosing among projects to be budgeted using a hospital's development fund
- Choosing among activities in the process of organizational improvement

Review of Guidelines

- Make a list of subjects/items
- Include importance and ease of implementation
- Build a focusing table
- Construct a focusing matrix
- Focus on the subjects/items in the upper right corner of the matrix
 - These are important and easy to implement

Use the Pareto Rule Carefully

- Underlying assumptions may be violated
- Remedies
 - Pareto analysis is only relevant in the presence of resource shortages
 - Use Pareto analysis where the relevant benefits or damages are on the y axis
 - Take advantage of the focusing matrix and focusing table

Management by Constraints

- Management by constraints is an innovative and effective approach developed by Goldratt and Cox (1992)
- Management by constraints is based on a seven-step process
 - Determine the system's goal
 - Establish global performance measures
 - Identify the system constraint
 - Decide how to exploit the constraint
 - Subordinate the rest of the system to the constraint
 - Elevate and break the constraint
 - If the constraint is broken, return to step 3

Determine the System's Goal

- The goal of an organization should guide every decision and action in the organization
- The goal of a business organization is to increase shareholders' value
- In not-for-profit organizations, the goal is determined by the mission of the organization
 - For example, the goal of a public health care organization is to maximize quality medical services provided to customers, subject to budgetary constraints

Determining Goals can be Tricky

- Make a quick decision on admission or discharge in an ED versus provide a comprehensive diagnostic workup
- Within an HMO, high-quality medical care versus long-term cost reduction
- Enhancing prestige within a large hospital versus increasing profits

Establish Global Performance Measures

- Performance measures serve as a guide towards the achievement of the organization's goal
 - E.g., the value of a company
- There is no single perfect performance measure
- But, we can define six basic performance measures
- We list them next

Six Basic Performance Measures

- Throughput (T)
- Operating expenses (OE)
- Inventory (I)
- Response time (RT)
- Quality (Q)
- Due-date performance (DDP)

Step 3: Identify the System Constraint

- The idea is to identify the causes that prevent the system from achieving its goal
- This involves a search for factors that restrict system performance
- Constraint: Any important factor that prevents an organization from achieving its goal
- Every system has a constraint

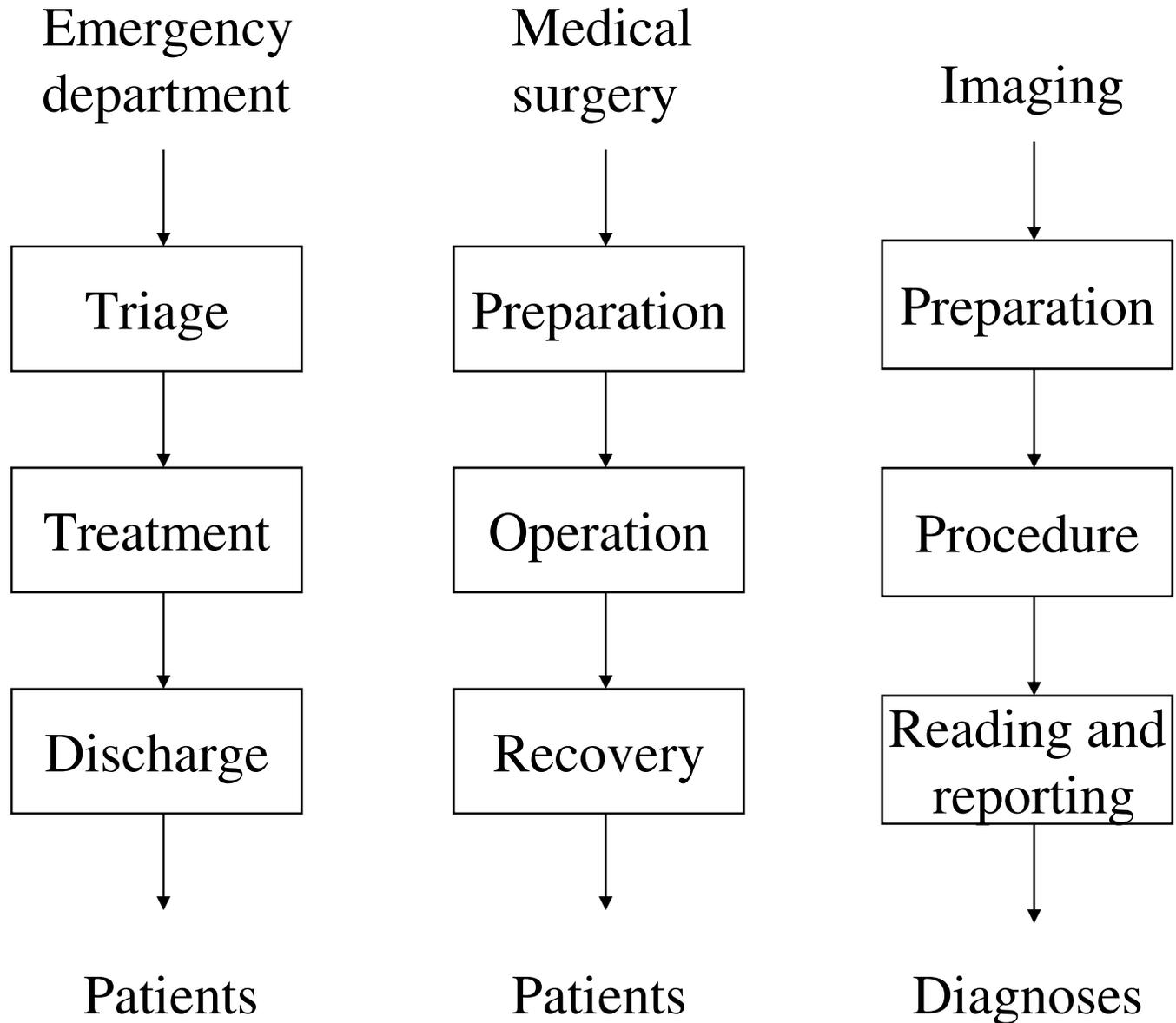
Identify the System Constraint

- If there were no constraints, unbounded performance would result
- In most cases, there are a small number of constraints
- Four types of constraints in a managerial system
 - Resource constraint
 - Market constraint
 - Policy constraint
 - Dummy constraint

Resource Constraint

- The resource constraint is often called the bottleneck
- This is the resource that constrains the performance of the entire system
- If only we had more of it
- On the next page, we see three managerial systems from the work processes point of view

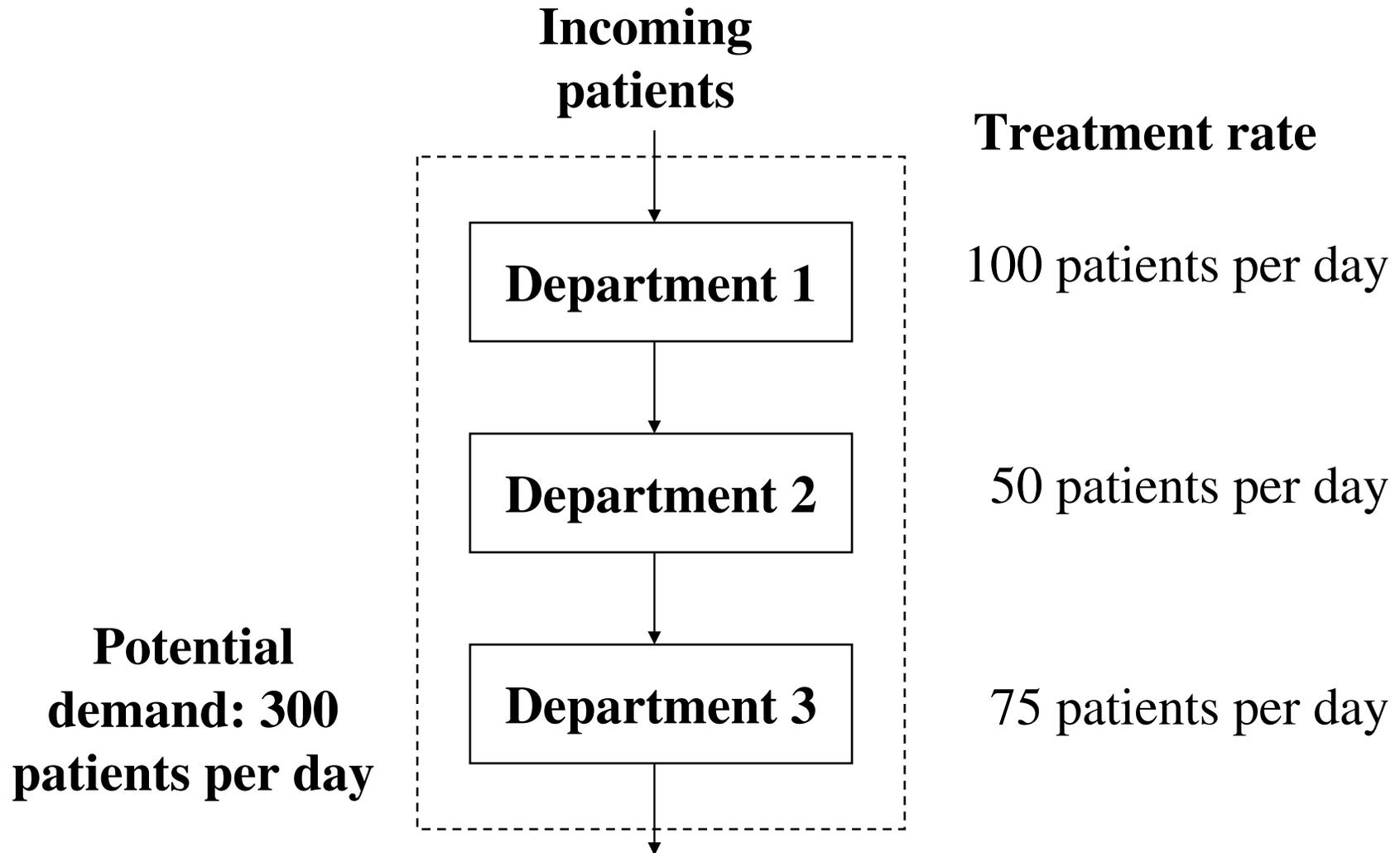
System Processes



The System Constraint

- On the next page, we see that each patient must be processed in each of three departments
 - Dept. 1 → Dept. 2 → Dept. 3
- The market wants to see 300 patients processed per day
- Departments 1 and 3 cannot handle the market demand, but they do not constrain the system
- Department 2 is the resource that is the system constraint
 - It can process only 50 patients per day

A System with a Resource Constraint



Lessons from the Example

- The system has a resource constraint
- Department 2 is the system bottleneck
- If we increase the daily capacity of Departments 1 and 3, the throughput (capacity) of the whole system will not change
- On the other hand, increasing the throughput of Department 2 will increase system throughput
- The bottleneck, Department 2, dictates the throughput for the whole system

Bottlenecks in Hospitals

- Bottlenecks exist in all areas of life
- Some hospital-related examples
 - In an OR at Hospital A, the bottleneck was the surgeon
 - In an OR at Hospital B, the bottleneck was the anesthetist
 - In an OR at Hospital C, the bottleneck was the room itself
 - In the ED of a hospital, the bottleneck was the emergency physician

Other Examples of Bottlenecks

- In a specialty clinic, an expensive technology (e.g., positron emission tomographic scan)
- In an office of health insurance claims, the lawyers who have to approve every settlement
- In a large HMO outpatient clinic, the physicians
- In a hospital obstetrics-gynecology ED, the imaging services
- At an airport, during peak times, the runways

Shortage of a Critical Resource

- A highly skilled surgeon in a hospital
- Highly skilled nurses in a hospital
- The cleaning crew of the ORs
- A magnetic resonance imager in a hospital
- It is not always easy to open/relieve a bottleneck quickly
 - It may require extensive capital
 - It may require long training periods

Permanent Bottlenecks

- Sometimes, a permanent resource constraint and permanent bottleneck exist
- Examples
 - Physicians with unique expertise
 - Anesthetists
 - ICU nurses
- In other situations, there is not a constant shortage of a critical resource

Peak Time Resource Constraints

- There are shortages at specific (peak) times
 - In a hospital ED after a bridge collapse
 - At UPS during the Christmas season
 - Airplane seats on Monday morning and Friday afternoon
- Resources are at excess capacity most of the time and at shortage during peak times
- The issue of peak time management may involve differential pricing of goods and services

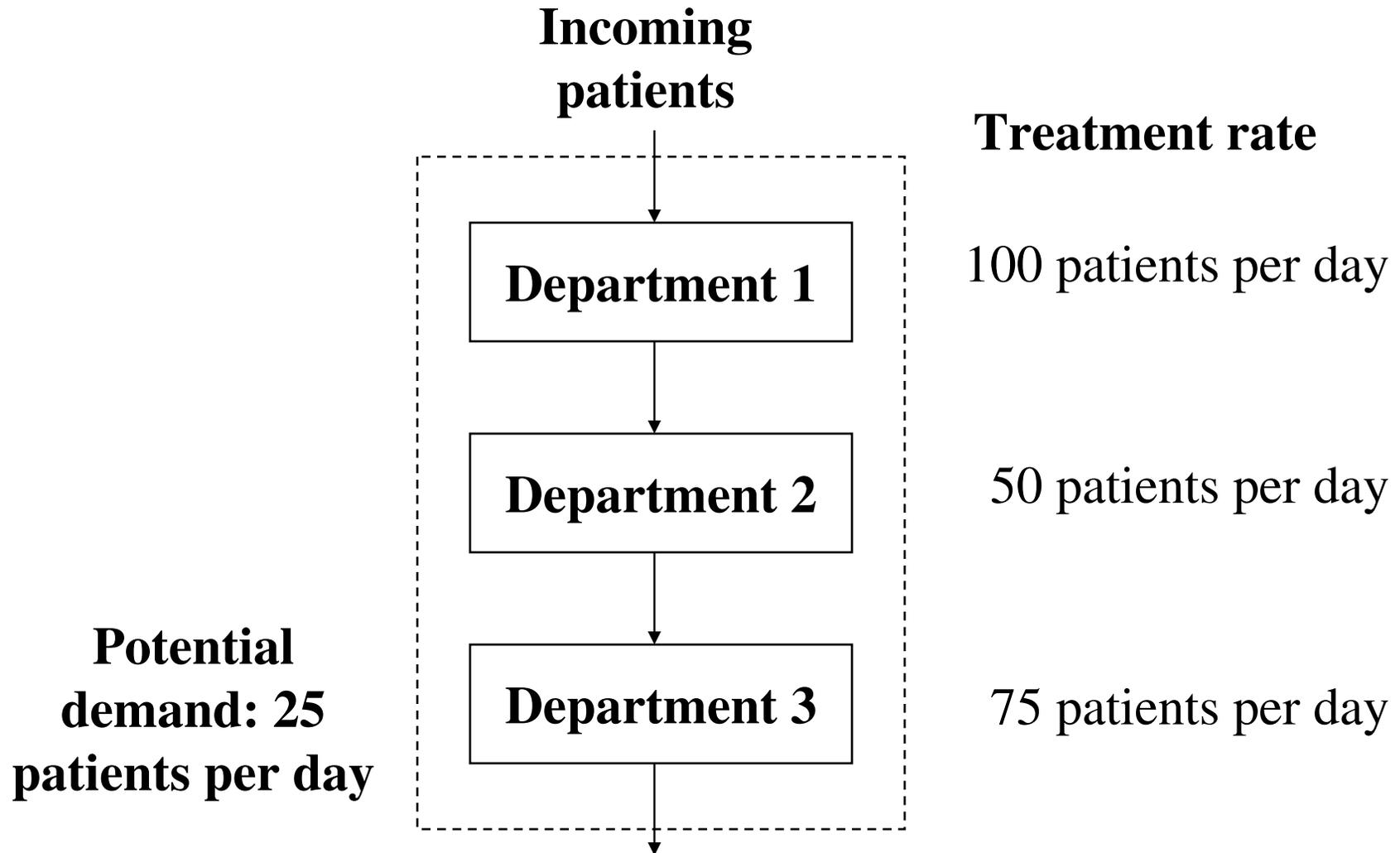
More on Constraints and Bottlenecks

- Some examples of peak time management relate to seasonality
 - Incidence of the flu in winter
- If we could increase the capacity of the resource identified as the bottleneck, would system throughput increase?
- If we reduce the capacity of this resource, will system throughput decrease?
- If the answer is yes, this resource is the system bottleneck

Market Constraint

- Definition
 - A situation where the market demand is less than the output capacity of each resource
 - Thus, market demand is the constraint that prevents the system from achieving its goal
- On the next page, we see that each one of the three resources has an excess capacity
- The market constrains the system here
- In the health care industry, we face both resource and market constraints

A System with a Market Constraint



Policy Constraint

■ Definition

- Adopting an inappropriate policy that limits system performance and achievement of goals and that may push in a direction that is against the organizational goal
- This is also known as a policy failure

■ An example

- A hospital is reimbursed by length of stay

The Impact of Policy: An Example

- A hospital is reimbursed by length of stay
 - As a result, there is less motivation to discharge patients early
 - Longer hospital stays result
 - Increased incidence of infections
 - It becomes difficult to handle as many patients as the hospital would like

A Second Example

- The hospital director forbids overtime work for hospital staff in order to contain costs
- This constrains the number of operations daily
- It increases the waiting time for surgery
- Some patients decide to go elsewhere
- This is an extremely negative outcome for the hospital

More Examples of Policy Constraints

- Setting standards that each employee must achieve
 - There is no incentive to exceed the standard
- Continuing to invest in a failing project because large amounts have already been invested in it
- Across-the-board personnel cuts of 10%
 - This may be counterproductive for the organization
- When is a policy constraint a system constraint?
 - When “breaking” the policy constraint \Rightarrow increased throughput

Dummy Constraint

■ Definition

- A situation where the system bottleneck is a relatively cheap resource compared with other resources in the system

■ An example

- A hospital OR used for coronary angiographies fell behind its schedule
- Surgeons, radiologists, nurses, surgical kits, etc. were available
- But, the OR was sometimes not being used

Dummy Constraint

- Why was the OR not in use?
 - The OR needs to be thoroughly cleaned between procedures
 - In order to cut costs, the hospital laid off one of two cleaners
 - The remaining cleaner had to clean both ORs and intensive care rooms
 - Thus, an inexpensive resource (the cleaning person) became a system constraint

Dummy Constraint

- An example from a hospital internal medicine ward
 - Blood specimens were placed in trays for transport to the lab
 - A shortage of trays \Rightarrow delays in collecting blood specimens \Rightarrow delays in receiving results \Rightarrow discharge delays
 - Again, an inexpensive resource prevented the ward from operating efficiently

Dummy Constraint

- An example from the ED
 - There was a shortage of clerical personnel for discharging patients
 - This led to discharge delays \Rightarrow overcrowding in the ED
 - A clerk is a relatively inexpensive resource
- Shortages in phone lines, fax machines, printers, blood pressure monitors, etc. are all dummy constraints

When is a Dummy Constraint a System Constraint

- If we could break the dummy constraint, could we increase throughput and enhance organizational value?
- If the answer is yes, then the dummy constraint is a system constraint
- Next, we discuss tools for identifying constraints in a health care system

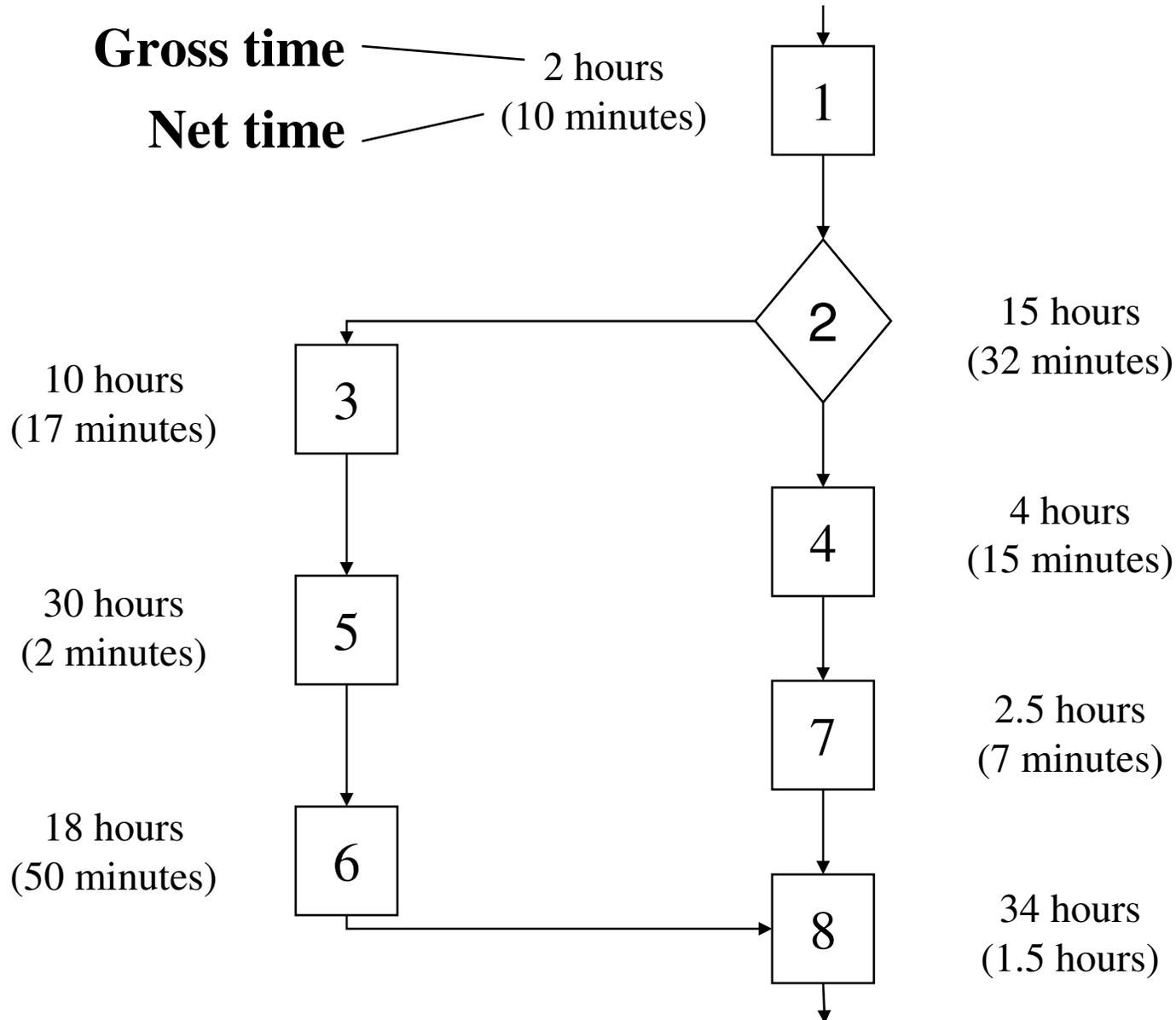
Tools for Identifying Constraints

- Ask workers in the field
- Ask the evening cleaning crew
- Tour the work area
- Several methodological techniques
 - Process flow diagram
 - Time analysis
 - Load analysis
 - Cost-utilization (CUT) diagram

Process Flow Diagram

- A process flow diagram is a basic flow chart
 - It describes the work flow
 - It includes the stages of the process and the decision nodes
- We try to simplify the process flow diagram
- Each work step includes actual processing time (net time) and total time (gross time)
- Gross time includes waiting time

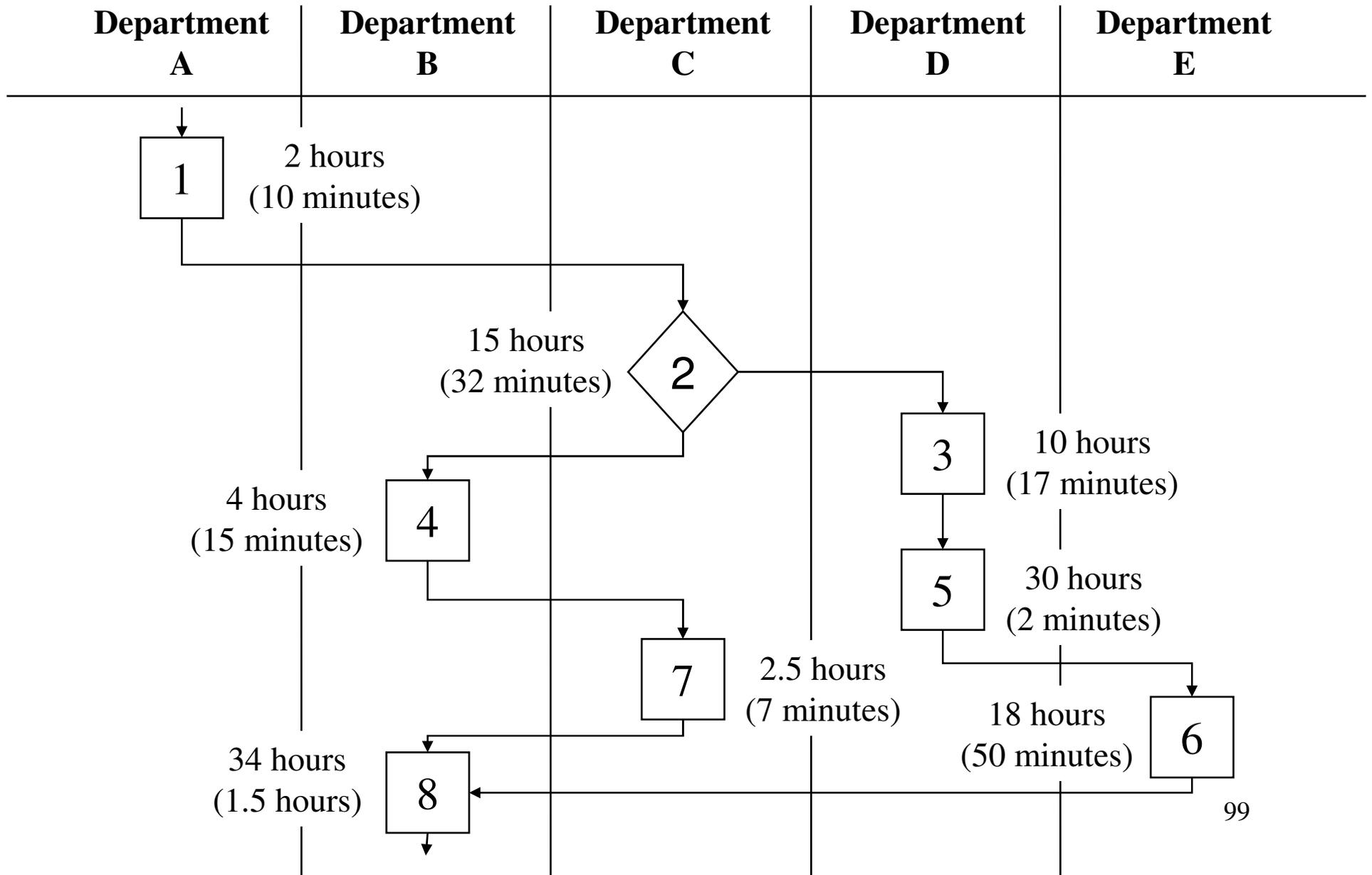
A Process Flowchart



Process Flow Diagram

- The basic process flow diagram can be expanded into a two-dimensional diagram
 - It presents various tasks performed by various departments
 - It helps people understand the work flow
 - Visual aids like this are always useful
- Analyzing the gross time spent by a patient in different parts of the hospital enables identification of the station where he/she spent the most time

A Two-Dimensional Process Flow Diagram



Process Flow Diagram

- The long (gross) time is usually due to waiting in line
- Note that step 8 has the longest (gross) duration of 34 hours
- Step 8 is suspected to be a system constraint
- Load analysis or capacity utilization is a simple tool for identifying the system bottleneck/most heavily used resource in the system

Load Analysis

- To determine the load on resources, we need
 - The total number of labor hours during the time period
 - The overall planned work
 - A table describing the effort in labor hours that each resource is required to invest in each product, customer, etc.

