OPIM101 Decision Analysis
Lecture Notes

Liang (Leon) Xu
Singapore Management University

Contents

1 Introduction to Decision Analysis and Linear Programming **1**
- 1.1 Learning Objectives . . . 1
- 1.2 Why Study Decision Analysis? . . . 1
- 1.3 Uncertainty Makes Decisions Hard . . . 2
- 1.4 What Is a Model? . . . 2
- 1.5 The Three Elements Of A Decision Model . . . 2
- 1.6 Can LLMs Formulate LP Models For Us? . . . 3
- 1.7 A Product-Mix Example: Lego Furniture . . . 3
- 1.8 Formulating The Linear Program . . . 4
- 1.9 Graphical Solution: Constraints And Feasible Region . . . 4
- 1.10 Graphical Solution: Iso-Profit Lines . . . 5
- 1.11 Corner Point Method . . . 6
- 1.12 Sensitivity Analysis . . . 7
- 1.13 Chapter Summary . . . 11

2 Formulation Practice and Excel Solver **12**
- 2.1 From Situation To Formulation . . . 12
- 2.2 Gita's Wrapping-Paper Business . . . 12
  - 2.2.1 Decision Variables . . . 13
  - 2.2.2 Objective Function . . . 13
  - 2.2.3 Constraints . . . 13
- 2.3 Graphical Check . . . 14
- 2.4 Common Formulation Mistakes . . . 14
- 2.5 Minimization Problems . . . 15
- 2.6 Four Special Cases In Linear Programming . . . 15
  - 2.6.1 Infeasibility . . . 15
  - 2.6.2 Unboundedness . . . 15
  - 2.6.3 Alternative Optimal Solutions . . . 15
  - 2.6.4 Redundant Constraints . . . 16
- 2.7 Why We Need Excel Solver . . . 16
- 2.8 A Good Solver Spreadsheet . . . 16
  - 2.8.1 SUMPRODUCT . . . 16
  - 2.8.2 Solver Setup . . . 17
- 2.9 Excel Shortcuts For Modeling . . . 17
- 2.10 Takeaways . . . 17

3 Sensitivity, Summation, and Blending **18**
3.1 Why Sensitivity Analysis Matters . . . 18
3.2 Two Ways To Do Sensitivity Analysis . . . 18
3.2.1 Trial And Error . . . 19
3.2.2 Sensitivity Report . . . 19
3.3 Gita's Wrapping-Paper Sensitivity Report . . . 19
3.4 Objective Coefficient Sensitivity . . . 19
3.5 RHS Sensitivity And Shadow Price . . . 20
3.6 Allowable Ranges . . . 20
3.7 Reduced Cost . . . 21
3.8 Why We Need Summation Notation . . . 21
3.9 One-Dimensional Sums . . . 21
3.10 Two-Dimensional Tables And Double-Index Variables . . . 22
3.11 Mussel Oil Blending Problem . . . 22
3.11.1 Supply Constraints . . . 23
3.11.2 Demand Constraints . . . 23
3.11.3 Objective Function . . . 23
3.11.4 Octane Constraints . . . 23
3.11.5 Sulfur Constraints . . . 23
3.11.6 Nonnegativity . . . 24
3.12 Solved Blending Plan . . . 24
3.13 Week 3 Summary . . . 24

4 Assignment and Transportation **25**
4.1 Why Network Models Matter . . . 25
4.2 Part I: The Assignment Problem . . . 26
4.2.1 Business Idea . . . 26
4.2.2 MachineCo Example . . . 26
4.2.3 Why The Binary Restriction Is Special . . . 27
4.2.4 Unbalanced Assignment Problems . . . 27
4.3 Part II: The Transportation Problem . . . 28
4.3.1 Business Idea . . . 28
4.3.2 Airport Construction Example . . . 29
4.3.3 Balanced Transportation Model . . . 30
4.3.4 Unbalanced Transportation . . . 30
4.3.5 Prohibited Routes And Big M . . . 31
4.3.6 Assignment As A Special Transportation Problem . . . 31
4.4 Managerial Takeaways . . . 32

5 The Transshipment Problem **33**
5.1 From Direct Shipping to Transshipment . . . 33
5.1.1 Three node roles . . . 33
5.1.2 Why use hubs? . . . 34
5.2 Flair Wheat: An Allowed-Route Model . . . 34
5.2.1 Business setting and network balance . . . 34
5.2.2 Decision variables and objective . . . 35
5.2.3 Write one constraint for each node . . . 35
5.2.4 Interpret a feasible solution . . . 36

5.3 A Complete-Table Formulation . . . 36
5.3.1 Why keep every node-to-node cell? . . . 36
5.3.2 Row sums are outflow; column sums are inflow . . . 37
5.3.3 General formulations . . . 37
5.3.4 Spreadsheet implementation . . . 38
5.4 Carbon-Aware Transshipment . . . 38
5.4.1 Sustainability has three dimensions . . . 38
5.4.2 Convert route movement into emissions . . . 39
5.4.3 Add carbon cost to the objective . . . 39
5.5 Tiger Beer's Singapore Supply Network . . . 40
5.5.1 Classify the seven nodes . . . 40
5.5.2 Build the extended cost matrix . . . 40
5.5.3 Formulate and interpret the model . . . 41
5.6 Common Modeling Mistakes . . . 41
5.7 Managerial Takeaways . . . 42
5.8 Self-Check . . . 42

6 Integer Linear Programming 43
6.1 When Divisibility Fails . . . 43
6.2 Harrison Electric: An Integer Product-Mix Model . . . 43
6.3 LP Relaxation . . . 44
6.3.1 Definition . . . 44
6.3.2 Harrison's relaxation and integer optimum . . . 44
6.3.3 Why simple rounding is unreliable . . . 45
6.4 Branch-and-Bound . . . 45
6.4.1 Harrison branch-and-bound calculation . . . 45
6.5 Binary Variables and Project Selection . . . 46
6.5.1 The knapsack structure . . . 46
6.5.2 Eight-item example . . . 46
6.6 Logical Constraints with Binary Variables . . . 47
6.6.1 At most, at least, and exactly . . . 47
6.6.2 Mutual exclusion . . . 47
6.6.3 Same choice . . . 47
6.6.4 Implication . . . 47
6.6.5 Two antecedents imply a third . . . 48
6.6.6 Either–or . . . 48
6.7 Panther Manufacturing: Fixed Costs and Conditional Production . . . 48
6.7.1 Fixed setup cost . . . 48
6.7.2 Deriving tight values of M_i . . . 49
6.7.3 Conditional minimum: zero or at least 1,000 . . . 49
6.8 Solver Workflow and Model Audit . . . 50
6.9 Common Modeling Errors . . . 50
6.10 Chapter Summary . . . 50
6.11 Self-Check Questions . . . 50

7 Probability, Expected Value, and Bayes' Theorem 52
7.1 Chapter overview . . . 52
7.2 Why probability matters . . . 52
7.2.1 From uncertain possibilities to measurable risk . . . 52
7.2.2 A motivating application: life insurance . . . 53
7.2.3 A good decision can still produce a bad outcome . . . 53
7.3 Experiments, outcomes, sample spaces, and events . . . 53
7.3.1 Random experiments . . . 53
7.3.2 Events . . . 53
7.4 Where probabilities come from . . . 54
7.4.1 Equally likely outcomes . . . 54
7.4.2 The frequency approach . . . 54
7.4.3 Subjective probability . . . 54
7.5 Operations on events . . . 54
7.5.1 Mutually exclusive and collectively exhaustive events . . . 55
7.5.2 The addition rule . . . 55
7.6 Conditional probability . . . 55
7.6.1 Information changes the reference group . . . 55
7.6.2 The multiplication rule . . . 56
7.7 Independence, dependence, and mutual exclusion . . . 56
7.8 Joint, marginal, and conditional probabilities . . . 56
7.9 Sequential experiments and probability trees . . . 57
7.9.1 Sampling with and without replacement . . . 57
7.9.2 Practice: color and label . . . 57
7.9.3 Does drawing first create an advantage? . . . 57
7.10 The birthday puzzle . . . 57
7.11 Prediction markets and expected profit . . . 58
7.12 Bayes' theorem . . . 58
7.12.1 Prior, evidence, and posterior . . . 58
7.12.2 Reading Bayes through a probability tree . . . 59
7.13 Case study: interpreting an HIV screening result . . . 59
7.13.1 A second independent test . . . 59
7.14 Bayes as a discipline of continuous learning . . . 60
7.14.1 When a prior becomes doctrine . . . 60
7.15 Strategic evidence: the Monty Hall problem . . . 60
7.16 Random variables and discrete probability distributions . . . 61
7.16.1 Mean, variance, and standard deviation . . . 61
7.17 Worked comparison: two investments . . . 61
7.18 Common mistakes . . . 62
7.19 Practice questions . . . 62
7.20 Chapter summary . . . 63

8 Decision Making under Risk and Uncertainty 64
8.1 Chapter overview . . . 64
8.2 Describe the decision before choosing . . . 64
8.2.1 Three decision environments . . . 64
8.2.2 Running example: Stephen Marts's product idea . . . 65
8.3 When probabilities are known: expected monetary value . . . 65

8.3.1 Sensitivity to the probability . . . 66
8.4 Risk preference, utility, and certainty equivalents . . . 66
8.4.1 Why a concave utility curve represents risk aversion . . . 66
8.4.2 Constant absolute risk aversion (CARA) . . . 67
8.5 When probabilities are unknown: five decision rules . . . 67
8.5.1 Maximax and maximin . . . 67
8.5.2 Criterion of realism and equally likely . . . 68
8.5.3 Minimax regret . . . 68
8.6 Thompson Lumber: expected value and sensitivity . . . 68
8.7 The newsvendor: construct the payoff, then choose . . . 69
8.8 What is perfect information worth? . . . 70
8.9 Common mistakes and practical checks . . . 70
8.10 Practice questions . . . 71
8.11 Chapter summary . . . 71

9 Decision Trees and the Value of Information 72
9.1 Chapter overview . . . 72
9.2 Structure a sequential decision . . . 72
9.2.1 Bill Sampra's job offers . . . 73
9.3 Solve a tree by rolling it backward . . . 73
9.3.1 Sensitivity and a reservation value . . . 74
9.4 When the future self may change the plan . . . 74
9.5 Thompson Lumber: a decision before the market is known . . . 75
9.6 A fallible market survey and Bayesian updating . . . 75
9.6.1 From likelihoods to signal probabilities . . . 75
9.6.2 Posterior probabilities after each signal . . . 76
9.7 Roll back the imperfect-information tree . . . 76
9.7.1 Why carrying exact probabilities gives a slightly different value . . . 76
9.8 The value of imperfect and perfect information . . . 77
9.8.1 Value of the fallible survey . . . 77
9.8.2 Perfect information as an upper bound . . . 77
9.9 Common mistakes . . . 78
9.10 Practice questions . . . 78
9.11 Chapter summary . . . 79

10 Simulation for Decisions under Uncertainty 80
10.1 Chapter overview . . . 80
10.2 Build a simulation around a decision . . . 81
10.3 Turn random numbers into business outcomes . . . 81
10.3.1 A discrete weekly-sales example . . . 81
10.3.2 Use RAND and a cumulative lookup in Excel . . . 82
10.4 Replicate, average, and interpret . . . 82
10.5 Insurance premiums: the Scottish Ministers' Widows' Fund . . . 83
10.5.1 Model inputs and an annual death count . . . 83
10.5.2 Claims, premium income, and the annual result . . . 83
10.5.3 Input quality is part of the decision . . . 84
10.6 Maria's Bakery: convert demand into profit . . . 84
10.6.1 A discount market for leftovers . . . 85

10.7 The newsvendor: compare order quantities . . . 85
10.8 A reliable spreadsheet workflow . . . 86
10.9 Common mistakes . . . 86
10.10 Practice questions . . . 87
10.11 Chapter summary . . . 88

11 The Psychology of Human Decisions **89**
11.1 Chapter overview . . . 89
11.2 Search for a workable choice . . . 89
11.2.1 Satisficing and bounded rationality . . . 90
11.2.2 Rational inattention: decide what to learn . . . 90
11.3 Fast responses, deliberate checks, and emotion . . . 90
11.4 Prospect theory: gains, losses, and chances . . . 91
11.4.1 Reference dependence and loss aversion . . . 91
11.4.2 Probability weighting . . . 92
11.5 Framing: equivalent facts, different focus . . . 93
11.6 Sunk costs and decisions from today forward . . . 94
11.7 Anchors, comparisons, and decoys in pricing . . . 94
11.8 Choice architecture, nudges, and friction . . . 95
11.9 Apply the ideas: redesign a study-app offer . . . 96
11.9.1 Test business and customer outcomes . . . 96
11.10 Use the ideas for your own decisions . . . 96
11.11 Common mistakes . . . 97
11.12 Practice questions . . . 97
11.13 Chapter summary . . . 98

Chapter 1

Introduction to Decision Analysis and Linear Programming

Source week: Week 1

1.1 Learning Objectives

By the end of this chapter, you should be able to:

1. Explain what decision analysis is and why uncertainty makes decisions difficult.
2. Describe the role of a model in structured decision making.
3. Formulate a simple linear programming model with decision variables, an objective function, constraints, and sign restrictions.
4. Interpret constraints graphically as boundary lines and feasible half-planes.
5. Identify a feasible region for a two-variable LP.
6. Use iso-profit lines and corner points to solve a small LP graphically.
7. Explain the basic idea of sensitivity analysis.

1.2 Why Study Decision Analysis?

Decision making is everywhere. A student decides which courses to register for. A commuter chooses between bus, train, taxi, and walking. A company decides how much inventory to hold, how to route deliveries, or how to price limited seats on a flight.

The swimming-pool rescue exercise gives a useful first intuition. Suppose someone is drowning and you must decide where to run along the pool edge and where to enter the water. The straight-line path may not be the fastest, because running and swimming have different speeds. The best decision is the path that minimizes total rescue time.

This example is related to the familiar statement that light follows a shortest-time path when it bends between materials. Light is not "in a hurry" and it is not making a conscious decision. Rather, the physical world is governed by laws that produce this path. In decision analysis, we are in the role of the decision maker: we can define possible paths, calculate their consequences, and choose the one that best meets the objective. This is the same modeling habit we will use

throughout the course: define the choice, define the objective, identify the constraints, and search systematically for the best feasible action.

Decision analysis is the application of scientific thinking to managerial and personal decisions. It uses models, data, probability, statistics, and optimization to make decision logic explicit.

The point is not to replace human judgment. The point is to improve it. A good decision model helps us see:

- what we can control,
- what we cannot control,
- what objective we are trying to improve,
- what trade-offs are unavoidable, and
- how sensitive the recommended decision is to uncertain assumptions.

1.3 Uncertainty Makes Decisions Hard

If we had a perfect crystal ball, many decisions would be easy. We could simply choose the alternative that leads to the best future outcome. In reality, future demand, costs, prices, travel time, customer behavior, and competitor actions are uncertain.

A tragic real-world example is the helicopter crash on January 26, 2020, involving Kobe Bryant, his daughter, and other passengers. In aviation, a pilot may need to decide whether to continue, turn back, or request a different clearance while weather and visibility are changing. The key challenge is that the decision must be made before the outcome is known. Decision analysis gives us language and tools for thinking about such situations: alternatives, information, uncertainty, consequences, and risk.

This course studies decisions both with and without uncertainty. Linear programming begins with decisions under certainty: the relevant numbers are treated as known. Later topics, such as probability, decision trees, value of information, and simulation, introduce uncertainty directly.

1.4 What Is a Model?

A **model** is a simplified representation of a real problem. It is not a perfect copy of reality. A useful model is realistic enough to answer the decision question, simple enough to solve, and clear enough to explain.

For business decisions, it is helpful to think in the sequence:

$$\text{Inputs and data} \longrightarrow \text{Model} \longrightarrow \text{Decision}$$

The model translates a messy real-world situation into a structured problem. The decision maker then interprets the model output and decides whether the recommendation makes practical sense.

1.5 The Three Elements Of A Decision Model

Most optimization models contain three basic elements.

Objective. What are we trying to maximize or minimize? Examples include profit, cost, waiting time, risk, or service level.

Decision variables. What can we choose or change? Examples include production quantities, shipment quantities, staffing levels, and investment amounts.

Constraints. What limits or requirements must the decision satisfy? Examples include limited labor, limited raw material, budget limits, demand requirements, and capacity restrictions.

1.6 Can LLMs Formulate LP Models For Us?

Large language models can help explain a problem, draft a formulation, and write solver code. But this does not remove the need to learn formulation. If you do not understand decision variables, objectives, constraints, and assumptions, you cannot reliably prompt the model, check the answer, or detect missing constraints.

This is especially important in optimization modeling. A model can look polished and still leave out a real business requirement. One paper we discuss in this course reports common optimization-modeling errors in large-model outputs. The largest category is low model completeness: the model understands part of the story but leaves out constraints or implicit conditions. This is exactly why students still need modeling judgment.

Modeling error type	Description	Proportion
Semantic misunderstanding	Misunderstands the problem to be solved.	13.40%
Objective/constraint translation error	Understands the requirement but formulates it incorrectly.	30.30%
Low model completeness	Ignores constraints or implicit conditions, oversimplifying the real problem.	56.30%

1.7 A Product-Mix Example: Lego Furniture

Suppose a company produces two products: tables and chairs. The company has limited labor, small blocks, and large blocks. The operations manager must decide how many tables and chairs to make.

Product	Labor hours	Large blocks	Small blocks
Table	3	2	2
Chair	2	1	2

The available resources are:

- 11 labor hours,
- 6 large blocks,
- 8 small blocks.