Load Analysis

- A manufacturer of surgical equipment (named MOSE) receives orders for surgical kits
- Can MOSE produce the entire order?
- MOSE has 190 monthly labor hours (190 monthly labor hours are available at each station)
- Each kit must visit each station, unless otherwise indicated
- Orders for next month
 - 100 kits of type A, 50 kits of type B, 25 kits of type C, 200 kits of type D

Load Analysis

- MOSE has four product stations
- See page 104 for more details
- Review page 105
- Observe that Station 3 cannot perform its task in 190 hours
- It is the most heavily utilized station and the system constraint

Labor Hours per Unit per Station

Labor Hours per Surgical Kit

Surgical Kit	Station 1	Station 2	Station 3	Station 4
A	0.60	0.15	0.73	—
B	0.35	0.72	1.18	0.50
C	1.60	—	1.36	2.56
D	0.20	0.06	0.44	0.41

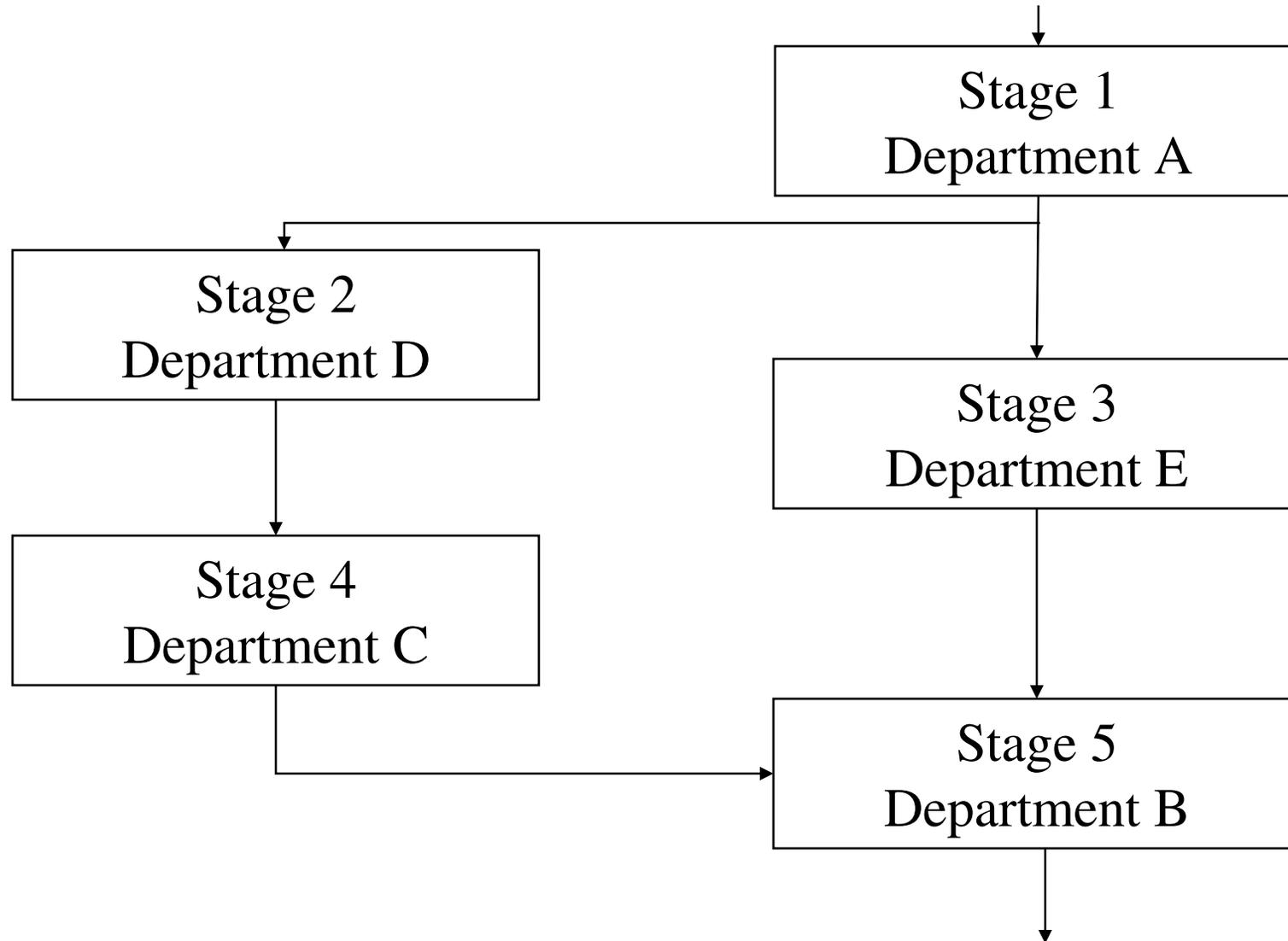
Load Analysis

Surgical Kit	Quantity	Station 1	Station 2	Station 3	Station 4
A	100	$100 \times 0.60 = 60$	$100 \times 0.15 = 15$	$100 \times 0.73 = 73$	—
B	50	$50 \times 0.35 = 17$	$50 \times 0.72 = 36$	$50 \times 1.18 = 59$	$50 \times 0.50 = 25$
C	25	$25 \times 1.60 = 40$	—	$25 \times 1.36 = 34$	$25 \times 2.56 = 64$
D	200	$200 \times 0.20 = 40$	$200 \times 0.06 = 12$	$200 \times 0.44 = 88$	$200 \times 0.41 = 82$
Total hours		157	63	254	171
Load (%)		83	33	134	90

Using a Cost-Utilization Diagram to Identify the Constraint

- A cost-utilization (CUT) diagram of a system is a bar graph where every bar represents a resource
- Bar height corresponds to resource utilization (in percent)
- Bar width corresponds to relative cost
- Relative costs can be defined in several ways
- The recommended approach is to use the marginal cost of each resource
- The order of bars on the horizontal axis is arbitrary

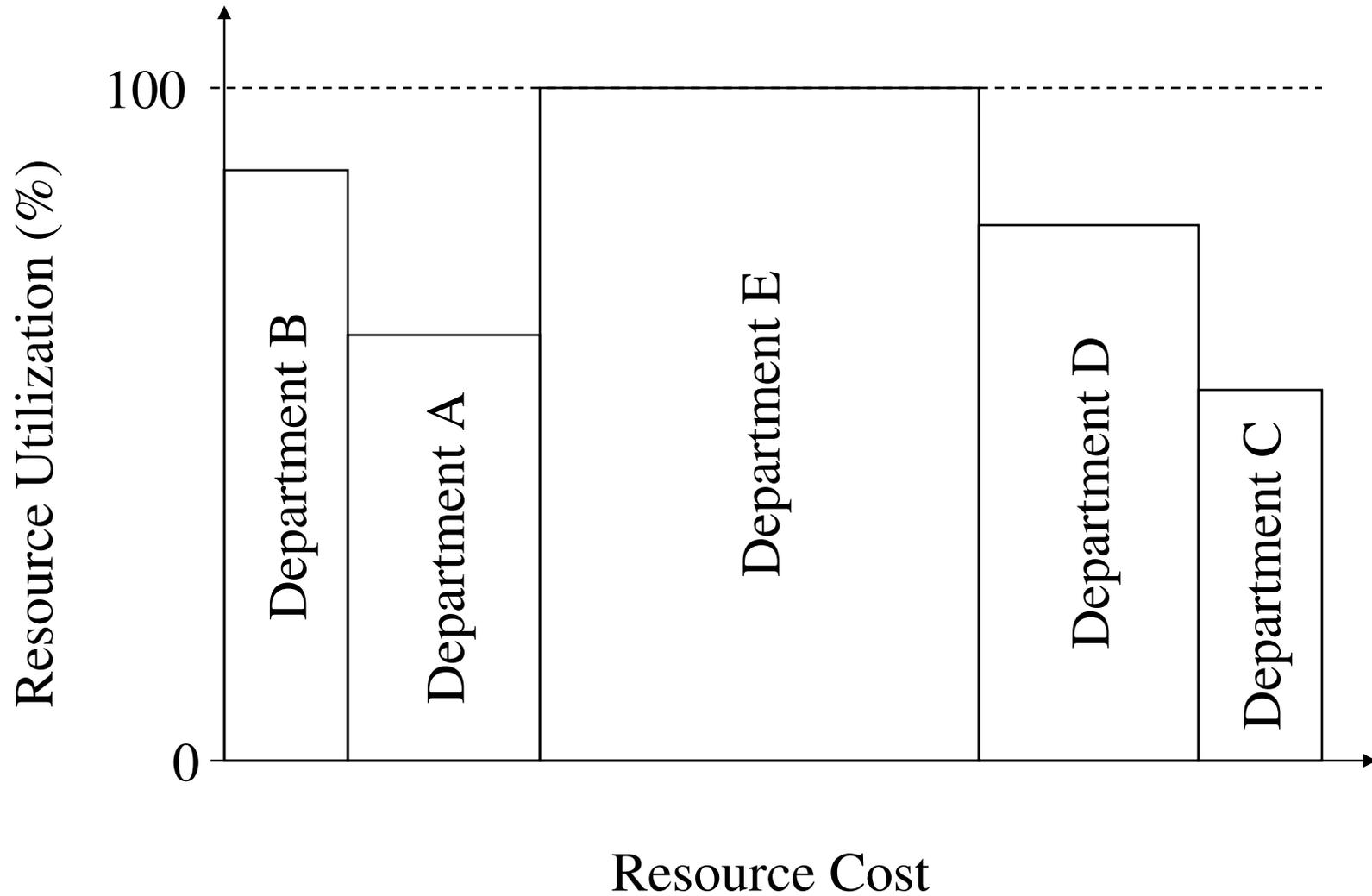
A System and its Work Process



Load Analysis with a Bottleneck

Resource	Load (%)	Cost of Resource (\$ thousands)
Department A	55	100
Department B	80	50
Department C	45	40
Department D	65	100
Department E	100	280

CUT Diagram of a System with a Resource Constraint



Using a CUT Diagram of a System

- On page 109, we see that the expensive bottleneck is the only fully used resource
- In an OR, the surgeon or anesthetist may be the bottleneck
- For an airline, the planes may be fully utilized and the crews only partially utilized
- In these cases, we note the following points

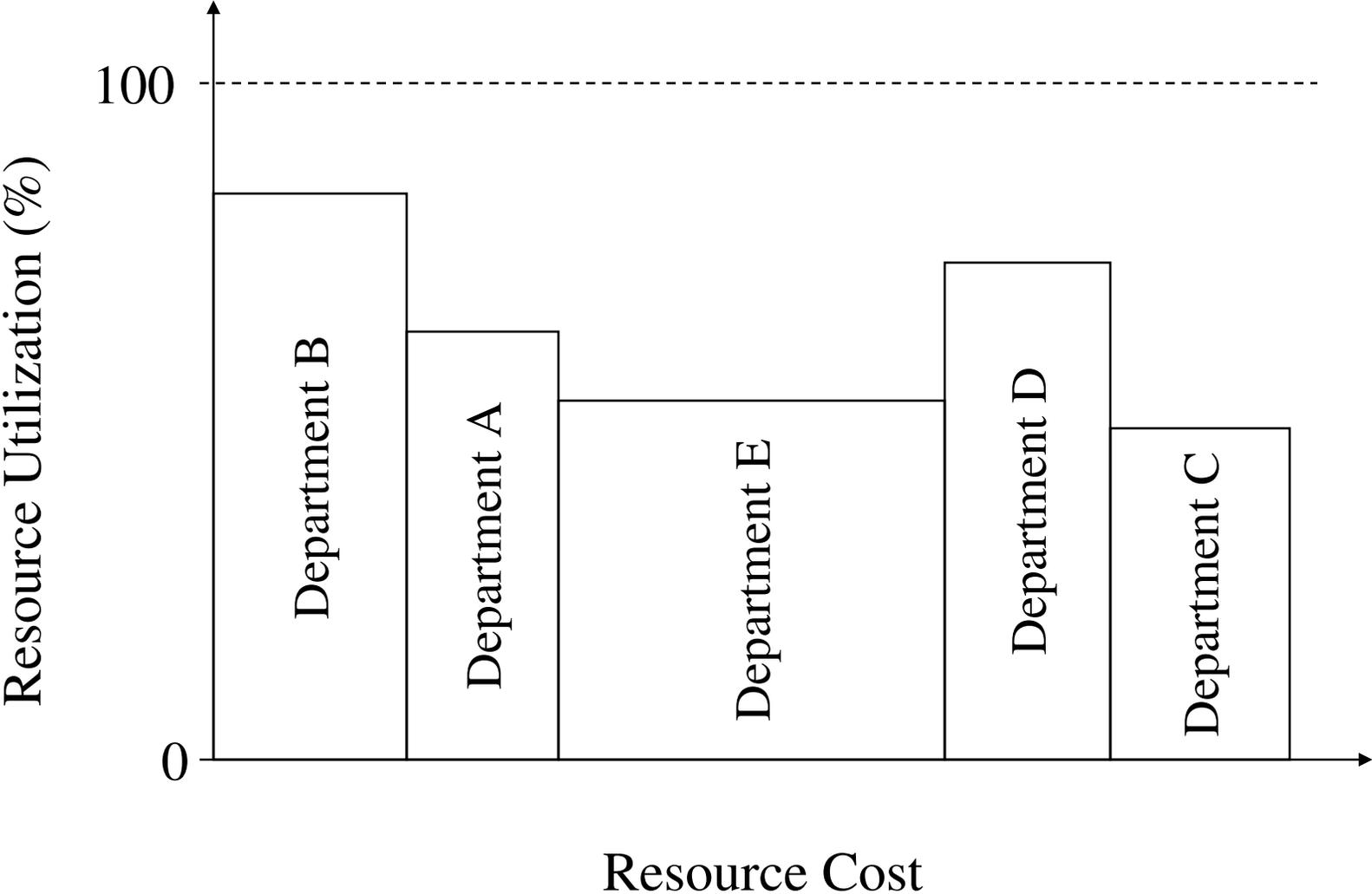
How to Deal with a Fully Utilized, Expensive Resource

- Is it possible/feasible to operate at excess capacity?
- If there is excess capacity in non-bottleneck resources, is it possible to rent/sell this excess capacity in an external market?
 - E.g., if a surgeon is the bottleneck in the OR, can the OR be rented to an external surgeon?
- Next, we consider another system
- The system is displayed on page 107, but the loads and costs are shown on page 112

Load Analysis in a System with a Market Constraint

Resource	Load (%)	Cost of Resource (\$ thousands)
Department A	65	10
Department B	80	17
Department C	45	15
Department D	70	15
Department E	55	28

CUT Diagram of a System with a Market Constraint



An Example of a Market Constraint

- The CUT diagram shows that the system has market demands that are lower than the capacity of each resource
- The system has a market constraint and has excess capacity
- The most utilized department is operating at 80% of capacity
- Is the market constraint temporary or permanent?

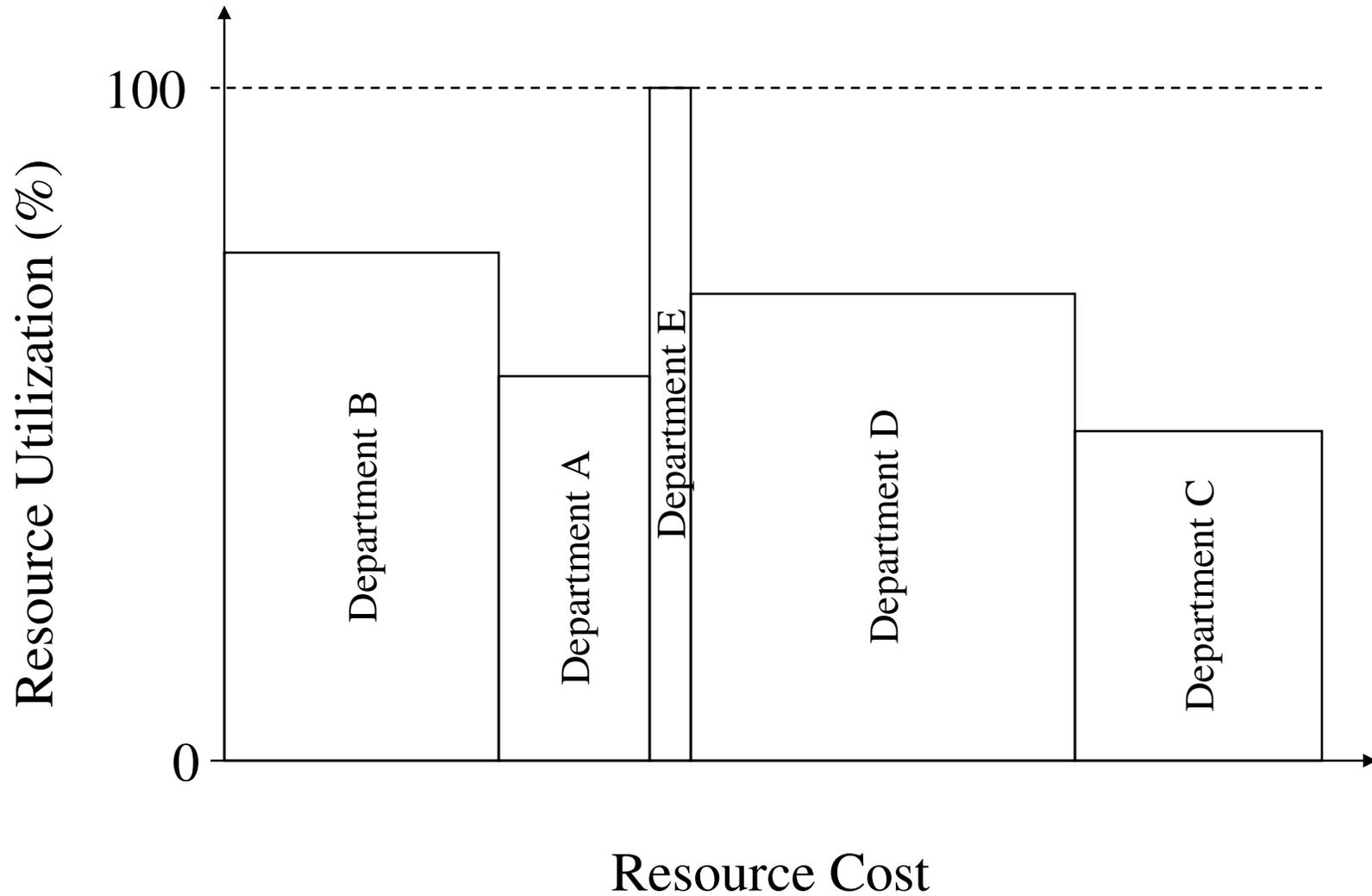
Market Constraints and Dummy Constraints

- Why is there a market constraint?
- Given the excess capacity, can management contract for additional work?
- Let's consider another system
- The system is displayed on page 107, but the loads and costs are given on page 116
- From pages 116 and 117, we can see that system output is constrained by an inexpensive resource
- This is a dummy constraint

Load Analysis for a System with a Dummy Constraint

Resource	Load (%)	Cost of Resource (\$ thousands)
Department A	65	200
Department B	75	350
Department C	40	380
Department D	70	590
Department E	100	10

CUT Diagram of a System with a Dummy Constraint



The CUT Diagram and Investment Decisions

- The CUT diagram serves as a tool for investment decisions
- Suppose we invest money to break the system constraint?
 - Is the additional throughput achieved worth the cost?
- If we double the capacity of the bottleneck resource, we don't always double the system throughput
- The bottleneck may shift from one resource to another

Make-or-Buy Decisions

- Following the investment, we must analyze the impact of the new capacity on the entire system
- CUT diagrams are useful in other settings
 - In deciding on whether new services or new products make sense
 - In choosing services or products to subcontract
 - In making decisions to end a service or product

Management by Constraints: A Short Review

- There are four types of constraints
 - Resource constraints (bottlenecks)
 - Market constraints (excess capacity)
 - Policy constraints
 - Dummy constraints
- Bottlenecks can be identified by using process flow diagrams, load analysis, and CUT diagrams

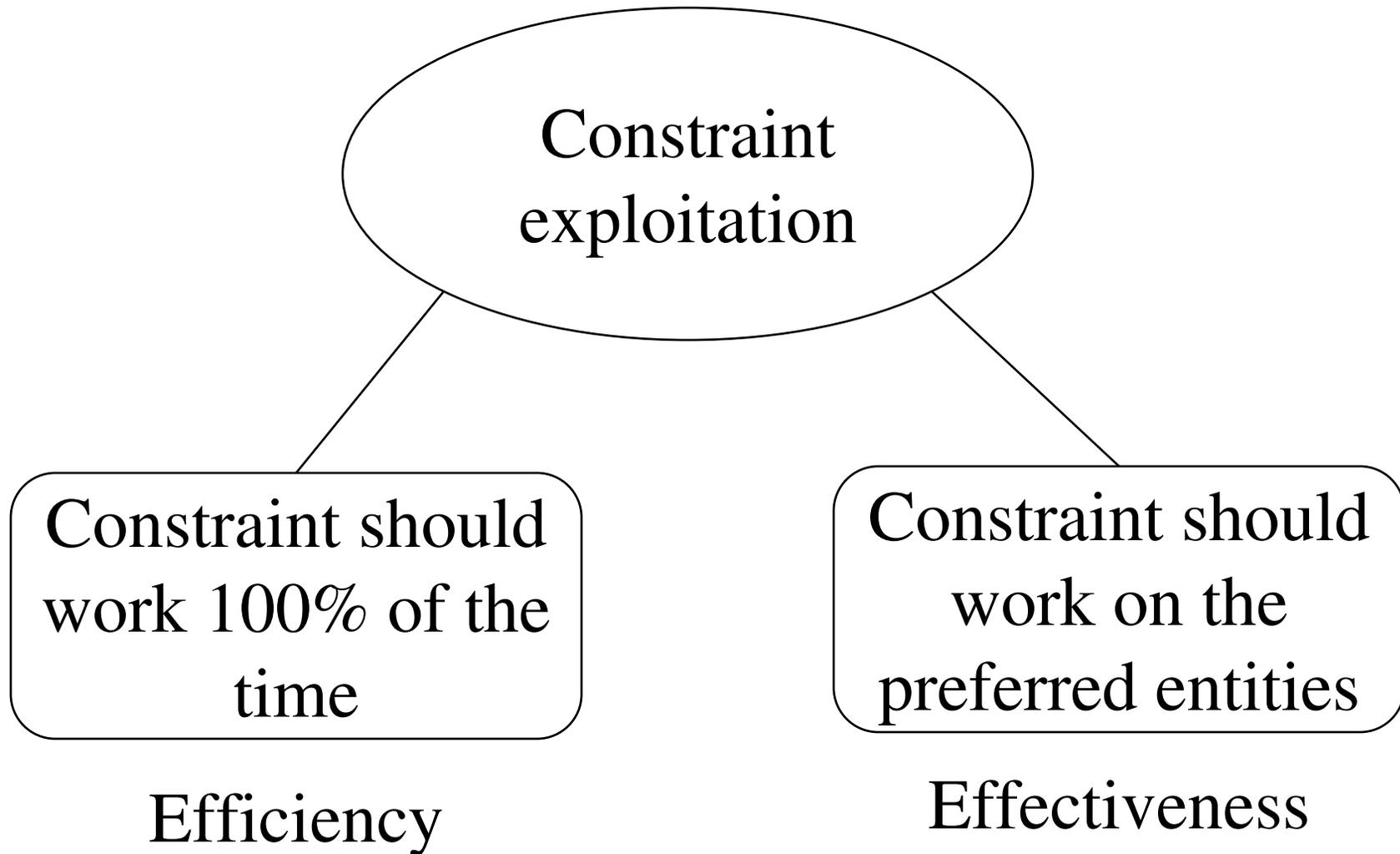
Management by Constraints in a Bottleneck Environment

- This section focuses on situations where system throughput is limited because of a resource constraint (bottleneck)
- Any improvement that adds effective capacity to the bottleneck constraint will increase throughput to the system
- Given a resource shortage (human or material), the inclination is to add personnel or acquire additional equipment
- The decision to increase resources should be postponed until after the improvement potential of the current bottleneck is fully exploited

Exploiting and Utilizing the Constraint

- Improvement via exploitation can be achieved relatively fast and is the most realistic improvement for the short term
- Exploitation is performed along two dimensions
 - *Efficiency*: Increasing bottleneck utilization to as close as possible to 100 percent
 - *Effectiveness*: Because the bottleneck cannot supply the entire demand, one must decide on the product or service mix of the bottleneck

Exploiting and Utilizing the Resource



Efficiency: Increasing Constraint Utilization

- The bottleneck determines system throughput
- An hour of bottleneck utilization is an hour of work for the entire system
- An hour lost in the bottleneck is an hour lost for the entire system
- Experience shows that we can significantly increase bottleneck throughput *without adding resources* by better focused management of the resources

Efficiency: Increasing Constraint Utilization

- For the bottleneck to work more efficiently, there are two options
 - Increase bottleneck capacity utilization to (or close to) 100%
 - Reduce bottleneck ineffective (garbage) time
- Bottleneck utilization may be increased by measuring its idle times and analyzing these times using the Pareto focusing method

Increasing Bottleneck Utilization

- In a large hospital, a bottleneck in patient processing was an expensive magnetic resonance imaging (MRI) machine
- It was found to be idle 32 percent of the time
- Idle times were handled as follows
 - Problem classification
 - Differential policy
 - Allocation of improvement resources

Increasing Bottleneck Utilization

- Problem classification: Pareto classification of problems revealed that 20 percent of problems (type A problems) account for 80% of idle time
- These problems are
 - Allocating blocks of time to wards that do not utilize their time
 - Concurrent lunch breaks of several technicians
 - Maintenance problems
- Differential policy: Management decided to focus mainly on type A problems