A table sells for 57. A chair sells for 37. A small block costs 5 and a large block costs 8. The profit margins are therefore:

$$\text{Table margin} = 57 - 2(8) - 2(5) = 31,$$

$$\text{Chair margin} = 37 - 1(8) - 2(5) = 19.$$

1.8 Formulating The Linear Program

Let:

$$T = \text{number of tables to produce}, \qquad C = \text{number of chairs to produce}.$$

The objective is to maximize total profit:

$$\max \ 31T + 19C.$$

The constraints come from the limited resources:

$$3T + 2C \leq 11 \qquad \text{labor},$$

$$2T + 2C \leq 8 \qquad \text{small blocks},$$

$$2T + C \leq 6 \qquad \text{large blocks}.$$

Finally, we cannot produce negative tables or chairs:

$$T \geq 0, \qquad C \geq 0.$$

Putting everything together:

$$\begin{aligned} \max \quad & 31T + 19C \\ \text{s.t.} \quad & 3T + 2C \leq 11 \\ & 2T + 2C \leq 8 \\ & 2T + C \leq 6 \\ & T, C \geq 0. \end{aligned}$$

1.9 Graphical Solution: Constraints And Feasible Region

With two decision variables, we can draw the LP. Each constraint creates a boundary line. For example, the labor constraint is:

$$3T + 2C \leq 11.$$

To draw the boundary, replace the inequality with equality:

$$3T + 2C = 11.$$

Two easy intercepts are:

$$T = 0 \Rightarrow C = 5.5, \qquad C = 0 \Rightarrow T = \frac{11}{3}.$$

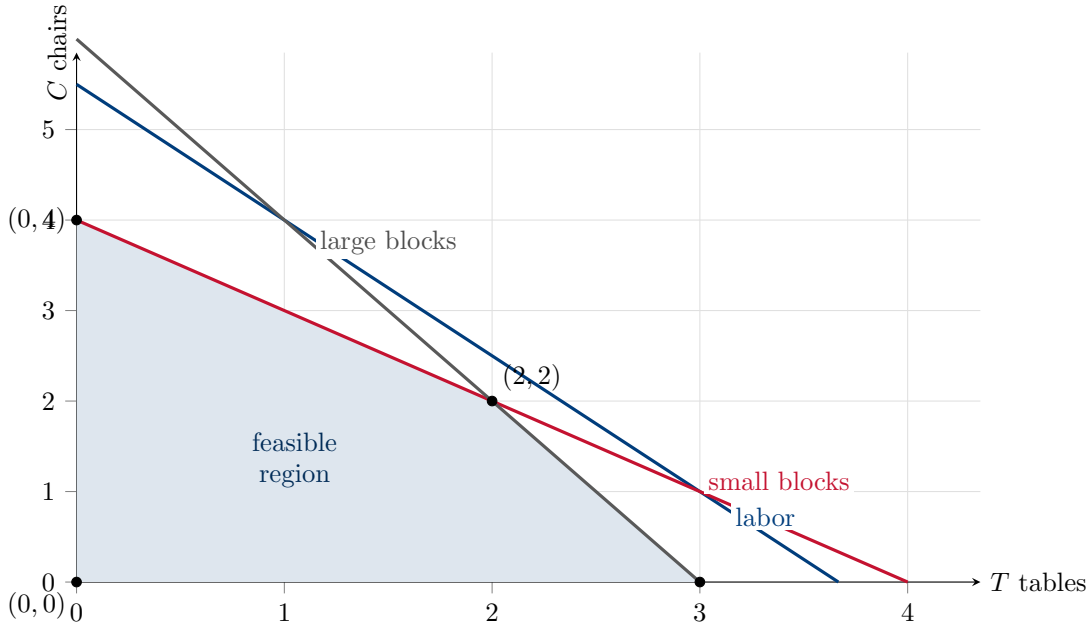
After drawing the line, test a point to decide which side is feasible. The origin is usually easiest:

$$3(0) + 2(0) = 0 \leq 11.$$

So the side containing the origin satisfies the labor constraint.
The same logic works for every linear constraint. To draw a constraint carefully:

1. Put the decision variables on the axes. In this example, put tables T on the horizontal axis and chairs C on the vertical axis.
2. Replace the inequality with equality. This gives the boundary line.
3. Find two points on the boundary line. Intercepts are usually easiest: set $T = 0$ to find the C-intercept, and set $C = 0$ to find the T-intercept.
4. Draw the boundary line through those two points.
5. Pick a test point, usually $(0, 0)$ if it is not on the boundary.
6. Substitute the test point into the original inequality. If it satisfies the inequality, shade the side containing the test point. If it does not, shade the other side.
7. Repeat for every constraint. The feasible region is the overlap of all shaded sides, including the nonnegativity restrictions.

The **feasible region** is the set of all points that satisfy every constraint at the same time.



1.10 Graphical Solution: Iso-Profit Lines

The objective function is:

$$31T + 19C.$$

An **iso-profit line** shows all combinations of tables and chairs that generate the same profit. For example:

$$31T + 19C = 31$$

and

$$31T + 19C = 62$$

are two parallel iso-profit lines. They have the same slope because the coefficients of the objective function have not changed.

To draw an iso-profit line, choose a profit level P and rewrite the objective as:

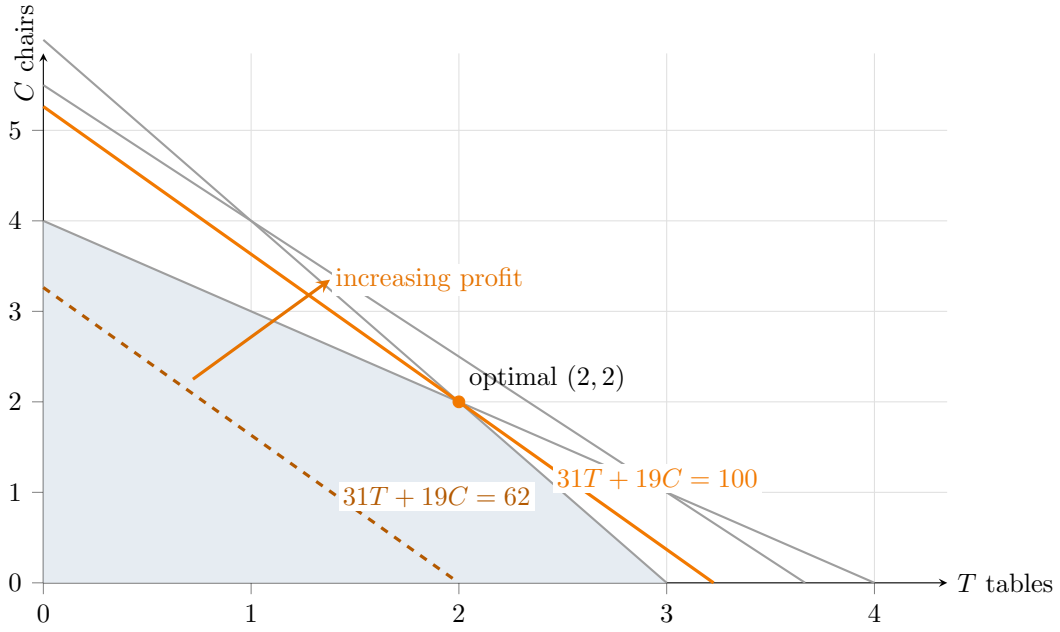
$$31T + 19C = P.$$

Solving for C gives:

$$C = \frac{P}{19} - \frac{31}{19}T.$$

This tells us two important things. First, the slope is $-31/19$, so every iso-profit line has the same slope. Second, changing P changes only the intercept, so higher-profit lines are parallel shifts of lower-profit lines.

Because this is a maximization problem, we slide the iso-profit line in the direction of increasing profit while keeping the same slope. One way to see the direction is to compare two lines: $31T + 19C = 62$ is better than $31T + 19C = 31$, and $31T + 19C = 100$ is better than both. Graphically, higher profit moves the line up and to the right.



The optimal point is the last point where the iso-profit line still touches the feasible region. In this problem, that point is $(T, C) = (2, 2)$. An objective line is not a constraint; we do not use it to define feasibility. We use it to rank feasible points.

1.11 Corner Point Method

A key LP theorem says: if an LP has an optimal solution, at least one optimal solution occurs at a corner point of the feasible region.

For the Lego furniture problem, the corner points are:

Corner	(T, C)	Profit $31T + 19C$
1	$(0, 0)$	0
2	$(3, 0)$	93
3	$(2, 2)$	100
4	$(0, 4)$	76

The best corner is:

$$T = 2, \qquad C = 2,$$

with profit:

$$31(2) + 19(2) = 100.$$

So the optimal production plan is to make 2 tables and 2 chairs.

1.12 Sensitivity Analysis

Sensitivity analysis asks what happens when the model inputs change. In a business setting, this is often more useful than simply knowing the optimal answer for one fixed set of numbers. Managers usually want to know whether the recommendation is stable, which resources are scarce, and which assumptions deserve attention.

In the Lego furniture problem, we can study three changes:

- What if we obtain two more small blocks?
- What if the profit margin for chairs increases?
- What if more labor is available?

Example 1: Increasing A Binding Resource

Suppose the number of small blocks increases from 8 to 10. The small-block constraint changes from:

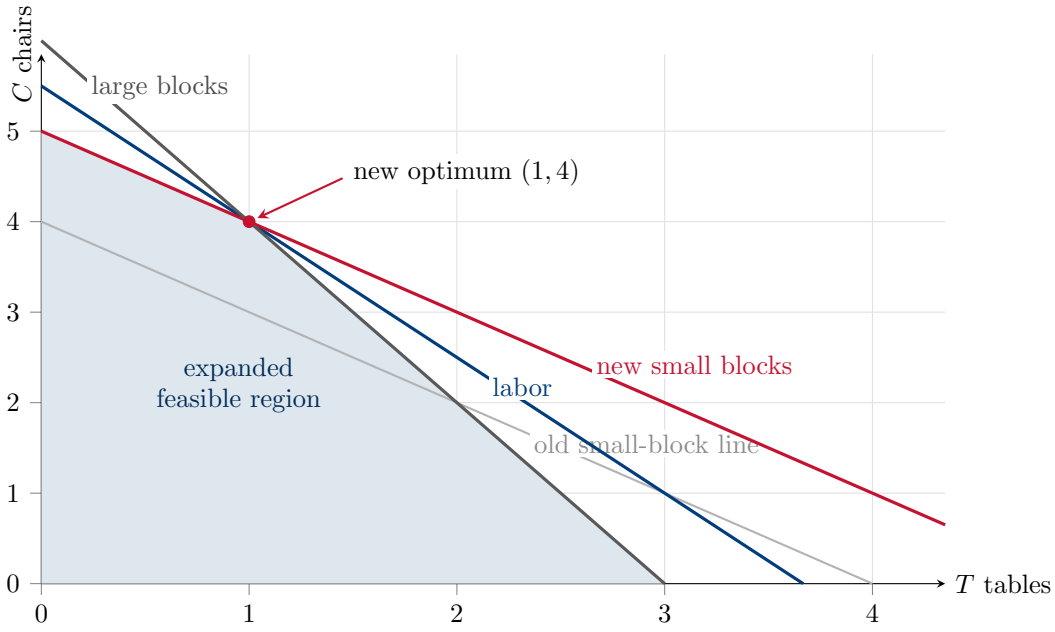
$$2T + 2C \leq 8$$

to:

$$2T + 2C \leq 10.$$

Only the right-hand side changes. Graphically, the small-block constraint shifts outward in a parallel way. The slope of the line does not change because the coefficients of T and C have not changed. The new LP is:

$$\begin{aligned} \max \quad & 31T + 19C \\ \text{s.t.} \quad & 3T + 2C \leq 11 \\ & 2T + 2C \leq 10 \\ & 2T + C \leq 6 \\ & T, C \geq 0. \end{aligned}$$



The new relevant corner points are $(0,0)$, $(3,0)$, $(1,4)$, and $(0,5)$. Their profits are:

Corner	(T, C)	Profit $31T + 19C$
1	$(0,0)$	0
2	$(3,0)$	93
3	$(1,4)$	107
4	$(0,5)$	95

The new optimal solution is:

$$T = 1, \qquad C = 4,$$

with profit:

$$31(1) + 19(4) = 107.$$

The original profit was 100, so two additional small blocks increase profit by 7. This gives a managerial interpretation: the most we should be willing to pay for two additional small blocks is 7 in total, or 3.50 per small block, if all other assumptions remain the same. If a supplier charges 7 per block, the extra blocks are not attractive because the total extra cost would be 14, which is larger than the extra profit of 7.

Example 2: Changing The Objective Coefficients

Now suppose the chair profit margin increases from 19 to 42. The constraints do not change, so the feasible region stays exactly the same. The objective becomes:

$$\max \; 31T + 42C.$$

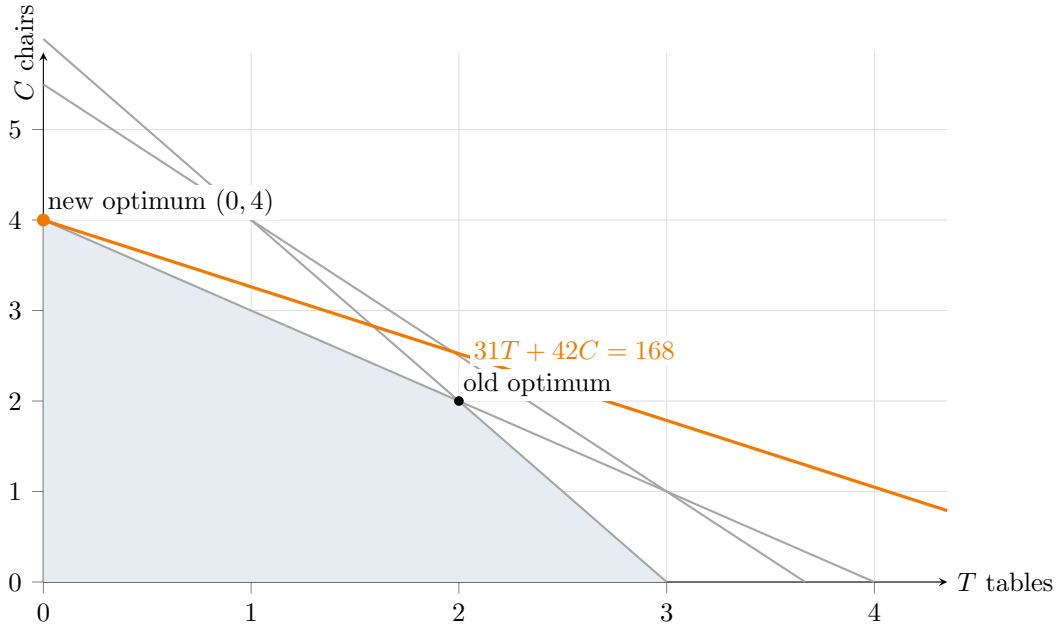
Changing the objective coefficients changes the slope of the iso-profit lines. Previously, the slope was:

$$-\frac{31}{19}.$$

Now the slope is:

$$-\frac{31}{42}.$$

The iso-profit line becomes flatter because chairs are now much more profitable relative to tables.



Evaluate the original corner points under the new objective:

Corner	(T, C)	New profit $31T + 42C$
1	$(0, 0)$	0
2	$(3, 0)$	93
3	$(2, 2)$	146
4	$(0, 4)$	168

The new optimal solution is:

$$T = 0, \qquad C = 4.$$

The key idea is that changing the objective coefficients does not change which points are feasible. It changes how we rank the feasible points. A production plan that was not best before can become best when the economic value of a product changes.

Example 3: Increasing A Non-Binding Resource

Finally, suppose labor availability increases from 11 to 12 hours. The labor constraint changes from:

$$3T + 2C \leq 11$$

to:

$$3T + 2C \leq 12.$$

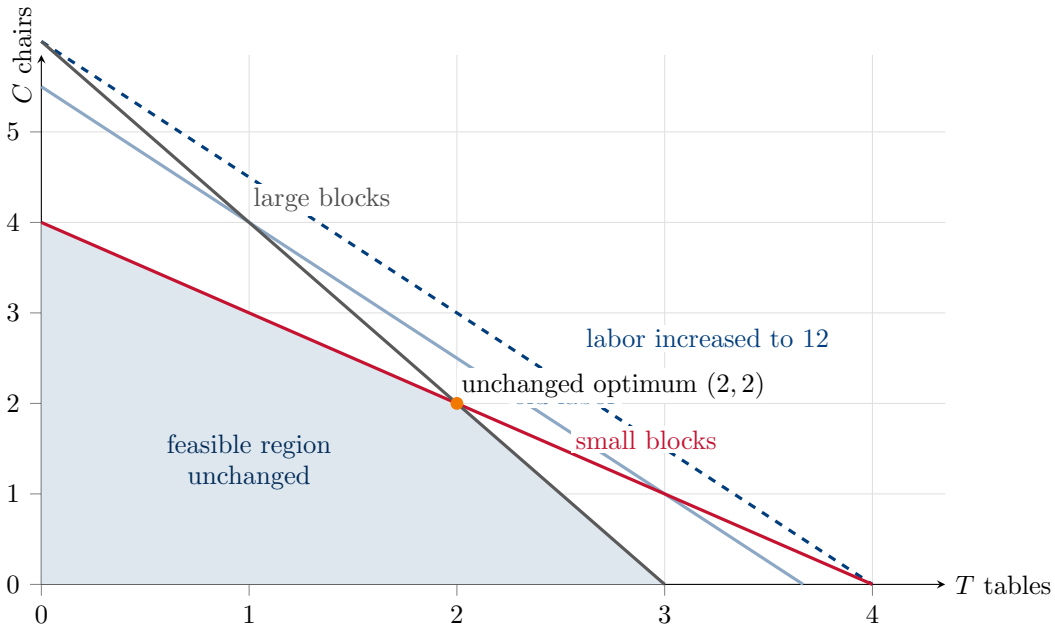
At the original optimal solution, $(T, C) = (2, 2)$, labor used was:

$$3(2) + 2(2) = 10.$$

So the original labor constraint already had slack:

$$11 - 10 = 1.$$

If labor capacity increases to 12, the labor line moves outward, but the feasible region does not expand in a useful way because the optimum is limited by other constraints.