Increasing Bottleneck Utilization

- Allocation of improvement resources: Most resources will be devoted to type A problems
- Management took the following steps
 - MRI blocks were eliminated, and imaging was scheduled by appointment or by emergent cases
 - Lunch breaks were staggered across three hours so that the bottleneck could operate at full capacity during lunch time
 - Maintenance problems were monitored and preventive maintenance was undertaken
 - The maintenance department was instructed to give the MRI top priority

Increasing Bottleneck Utilization

- The ORs in a public hospital were a bottleneck and were idle 42 percent of the time
- The main causes were
 - The wait for the cleaning crew (dummy constraint)
 - The anesthetist cancelled operations because patients had not had all prerequired tests

Increasing Bottleneck Utilization

- Another cleaning crew was assigned to the OR area
- A preoperative clinic made sure that a “complete kit” was created about one week before the scheduled surgery
- As a result, idle times in the ORs decreased dramatically

Reducing Ineffective (Garbage) Time

- Ineffective time may vary in different ways
- Garbage time: When the bottleneck is devoted to activities that do not add value to the customer, the service, or the product, or to activities it should not perform
- This is the ineffective time of the bottleneck
- Several examples follow

Reducing Ineffective Time

- In a group dental practice, the dentists spent time typing reports and scheduling patient appointments
 - This can be done by a secretary
- The bottleneck in the office of Minnesota State Claims was the attorney who had to sign off on every claim
- This created a backlog and delays in claim processing
 - Authority was delegated to claims specialists and the delays were dramatically reduced

Reducing Ineffective Time

- The sales personnel of a large multinational pharmaceutical firm estimated that 50 percent of their time was ineffective
- Classifying causes: The causes were classified in A, B, and C groups
- Group A included 20 percent of causes and accounted for 80 percent of the garbage time
- Group A included
 - Working with an incomplete kit (the sales force approached customers without understanding their needs)

Reducing Ineffective Time

- Failure to correctly identify the actual decision maker
- Dealing with administrative and logistic problems of the customer
- Differential policy: The firm decided to treat only the above causes of garbage time
- Resource allocation: Substantial management resources were devoted to addressing the above three problems

Reducing Ineffective Time

- As a result, the garbage time of the sales force was reduced from 50 percent to 40 percent
 - This is equivalent to increasing the sales force by 20 percent
 - Profit increased as a result
- In a large hospital, the chief nurse is expected to manage the facility and mentor junior nurses
 - 30 percent of her time is wasted on handling the paperwork of newly admitted patients
 - This paperwork should be handled by another experienced nurse

Reducing Ineffective Time

- In the surgical department of a hospital, the bottleneck was the anesthetist
- About 30 percent of his time was ineffective
 - 10 percent due to lack of synchronization with other OR staff
 - 10 percent due to incomplete kits
 - 10 percent between the end of one surgery and the start of the next one

Effectiveness

- Because the bottleneck cannot supply the entire demand, one must decide on the product or service mix or the projects or customers on whom the bottleneck will operate
- Strategic gating: A process of prioritization that
 - Defines the value of the different tasks, products, services, projects, or customers that are valuable to the organization
 - Decides which will be carried out and in which priority
 - Decides which will not be carried out

Strategic Gating

- In a large firm that produces medical devices, the R & D department was working simultaneously on four new products
 - Each product was a potential breakthrough in its area
 - Given the workload in the development department and the demands in the market, a strategic gating decision was made to halt the development of two products
 - This decision resulted in a competitive time to market for one of the devices, enhancing the firm's value

Strategic Gating

- An HMO was planning a campaign to increase membership
 - 75 big firms and 20 small ones were identified
 - There was a small time window in which workers were allowed to change carriers
 - The HMO prioritized firms based on firm size and ease of attracting their workers
 - They decided to focus on 30 large firms and 10 small ones
- This strategic gating resulted in a 65% success rate

Prioritization Methods: Strategic Gating

- There are several methods for prioritization
 - Use a Pareto diagram
 - Use a focusing table (easy-important) and a focusing matrix
 - Use specific contribution
- A Pareto diagram can be drawn for the potential contribution to the system
 - But it does not take into account the bottleneck time needed for each activity

Strategic Gating

- A focusing table and a focusing matrix display the importance and ease of implementation for every task
- One measure of ease of implementation is the number of hours needed by the bottleneck resource
- An IT example is presented on the next page

Strategic Gating

- The IT department in a large private hospital was the bottleneck for many activities
 - Every department and ward wanted the development of IT applications
- Management had to prioritize these proposed projects
 - Ease of development = number of person hours
 - Importance = contribution to hospital profits over the next three years

Strategic Gating

- The tool for ranking the most valuable products, jobs, and customers is their specific contribution
- The specific contribution (the contribution per unit of resource) of a product, task, service, or customer is the expected contribution divided by the time investment of the resource

$$\text{Specific contribution} = \frac{\text{contribution}}{\text{time invested by bottleneck}}$$

Strategic Gating

- In strategic gating, we calculate the specific contribution for every product, service, task, or customer
- We then choose the items with the highest specific contribution until the capacity constraint is reached
- The sales department can use this to identify the best customers
- The marketing and development departments can use this for new product decisions

Strategic Gating

- Many organizations prefer to perform strategic gating using the focusing table and focusing matrix
 - The visual representation makes it easier to decide
- Example: The R&D department of an electronics firm that specializes in imaging technology was considering four projects
- See page 146 for details
- Prioritization based on specific contribution:
MRI 1, MRI 2, CT scan 1, CT scan 2

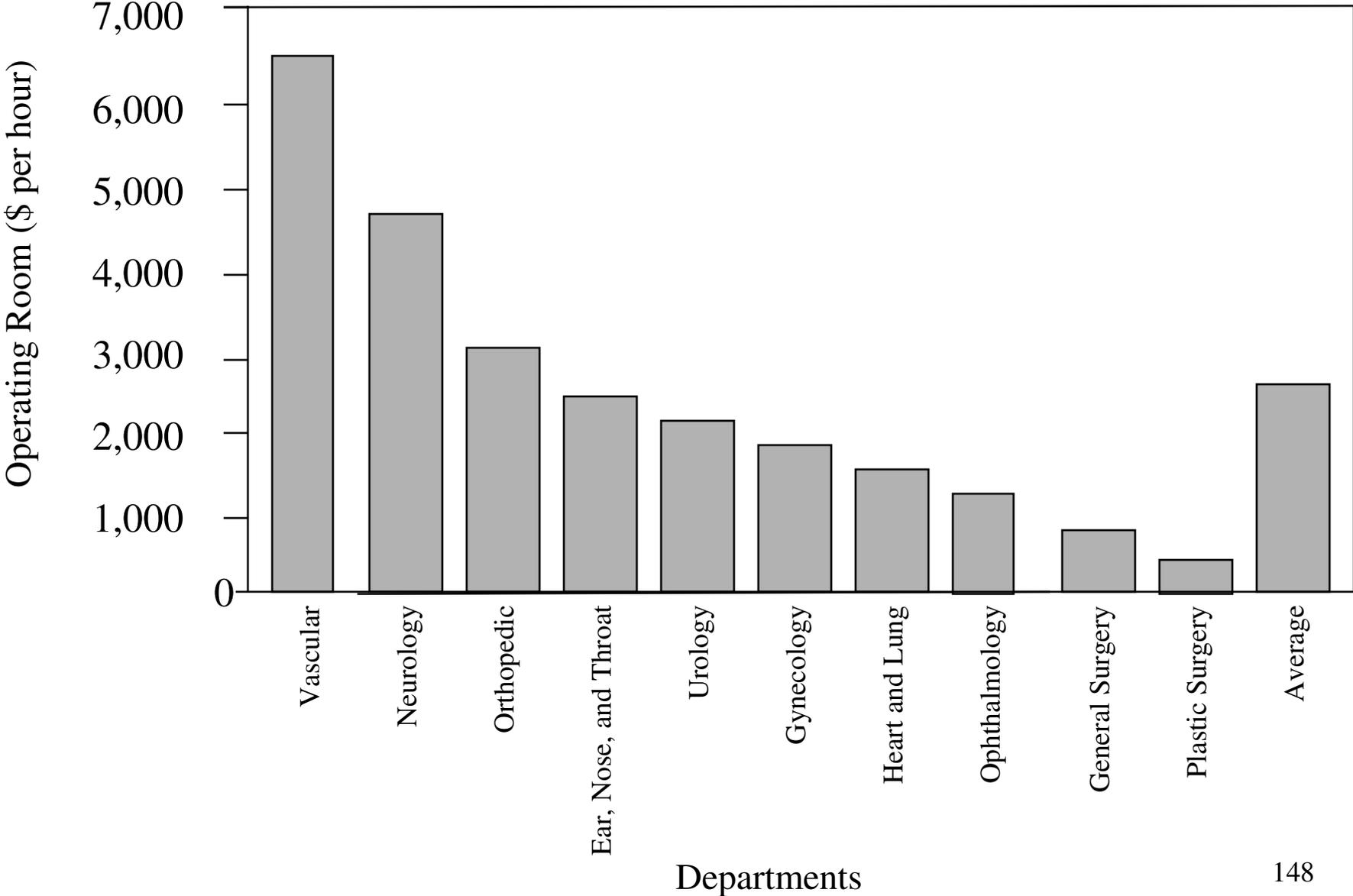
Imaging Projects

Project	Contribution (\$ thousands)	Development Effort (person- years)	Specific Contribution (\$ thousands per person-year)
CT scan 1	56	0.5	112.0
MRI 1	2,470	1.0	2,470.0
CT scan 2	345	5.0	69.0
MRI 2	1,250	2.0	612.5

Specific Contribution

- A private hospital provides surgical services where external surgeons bring in external patients
- The ORs are the system bottleneck
- A specific contribution analysis was performed
- Operations were ranked based on financial contribution per hour of OR
- Vascular surgery is most profitable (see next page)

A Pareto Diagram of Specific Contribution



Strategic Gating

- The graph on the previous page can help the hospital prioritize the types of operations they want to focus on
- Strategic gating decisions are difficult
- A decision on *what to produce* and *what to focus on* implies *what to give up*
- There is always the fear that the product rejected through prioritization could have been a winner
- Still, indecision is a dangerous alternative

The Global Decision-Making Method

- The specific contribution is only one part of tactical and strategic prioritization and screening of tasks, services, or products
- The global decision-making process includes three steps
 - Make a global economic decision from the CEO's perspective (specific contribution may be used here)
 - Account for strategic considerations
 - If necessary, change local performance measures

Breaking Policy and Dummy Constraints

- Bottlenecks must be exploited efficiently and effectively
- Dummy and policy constraints must be broken
 - The dummy constraint of the need for an additional cleaning person in the OR can be immediately remedied
 - A detailed cost-benefit analysis is not necessary
- Policy constraints are more difficult to deal with
- New policies must be considered

Subordinate the System to the Constraint

- Once we focus on the constraint (bottleneck) and improve its management, we need to manage and operate noncritical resources
- The remaining resources should *serve* and *assist* the bottleneck
 - In a group dental practice, the dentists (and their time) are the system constraints
 - The other workers (hygienists, assistants, secretary) should assist the dentists
- The noncritical resources must be available to assist system constraints, especially at peak times
 - Suppose one dentist in the above group practice has just given birth

Subordinate the System to the Constraint

- In an OR in a hospital, the bottleneck could be the anesthetists, nurses, or the OR capacity
- Management must identify the bottleneck and
 - Subordinate all other resources to serve and assist the bottleneck
- Implementing the subordination phase may be difficult
- When the bottleneck is the senior surgeon, other surgeons can often fill in

Subordinate the System to the Constraint

- When the anesthetist is the bottleneck, it is difficult to subordinate the surgeons to the specific timetable of the anesthetist
- Another example
 - If efficient use of a hospital lab calls for batches of a hundred specimens at a time, then all wards and logistics should adhere to this constraint

Subordination Mechanisms

- Noncritical resources can be subordinated to an organizational constraint using the following mechanisms
 - Tactical gating
 - The drum-buffer-rope (DBR) mechanism
- Tactical gating means the controlled release of tasks (jobs) to the system
- The tactical gating mechanism employs the following policies

Tactical Gating Mechanism

- All tasks will be released for work in the *right batch size*
- Only tasks screened by the gating process will be released for workup
- All tasks will be released only through the body or person in charge of the gating
- All tasks will enter the system with a complete kit
- All tasks will enter according to the DBR scheduling mechanism

The Drum-Buffer-Rope Mechanism

- The DBR mechanism is a scheduling mechanism for entering tasks into the system
- The drum provides the rhythm for the flow of tasks through the system
- The system constraint determines the rate at which tasks should enter the system and flow through it
- In the presence of a resource constraint, the drum will be the *work rate of the bottleneck*

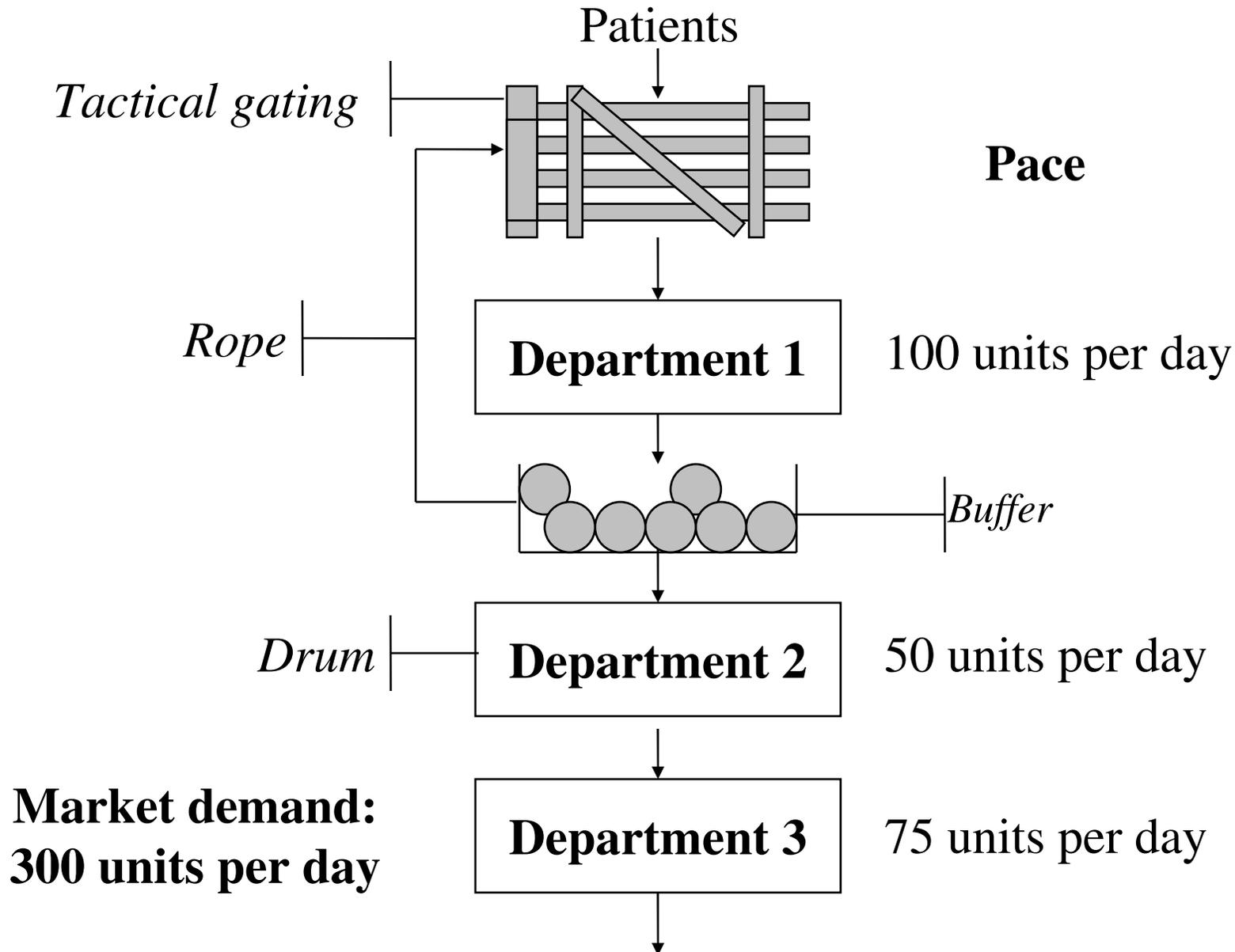
The Drum-Buffer-Rope Mechanism

- In the case of a market constraint, the *rate of market demand* will dictate the pace for the whole system
- The buffer refers to a controlled quantity of tasks that accumulate before the bottleneck to assure it is fully utilized
- This protects the system against fluctuations
- Typical causes of fluctuations are given next

Causes of Fluctuations

- Worker absenteeism
- Technical problems/human error in the service or production process
- Low quality of information or materials
- Patients don't show up for appointments
- Delays in the supply of raw materials and components

The Drum-Buffer-Rope Mechanism



The Drum-Buffer-Rope Mechanism

- The *buffer* protects the bottleneck output from various fluctuations by assuring continuity of its operation
- The size of the buffer is based on the various fluctuations in the state of the system and the time to correct them
- The *rope* of the mechanism transfers (pulls) information on the size of the buffer to the tactical gating mechanism

The Drum-Buffer-Rope Mechanism

- The rope allows for the timely release of tasks into the system
- Example from an orthodontics clinic
 - The orthodontists and clerical staff schedule patients so that the orthodontists (the bottleneck) are always busy
 - They create a 15-minute-wait (two patients) buffer for each orthodontist

The Drum-Buffer-Rope Mechanism

- Orthodontists clinic example, continued
 - If the buffer were larger, patients would complain
 - In this clinic, each orthodontist works in three rooms in parallel
 - In addition to the two buffer rooms, two additional patients are waiting in the waiting area
 - On average, wait time in the waiting area is 15 minutes

The Drum-Buffer-Rope Mechanism

- Orthodontics clinic example, continued
 - To reduce fluctuations in patient arrivals, patients are called the day before their appointment to verify their arrival time
- An OR example
 - The duration of surgical procedures is uncertain
 - Here, the drum is the system bottleneck, i.e., the surgeon's rate of performing surgery

The Drum-Buffer-Rope Mechanism

- OR example, continued
 - A buffer of patients will be waiting in the preoperative room
 - When a patient is wheeled into the OR, the rope will signal the ward to send another patient
- An airport example
 - In airports, the runways are a key bottleneck

The Drum-Buffer-Rope Mechanism

- Airport example, continued
 - When at close to capacity, arriving planes circle above the airport and serve as buffers
 - The buffer size depends upon safety concerns
 - When the buffer is full, approaching aircraft are told to slow down, circle in wide circles, or are diverted
 - The control tower is responsible for gating

The Drum-Buffer-Rope Mechanism

- Airport example, continued
 - When a plane lands, it is immediately directed to a side runway to allow for another plane to land and another plane to enter the buffer
- Up to this point, we have tried to increase the output of a given system without any changes to the system itself

Elevate and Break the Constraint

- Now, we consider structural changes to the system to increase the effective capacity of the bottleneck
- Increasing this capacity will increase the throughput of the whole system
- Elevating and breaking the constraint can be achieved in two ways
 - Using capital investment
 - Use of offloading

Elevating Using Capital Investment

- Increase the capacity of the constraint by
 - Recruiting additional staff
 - Purchasing additional equipment
 - Working additional shifts
 - Working overtime
 - Hiring subcontractors
 - Outsourcing
 - Recruiting distributors and value-added retailers for the sales force

Examples of Capital Investment

- In an ED of a hospital, the consulting surgeons were the bottleneck
- Management decided to hire some retired surgeons (who no longer operate, but have great diagnostic skills) to elevate the system's constraint
- A large HMO had to deal with many malpractice suits
- Their own attorneys became a bottleneck
- They started using the services of outside law firms

The Offloading Mechanism

- *Offloading*: Relieving the load from the bottleneck by transferring some of the workload to noncritical resources
 - In many HMO clinics, physicians and nurse practitioners form teams so that the physicians can offload some of their tasks to the nurses
- A more detailed example follows

An Example of Offloading

- In a large HMO, a senior VP had to sign every request for surgery in nonaffiliated hospitals
- She was the bottleneck responsible for delays
- She had to verify need, justify surgery in a specific hospital, and determine the level of copayment
- She was very knowledgeable and was a scarce resource
 - Clerks were trained to handle the routine requests

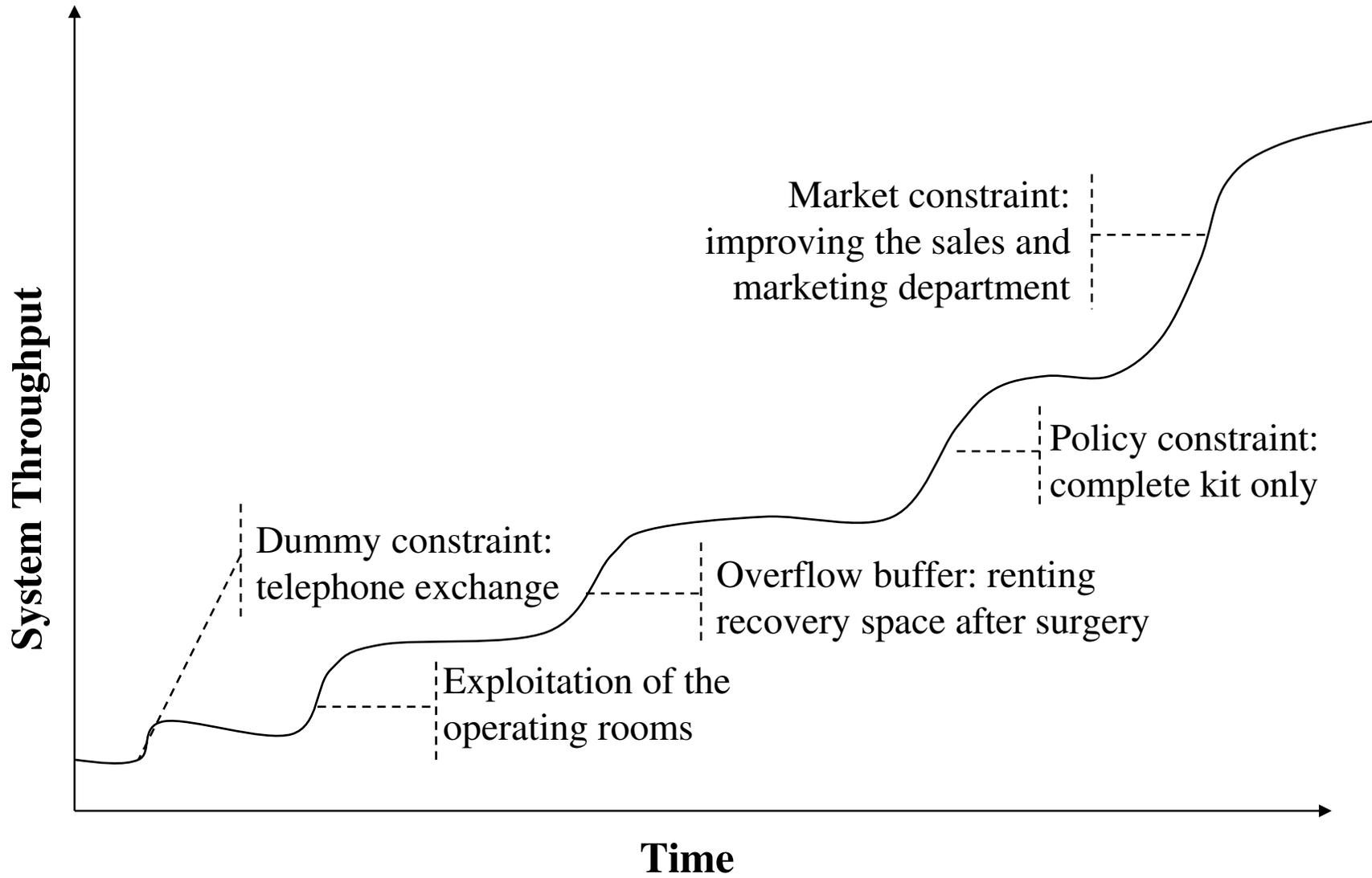
Additional Examples of Offloading

- A dental hygienist relieves the burden from the dentist by performing some of the dentist's tasks
- In university hospitals, teaching and research assistants serve as offloads for the expensive resource of the professors and senior researchers who do both clinical and academic work
- In many complex surgeries, both the beginning and the “closing” are performed by junior surgeons
- In a supermarket during peak times, the number of cashiers is the bottleneck
 - Adding a bagger to help each cashier serves as an offload

Suppose a Constraint is “Broken”

- We must return to the step of identifying the new system constraint
- There is always a constraint (bottleneck) in a system
- The task is to identify constraints, manage them, break them, and face a new constraint
- By moving from constraint to constraint, system output increases
- See the next page for an illustration

Continuous Improvement in a Surgical Department



Suppose a Constraint is “Broken”

- In routine processes, constraints do not change frequently and are rather stable
- In one-time processes (e.g., projects, sales campaigns) constraints change rapidly and bottlenecks can move from one place to another
- An organization accustomed to a constraint being in one place may behave as if the constraint is still there
 - Management must overcome inertia of this sort
- An example from an emergency department follows

An Emergency Department Example

- In an ED, the bottleneck was radiology services
- By using existing equipment and breaking some policy constraints, management was able to operate the two x-ray rooms as follows
 - They divided patients into two groups: ambulatory patients and recumbent patients
 - The “walkers” walked into one room for service
 - The others were wheeled into the second room in specially purchased beds, with no bed transfer required
- Thus, the radiology constraint was broken

An Emergency Department Example, continued

- The next constraint was a lack of expert surgical consultants
- A consultant was called when two or three patients required it (unless there was an emergency)
- The hospital assigned a special surgeon who was continuously available in the ED
- With the surgical consultant constraint broken, the system faced a policy constraint in the discharge process
- A team of nurses, doctors, and administrators redesigned the discharge process
- In sum, average waiting time was reduced by 40 percent

Expanding Effective Capacity

- The use of techniques from management by constraints frequently results in a
 - Rapid increase in throughput
 - Significant decrease in response times
 - Improvement in quality
- The organization usually becomes more attractive to customers and patients
- Management must be careful not to take on commitments beyond its new effective capacity

Suppose the Market is the Constraint

- A system always has a bottleneck
- Previously, we discussed situations involving scarce resources where the constraints were these resources
- In other cases, there is excess operational capacity
- The market becomes a system constraint
- Even private or budgeted hospitals and HMOs must realize that they may have excess capacity and need a strong marketing function

Suppose the Market is the Constraint

- Rather than lay off people in slow times, the hospital should consider taking on outside tasks to fill this excess capacity
- A small example follows
 - A hospital buys an expensive imaging machine
 - The machine is needed for some diagnoses
 - But, there is not enough demand for its use
 - The hospital can sell imaging services to HMOs and private physicians

Suppose the Market is the Constraint

- A technological change can cause excess capacity at bottlenecks
- E.g., cardiac angiography with the introduction of stents
 - Many cardiologists adopted this practice
 - More patients were sent for this procedure
 - Fewer coronary bypass surgeries were performed
 - Excess capacity for cardiac surgeons resulted