The optimal solution remains:

$$T = 2, \qquad C = 2,$$

and the profit remains:

$$31(2) + 19(2) = 100.$$

The labor utilization before the change is:

$$\frac{\text{labor used}}{\text{labor available}} = \frac{10}{11} = 90.9\%.$$

After increasing available labor to 12, labor used is still 10, so utilization becomes:

$$\frac{10}{12} = 83.3\%.$$

This is an important managerial lesson. Increasing a non-binding resource does not necessarily improve output or profit. It may simply lower utilization and create more idle capacity.

Bottlenecks And Managerial Interpretation

A **bottleneck** is the resource or constraint that limits performance. In an LP model, a binding constraint is often a bottleneck because it is tight at the optimal solution. A non-binding constraint is not currently limiting the decision.

For the original Lego furniture solution $(2, 2)$:

- Small blocks are fully used: $2(2) + 2(2) = 8$ out of 8.
- Large blocks are fully used: $2(2) + 2 = 6$ out of 6.
- Labor is not fully used: $3(2) + 2(2) = 10$ out of 11.

Therefore, labor is not the bottleneck in the original model. Adding more labor does not help unless another part of the system also changes. The bottlenecks are the resources that prevent us from moving to a better feasible solution. Sensitivity analysis helps managers avoid spending money on resources that are not actually constraining performance.

1.13 Chapter Summary

In this chapter, we introduced decision analysis as a structured way to improve decision making. We then built our first linear programming model. The Lego furniture example showed how to define decision variables, write an objective function, express constraints, draw a feasible region, and solve an LP using iso-profit lines and corner points.

Check Your Understanding

1. What is the difference between a decision variable and a constraint?
2. Why does the feasible region contain only solutions that satisfy all constraints?
3. In the Lego furniture problem, why is $(2, 2)$ better than $(3, 0)$ even though both are feasible?
4. What does it mean for a constraint to have slack?
5. Why do iso-profit lines have the same slope when only the profit target changes?

Chapter 2

Formulation Practice and Excel Solver

Source week: Week 2

Week 1 introduced the logic of linear programming through a two-variable graphical model. Week 2 has a different emphasis. Before a solver can give an answer, someone must translate a messy business situation into a mathematical model. That translation is often the hardest and most valuable part of decision analysis.

This week has three goals:

1. practice linear-programming formulation with paper and pencil;
2. understand four special cases that can occur in LP models;
3. implement and solve an LP model in Excel Solver.

2.1 From Situation To Formulation

A linear-programming formulation always answers three questions:

1. What are the **decision variables**?
2. What is the **objective function**?
3. What are the **constraints**?

The order matters. Do not start with equations. Start with the managerial decision. Equations should come after the story is understood.

2.2 Gita's Wrapping-Paper Business

Gita has started a gift-wrapping paper printing business. She has limited resources:

- 120 meters of paper,
- 480 units of red paint,
- 300 units of blue paint,
- 300 units of yellow paint.

She can produce two designs. Design 1 sells for \$20 per meter and uses 3 units of red paint and 3 units of yellow paint per meter. Design 2 sells for \$25 per meter and uses 3 units of red paint, 3

units of blue paint, and 2 units of yellow paint per meter. Each design uses 1 meter of paper per meter produced.

2.2.1 Decision Variables

Let

$$D_1 = \text{meters of Design 1 wrapping paper to produce}$$

and

$$D_2 = \text{meters of Design 2 wrapping paper to produce.}$$

The units are important. If $D_1 = 20$, that means Gita produces 20 meters of Design 1. It does not mean 20 rolls, 20 dollars, or 20 units of paint.

2.2.2 Objective Function

Gita wants to maximize revenue. The revenue per meter is \$20 for Design 1 and \$25 for Design 2, so

$$\text{Revenue} = 20D_1 + 25D_2.$$

The objective is

$$\max 20D_1 + 25D_2.$$

2.2.3 Constraints

Each constraint converts a resource limit into an inequality.

Resource	Design 1 use	Design 2 use	Available
Paper	1	1	120
Red paint	3	3	480
Blue paint	0	3	300
Yellow paint	3	2	300

The constraints are:

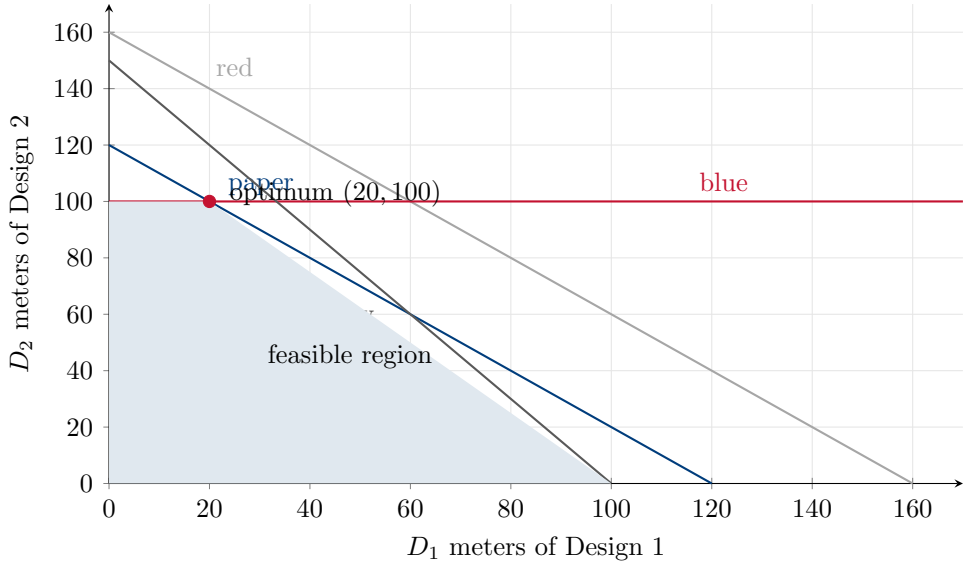
$$\begin{aligned} D_1 + D_2 &\le 120 && \text{paper,} \\ 3D_1 + 3D_2 &\le 480 && \text{red paint,} \\ 3D_2 &\le 300 && \text{blue paint,} \\ 3D_1 + 2D_2 &\le 300 && \text{yellow paint,} \\ D_1, D_2 &\ge 0 && \text{nonnegativity.} \end{aligned}$$

Putting everything together:

$$\begin{aligned} \max \quad & 20D_1 + 25D_2 \\ \text{s.t.} \quad & D_1 + D_2 \le 120 \\ & 3D_1 + 3D_2 \le 480 \\ & 3D_2 \le 300 \\ & 3D_1 + 2D_2 \le 300 \\ & D_1, D_2 \ge 0. \end{aligned}$$

2.3 Graphical Check

Even when we plan to use Excel Solver, a small two-variable model is worth checking graphically. A graph helps us see whether the formulation makes sense.



The optimal solution from the graphical method and from Excel Solver is

$$D_1 = 20, \qquad D_2 = 100, \qquad \text{revenue} = 2900.$$

At this point, paper and blue paint are binding. Red and yellow paint have slack.

2.4 Common Formulation Mistakes

Students often make mistakes before calculation begins:

- choosing vague decision variables, such as "profit" instead of production quantities;
- forgetting a resource constraint;
- mixing units, such as dollars, meters, and paint units in one equation;
- treating a resource limit as an equality when unused capacity is allowed;
- using a solver before checking whether the model reflects the story.

A useful discipline is to read every constraint aloud. For example,

$$3D_1 + 2D_2 \leq 300$$

should read as: "yellow paint used by Design 1 plus yellow paint used by Design 2 cannot exceed 300 units."

2.5 Minimization Problems

Not every LP maximizes profit or revenue. Some models minimize cost, distance, time, waste, or risk. The logic is the same:

- an iso-profit line becomes an **iso-cost line**;
- instead of sliding toward higher profit, we slide toward lower cost;
- the optimum still occurs at a corner point when a finite optimum exists.

The important idea is not the word "maximize." The important idea is **optimize**: choose the best feasible solution according to the objective.

2.6 Four Special Cases In Linear Programming

LP models do not always produce one ordinary optimal corner. Solver messages are part of the learning, not just computer output.

2.6.1 Infeasibility

An LP is **infeasible** when no point satisfies all constraints.

For example, suppose a one-variable model includes

$$x \geq 5 \quad \text{and} \quad x \leq 3.$$

There is no value of x that satisfies both. In a business model, infeasibility often means that the requirements are too ambitious for the available resources.

2.6.2 Unboundedness

An LP is **unbounded** when the objective can improve without limit. For example,

$$\max x \quad \text{s.t.} \quad x \geq 0$$

has no finite optimum. We can always increase x and improve the objective.

Unboundedness often signals a missing constraint. In a real business, there is usually some limit on demand, capacity, budget, or time.

2.6.3 Alternative Optimal Solutions

An LP has **alternative optimal solutions** when more than one feasible solution gives the same optimal objective value. Graphically, this happens when the objective line is parallel to a binding edge of the feasible region.

This is not a failure. It means the manager has flexibility. If two production plans earn the same profit, the manager can choose using secondary criteria such as risk, fairness, customer service, or ease of implementation.

2.6.4 Redundant Constraints

A **redundant constraint** does not change the feasible region. Removing it would not change which points are feasible.

Redundancy is not always bad. A redundant constraint might represent a policy, safety rule, or reporting requirement. But when teaching modeling, redundancy is useful because it forces us to ask which constraints actually shape the decision.

2.7 Why We Need Excel Solver

The graphical method is useful for teaching, but it is limited:

- it only works visually with two decision variables;
- real models may have dozens, hundreds, or thousands of variables;
- drawing every constraint is slow and error-prone.

Excel Solver lets us keep the formulation logic but solve larger models with software.

2.8 A Good Solver Spreadsheet

A Solver spreadsheet should make the model easy to audit. The source slides give four practical guidelines:

1. Organize the data, then build the model around the data.
2. Do not embed parameters directly in formulas.
3. Things that are logically related should be physically related.
4. Use formulas that can be copied.

For the Gita model, a clean spreadsheet has:

Model element	Spreadsheet location idea
Decision variables	cells for D_1 and D_2 quantities
Objective cell	one cell with total revenue formula
Constraint LHS cells	one calculated cell per resource used
Constraint RHS cells	available resource limits
Parameters	prices and resource-use coefficients in visible cells

2.8.1 SUMPRODUCT

The Excel function

```
SUMPRODUCT(array1,array2)
```

multiplies corresponding entries and adds the products.

For a resource row, this is exactly what we need:

$$\text{resource used} = (\text{decision quantities}) \cdot (\text{unit resource use}).$$

For example, the paper-used formula can be written as:

```
=SUMPRODUCT($B$4:$C$4,B9:C9).
```

The dollar signs lock the decision-variable cells when the formula is copied down to other resource rows.

2.8.2 Solver Setup

For the Gita model:

- Set Objective: total revenue cell.
- To: Max.
- By Changing Variable Cells: D_1, D_2 quantity cells.
- Subject to: each amount-used cell is less than or equal to its available resource cell.
- Select: Make Unconstrained Variables Non-Negative.
- Solving Method: Simplex LP.

For the Lego model, the structure is the same. Only the parameters and formulas change.

2.9 Excel Shortcuts For Modeling

Shortcuts are not the main point of the course, but they reduce friction. Useful shortcuts include:

Task	Windows	Mac
Edit active cell	`F2`	`F2`
Make reference absolute	`F4`	`F4`
Show formulas	`Ctrl + ‘`	`Ctrl + ‘`
Move between sheets	`Ctrl + PageUp/PageDown`	`Ctrl + PageUp/PageDown`
AutoSum	`Alt +=`	`Shift + Cmd + T`
Select current region	`Ctrl + A`	`Ctrl + A`

The most important shortcut concept is not speed. It is auditability. A student who can show formulas, trace dependents, and inspect references can find modeling errors more easily.

2.10 Takeaways

- Formulation is the bridge from a business story to a solvable model.
- Decision variables must be precise and have units.
- Constraints should be read as resource-use or requirement statements.
- Infeasible, unbounded, alternative-optimum, and redundant-constraint cases are meaningful model diagnostics.
- Excel Solver requires a clean spreadsheet: objective cell, variable cells, constraint LHS cells, and RHS limits.

Chapter 3

Sensitivity Analysis, Summation Notation, and Blending Models

Source week: Week 3

Week 1 taught us how to solve a small LP graphically. Week 2 taught us how to formulate and solve an LP in Excel Solver. Week 3 asks two deeper questions.

1. Once Solver gives an optimal solution, how stable is that answer if the inputs change?
2. How do we write larger models without listing every equation one by one?

The first question is answered by **sensitivity analysis**. The second requires good notation, especially summation notation and double-index decision variables. Both ideas are essential for moving from small classroom examples to larger business models.

3.1 Why Sensitivity Analysis Matters

An LP model treats the input data as known: prices, costs, resource availability, and technology coefficients are all entered as numbers. In practice, those numbers move. Market prices change. Supply availability changes. Demand forecasts change. Even if the mathematical model is correct today, a manager still needs to know whether the recommendation is fragile.

sensitivity analysis asks "what if" questions around the optimal solution:

- What if a selling price changes?
- What if a resource becomes more or less available?
- What if a constraint becomes tighter?
- Which resource is acting as a bottleneck?

The important habit is to distinguish between two effects. A parameter change may change only the objective value, or it may also change the optimal decision. Those are very different managerial situations.

3.2 Two Ways To Do Sensitivity Analysis

There are two common approaches.

3.2.1 Trial And Error

In the trial-and-error approach, we change the data, resolve the LP, and compare the old and new solutions. This is flexible and easy to understand, but it can be slow when there are many possible changes to test.

3.2.2 Sensitivity Report

Excel Solver can generate a sensitivity report after solving an LP with the Simplex LP method. This report summarizes how much certain parameters can change before the current basis changes.

The report is compact, but students must read it carefully. It answers specific questions under specific assumptions. In particular, most entries assume one parameter changes while the rest of the model remains unchanged.

3.3 Gita's Wrapping-Paper Sensitivity Report

Recall Gita's solved model from Week 2:

$$\begin{aligned} \max \quad & 20D_1 + 25D_2 \\ \text{s.t.} \quad & D_1 + D_2 \le 120 \\ & 3D_1 + 3D_2 \le 480 \\ & 3D_2 \le 300 \\ & 3D_1 + 2D_2 \le 300 \\ & D_1, D_2 \ge 0. \end{aligned}$$

The optimal solution is

$$D_1 = 20, \qquad D_2 = 100, \qquad \text{revenue} = 2900.$$

At this solution, paper and blue paint are binding. Red and yellow paint have unused capacity.

3.4 Objective Coefficient Sensitivity

An **objective coefficient** is the number multiplying a decision variable in the objective function. In Gita's model, the objective coefficients are 20 for Design 1 and 25 for Design 2.

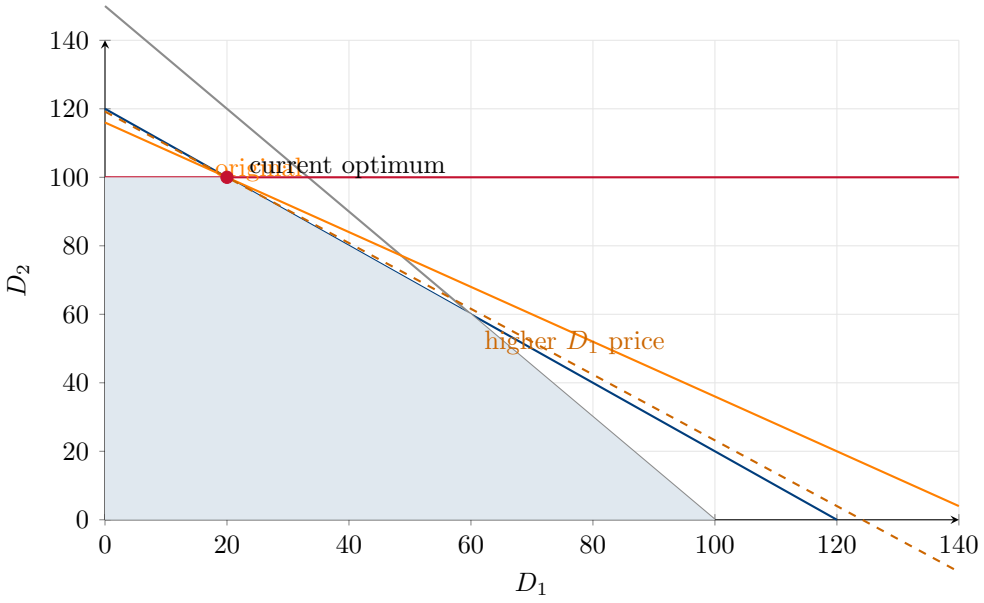
Changing an objective coefficient changes the slope of the iso-profit line. It does not change the feasible region, because the constraints are unchanged.