Suppose the Market is the Constraint

- Treating a market constraint is more difficult than treating a resource constraint
- Managing an internal bottleneck offers management an opportunity to assert more control over the organization
- Some of the factors involved in dealing with a market constraint are beyond management's control
- Most non-health care organizations are faced with a market constraint

Suppose the Market is the Constraint

- In today's business environment, it is wise to have excess capacity in the production and service resources
- Why should we strive for this excess capacity?
- Because, if the marketing and sales force can obtain additional orders, it would be an organizational sin for operational constraints to prevent their satisfaction
- Managing a market constraint is a key to the success of many firms and businesses

Exploiting a Market Constraint

Exploiting
the market
constraint

Strategic market effectiveness

- Gating of markets, products, services, and customers
- Excellence in quality and time to market (TTM)
- Prioritization in bid-no bid process

Operational efficiency of sales and marketing

- Exploitation of marketing and sales people
- Improvement of market response time
 - Cutting quotation and response time
 - Cutting TTM of products and services
 - Improving quality of products and services
- Cost reduction

Marketing and Sales Efficiency

- There are two ways to achieve marketing and sales efficiency
 - Exploit marketing and sales personnel
 - Improve response to market needs
- The marketing and sales personnel are usually a permanent bottleneck
- The *garbage time* of the sales force must be reduced
- The sales and marketing operation should be *structured* with stable and simple processes
 - E.g., new products and services must be launched in a systematic way

Improving Response to Market Needs

- Reduce response time in bids, sales, and marketing
 - Improve work processes
 - Cluster customers with similar features

- Shorten the time to market (TTM) of services and products
 - This helps to reach customers faster than the competition
 - This conveys a positive image to customers
 - It also increases the contribution from sales to the organization

Improving Response to Market Needs

- Improve service or product quality
- Reduce costs
 - Operational efficiency can improve response times, quality, and the contribution of sales
 - Cost savings can sometimes permit price reduction which may improve competitiveness

Strategic Marketing Effectiveness

- Earlier, we introduced strategic gating, which screens and prioritizes customers or products based on their
 - Specific contribution
 - Positioning in the focusing matrix (easy-important)
- Even when an organization is at excess capacity and facing a market constraint, the sales and marketing personnel are a permanent bottleneck
- There are always more potential customers, initiatives, and services than time permits 189

Strategic Marketing Effectiveness

- The screening procedure is, therefore, extremely important
- One should not forgo sales of services and products to customers where the sales personnel are *minimally* involved, even if the contribution volume is small
 - Repeat sales to existing customers may be very easy
- For sales of services or products to customers which require substantial investments in time, we need to perform a prioritization process of strategic gating
- See the next page for an example

The Specific Contribution of Customers/Sales Effort

Customer	Contribution to Profit (\$ thousands)	Sales Effort (days)	Specific Contribution (\$ per day)
HMO 1	2,345	120	19,500
HMO 2	5,600	30	186,700
Group practice	575	80	7,200
Hospital	12,650	500	25,300

An Example of Strategic Gating

- The salesperson should first approach HMO 2, because the specific contribution is highest
- Next, the salesperson should focus on Hospital as a customer
- And so on, until 100 percent of the salesperson's time is utilized
- Note the error on page 105 in the textbook
- Strategic gating will also be performed on the variety of services and products that the organization sells

Strategic Marketing Effectiveness

- Sometimes, it is possible to achieve a higher volume of contributions by focusing on a smaller number of services and products
 - This removes *distracting* services and products
- Occasionally, it makes sense to cease the sales of services and products when their contribution volume is too small
 - Option 1: Discontinue offering them to the market
 - Option 2: Raise the price, maybe demand will persist

Excellence in Response Times and Quality

- Japanese firms have captured the world market in cars and home electronics by emphasizing quality
- Their operational skills have enabled them to achieve high quality, low production costs, fast response times, and low inventories
- “Made in Japan” signals high quality to the customer who is willing to pay a premium for products with this label

Excellence in Response Times and Quality

- The Mayo Clinic is in Rochester, Minnesota
- It is in the middle of nowhere (slight exaggeration)
- It is regarded as one of the leading hospitals in the world
- Its perceived superior quality attracts patients from all over the world
- Moreover, patients are willing to pay a premium
- The computer manufacturer Dell is another example
 - Fast response times and low inventory costs lead to a large market share of personal computers

Prioritizing in Bid-No Bid Processes

- In some companies, most sales are handled via bids
- An important step in this process is deciding whether to bid or not
- Preparing a bid can be expensive and time-consuming
- Issues to take into account
 - Compatibility with company strategy
 - Likelihood of winning
 - Potential contribution to profits
- A policy where an organization responds to bids at every request becomes a policy constraint

Breaking Policy and Dummy Constraints

- To increase the contribution volume of sales, we must identify policy and dummy constraints that impede the organization's efforts to exploit the market

- Some helpful general rules follow
 - Be flexible in pricing products and services
 - Don't give up on small customer orders
 - Avoid selling only complete sets, sell what the customer wants
 - Avoid selling a “complete product line or nothing” (selling a partial basket can yield nice profits)

Breaking Policy and Dummy Constraints

- Additional helpful general rules
 - Avoid rewarding sales personnel by sales volume (look at contribution to profits)
 - Avoid rewarding sales personnel regardless of returns and cancellations (subtract returns and cancellations)
 - Avoid dummy constraints (buy a fax machine or a cell phone, if necessary)

- Breaking such policy and dummy constraints as illustrated above helps to increase the contribution of sales to the organization

Subordinate the System to the Constraint

- There is a clear advantage for an organization that satisfies its customers' needs more than the competition
- With a market constraint, the organization must subordinate the whole system to the market
- This subordination unfolds in the following ways
 - Customization to customer demand and needs (when this requires some development, strategic gating can help decide how to best utilize the constrained development resources)
 - Fast response to market needs

Subordinate the System to the Constraint

- This subordination unfolds in the following additional ways
 - Direct link to the end customer
 - ❖ Make an effort to establish a connection with the end customers by bypassing the hurdles of the distribution channels and agents
 - ❖ If a hospital has referrals from HMOs, the hospital should establish a direct link with the insured to fully understand their needs
 - Adjusting the organizational structure to customer types or customer needs
 - ❖ Examples: Young vs. old, chronically ill vs. acutely ill, male vs. female

Subordinate the System to the Constraint

- Subordinating technology to market needs
 - In high-tech organizations, there is an inherent conflict between marketing and development
 - Development and technology must be subordinated to the market via the marketing personnel
- Market segmentation and product differentiation
 - In the case of excess capacity, employ product and price differentiation

Elevate and Break the Constraint

- In situations with a market constraint, organizations are tempted to cast a wide net
- However, a focused policy targeting the main objectives and focusing the efforts on these targets is usually wiser
- To elevate and break the market constraint, the following steps are recommended

To Elevate and Break the Market Constraint

- Establish a focused strategy
- Enter a new product or market
- Add marketing and sales channels
- Market segmentation and product differentiation
- Stretch the brand name
- See the next page for additional steps

To Elevate and Break the Market Constraint

- Create added value for the customer
 - Create a package of services and products that complement one another
 - Manage the customer's facilities, shift to outsourcing of customer activities
 - Manage inventory for the customer
 - Build customer loyalty (customer clubs)
 - Customer relationship management
- Cooperation and strategic alliances
- When the constraint is broken, we must look for the next constraint

Managing Marketing and Sales

- Focus on most valuable customers (MVCs)
- MVCs are type A customers in the Pareto analysis of the organization
 - About 20% of customers (existing and potential) who contribute 80% of contribution volume
- An example from a multinational pharmaceutical company
 - Old policy: Every salesperson handled 40 existing customers and about 100 potential ones
 - New focused policy: Every salesperson handles sales and retention for 10 existing customers and approaches 10 new potential customers

Managing Marketing and Sales

- An increase in throughput resulted
- Focusing enables a better response to the needs of the MVCs
 - Establish close personal contacts with key people in the customer organization
 - Collect and analyze data on the customer and his relevant markets
 - Analyze MVC needs and identify channels for providing more value to the customer
 - Treat MVCs as VIPs

Peak Management

- How do we reduce the effect of peak-time loads?
- Differential pricing of products or services during different times
- The reduction of temporary loads during peak times can be achieved using the following strategies
 - Stretch peak times, e.g., widen hours in free clinic, keep OR open from 6 am to 11 pm

Peak Management

- Plan capacity, i.e., adjust the level of the labor force by time of day, season, etc.
- Transfer load to low-load periods, i.e., appropriate pricing/rewards can divert some demand to low-load periods (e.g., airlines, phone companies)
- Use temporary help (e.g., tax firms during tax season, UPS during Christmas season)

Peak Management

- In a work environment with obvious peak times, some resources are at excess capacity most of the time and in short supply during peak times
- Management must be prepared to handle both situations
 - Excess capacity
 - Shortage of resources
- In hospital EDs, there is often excess capacity by day and resource constraints by night

Where Should the System Constraint be Located?

- Where should the constraint be?
- Where is the constraint now?
- How can we transfer the constraint to the proper place?

Where Should the Constraint Be?

- Does the constraint have to be an internal one (resource constraint) or an external one (market constraint)?
- If the constraint is a resource constraint, which resource has to be the constraint?
- There are clear advantages to an internal bottleneck
 - Control over the system and cost of the system

Where Should the Constraint Be?

- On the other hand, the firm misses out on business opportunities because of lack of capacity and allows competitors to enter the market
- Choosing a resource constraint is appropriate for firms that have a critical and expensive resource, whose capacity is difficult to increase
- In the above case, the constraint should be the most critical or most expensive resource in the system

Where Should the Constraint Not Be?

- Operations should *not* be the system constraint
- Operations can usually be expanded
 - E.g., its tasks can be given to subcontractors
- The Head of Operations should see to it that the needed quantities will be produced on time and at the required level of quality
- To achieve this, operations should have protective capacity and may be at excess capacity
 - In Japan, production lines have an excess capacity of about 30%

The Impact of a Market Constraint

- With a market constraint, there is excess capacity
- A firm with excess capacity does not give up on orders and on possibilities for increasing output
- This has strategic importance
 - Giving up market share because of a resource constraint allows competitors to emerge and grow
 - This poses a long-run threat

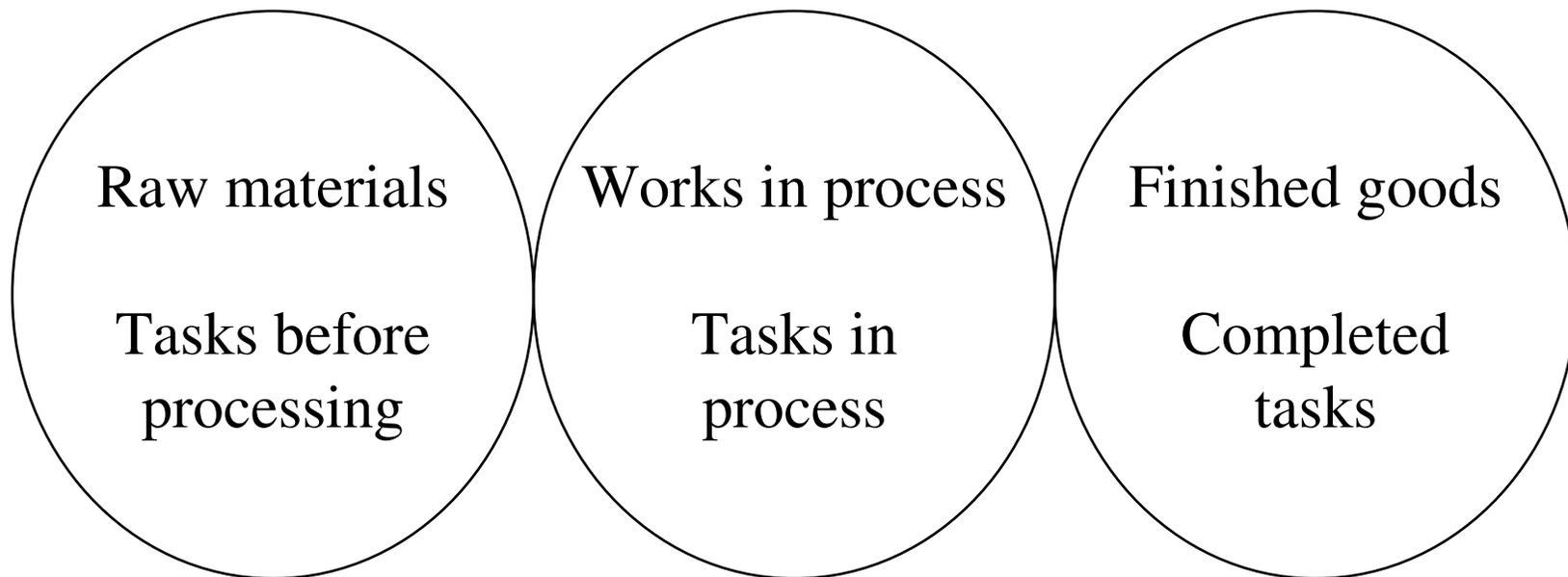
The Impact of a Market Constraint

- Excess capacity can also allow the firm to deal with occasional demands, large contracts, and emerging market opportunities
- Having a market constraint carries a higher price tag because the investment in resources costs more
- System constraints are not necessarily permanent
 - The market continues to evolve

The Evils of Long Response Times

- Long response times for service, production, and development are a major concern today
- To shorten response times, we examine the relationship between response time and amount of work in process (WIP)
- The management of physical, nonphysical, and human inventories plays a key role in health care systems
- There are three types of inventories

Types of Inventories



Types of Inventories

- Raw materials: Materials, components, information, tasks, and the like before they enter processing in the system
 - Tasks that have not yet been handled
- Works in process: Intermediate products or tasks whose handling has been started, but not yet completed
 - Tasks in process
- Finished goods: Completed tasks

Examples of WIP

- Patients being treated in an emergency department (ED)
- Development of software and hardware that has not yet been completed
- Assemblies on an assembly line
- Purchase orders in the negotiating process or in the stage of approval
- Receipts handled in the accounting department
- Business information processing
- Equipment and instruments being repaired in the maintenance department

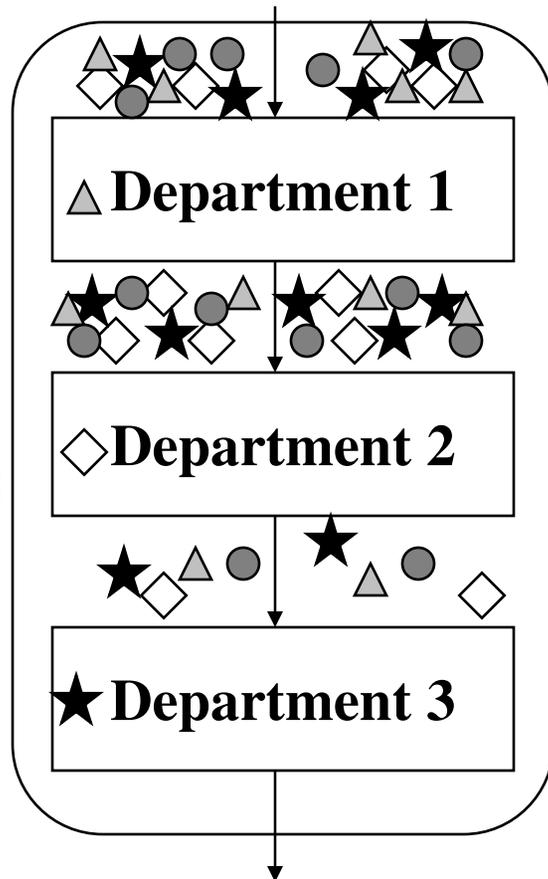
The Evils of WIP

- Consider two imaging and testing centers of two competing HMOs, call them *A* and *B*
- The clinics are similar and the testing centers are similar
- Both operate in the same market
- Both the arrival and the departure rates in each center are 20 patients per hour
- The difference is the amount of WIP
- *A* has 60 patients and *B* has 20 patients

A High-WIP Center versus a Low-WIP Center

Center A

20 patients per hour

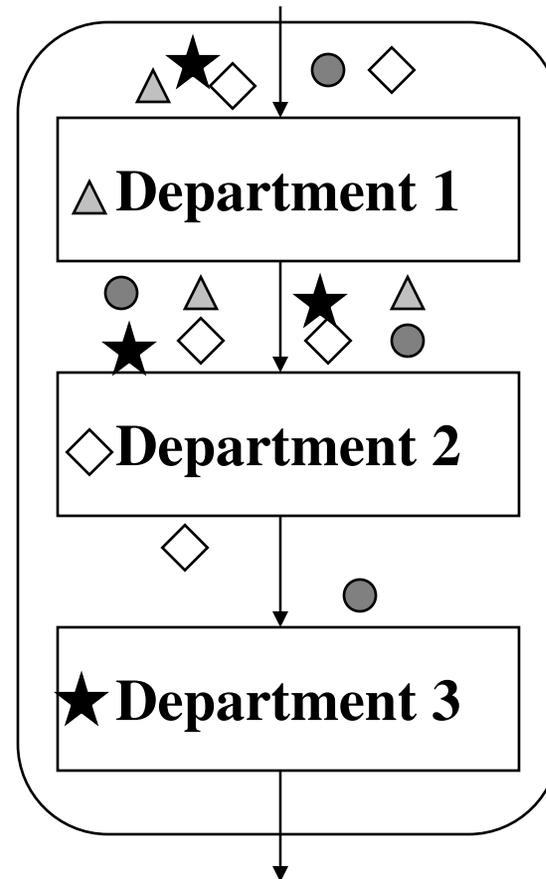


WIP:
60 patients

20 patients per hour

Center B

20 patients per hour



WIP:
20 patients

20 patients per hour

The Evils of WIP

- These numbers include patients being seen or treated in each department and patients waiting before each department
- Ignore the reasons for the difference, for now
- Does the higher level of WIP in center A improve or harm this center's performance?

Relation Between Response Time and WIP

- High levels of WIP lead to long response times
- The average response time of the system is proportional to the WIP level
- The average time spent by a patient in center *A* is three hours compared with one hour in center *B*
- On average, the throughput time (time from entry until departure) of a patient arriving at *A* is $60/20 = 3$ hours
- The average throughput time at *B* is $20/20 = 1$ hour
- In general, average response time = $\frac{\text{WIP}}{\text{system throughput rate}}$

Relation Between Response Time and WIP

- Next, we examine the gross and net response time of the system
- In general, most of the response time will be caused by the waiting times at the various workstations
- Treatment or processing time is rather short
- In service and industry organizations, net processing time is only 5 to 10 percent of the total response time
- In many cases, work time is less than 1% of total response time

Relation Between Response Time and WIP

- Two examples
 - The average gross time for a patient in a hospital ED is about four hours
 - The total treatment time is a few minutes
 - The response time of a contractor for fixing a medical instrument is two weeks
 - The net time spent fixing the instrument is about two hours

Relation Between Response Time and WIP

- The gross response time (throughput time) of the process equals the sum of processing times (net times) of the individual unit (patient) in the various stations plus the sum of the waiting times of the unit (patient) before the various stations

- In other words,

$$\text{throughput time} = \text{processing time} + \text{waiting time}$$

- The response time of a process is proportional to the amount of WIP

Relation Between Response Time and WIP

- If WIP is reduced, the response time will be reduced
- High levels of WIP \Rightarrow long response times \Rightarrow damage to the performance of an organization
- Undesirable consequences of long response times
 - Missed opportunities
 - High operating expenses
 - Diminished quality
 - Diminished control
 - Diminished flexibility to market and technological changes
 - Diminished cash flow

Undesirable Consequences of Long Response Times

- Diminished motivation of managers and workers
 - Missed deadlines
 - Lack of customer satisfaction
 - Diminished forecasting capability
 - Diminished throughput
 - Acquired infections in hospitals
- We discuss each of these undesirable consequences in detail next

Undesirable Consequences of Long Response Times

■ Missed opportunities

- Windows of opportunity for new services or products open for short periods
- An organization with fast response times can take advantage of these opportunities
- E.g., express care at UMMC
- The faster an organization comes out with new services or products and makes these available to customers, the higher the prices they can charge
- Often, being first to market enables an organization to capture a significant market share

Undesirable Consequences of Long Response Times

- High operating expenses
 - Accumulation of WIP and long response times leads to high operating expenses (inventory carrying costs, maintenance and control)
- Diminished quality
 - The longer the response time and the longer work is in process, the more the damage to quality
 - Materials that have been delayed in the process are left unprotected and vulnerable to damage
 - In service units, the quality of decisions/treatment diminishes because of the time lag

Undesirable Consequences of Long Response Times

- Diminished quality, continued
 - The longer the response time, feedback on earlier mistakes arrives later to the station responsible for the mistake
 - The longer a patient spends in the hospital, the more he/she is exposed to dangerous infections
- Diminished control
 - A high amount of WIP allows workers more freedom in selecting preferred tasks to work on
 - As a result, managerial control diminishes as well as the ability to dictate priorities

Undesirable Consequences of Long Response Times

- Diminished flexibility to market and technological changes
 - Long response times and large WIP make it difficult to introduce changes in services or products
 - Changes may be necessary as the market and technology evolve
- Diminished cash flow
 - Companies with slow response times and large inventories may face a crisis caused by unfavorable cash flow
 - They are usually required to pay for raw materials within a short time, whereas payment will be received later because of a long response time

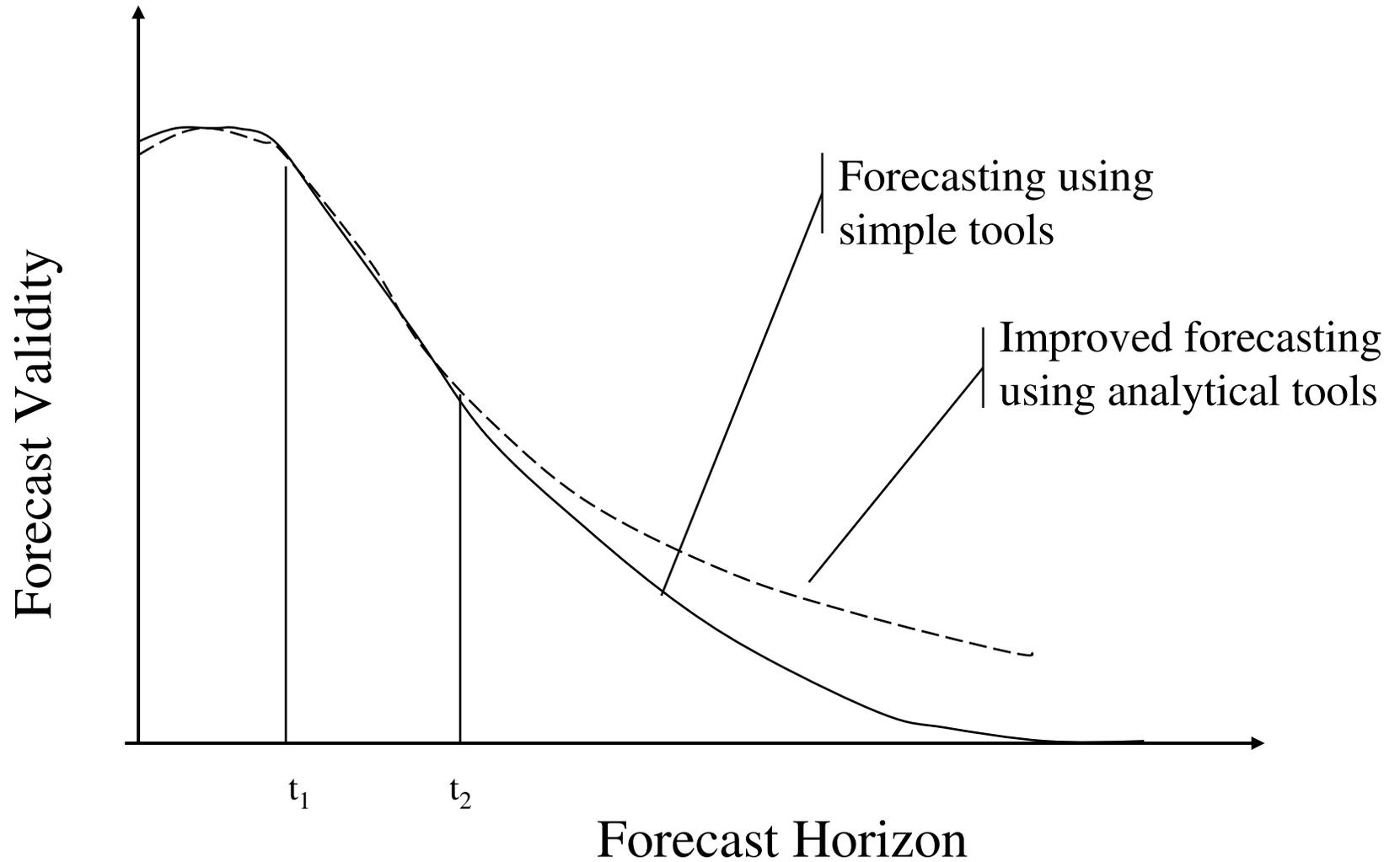
Undesirable Consequences of Long Response Times

- Diminished motivation of managers and workers
 - Managers and workers in a high WIP environment with long response times experience frustration resulting from work pressure and frequent shifting from task to task
- Missed deadlines
 - The accumulation of WIP and long response times diminishes an organization's ability to meet deadlines and adhere to timetables

Undesirable Consequences of Long Response Times

- Lack of customer satisfaction
 - The service, treatment, or products that are delivered late or with inferior quality lead to customer dissatisfaction
- Diminished forecasting capability
 - Forecasting is important for planning the human resources, raw materials, inventory, marketing, sales, and cash flow
 - Forecasting ability is a function of the forecasting horizon (see the next page for details)

Effect of the Forecasting Horizon on Forecasting Quality



Undesirable Consequences of Long Response Times

- Diminished forecasting capability
 - Reducing system response time will significantly improve forecasting ability
 - If the system response time is t_2 (days or weeks), then one must forecast the market demand at that time
 - If the response time is reduced to t_1 , the quality of the forecast will improve
 - Reduce WIP and you reduce response time

Undesirable Consequences of Long Response Times

- Diminished throughput
 - High levels of WIP cause system inefficiency, which diminishes throughput
- Acquired infections
 - The longer patients stay in a medical environment and the more patients who are there at any given time increase dramatically the chance of acquiring a host of infections

An Example of Undesirable Consequences

- A firm specializing in advanced medical testing equipment won a large contract
- The firm experienced delays in delivery orders, high production costs, rejections due to poor quality, and problems in management and control
- After careful analysis, they found that large WIP \Rightarrow long response times
- Some components in the WIP inventory when exposed to oxygen for too long caused component and equipment failure
- Once WIP inventory and response times were reduced, the firm was able to produce at high throughput with few failures