For Design 1, the sensitivity report says that the coefficient 20 has allowable increase 5 and allowable decrease 20. That means:

- If the price of Design 1 increases from 20 to 23, the current product mix remains optimal.
- The profit increases by $3 \times 20 = 60$, because Gita currently produces 20 meters of Design 1.
- If the price increases from 20 to 26, the change exceeds the allowable increase. The old product mix may no longer be optimal, so the model should be resolved.

The allowable range tells us when the current decision remains optimal. It does not say that profit is unchanged. If the product mix remains the same, the objective value changes by:

$$\text{change in coefficient} \times \text{current value of the variable}.$$



3.5 RHS Sensitivity And Shadow Price

The **right-hand side** (RHS) of a constraint is the limit or requirement on the right side. In Gita's paper constraint,

$$D_1 + D_2 \leq 120,$$

the RHS is 120 meters of paper.

The **shadow price** is the change in the objective value for a one-unit increase in the RHS, as long as the change stays within the allowable range.

For example, if the shadow price of paper is 20, then one more meter of paper increases revenue by 20, provided the change is within the valid range. If Gita gets 10 more meters of paper and the allowable increase is at least 10, the estimated profit increase is:

$$10 \times 20 = 200.$$

A non-binding constraint has slack at the optimum. Increasing a resource that is already unused usually has shadow price zero. This is the bottleneck idea: only scarce resources that block improvement have positive marginal value.

3.6 Allowable Ranges

Sensitivity report numbers are local. They are valid only within the allowable increase and allowable decrease.

- Inside the allowable range, the report can answer the "what if" question directly.
- Outside the allowable range, the basis may change. The model should be resolved.
- If multiple RHS values change together, the 100% rule can sometimes check whether the shadow prices remain valid.

The **100% rule** for simultaneous RHS changes says:

$$\sum \frac{\text{amount of change}}{\text{allowable change in that direction}} \leq 1.$$

If the total is at most 1, the current shadow prices remain valid. If the total exceeds 1, we cannot rely on the report without resolving the model.

3.7 Reduced Cost

The **reduced cost** matters most when a decision variable is zero in the optimal solution. For a maximization problem, it tells how much the objective coefficient of that variable would need to improve before the variable could become positive in the optimal solution.

If a variable has a positive final value, its reduced cost is usually zero in a standard LP sensitivity report. If a variable is zero, the reduced cost is a signal that the activity is currently unattractive under the current data.

3.8 Why We Need Summation Notation

For two products, we can write each equation by hand. For three crude oils and three gasoline types, equations multiply quickly. For 50 sources and 100 destinations, listing every term becomes impossible.

Summation notation is a compact way to say:

> "Repeat the same idea over a collection of indexed terms."

Students should not treat the sigma symbol as decoration. It is an instruction. It tells us which index moves and which index is fixed.

3.9 One-Dimensional Sums

Suppose q_1, q_2, q_3 are production quantities. Then

$$\sum_{j=1}^{3} q_j = q_1 + q_2 + q_3.$$

The index j moves from 1 to 3. The expression inside the sum is repeated once for each value of j.

If p_j is the unit price of product j, total revenue is:

$$\sum_{j=1}^{3} p_j q_j.$$

This means price times quantity for each product, then add the results.

3.10 Two-Dimensional Tables And Double-Index Variables

In a blending model, we often need a decision for each pair:

$$x_{ij} = \text{barrels of crude oil } i \text{ used to produce gasoline } j.$$

The first index i identifies the crude oil. The second index j identifies the gasoline type.

	Gasoline 1	Gasoline 2	Gasoline 3
Crude 1	x_{11}	x_{12}	x_{13}
Crude 2	x_{21}	x_{22}	x_{23}
Crude 3	x_{31}	x_{32}	x_{33}

To sum across a row, fix i and vary j. For example, total use of crude oil 1 is:

$$\sum_{j=1}^{3} x_{1j} = x_{11} + x_{12} + x_{13}.$$

To sum down a column, fix j and vary i. Total production of gasoline 2 is:

$$\sum_{i=1}^{3} x_{i2} = x_{12} + x_{22} + x_{32}.$$

To sum all cells, vary both indices:

$$\sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij}.$$

3.11 Mussel Oil Blending Problem

Mussel manufactures three gasoline types by blending three crude oils. It can buy up to 5000 barrels of each crude oil per day. Gasoline types differ in selling price, required octane, sulfur limit, and demand.

	Gasoline 1	Gasoline 2	Gasoline 3
Selling price	70	60	50
Demand limit	3000	2000	1000
Minimum octane	10	8	6
Maximum sulfur	0.01	0.02	0.01

	Crude 1	Crude 2	Crude 3
Cost	45	35	25
Octane	12	6	8
Sulfur	0.005	0.020	0.030
Available	5000	5000	5000

Let x_{ij} be barrels of crude oil i blended into gasoline j.

3.11.1 Supply Constraints

For each crude oil i, total use across all gasoline types cannot exceed 5000:

$$\sum_{j=1}^{3} x_{ij} \leq 5000, \qquad i = 1, 2, 3.$$

3.11.2 Demand Constraints

For each gasoline type j, total production cannot exceed the demand limit d_j:

$$\sum_{i=1}^{3} x_{ij} \leq d_j, \qquad j = 1, 2, 3.$$

3.11.3 Objective Function

Total revenue depends on gasoline type. Total cost depends on crude oil type. The profit objective is:

$$\max \sum_{j=1}^{3} p_j \left(\sum_{i=1}^{3} x_{ij} \right) - \sum_{i=1}^{3} c_i \left(\sum_{j=1}^{3} x_{ij} \right).$$

Equivalently, each x_{ij} earns a margin $p_j - c_i$:

$$\max \sum_{i=1}^{3} \sum_{j=1}^{3} (p_j - c_i) x_{ij}.$$

3.11.4 Octane Constraints

Average octane must meet a minimum. For gasoline j:

$$\frac{\sum_{i=1}^{3} o_i x_{ij}}{\sum_{i=1}^{3} x_{ij}} \geq r_j.$$

This expression looks nonlinear because of the fraction, but we can multiply both sides by total production:

$$\sum_{i=1}^{3} o_i x_{ij} \geq r_j \sum_{i=1}^{3} x_{ij}, \qquad j = 1, 2, 3.$$

This is linear because r_j is data, not a decision variable.

3.11.5 Sulfur Constraints

Average sulfur must not exceed a maximum. For gasoline j:

$$\frac{\sum_{i=1}^{3} s_i x_{ij}}{\sum_{i=1}^{3} x_{ij}} \leq u_j.$$

The linear form is:

$$\sum_{i=1}^{3} s_i x_{ij} \leq u_j \sum_{i=1}^{3} x_{ij}, \qquad j = 1, 2, 3.$$

3.11.6 Nonnegativity

$$x_{ij} \geq 0 \qquad \text{for all } i = 1, 2, 3 \text{ and } j = 1, 2, 3.$$

3.12 Solved Blending Plan

The solved workbook gives the following optimal blending plan:

	Gasoline 1	Gasoline 2	Gasoline 3
Crude 1	2400	800	800
Crude 2	0	0	0
Crude 3	600	1200	200

Total production is 3000, 2000, and 1000 barrels of the three gasoline types. Total revenue is 380000, total crude cost is 230000, and total profit is 150000.

The key modeling lesson is not the final number. The key lesson is that the table of variables tells us which summation to use:

- row sums correspond to crude oil use;
- column sums correspond to gasoline production;
- weighted column sums handle quality constraints;
- weighted row sums handle crude costs.

3.13 Week 3 Summary

- Sensitivity analysis helps managers understand whether an optimal solution is robust.
- Objective coefficient changes affect the iso-profit line; RHS changes affect the feasible region.
- Shadow price measures the marginal value of one more unit of a resource within an allowable range.
- Summation notation is a compact way to describe repeated table logic.
- In blending models, double-index variables connect inputs to outputs.
- Average quality requirements can often be written as linear constraints by multiplying through by total production.

Chapter 4

Network Models: Assignment and Transportation

Source week: Week 4 lecture: assignment and transportation models

4.1 Why Network Models Matter

Many important business decisions are allocation decisions. A firm may need to assign employees to projects, machines to jobs, drivers to passengers, shipments to routes, or products to customer regions. These decisions are not just arithmetic exercises. They shape service quality, cost, responsiveness, and fairness.

Network models give us a disciplined way to describe these decisions. A **network** has nodes and arcs. Nodes represent places, people, jobs, facilities, or demand points. Arcs represent possible connections: a driver can serve a passenger, a machine can process a job, or wheat can move from Memphis to Chicago. A decision variable is placed on each possible connection.

This chapter studies two classic network models:

Model	Core question	Typical applications
Assignment	Which resource should be matched to which task?	Job-machine matching, staff-project assignment, driver-passenger assignment on platforms such as Grab
Transportation	How much should move from each source to each destination?	Supply chain distribution, warehouse-to-store shipping, moving sand from bay locations to an airport site

The two models are closely connected. The assignment problem is a special case of the transportation problem: each resource supplies one unit and each task demands one unit. Chapter 5 extends this logic to transshipment networks in which flow may pass through intermediate hubs.