An Example of Undesirable Consequences

- In a medical staffing firm, the process of assigning workers to hospitals was long
- Many potential workers managed to find assignments on their own or through agencies
- This is sometimes referred to as *evaporating inventory*
- Shortening the process response time greatly improved the situation

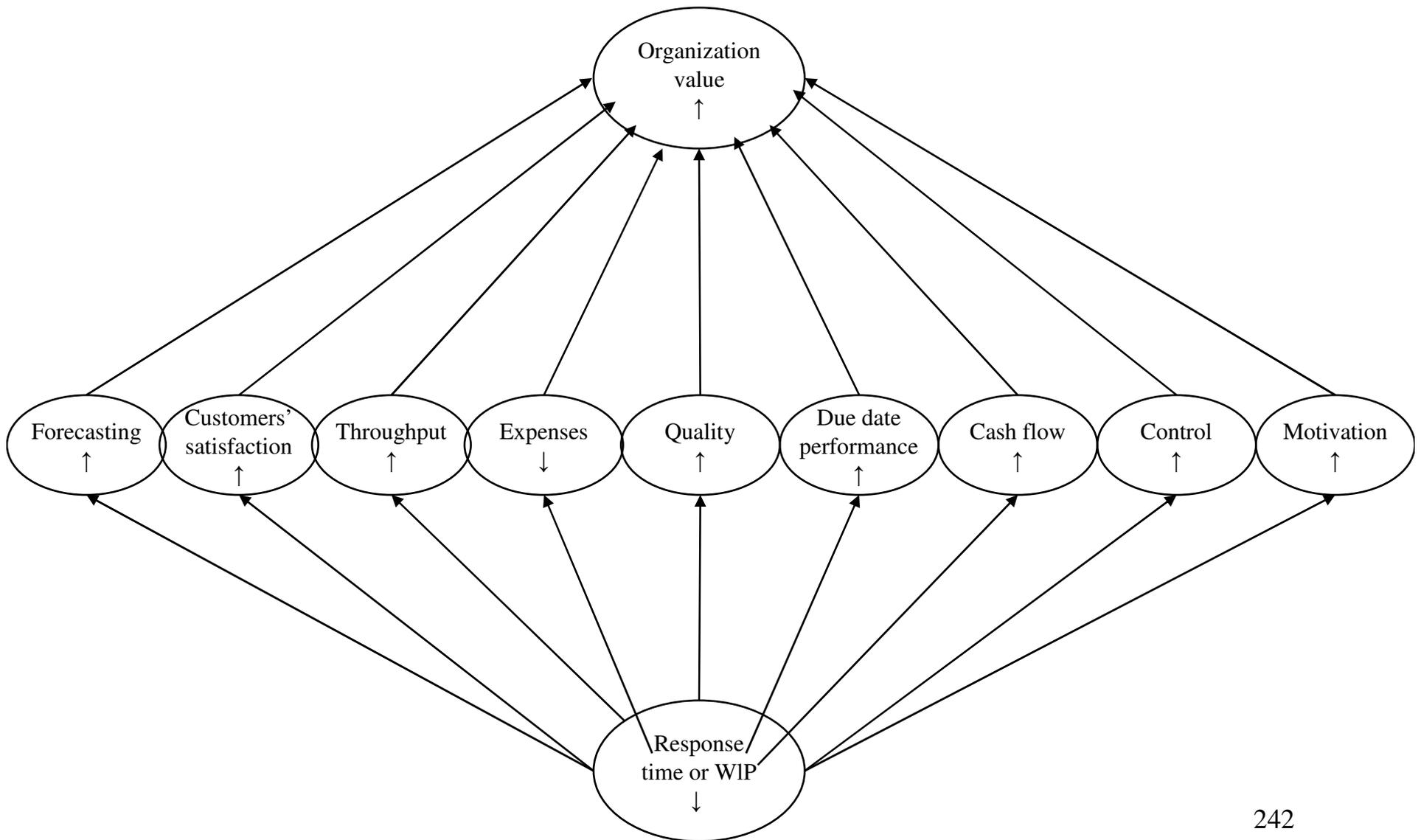
An Example of Undesirable Consequences

- A community hospital reported a high rate of cross-infections in patients hospitalized in the internal medicine wards
- Careful study revealed that this hospital had an average length of stay that was two days longer than others
- Some of the causes were administrative
 - The discharge policy required too many signatures
 - Processing of information was slow
- Administrative policies were changed ⇒ length of stay fell by 1.2 days ⇒ number of cross-infections dropped

Causes of Excess WIP

- Efficiencies syndrome
- Ignorance
- Viewing inventory as assets
 - Looking at an organization from the narrow perspective of financial accounting can wrongly encourage management to increase inventory to show larger assets on the balance sheet
- WIP summary: Shortening response time and reducing the level of WIP serve to bring about significant improvements in organizational performance

Effect of Reducing WIP on Organization Function and Value



Reducing Response Times

- Short response times have strategic and practical importance
- At the strategic level, short response times allow an organization flexibility in moving from one product to another, from one service to another, and from one product or service line to another to better respond to market changes
- At the tactical level, fast response times enable an organization to work with low levels of WIP, reduce costs, and increase throughput
- Both response times and WIP can be reduced by the methods listed next

Methods for Reducing WIP and Response Time

- Strategic gating
- Tactical gating
- Working with a complete kit
- Managing the bottlenecks
- Working by just-in-time (JIT) rule 1
- Working by JIT rule 2
 - Working with small and appropriate work batches and work packages
 - Working with small transfer batches

Methods for Reducing WIP and Response Time

- Working by JIT rule 3
- Measurement and control
- Implementing the drum-buffer-rope (DBR) mechanism
- Quality improvement and reduction of the “garbage plant” in the process
- Avoiding bad multitasking
- Implementing group technology
- Working in parallel rather than serially

Strategic Gating

- Strategic gating is a managerial tool that screens out tasks that should not consume the scarce time of the bottleneck
- Organizations that routinely apply strategic gating (in development, marketing, and sales) reduce workloads by up to 25%
 - Thus, reducing response time for the important tasks
- The time to market is also reduced
 - This contributes to competitiveness

Hospital and Strategic Gating

- Sometimes, strategic gating does not apply
- EDs do not decide that some patients will not be admitted
- Private hospitals do practice strategic gating
 - A hospital may decide not to perform certain kinds of surgeries
 - Many surgical centers don't perform gynecological surgery because of frequent malpractice suits
- Some hospitals specialize in specific procedures, screening out others

Tactical Gating

- Tactical gating is the controlled release of tasks into the system and is appropriate for all organizations
- The tasks that pass the strategic-gating screening must go through tactical gating
- The roles of tactical gating are
 - Releasing only tasks with a complete kit
 - Releasing tasks in small and appropriate batches
 - Assuring that all tasks are released by one source that is responsible for the tactical gating
 - The timing of task release will be determined by the bottleneck capacity while maintaining an appropriate buffer in front of the bottleneck

Tactical Gating

- A triage nurse sets priorities and channels patients to various destinations
- EDs are divided into surgical emergencies, internal medicine, gynecology, and trauma
- Working with complete kits
 - Helps reduce response times
 - We will discuss this later

Managing Bottlenecks

- We now discuss how JIT helps to reduce response times
- The JIT method is of utmost importance in reducing response times
- JIT emerged from industrial plants in Japan
- About 70% of the success attributable to JIT is based on universal concepts
- The remaining 30% contribution is culture-dependent and we do not address it here
- JIT is a method that stands on its own and can be applied as is, but it can be combined with management by constraints

JIT Theory

- Three basic and simple rules
 - Rule 1: Work only as needed in terms of time, quantity, and specifications
 - Rule 2: Work in small, appropriate, and smart batches
 - Rule 3: Avoid waste and activities that do not add value to the organization
- JIT Rule 1
 - A product or service should not be delivered earlier or later than the target time
 - One should not produce more or less than the required quantity

Managing Bottlenecks

- JIT Rule 1
 - A product should be neither underdesigned nor overdesigned with regard to specifications
 - We can identify two types of managerial deviations: shortage deviation and surplus deviation
- Shortage deviation
 - Say demand = 10 and actual supply = 8
 - This is a serious deviation
- However, routine control mechanisms of the organization will usually close the gap

Mechanisms to Close the Gap

- The worker, knowing that he or she will not meet demand, will meet the shortage within a short time
- The worker's superior will monitor the demanded quantities and make sure the deviation is corrected
- Sales and marketing managers will work to close the gap if others have not done so
- The financial manager will point out situations with cash-flow gaps
- The customer will approach management with a request to correct the deviation

Surplus Deviations

- Say demand is for 10 units, but 12 units have been provided
- The organization typically does not address this in the short term
 - The worker feels like an overachiever
 - Managers are too busy with real crises
 - Sales and marketing don't know about the surplus
 - The financial manager will deal with this at the end of the quarter (or year)
 - The accountant will then write down the inventories
 - The customer may not know about the surplus

Surplus Deviations

- Surplus deviations are not routinely dealt with
- But, they can cause damage
 - If the resource is a bottleneck, then generating a surplus of two units (20%) translates into a 20% wastage of the resource
 - Generating the surplus requires using materials beyond those planned for, thus precluding their use for other products
 - WIP is increased
 - It creates unneeded inventory
- Surplus deviation is as bad as shortage deviation

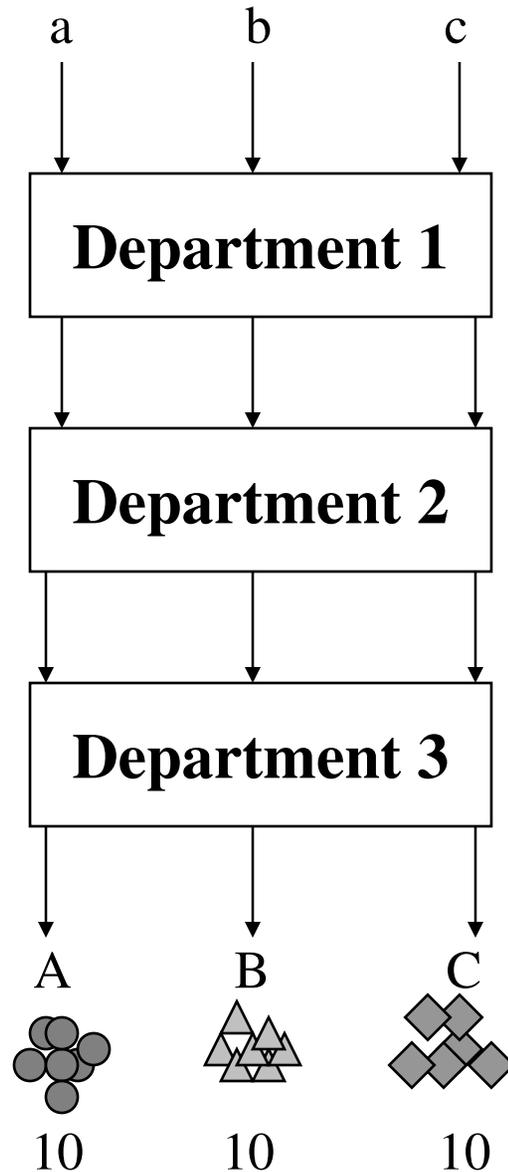
Implementing JIT Rule 1 in Maintenance and Scheduling

- Maintenance of equipment should not be done less than needed nor more than needed
- JIT Rule 1 is also applicable to scheduling of tasks and meetings
 - A surgeon wants to start operating at 10 am
 - He tells the OR staff he wants everything ready at 9:30 am
 - The OR staff requests that the patient be there at 9 am
 - The nursing staff asks an orderly to transport the patient at 8 am

Violating JIT Rule 1

- On the next two pages, we demonstrate what can happen when JIT Rule 1 is violated
- The 1-2-3 pharmaceutical production process has a demand of ten units per month for each of three drugs, A, B, and C
- Demand can be met with current resources
- Suppose the production manager decides to work on batches of 20 units of every product
- Let's see what happens

A 1-2-3 Production System with Three Products



Raw materials for production

50 units per month

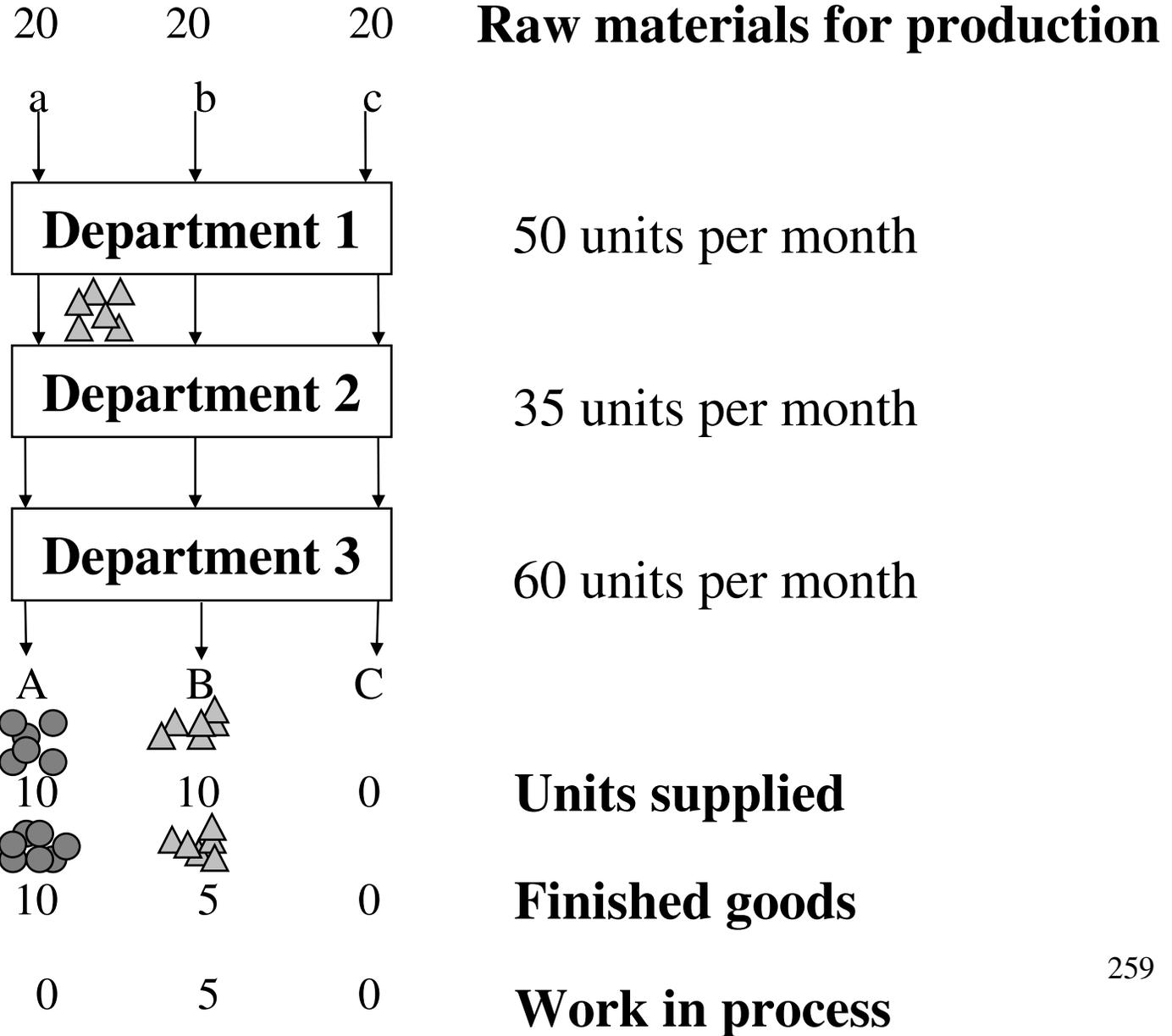
35 units per month

60 units per month

Products

Monthly Demand

Violation of JIT Rule 1



Violating JIT Rule 1

- Twenty units of product A were put into production
- Ten units were delivered to customers and ten were stacked in finished goods inventory
- The 20-unit batch of product B became stuck in department 2
 - 15 units finished all three processing stages
 - 5 units were left as WIP inventory
 - Of the 15, 10 were delivered to customers and five were left in finished goods inventory
- Units of product C were never delivered to customers

Violating JIT Rule 1

- The company produced 35 units rather than 30
- Product C was not produced at all
- Examples
 - In the U.K., the law requires EDs to process a patient in under 4 hours
 - To avoid breaking the law, ambulances arriving at the ED, when full, were asked to wait outside until the ED was ready to process the patient
 - In an ED, a nonurgent patient is sometimes treated at the expense of an urgent one who is waiting unnecessarily

The 40-20-40 Phenomenon

- Ignoring JIT Rule 1 results in the 40-20-40 phenomenon in organizations
 - 40% of demand is supplied ahead of time
 - 20% is supplied on time
 - 40% is supplied late
- Observing JIT Rule 1 “balances the line”

JIT Rule 1

- JIT Rule 2: Work in small, appropriate, and smart batches
- We refer to a batch (lot) as several units which are processed sequentially
- There are working (production) batches and transfer batches
- A working batch reflects several units that are processed at a work center in between two setups
- The determination of batch size is an important issue in planning a service or production
 - In pharmaceuticals, if the work batches are too big, we may lose products due to expiration dates

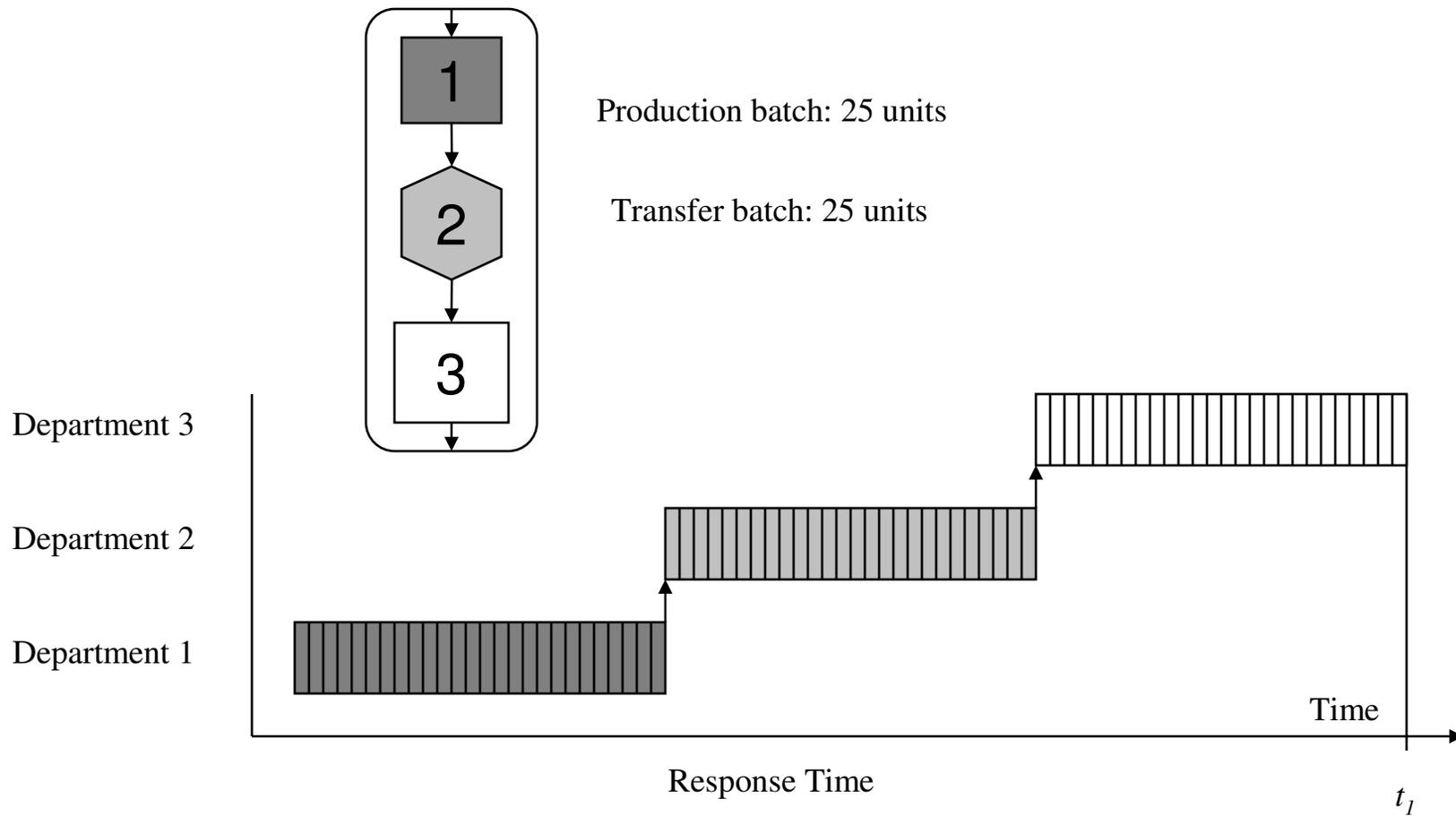
Batch Size May Vary as the Process Unfolds

- The purchase batch for production is 10,000 units
- The shipment batch from the material's supplier is 1,000 units
- Batch size when inspecting a shipment is 500 units
- Production batch is 100 units
- Transfer batch between workstations is 50 units
- Batch size for delivery to customer is 250 units

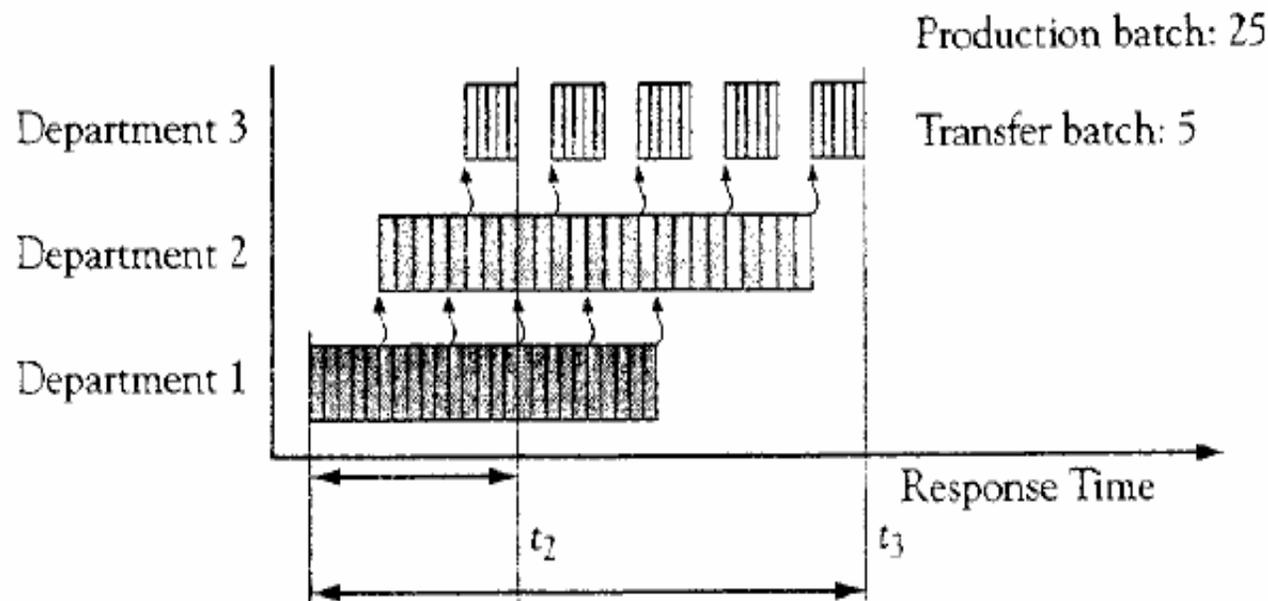
Transfer Batches

- A transfer batch can relate to
 - How frequently a consultant-specialist visits the ED
 - How frequently blood specimens are transferred from the ward to the laboratory
- Transfer batch: The number of units, number of work hours, or frequency of transfer between one workstation and another
- The transfer batch can be bigger than, smaller than, or equal in size to the work batch
- The desire is to make the transfer batches as small as possible
- The next two pages illustrate the effects of transfer batches

Response Time of a 1-2-3 System with a Transfer Batch of 25 Units



Response Time of a 1-2-3 System with a Transfer Batch of 5 Units



Transfer Batches

- Both figures have production batches of 25 units, but different transfer batches
- On page 266, we see
 - The 25 units of the batch are transferred from station 1 to station 2 once the full batch is processed in station 1
 - Once fully processed in station 2, the batch is transferred to station 3
 - After being processed in station 3, the batch is sent to the customer

Transfer Batches

- The system response time here is long, t_1
- There is also a concern that the bottleneck (e.g., station 2) may have idle time when waiting for the finished batch from station 1
- On page 267, we see
 - The production batch is still 25, but the transfer batch has been cut to 5 units
 - Once 5 units have been processed at a given station, they are transferred to the next station
 - Note that the response time has been reduced to t_3 units

Transfer Batches

- To fully exploit this improvement, the customer needs to accept partial shipments
- If so, the first units can be sent to the customer after t_2 units of time
- In many situations, customers are happy to receive some units early, with the rest to follow shortly thereafter
- We discuss an example that comes from the ED of a large hospital next

An Example of Transfer Batches

- In the ED, some patients have their blood drawn for various tests
- In the past, the staff would wait approximately an hour for numerous specimens to collect on the tray before transferring them to the lab
 - This caused an average wait of ½ hour for each specimen
- Today, each blood specimen is transferred to the lab immediately
- The shorter response time of test results yields a shorter patient stay in the ED and quicker diagnoses and treatment

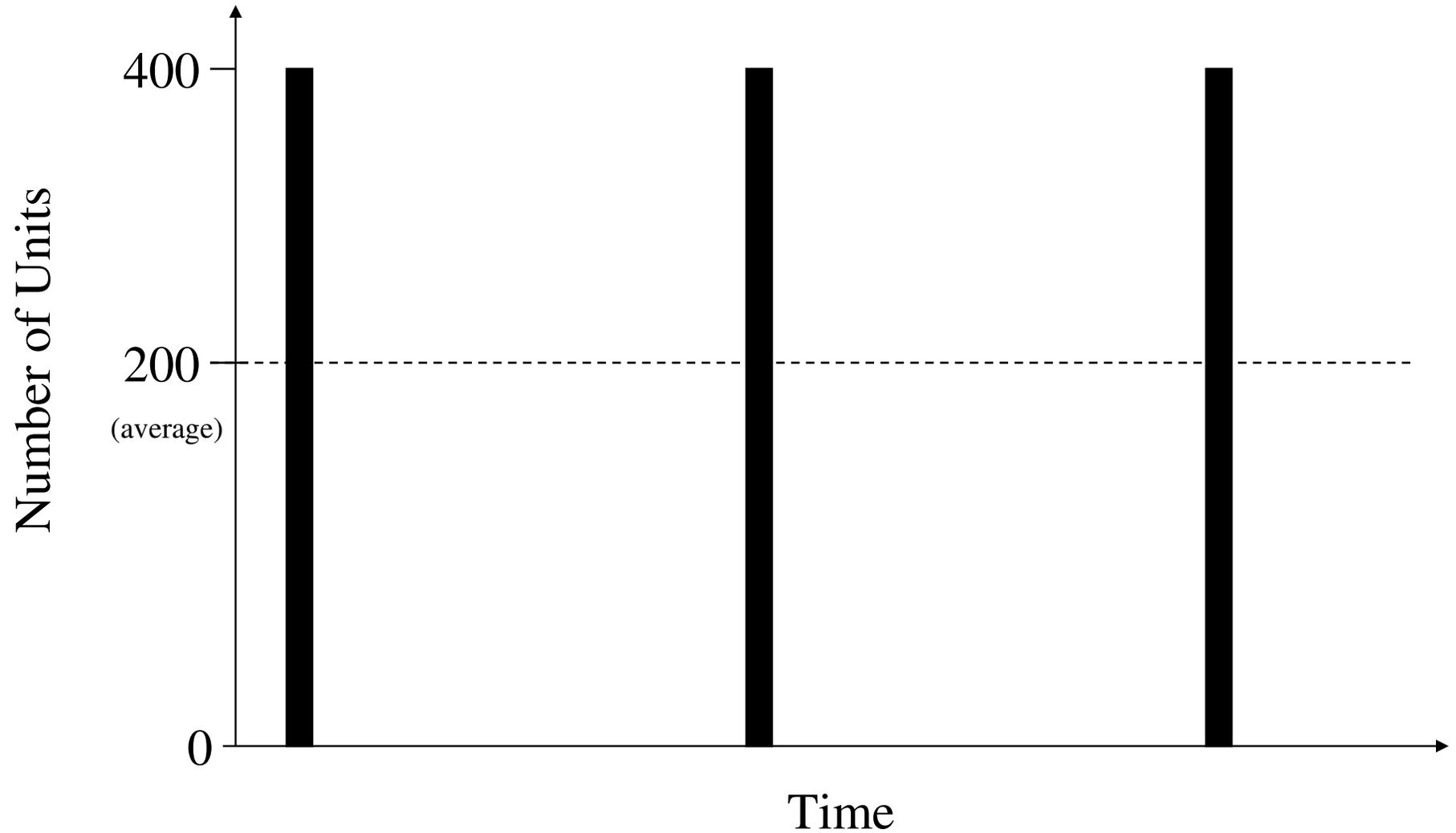
Transfer Batches

- In production and maintenance processes, the transfer batch is measured by the number of units
- In service organizations, the transfer batch is measured in terms of the frequency of transfer (e.g., one hour)
- In a large hospital, the response time of returning lab results to the wards was reduced by a factor of 10 simply by reducing the size of the transfer batch
 - The transfer batch was determined by the size of the transfer trays
 - Large trays were replaced with smaller ones
 - The frequency of transfers increased and response times were shortened

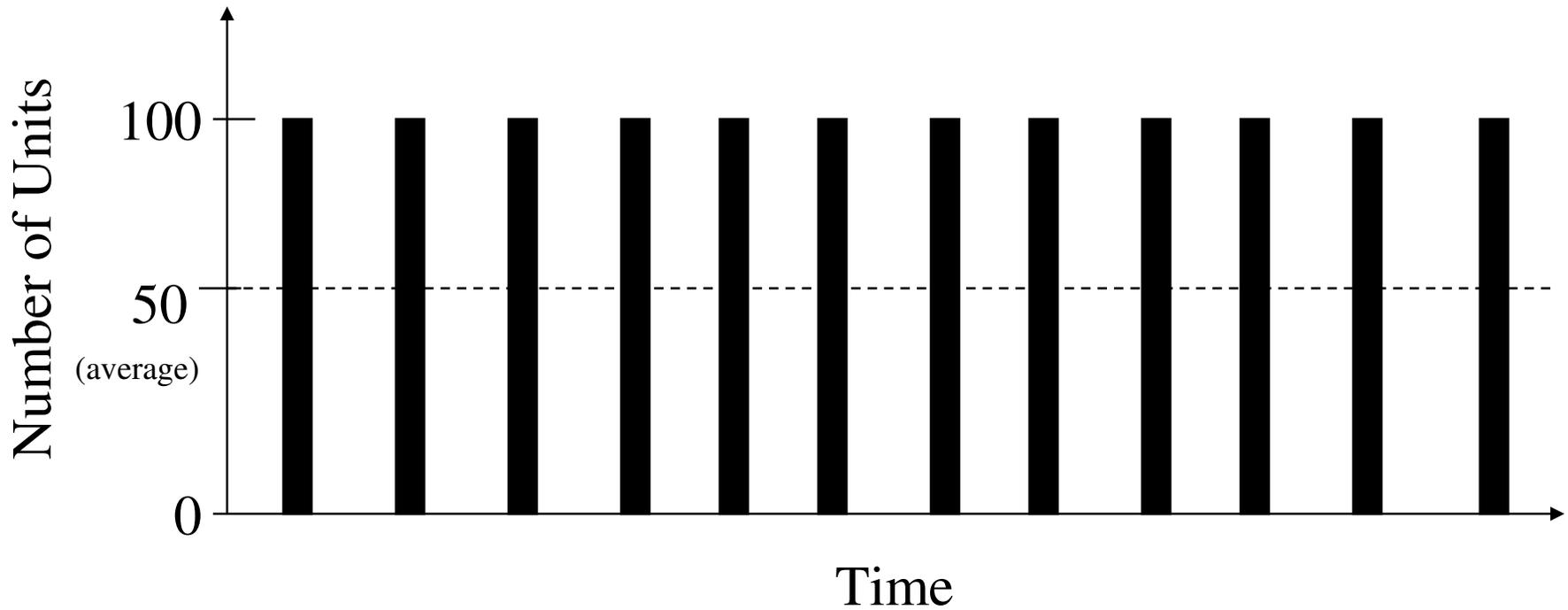
Working (Production) Batches

- A new setup is initiated only when the work on the current batch is completed and we are ready to work on the next batch
- Working batch (or production batch): The number of units (or labor hours) that are worked on continuously at a workstation
 - This is the amount of work between one setup and the next
- Let's examine the effect of shifting from large working batches to smaller ones

Working with Large Working Batches



Working with Smaller Working Batches



Shift from Large to Small Working Batches

- Shifting to small working batches is a major contribution of Japan to management
- On page 274, we see monthly batches
- On page 275, we see weekly batches
- A shift from monthly batches to weekly ones reduced the amount of WIP from 200 to 50 units
- The reduction of the working batch has two positive outcomes
 - Reduction of WIP and response times
 - Improved quality and service to the customer

The Benefits of JIT Rule 2

- In a billing department of a large hospital, preparing and mailing bills occurred once a month
- When billing took place, the department was under pressure
 - Problems and mistakes resulted
- As a remedy, the tasks were performed three times a month
 - The pressure decreased as did billing errors
 - Better customer service resulted
 - It reduced the number of customer calls during peak hours

The Benefits of JIT Rule 2

- In a medical electronics firm that produced expensive imaging equipment, systems were manufactured in batches of 6 months' supply
- While assembling the first units, engineers discovered problems with the electronic boards
- The firm moved to producing smaller batches of 2 months' supply
 - The cycle time was reduced
 - The number of rejects was reduced
- JIT Rule 2 means working with small, appropriate, and smart batches

The Benefits of JIT Rule 2

- Appropriate batches are batch sizes that are congruent with the supply rate expected by a customer
- Smart batches involve using common sense in considering the special needs of the organization
- The effects of reducing batch size (both working batches and transfer batches) are immediate, are relatively easy to implement, and help reduce system response times

Strategic Importance of Reducing Response Times and Working with Small Batches

- Working with small batches allows firms to produce a large variety of products with short response times
- Working with big batches implies longer response times
 - This requires organizations to carry larger finished goods inventory
 - Higher costs result from higher inventory carrying costs and the need to occasionally sell unneeded inventory at low prices

What Prevents us from Working with Small Batches?

- Fear of increasing the number of setups
- Economies-of-scale thinking
- Fear of more complex control
- Fear of increasing cost per unit
- We discuss each of these in detail, next

Fear of Increasing the Number of Setups

- The setup time is nonproductive time
 - During setup, one cannot produce or provide service
- In the past, the setup process was complex
- The modern approach to setups is to avoid long and disorderly setups
- Reducing setup time by 50% allows a similar reduction in the working batch while maintaining the same ratio between productive time (working on the batch) and nonproductive time (setup)

Fear of Increasing the Number of Setups

- A key point is that additional setups in resources that are not bottlenecks do not cost money
- On the next page, we see the relation between the capacity utilization and the reduction of 50% in the batch size in a work center with a 60% capacity utilization
- As long as the resource is not a bottleneck, working with small batches only moderately impacts the capacity utilization of this resource

Effect of Batch Size on the Load of Noncritical Resources

Load	Large Batch	Small Batch
Percentage of productive work	50	50
Percentage of setup time	10	20
Percentage of total capacity utilization	60	70

Note: Small batch is half the size of the large batch.

Economies-of-Scale Thinking

- People are used to the concept of economies of scale where large quantities imply savings
- When batches are processed in a system with excess capacity, economies of scale are not important
- In fact, working with large batches increases response time and other undesirable consequences of WIP
- When a workstation is a bottleneck, setup time should be shortened as much as possible

Fear of More Complex Control

- Working in small batches and splitting a batch into several smaller transfer batches seemingly leads to more complex control as a result of needing to control more batches
- It is true that shifting to smaller working and transfer batches leads to handling more batches
- However, with shorter times for each batch, the number of batches in the system at any given time is smaller rather than bigger

Increasing Cost per Unit

- A manager who works in a system that is using the classical costing approach may resist shifting to work involving smaller batches, fearing this may negatively affect the cost-per-unit measure
- For example, in a station that is not a bottleneck, setup time is 1 hr. and the processing time of each unit is $\frac{1}{4}$ hr.
- The cost per unit in a batch of 100 is

$$T(100) = \frac{1 + (0.25 \times 100)}{100} = \frac{1 + 25}{100} = \frac{26}{100} = 0.26 \text{ hours per unit}$$

Increasing Cost per Unit

- The cost per unit in a batch of 10 is

$$T(10) = \frac{1 + (0.25 \times 10)}{10} = \frac{1 + 2.5}{10} = \frac{3.5}{10} = 0.35 \text{ hours per unit}$$

- The measurement of cost per unit is a local view that can cause suboptimization, especially in a workstation that is not a bottleneck
- A global view requires consideration of the benefits and costs associated with smaller batches

JIT Rule 3

- Waste includes activities, processes, or use of capital that do not contribute added value to the organization, the customer, the process, or the product
- Examples of waste
 - Overproduction
 - Waiting times
 - Unnecessary conveyance
 - Rejected products in processing
 - Surplus stock

JIT Rule 3

- More examples of waste
 - Poor quality
 - Unnecessary space
 - Capital surplus
 - Overspecification and overdesign of the product or service
 - Unnecessary steps and processes
- JIT Rule 3: Avoid waste and activities that do not add value to the organization

JIT Rule 3

- Measurement and control
 - Measurement, by itself, results in real improvement
- Implementing the DBR Mechanism
 - Implementing DBR in marketing, sales, production, development, and service leads to significant reduction of the work process
- Quality Improvement
 - Poor quality in processes of sales, marketing, development, production, or service results in much rework, thus increasing response times and the amount of WIP
 - We discuss quality improvement later

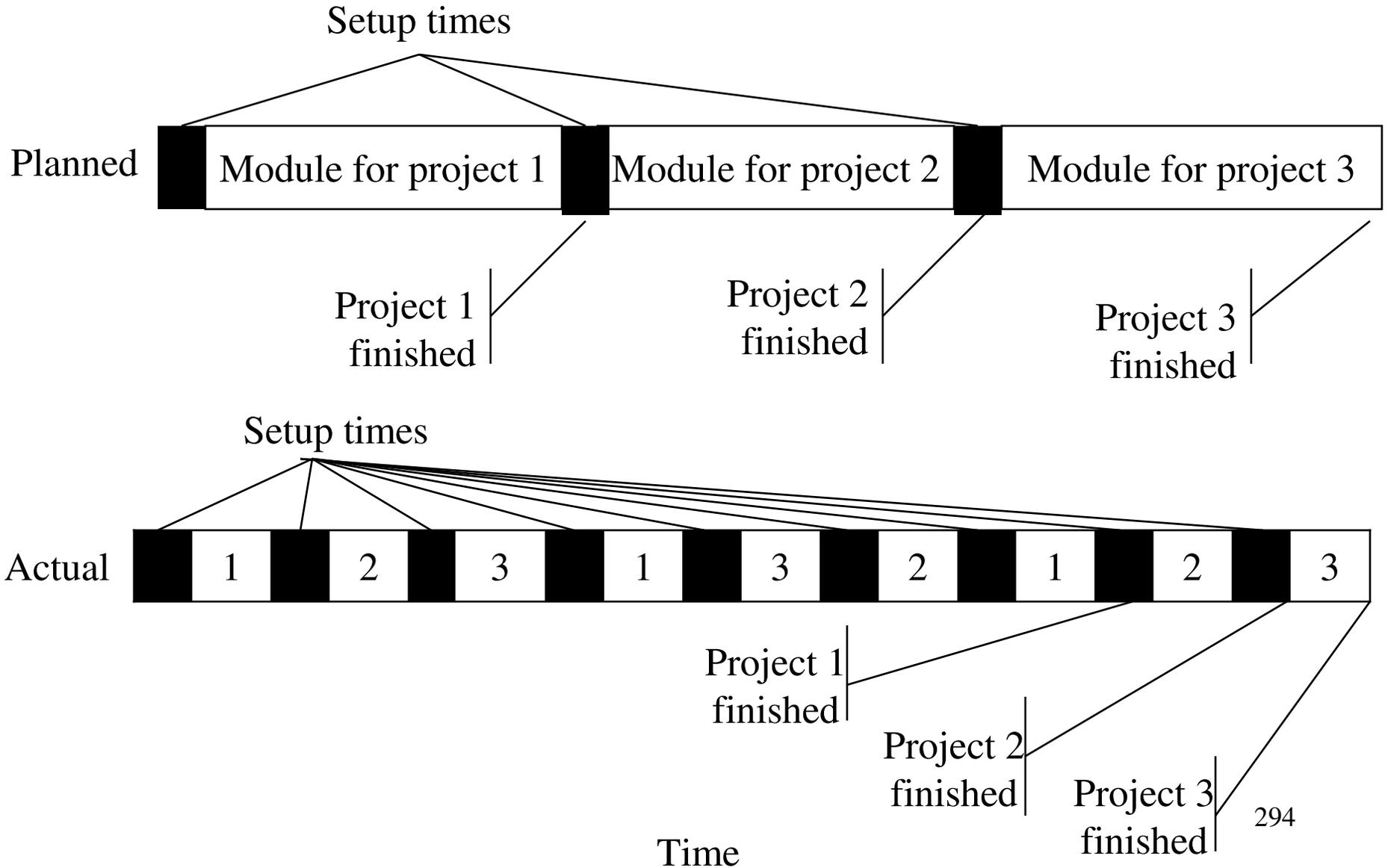
JIT Rule 3

- **Avoiding Bad Multitasking**
 - This is the phenomenon of jumping among many open tasks that are waiting to be processed
 - Bad multitasking leads to a decrease in throughput, longer response times for finishing tasks, late deliveries, reduced work quality, and increased WIP
- **Now we demonstrate the negative effect of bad multitasking**
 - Suppose 3 software development projects and 3 different project managers, but one software engineer
 - Projects 1,2, and 3 are due in 2, 4, and 6 weeks 292

JIT Rule 3

- In reality, things happen differently
 - See the figure on the next page
- Suppose each project manager pressures the software engineer to work on his project
- In the end, none of the projects is delivered on time
 - More time is spent on setup and relearning
 - Work quality may also suffer

Negative Effects of Bad Multitasking



Bad Multitasking

- Bad multitasking can be reduced by
 - Teach and explain
 - Apply strict control using tactical gating
 - Measure and control the number of tasks assigned to each worker
- A significant reduction of bad multitasking will reduce response times and increase throughput and quality

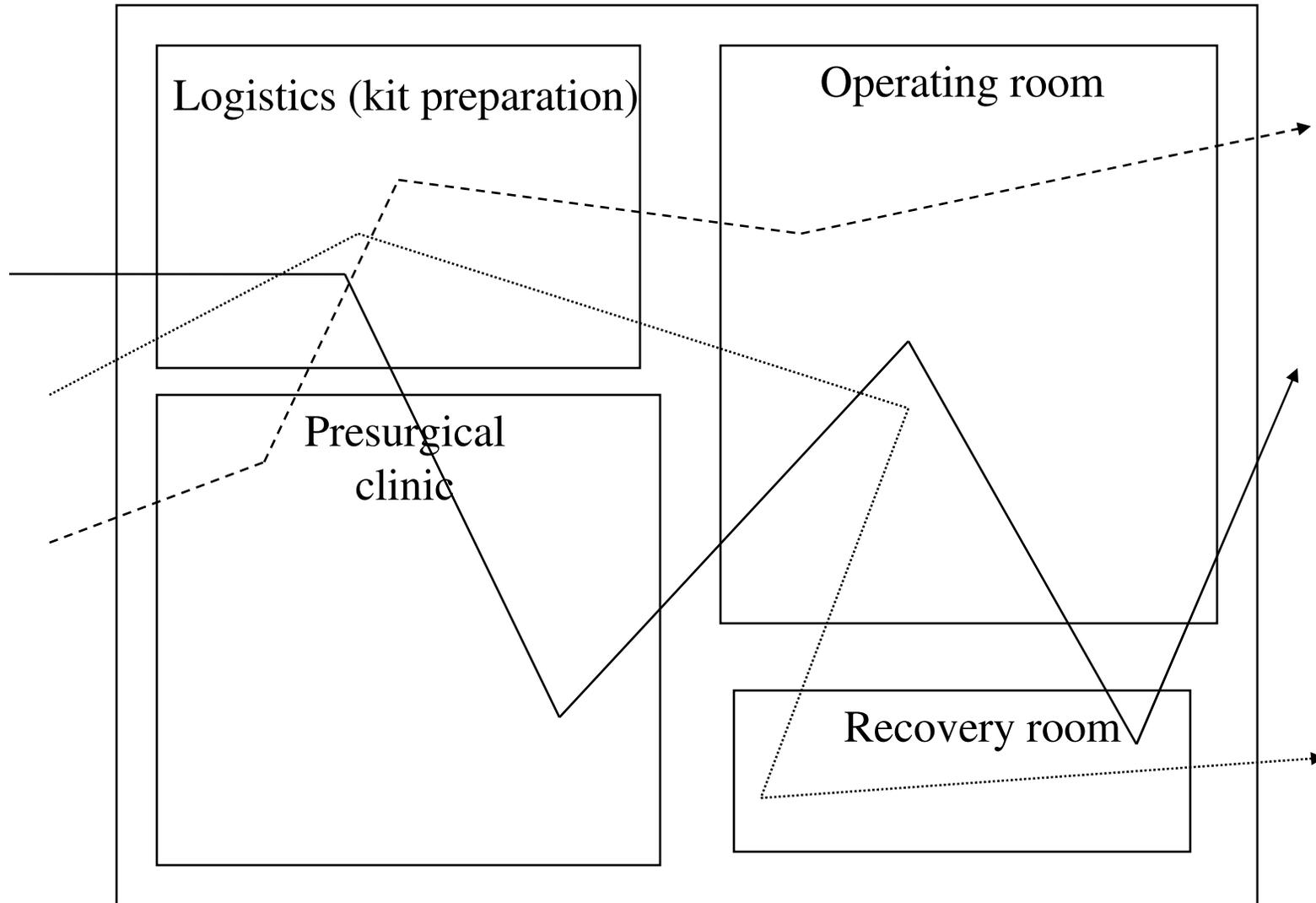
Implementing Group Technology

- Group technology is an approach where similar tasks are grouped and aggregated under specialized work groups
- A hospital ED may be separated into a surgical ED, internal medicine ED, pediatric ED, and gynecological ED
 - Within each specialized ED, the variance among patients is small
 - Each specialized ED can thus offer more professional, structured, and uniform medical diagnosis and treatment because each employs the relevant specialists

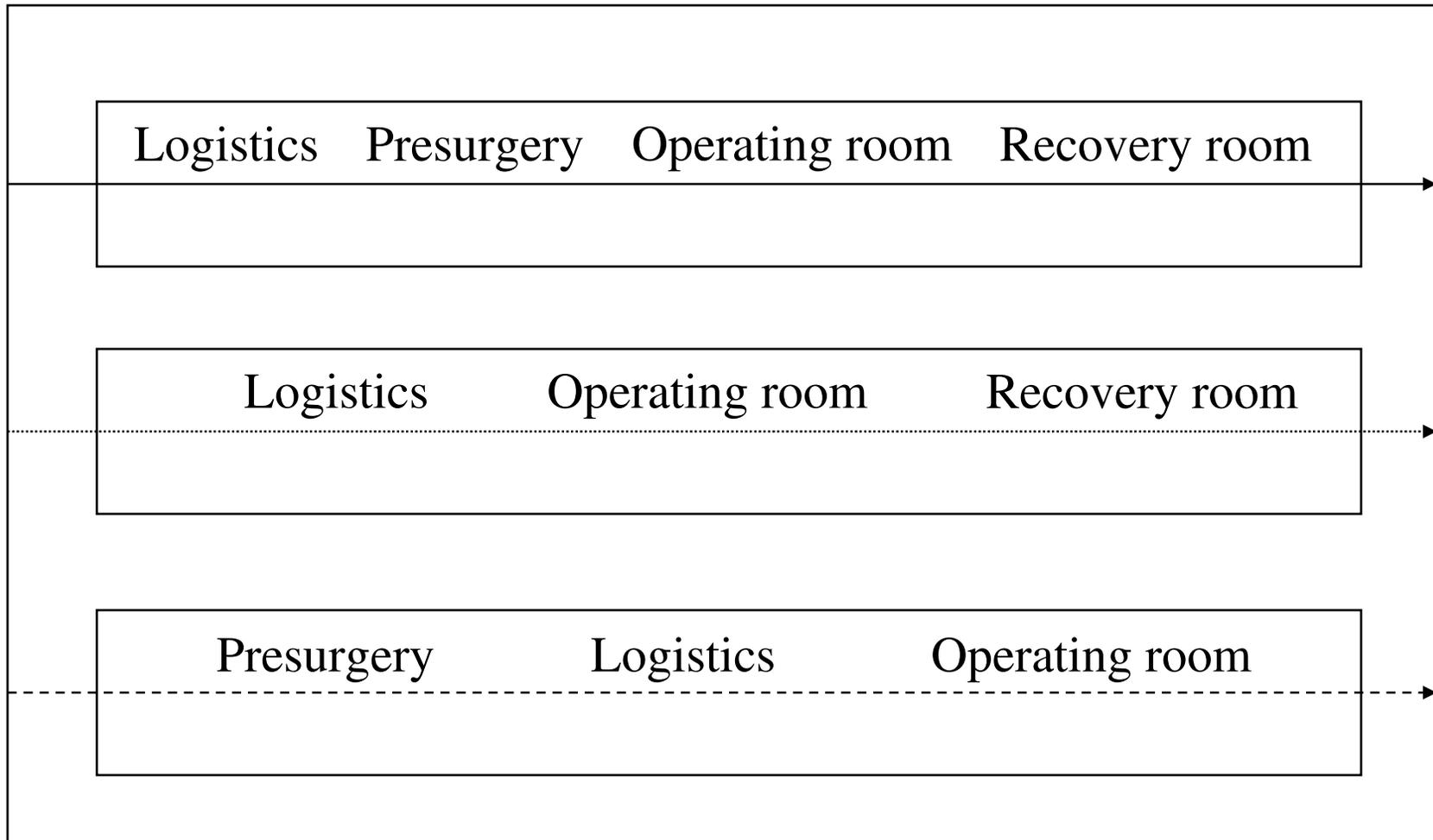
Implementing Group Technology

- The classical approach in OM is to create teams that specialize functionally, where the customer or product moves among the groups
 - See page 298 for an illustration of this
- In the group technology approach, integrative teams provide the entire service needed by the customers in one place
 - See page 299 for an illustration of this

System Arrangement with Functional Structure



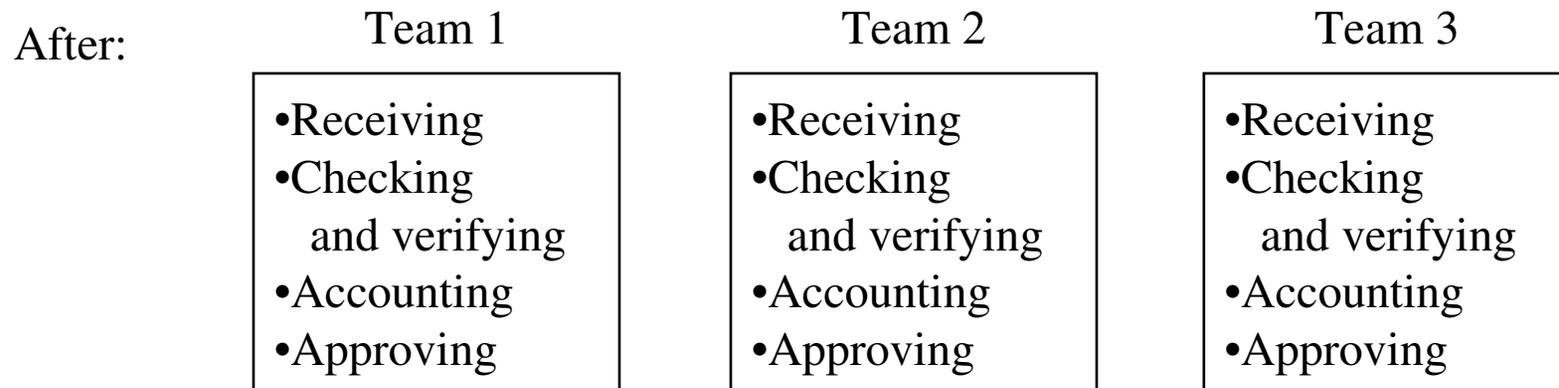
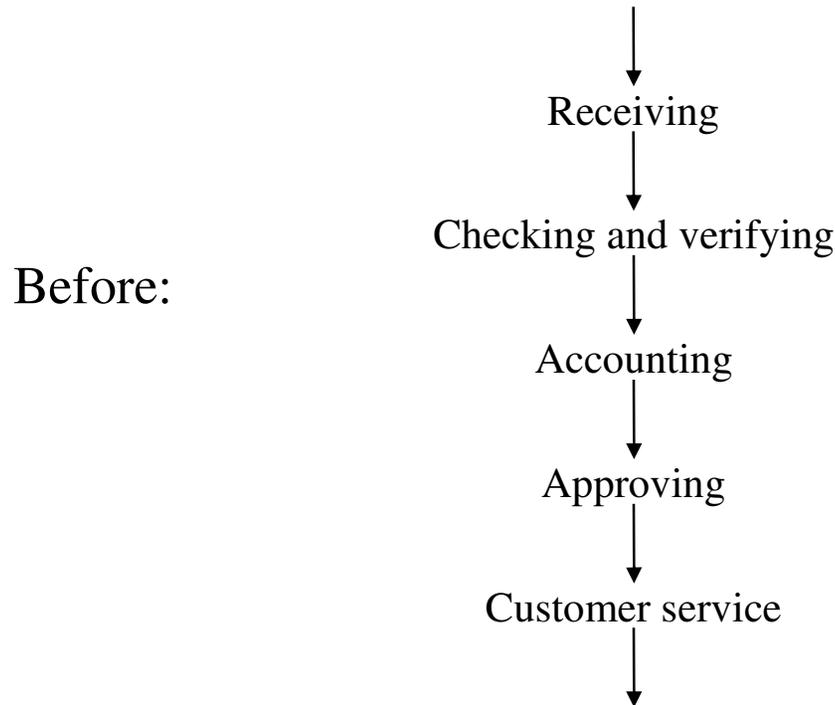
System Arrangement Using Group Technology



Implementing Group Technology

- The two previous pages demonstrate how a complex work flow under the functional structure becomes simpler under the group technology approach
- The group technology approach generates a collective responsibility and accountability of the group to the customer or the product
- It yields faster response times, a reduction in WIP, and increased output
- In a claims department of a health insurance company, every customer had to go through five steps
 - See next page for before and after

The Claims Department Before/After Group Technology



Implementing Group Technology

- A full (100%) group technology can only rarely be achieved
- There are usually some common resources shared by several groups
 - E.g., imaging services can be common for all the specialized EDs in the earlier example
- Group technology can succeed where
 - Work volume justifies a specialized group
 - The resources shared by the various teams are not bottlenecks
 - Team spirit can be created

What is a Complete Kit?

- A complete kit in health care is the set of components and materials, medical documents, laboratory results, and other information needed to complete a given procedure, medical process or task
- The in-kit of a given task is the matter and data required as an input to an operation or medical procedure
- The out-kit of a given task is all the material and data required as an output of an operation or medical procedure

Drawbacks of an Incomplete Kit

- To appreciate the value of a complete kit, it is important to understand the disadvantages of working with an incomplete kit
- We will demonstrate these in three areas of the health care environment
 - The ED
 - The OR
 - Pharmaceutical purchasing
- Any specific task in a medical environment defines a process

More Work in Process

- For the ED, the process includes
 - Taking the medical history of a patient
 - Physical examinations by nurses and doctors
 - Medical tests
 - The conclusion reached by the doctor (discharge, admit, or transfer)
- WIP in the ED means people waiting for completion of the admission process
- Using an incomplete kit leads to more WIP, i.e., more people waiting in the ED

More Work in Process

- An ED example of an incomplete kit
 - Suppose a specialist consultation is requested for a patient in the ED
 - The consultant begins, but not all lab results, imaging results, and an electrocardiogram are available
 - The specialist may need another visit later on
 - The patient must wait
 - The consulting physician may become busy elsewhere and may become a bottleneck

Longer Response Time

- In the OR, the response time would be the time from when a patient is moved to the OR until he or she is moved back to the ward or to the intensive care unit
- More incomplete kits cause more WIP and hence a longer response time
- Suppose a patient arrives in the OR without a complete kit
 - E.g., electrolyte data not available
- The anesthetist may decide to wait until the data becomes available (after a blood test) or may postpone surgery \Rightarrow longer response time

High Variance of Quoted and Planned Response Times

- When a patient arrives with a complete kit for a routine procedure or test (e.g., colonoscopy), it is easy to predict the procedure response time
 - The response time deviation is relatively small
- When a patient arrives with an incomplete kit, the response time variance increases
 - Prediction, scheduling, and planning becomes difficult
 - This leads to inefficient performance

Poor Quality and More Reworking

- Patients arriving with incomplete kits tend to wait too long in inadequate facilities
- When missing items or information arrive, doctors must review them ⇒ poor-quality service
- The clinical outcome may be impaired due to the delay
- Since several different physicians from different shifts may need to interface with the patient, delays may be exacerbated

Impact on an Operating Room

- The number of patients being operated on in a specific time interval is the throughput of an OR
- Using an incomplete kit in the OR causes a decline in throughput
- Using an incomplete kit increases the required time per patient
 - Due to duplicate handling
- In related studies, researchers found the inefficiency factor associated with an incomplete kit to be about 80%
 - A process that takes 1 hour with a complete kit may require 1.8 hours with an incomplete kit

More Drawbacks of an Incomplete Kit

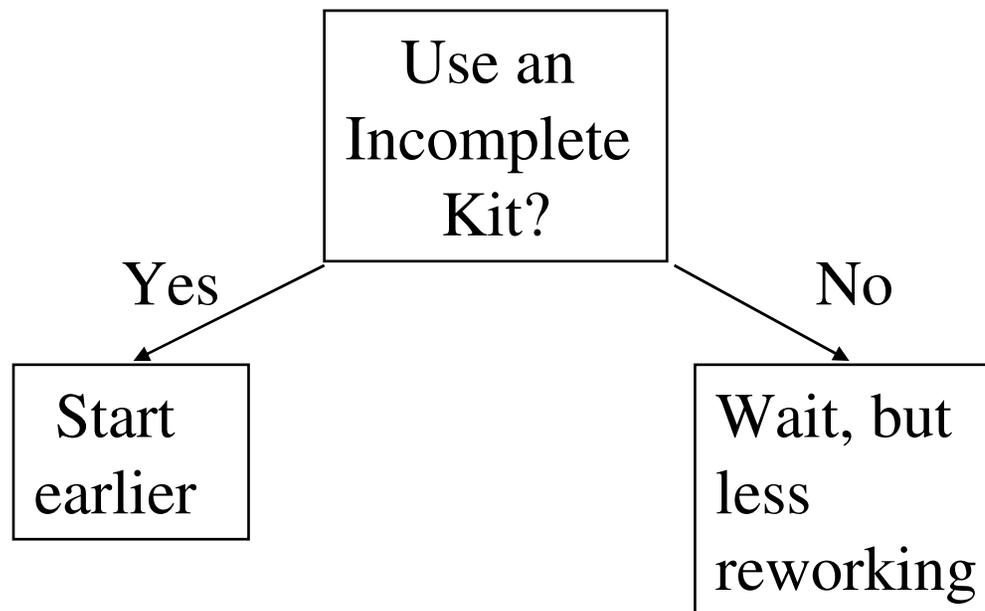
- A high WIP causes higher operating expense due to more holding costs, more scrap, and more work put into the task
- Decline in staff motivation
 - Using an incomplete kit goes against the grain
 - E.g., consider a physician on duty in the ED needing to see many patients unnecessarily two or more times because a complete kit is not available the first time

What Stops People from Using a Complete Kit?

- The efficiencies syndrome
 - This is the urge to have resources utilized as much as possible
 - The basic remedy is a major change in the management of the organization, incorporating the complete kit approach into the overall concept
- Pressure for an immediate response
 - E.g., sometimes a physician in the ED starts treating a patient with an incomplete kit due to pressure for an immediate response

Staff Eagerness to Show Goodwill

- In response to pressure from management, nurses and physicians express their goodwill by releasing incomplete kits to the medical or surgical ward
- The conflict is represented below



The Complete Kit Concept in Health Care

- General rule: A gatekeeper should be designated as the only person authorized to release jobs
- A colonoscopy is one of the most significant diagnostic and therapeutic applications of endoscopy
 - It can diagnose potentially curable colonic cancers that are missed by other techniques
 - At least 24 hours of patient preparation is required
 - Patients are given a checklist
 - But, some cancel and others show up not-fully prepared

Incomplete Kits and Colonoscopies

- It is important to instruct the staff not to schedule patients for colonoscopy who have not undergone full preparation
- What can be done?
 - Three days before the scheduled date, each patient should be phoned and reminded of the instructions
 - If a patient cancels at this time, another can be scheduled
 - If the cancellation rate is (still) at 10%, one should overbook by 10%

Incomplete Kits and Hernia Surgery

- One of the authors of this book was scheduled for hernia surgery
- Upon checking in to the hospital, he was asked what medications he was taking
- He was taking vitamins and low-dose aspirin
- He should have been instructed to stop taking aspirin 10 days before surgery, but he had not
- The admission was cancelled and the surgery postponed
- Adverse effects: Empty bed in the ward, vacant slot in OR schedule, patient's time was wasted, etc.

Implementing the Complete Kit in Health Care

- The introduction of a complete kit process has to be part of a major change in the organization
- Top management has to be involved in the process
- One person in each department has to be appointed to take charge
- The process must be monitored
- Employees and internal customers need to be informed of the change

Implementing the Complete Kit in Health Care

- External customers (e.g., HMOs) should be notified that they will get better due-date performance and response times if they submit a complete kit
- Components and materials should be ordered in complete kits
- All activities should be synchronized, ensuring that the out-kit of the current activity is the in-kit of the next one
- Components and procedures need to be standardized whenever possible

Implications for Health Care MIS Departments

- The management information system (MIS) department in a health care organization has to support activities with certain tools to enable a kit
- Most of the items in a medical kit are medical information and data
- An MIS plays a major role in reinforcing and implementing the complete kit paradigm and should be designed accordingly
- Using an intranet with an automatic checklist is a good starting point

Implications for Medical Purchasing/Logistics Departments

- The purchasing department would change its procedures
- It would purchase complete kits
 - The orders would be in kits, rather than components
 - The complete kit of a colonoscopy includes an endoscope, various medications, a resuscitation cart, etc.
- Purchasing departments would work with fewer suppliers and purchase more items from each supplier
- Suppliers would be evaluated based on price, response time, quality, and completeness of kits

Complete Kit Summary

- A study of health care organizations that have implemented the complete kit concept shows that it has reduced WIP and cut response time by a factor of three
- Major improvements are possible in the following areas of health care
 - EDs
 - ORs
 - Outpatient clinics
 - Radiology departments

Performance Measures and Managerial Control

- As we have seen, local performance measures may lead to suboptimization
 - A measure such as cost per unit may result in the production of large batches
 - A measure of the number of patients seen in an ED may lead to early discharge and suboptimal treatment
- Thus, we need global performance measures that will yield the benefits described next

Benefits of Global Performance Measures

- Voicing management policy regarding an organization's goals
 - People behave according to how they are measured
 - Align individual behavior with organizational goals
- Aid decision making
 - The performance measures are an intermediary between decision makers
 - In managing the introduction of a new drug, if time to market is an important performance measure, the project manager will seek to reduce the time to market

Benefits of Global Performance Measures

■ Control

- Performance measures enable an organization to monitor whether the organization is moving toward achieving its goals in the short and long term
- What corrective actions (if any) are needed

■ Reward and evaluation

- Good performance measures should be linked to evaluation and reward

Benefits of Global Performance Measures

- Ability to decentralize
 - Appropriate performance measures allow decentralized decision making to work
 - These performance measures have the following characteristics
 - Global and effective
 - Simple and clear
 - Based on satisficer approach
 - Linked to easy and simple data collection
 - Customized to the organization

The Six Global Performance Measures

- Throughput (T)
- Operating expenses (OE)
- Inventory (I)
- Response time (RT)
- Quality (Q)
- Due date performance (DDP)

Throughput

- Throughput is the effective output of an organization
- A hospital OR should be interested in the income from procedures and surgeries
 - Throughput is the monetary contribution of the procedures
- In business, throughput is defined as total *actual sales* minus *real variable costs* of those sales
 - *Actual sales*: Sales that were actually carried out, minus returns and cancelled sales
 - *Real variable costs*: Raw materials, components, subcontractors, commissions, and so forth

Throughput

- See the table on the next page
- Suppose the company decides to produce 200 units of A, 150 units of B, and no units of C in order to maximize the number of units produced
- The throughput is given by

$$\begin{aligned} T &= 100(10 - 4) + 100(10 - 4) + 0 = \\ &100(6) + 100(6) + 0 = 600 + 600 + 0 = \$1,200 \end{aligned}$$

- If the company had produced to meet demands, the throughput would have been \$1800

Calculation of Throughput

	Product A	Product B	Product C
Demand (units)	100	100	100
Selling price per unit (\$)	10	10	10
Real variable costs per unit (\$)	4	4	4
Actual production (units)	200	150	0

Throughput

- In many places, middle management is occasionally measured by the number of units produced
 - This can lead to accepting orders that will result in losses
- Measuring throughput in monetary units (dollars) sends a message throughout the organization that *we mean business*

Operating Expenses

- Operating expenses include direct labor, indirect labor, rent, and other fixed expenses
- These expenses are fixed expenses of the organization in the short term and the medium term
- The manager should evaluate any suggestion for change or improvement to determine if operating expenses will increase or decrease and the effect on throughput

Inventory

- Inventory is classified into three categories
 - Raw materials inventory
 - WIP inventory
 - Finished goods inventory
- The value of the three inventory types will be measured in the cost of *raw materials only*
- In health care organizations, the WIP inventory reflects patients in the system, whether waiting or being cared for
- Patients waiting to enter the system (e.g., patients waiting for organ transplants) represent the raw materials inventory

Response Time

- An appropriate measure of response time is one that looks at the process from the perspective of the customer
- For example, patients are concerned with the total time they spend in the system, including various waits
- They do not care who is responsible for a longer-than-usual wait

Quality

- Every organization must define its relevant quality measures
 - Percentage of products or services achieved correctly the first time
 - Costs of “nonquality”
 - Customer satisfaction
 - Percentage of rehospitalizations within two days after discharge
 - Number of customer complaints
- In health care, we focus on two types of quality: service quality and clinical quality

Due Date Performance

- Due date performance reflects an organization's reliability in meeting deadlines
- Due date performance can be measured in several ways
 - Percentage of on-time performance
 - Delayed revenue collection as a result of not being on time (back orders)- simple, useful, and clear
 - Dollar days
- Dollar days calculates the due date performance as the sum of the cash value of orders multiplied by the number of days of delay

Due Date Performance Using the Dollar-Days Measure

Order Number	Order Value (\$ thousands)	Delivery Date	Days Late	Dollar Days for June 1 (millions)
351	100	January 1	150	15
352	200	February 1	120	24
353	300	March 1	90	27
Total dollar days for delayed orders				66

Calculating Profit

- We can conceptually define the profit of a firm to be

$$\text{Profit (P)} = \text{Throughput (T)} - \text{Operating Expenses (OE)}$$

- The importance of this relation is that it forces decision makers in an organization to think globally and ask questions such as
 - Will organizational output increase as a result of the decision?
 - Will the operating expenses decrease as a result of the decision?
 - Will the decision increase the difference of throughput minus operating expenses?

Adapting Global Performance Measures to an OR

- **Throughput:** The throughput of the OR can be measured by monetary contributions (revenues for surgery minus real variable costs)
- **Operating expenses:** Rent for the OR and payment for surgeons, anesthesiologists, nurses, and administrative and cleaning staff
- **Inventory:** The number of patients waiting in the OR itself and the number waiting in the preparation room

Adapting Global Performance Measures to an OR

- Response time
 - Time from summoning a patient from the ward to the end of surgery
 - Time from when a patient enters the OR until leaving for the recovery room
 - Time from finish with one patient until the first incision on the next patient
- Quality: clinical quality and service quality
- Due date performance: Meeting the OR schedule

Adapting Global Performance Measures to an OR

- The six global measures form a good starting basis
- In the OR example, we can drop the measure of due date performance and replace it with two other measures
 - Percentage of patients rejected for surgery
 - Percentage of patients arriving in the OR with an incomplete kit
- Begin measuring right away
 - Start with two or three reasonable measures
 - Refine and add measures over time

The Measurements Profile and Global Decision Making

- The measurements profile is a tool for aiding in global decision making
- Its use examines the alternatives with respect to the six global measures
- This profile presents a succinct picture of each alternative and thus enables an easier comparison of the effect of each alternative on the various dimensions of organizational performance

The Measurements Profile

Performance Measure	Alternative A	Alternative B
---------------------	---------------	---------------

Throughput (T)

Operating expenses (OE)

Inventory (I)

Response Time (RT)

Quality (Q)

Due date performance (DDP)

Fluctuations, Variability, and Uncertainty

- We refer to the uncertainty related to malfunctions, faults, and disruptions as *fluctuations*
- Sources of fluctuations
 - Fluctuations in demand
 - Fluctuations in capacity
 - Fluctuations in quality
 - Fluctuations in the availability of materials and parts

Sources of Fluctuations

- Fluctuations in demand
 - Seasonality
 - Technological changes and preference changes
- Fluctuations in capacity
 - Variation of work rate at different stations
 - Variation in setup times
 - Malfunctions
 - Employee absenteeism
 - Scheduling and timing problems
 - Incomplete kits

Sources of Fluctuations

- Fluctuations in quality
 - Unexpected defects
- Fluctuations in the availability of materials and parts
 - Problems with quality of materials and components
 - Delays in supply
 - Supply problems (less than ordered, different from what was ordered, and so forth)

ED Fluctuations

- Fluctuations occur in the ED of a hospital due to
 - Uncertainty in the diagnosis
 - Unavailability of consultant-specialists
 - Unavailability of operating rooms
 - Unavailability of critical resources
 - Uncertainty of demand

Evolution of Fluctuations in a Process

- Consider the planning of a new line of services, products, or development (see the next page)
- Every task that enters the system (e.g., a patient receiving treatment) must go through stations 1, 2, and 3
- Let's assume that \$15 million has been targeted for creating this process
- The developers are faced with three alternatives for buying equipment, each costing \$15 million

The Planned Process

**Incoming
patients or jobs**



Department 1



Department 2

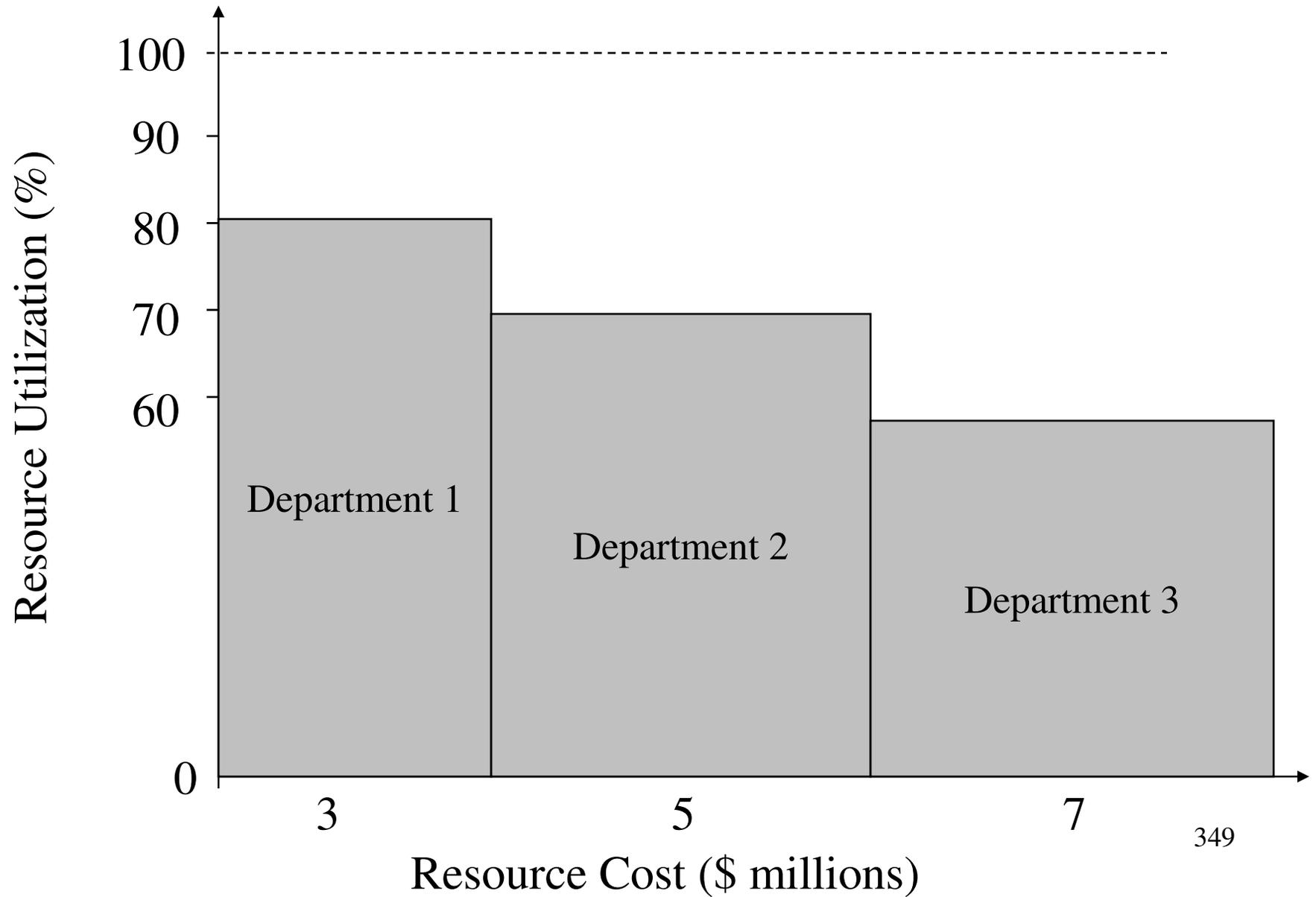


Department 3

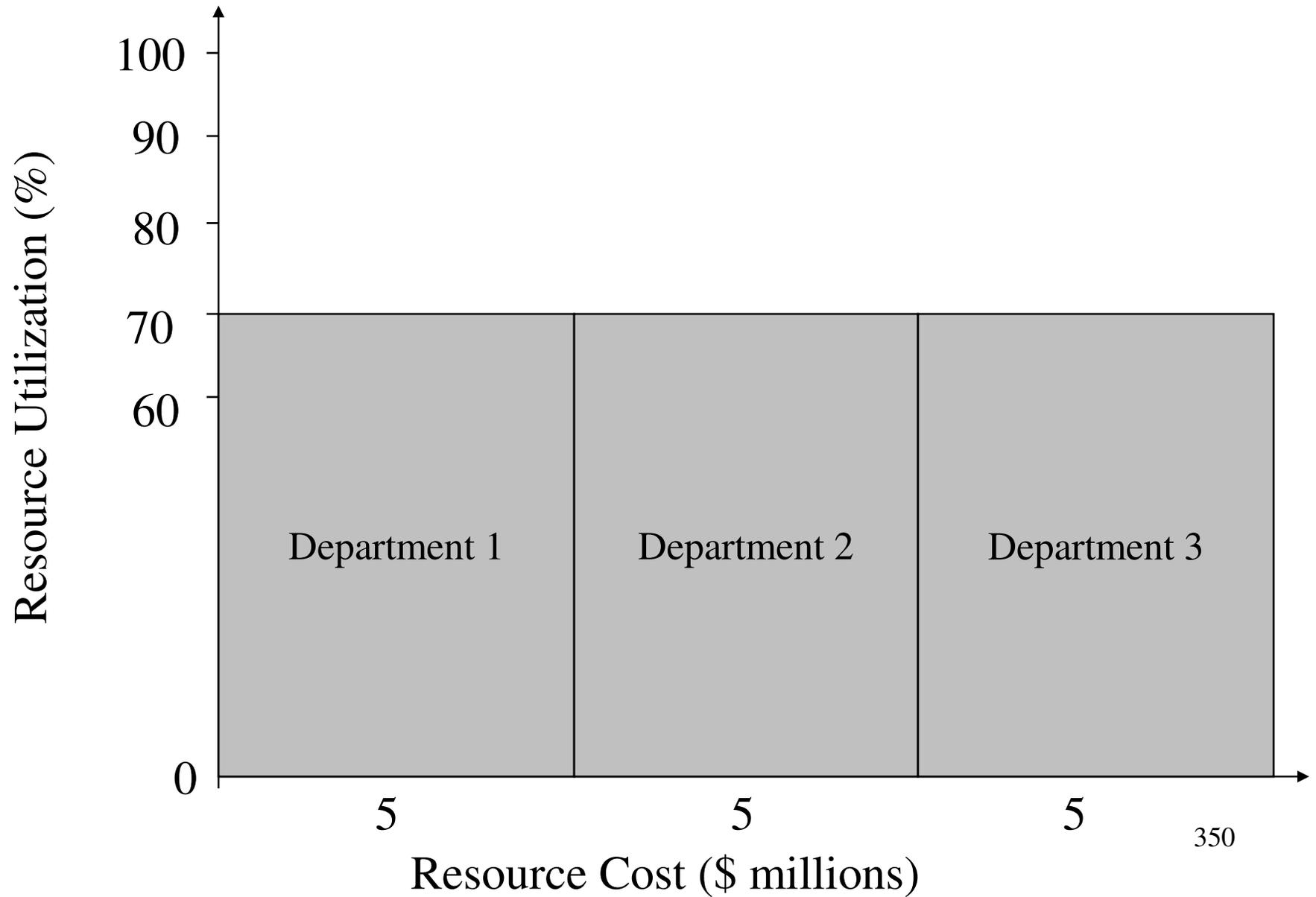


**Patients or jobs
after treatment**

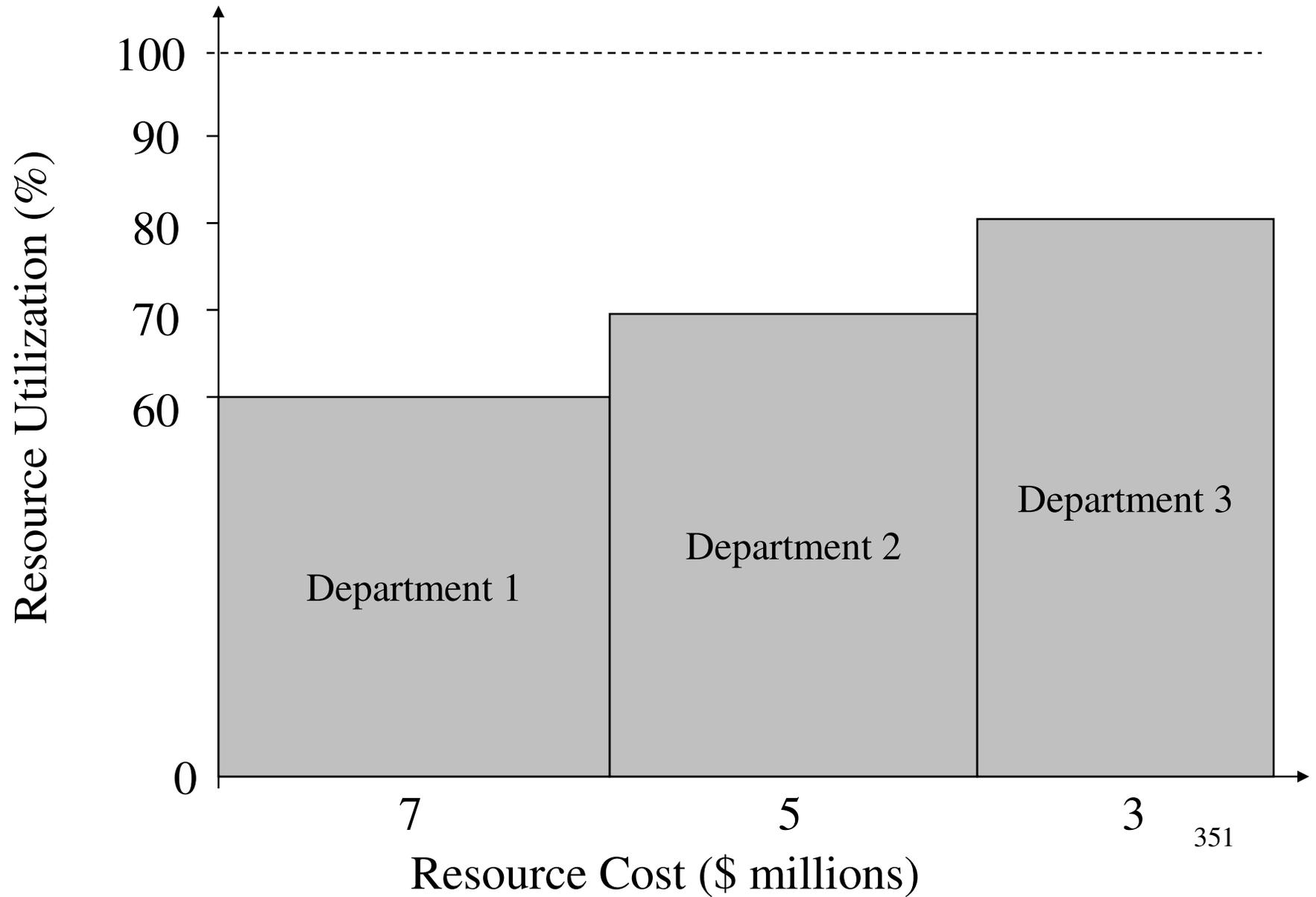
A CUT Diagram for Alternative A



A CUT Diagram for Alternative B



A CUT Diagram for Alternative C



Evolution of Fluctuations in a Process

- Each of the three alternatives is presented using a cost-utilization (CUT) diagram
- The CUT diagram presents the average utilization of the resources
- Which alternative is best for the organization?
- All seem to have excess capacity
- But, the CUT diagram presents only the average utilization of the resources
- This alone can lead to poor decisions

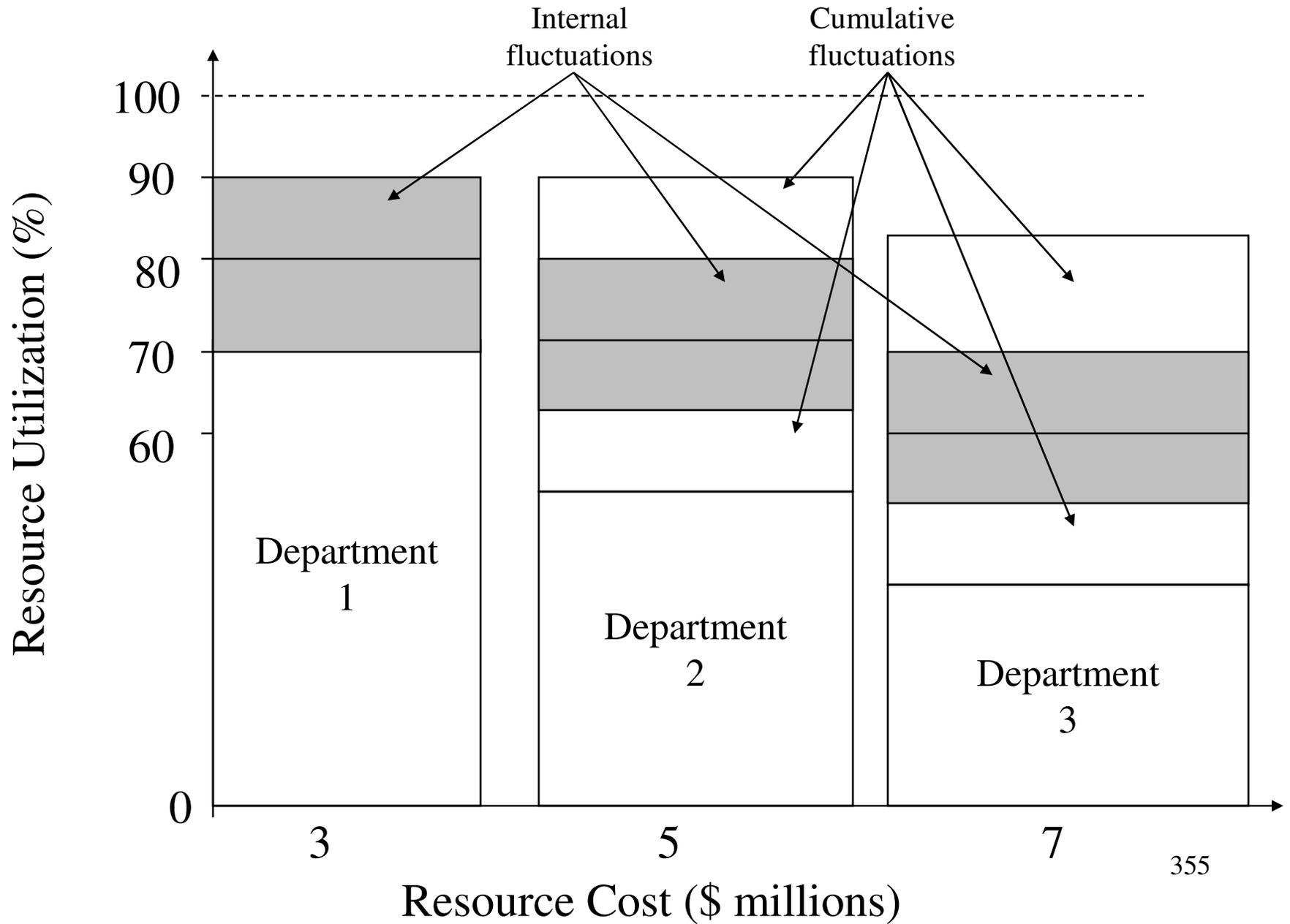
Which Alternative is Best?

- There are two types of fluctuations
 - Internal fluctuations
 - Cumulative fluctuations
- Internal fluctuations come from the station itself
 - A computer crash at a station
 - Worker absenteeism
 - Quality problems
- Cumulative fluctuations result from performance in preceding stages of the process

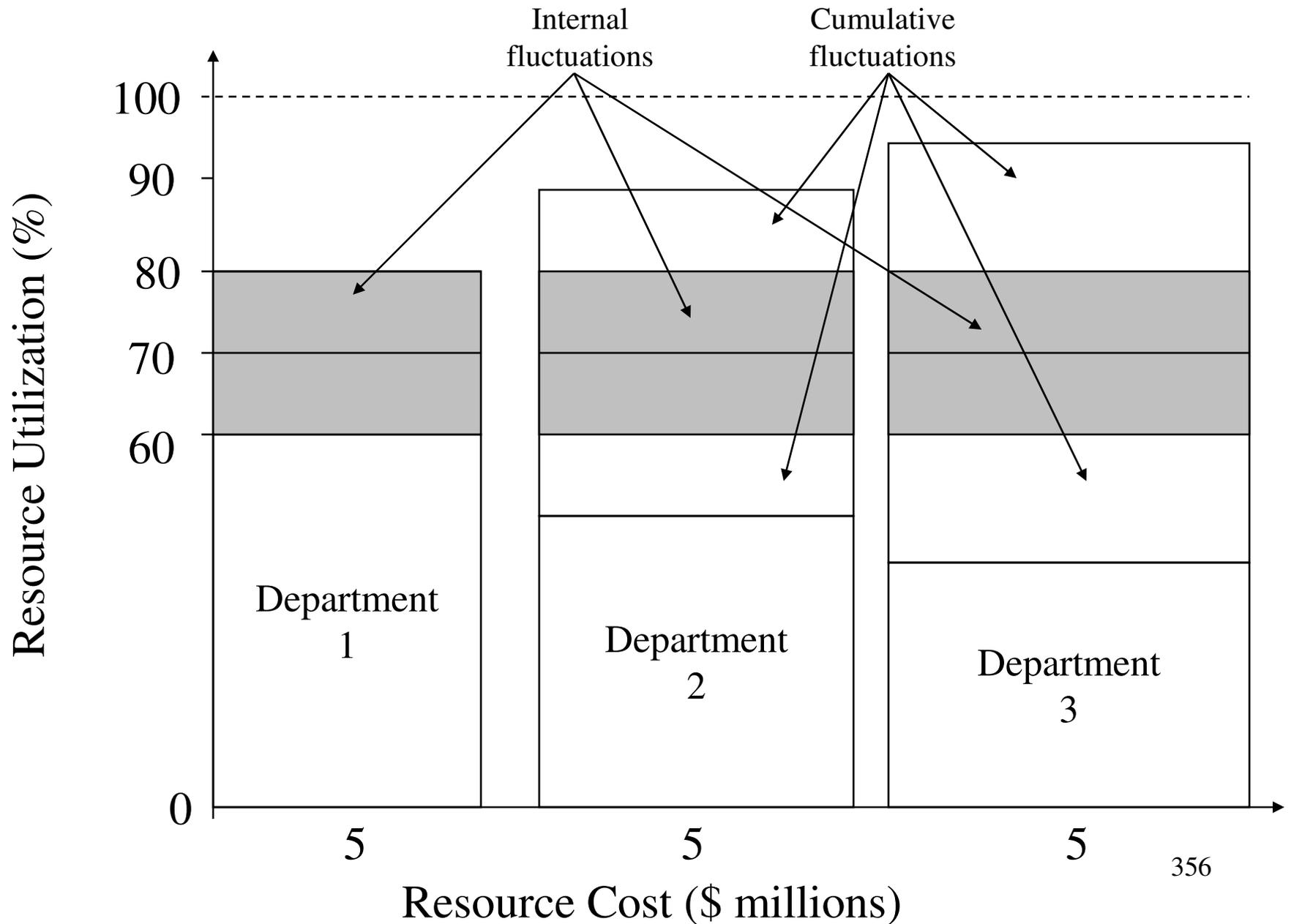
Cumulative Fluctuation

- Think of a group of hikers marching in a single file
- Those hikers at the end will occasionally have to run to keep up and sometimes have to slow down as a result of the cumulative fluctuations of the hikers ahead of them
- The standard deviation of the speed of a walker at the end is bigger than at the front of the line
- Let us now review the CUT diagrams of the three alternatives where the diagrams show the percent utilization on one of the busiest days of the year

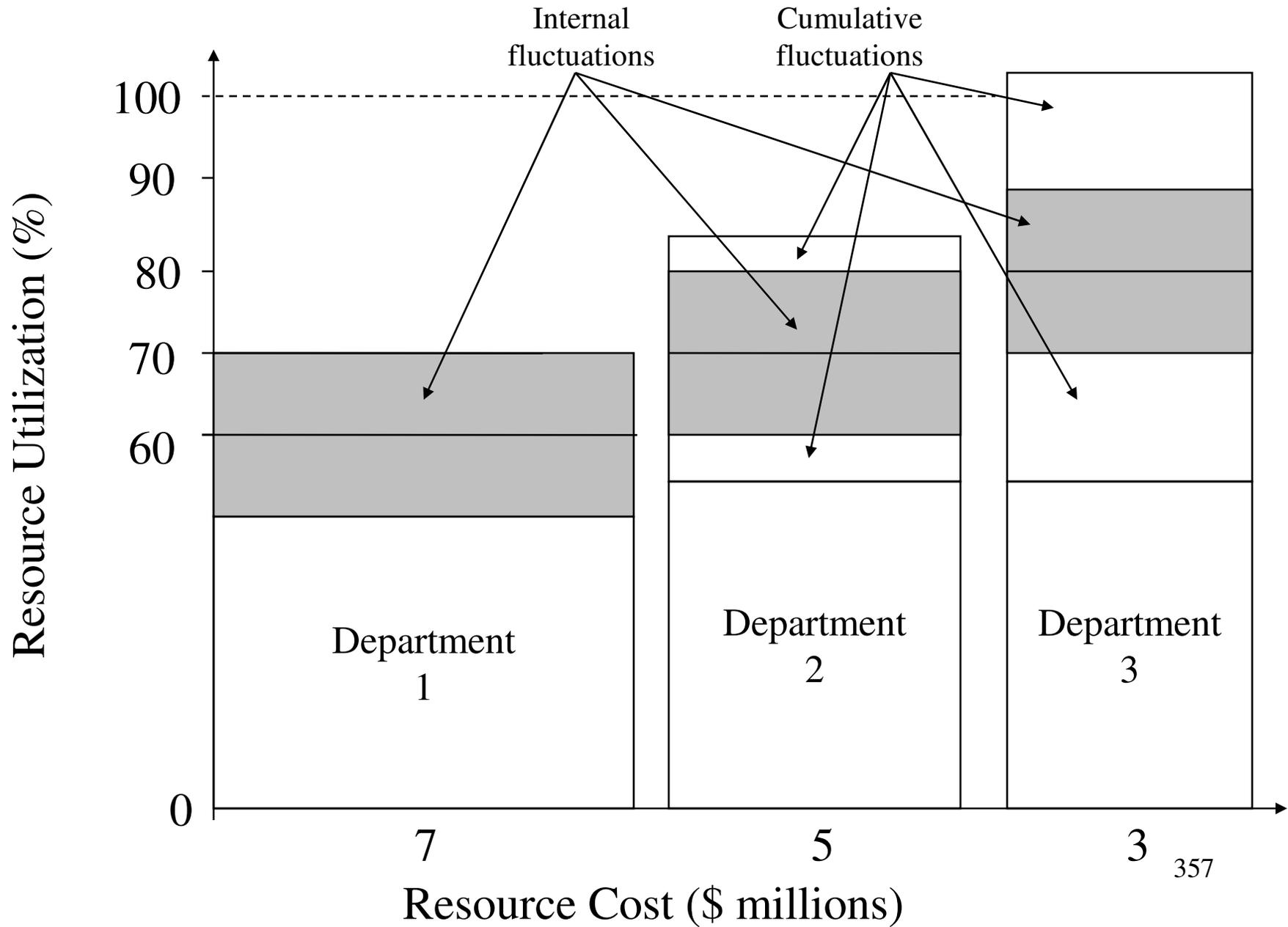
Alternative A with Internal and Cumulative Fluctuations



Alternative B with Internal and Cumulative Fluctuations



Alternative C with Internal and Cumulative Fluctuations



Which Alternative is Best?

- Looking at these three figures, it is clear that alternative A is the preferred choice
 - Even at peak utilization, there is no bottleneck
 - With alternative B, there is nearly a bottleneck
 - The cumulative fluctuations create, some of the time, a bottleneck at department 3 of alternative C
- Even though on the average a department has adequate capacity, when a “wave” of work arrives, it might become a bottleneck
 - We deal with this next

Elements of Capacity

- The maximal (theoretical) capacity of a resource can be broken down into three parts
 - Nominal capacity of the resource
 - Protective capacity of the resource
 - Excess capacity of the resource
- The nominal capacity of a resource is the average capacity utilized for performing its tasks
- The protective capacity of a resource is that which is targeted to overcome the internal and cumulative fluctuations in load

Elements of Capacity

- The protective capacity protects the system against routine fluctuations
- It is not intended to protect against rare, unexpected fluctuations (e.g., due to a hurricane)
- The excess capacity of the resource is that part of the maximal capacity of the resource that is not used
- There are lessons to be learned from the above

Lessons Regarding Capacity

- During the planning stage, it is not enough to consider the nominal capacity of a resource
 - E.g., in designing an ED, two separate EDs should be considered, a day ED and a night ED
- Where there are large differences between normal (off-peak) and peak times, the system can be planned to satisfy peak loads and to sell the excess capacity during off-peak times
 - E.g., a restaurant whose primary business is during evening hours can try to offer special *business lunch* menus

Lessons Regarding Capacity

- A system needs protective capacity
 - It assures achieving maximal output for the whole system
- Resources with a capacity utilization exceeding 85% are considered bottlenecks and must be managed accordingly
- While planning a service or production, we need to arrange sufficient protective capacity
 - An average utilization of about 70% is good

Traditional Approaches to Managing Fluctuations in Demand

- Traditional management solves the problem by carrying a large finished goods inventory
 - This is usually an expensive solution
- When there is a decline in sales or with a large inventory, special discounts are offered
 - The sales staff focuses on selling what they have as opposed to what the market demands

Traditional Approaches to Managing Capacity Fluctuations

- To overcome capacity fluctuations and concurrently meet market requirements for quick response, an organization builds up high levels of WIP inventory
- It is not uncommon to see organizations buy excess capacity to meet demand fluctuations
 - This can be expensive
- Expediting orders can solve local problems for some customers

Traditional Approaches to Managing Fluctuations in Quality

- Overproduction to protect against an unexpectedly high number of rejects
- Creation of repair teams and repair stations
- At the end of each process, one establishes a sorting station to sort high-quality products from poor-quality products
 - This is a wasteful approach
 - It does not prevent continued production of inferior products

Traditional Approaches to Managing Fluctuations in Availability of Materials and Parts

- Many organizations carry high levels of raw materials and components inventory
- In situations where a customer cannot depend on the quality of work of suppliers, there is a need to establish a testing center to sort and screen the incoming shipments
- The approaches of traditional management are usually based on responses to fluctuations that have already happened

The Focused Management Approach to Managing Fluctuations in the Health Care System

- Fluctuations must be managed and not responded to when they happen
 - In an effective manner
 - With minimal cost
- We discuss two ways of managing and dealing with fluctuations
 - Protecting against fluctuations
 - Reducing fluctuations

Protecting Against Fluctuations

- Building a buffer and managing it using the DBR approach
 - This helps to increase output and reduce response times
- Creating protective (or excess) capacity
 - This can be achieved via subcontractors or the purchase of additional resources
 - This approach is relatively expensive
- Building a buffer of finished goods inventory
 - This will cushion the demand fluctuations in the event of a market constraint

Protecting Against Fluctuations

- Building an overflow buffer after the bottleneck
 - This will allow the bottleneck to keep working when there are delays in stations that follow it
- Building a buffer of raw materials and parts
 - This protects against possible fluctuations in supply
- The problem is that the improvement from the above mechanisms, as big as it may be, is a one-time improvement

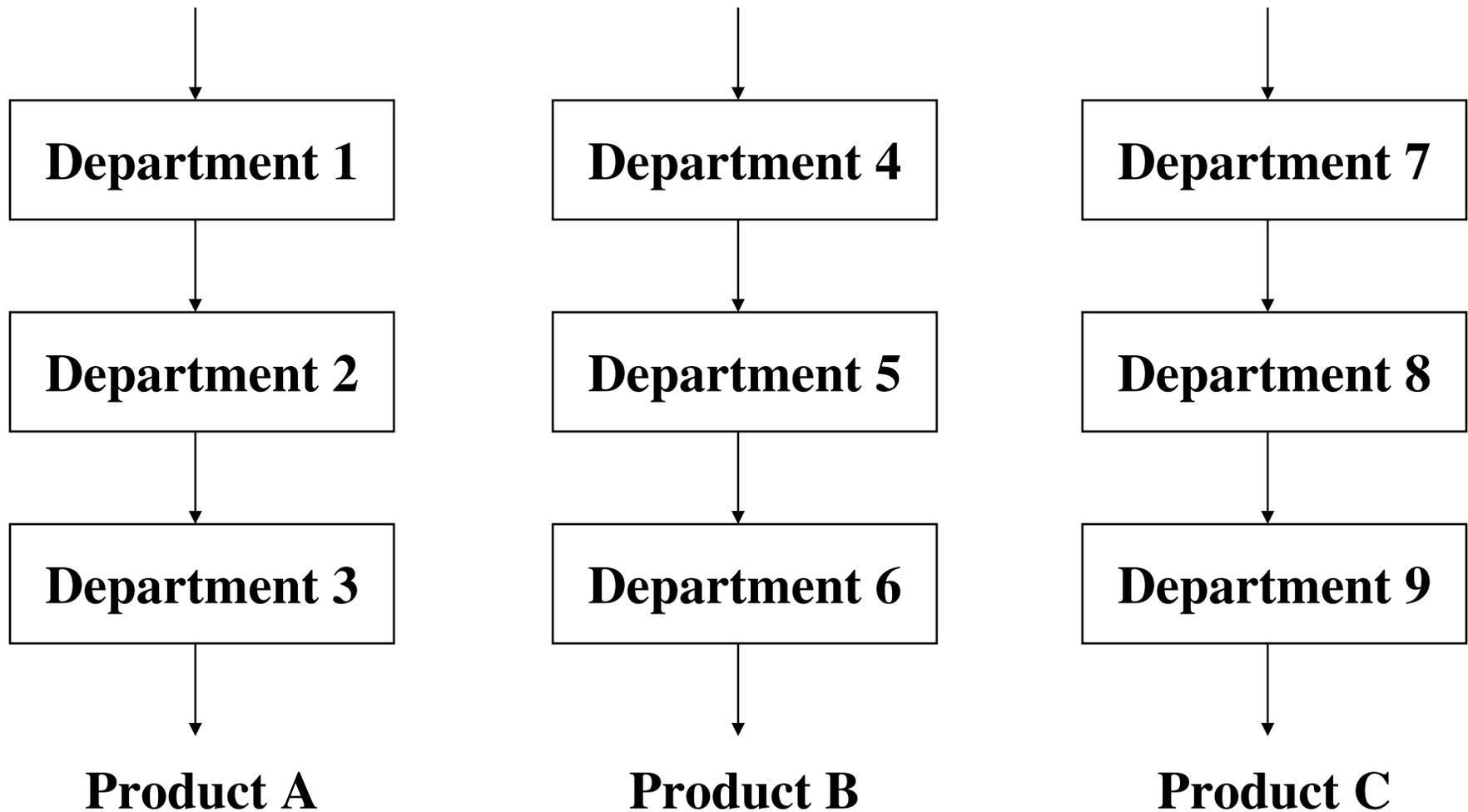
Reducing Fluctuations

- There are several mechanisms to reduce fluctuations
 - The reduction of fluctuations is usually a slow process
- Reducing response time
 - This allows for better prediction of demand ⇒ a reduction of fluctuations
- Sharing information with customers and suppliers
 - A customer providing his/her supplier with the demand forecast reduces the supplier's uncertainty and allows the supplier to obtain raw materials, plan personnel needs, etc.

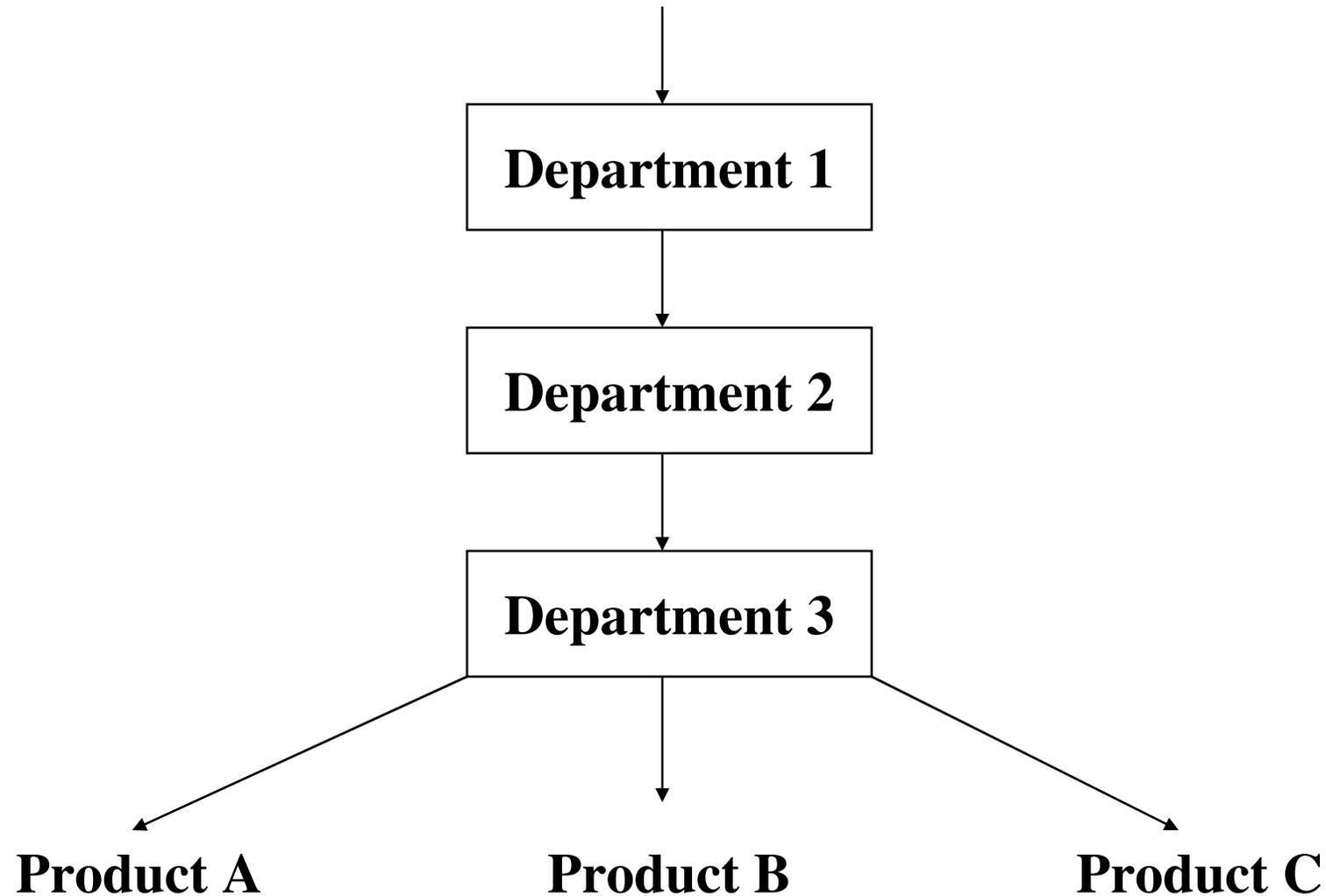
Reducing Fluctuations

- Creating a common core for several products or services (mushroom effect)
 - Producing or providing service for every product or service separately (see page 372) requires separate planning of resources (and raw materials) for every line
 - Creating a common core for several products (see page 373) allows the variation in the planning of resources and raw materials to be reduced

A Separate Line for Every Product



A Common Core for Several Products (Mushroom Effect)



Reducing Fluctuations

- Standardizing components and raw materials
- Uniting finished goods warehouses
- Using a common buffer for all project activities
- Reducing overspecification and overdesign
- Improving quality and process control
- Working with small working batches
- Bottom line: Fluctuations that are not properly managed may turn noncritical resources into bottlenecks