The practical modeling habit is the same throughout:

1. Define what each decision variable means.
2. Write an objective that adds up total cost, time, or distance.
3. Write one clear balance constraint for each row, column, or node.
4. Check whether the model's mathematical assumptions match the business meaning.

4.2 Part I: The Assignment Problem

4.2.1 Business Idea

An **assignment problem** matches resources to tasks on a one-to-one basis. Each resource should be assigned to one task, and each task should receive one resource. The objective is usually to minimize total cost, total time, or total distance.

Examples include:

- assigning machines to production jobs,
- assigning consultants to client projects,
- assigning contracts to bidders,
- assigning delivery drivers to customer orders,
- assigning ride-hailing drivers to passengers.

The key feature is that the decision is yes-or-no. Either machine 1 is assigned to job 3, or it is not. Either a driver is assigned to a passenger, or not.

4.2.2 MachineCo Example

MachineCo owns four machines and must complete four jobs. Each machine must be assigned to exactly one job, and each job must be assigned to exactly one machine. The setup time for each machine-job pair is:

	Job 1	Job 2	Job 3	Job 4
Machine 1	14	5	8	7
Machine 2	2	12	6	5
Machine 3	7	8	3	9
Machine 4	2	4	6	10

The goal is to minimize total setup time.

Define the decision variable

$$x_{ij} = \begin{cases} 1, & \text{if machine } i \text{ is assigned to job } j, \\ 0, & \text{otherwise.} \end{cases}$$

Let c_{ij} be the setup time for machine i on job j. The objective is

$$\min \sum_{i=1}^{4} \sum_{j=1}^{4} c_{ij} x_{ij}.$$

There are two families of constraints. It is useful to write them first without summation notation, because this shows what the summation is doing.

First, each machine must be assigned to one job:

$$\begin{aligned} x_{11} + x_{12} + x_{13} + x_{14} &= 1 \quad \text{Machine 1}, \\ x_{21} + x_{22} + x_{23} + x_{24} &= 1 \quad \text{Machine 2}, \\ x_{31} + x_{32} + x_{33} + x_{34} &= 1 \quad \text{Machine 3}, \\ x_{41} + x_{42} + x_{43} + x_{44} &= 1 \quad \text{Machine 4}. \end{aligned}$$

Each line adds across one row of the assignment table. The compact summation form is

$$\sum_{j=1}^{4} x_{ij} = 1 \quad \text{for each machine } i = 1, 2, 3, 4.$$

Second, each job must receive one machine:

$$\begin{aligned} x_{11} + x_{21} + x_{31} + x_{41} &= 1 \quad \text{Job 1,} \\ x_{12} + x_{22} + x_{32} + x_{42} &= 1 \quad \text{Job 2,} \\ x_{13} + x_{23} + x_{33} + x_{43} &= 1 \quad \text{Job 3,} \\ x_{14} + x_{24} + x_{34} + x_{44} &= 1 \quad \text{Job 4.} \end{aligned}$$

Each line adds down one column of the assignment table. The compact summation form is

$$\sum_{i=1}^{4} x_{ij} = 1 \quad \text{for each job } j = 1, 2, 3, 4.$$

Finally,

$$x_{ij} \in \{0, 1\}.$$

The first family of constraints is a row-sum requirement. The second family is a column-sum requirement. This row-and-column structure is the reason assignment models are easy to audit in a table or spreadsheet.

4.2.3 Why The Binary Restriction Is Special

At first glance, the assignment model is not a linear program because x_{ij} is binary. The source slides emphasize an important result:

> For the standard balanced assignment problem, the binary restrictions can be relaxed and the problem can still be solved as an LP.

That means we may replace

$$x_{ij} \in \{0, 1\}$$

with

$$0 \leq x_{ij} \leq 1.$$

For this special structure, an optimal LP solution will still assign whole machines to whole jobs. The intuition is that the row and column equalities are very restrictive. The corner points of the feasible region correspond to assignment patterns.

This result should not be overgeneralized. If we add extra business rules, such as “machine 1 and machine 2 cannot both be used” or “at least two priority jobs must use senior workers,” the special assignment structure may be broken. Then the binary restrictions may matter again.

4.2.4 Unbalanced Assignment Problems

The assignment problem is **balanced** when the number of resources equals the number of tasks. If the numbers do not match, the source slides show two modeling approaches.

Direct unbalanced formulation. If there are more machines than jobs, some machines may be unused. Keep the job constraints as equalities, but change the machine constraints to inequalities:

$$\sum_j x_{ij} \leq 1 \quad \text{for each machine } i,$$

$$\sum_i x_{ij} = 1 \quad \text{for each job } j.$$

If there are fewer machines than jobs, some jobs may be unassigned unless additional resources are available. Keep the machine constraints as equalities, but change the job constraints to inequalities:

$$\sum_j x_{ij} = 1 \quad \text{for each machine } i,$$

$$\sum_i x_{ij} \leq 1 \quad \text{for each job } j.$$

Balanced-form trick. We can also keep the balanced assignment format by adding a dummy row or dummy column.

Situation	Add this dummy item	Business meaning
More machines than jobs	Dummy job with zero setup time	A machine assigned to the dummy job is idle
More jobs than machines	Dummy machine with zero setup time	A job assigned to the dummy machine is not performed by a real machine

The dummy item is not real. It is a modeling device that lets us use the same row-and-column equality structure.

Why do we bother making the problem balanced? One reason is computational. The standard balanced assignment problem can be solved by special-purpose algorithms rather than by a general LP solver. The best-known example is the **Hungarian algorithm**, also called the Kuhn–Munkres algorithm. It works on a square cost matrix and searches for a minimum-cost perfect matching: one selected cell in each row and one selected cell in each column. The original version was already polynomial-time, and modern versions solve an $n \times n$ assignment problem in about $O(n^3)$ time. Adding a dummy row or column turns a rectangular problem into the square format expected by this algorithm.

4.3 Part II: The Transportation Problem

4.3.1 Business Idea

A **transportation problem** decides how much quantity should move from each supply point to each demand point. The decision is no longer yes-or-no. It is a nonnegative quantity.

The source slides motivate this with supply chain examples: chocolate is consumed in one set of countries and produced in another; flowers are produced in one region and sold globally; Singapore imports food from many places. These examples all require the same managerial question:

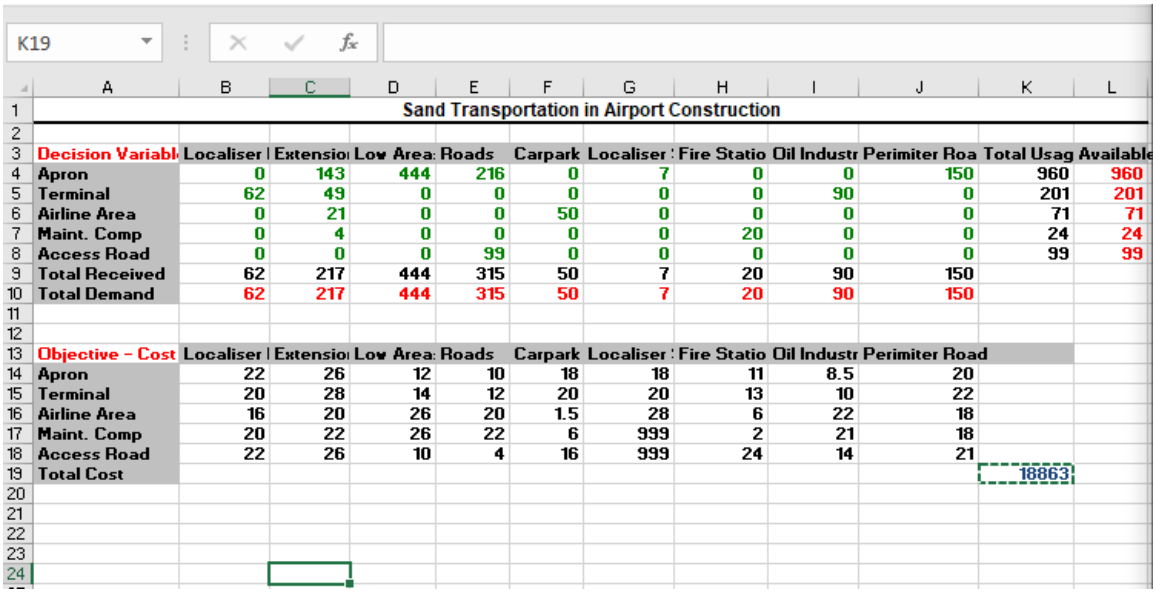
How should limited supply be allocated across demand locations at minimum total cost?

The transportation problem has:

- **sources**, which have available supply,
- **destinations**, which have demand requirements,
- **unit costs**, one for each source-destination route.

4.3.2 Airport Construction Example

The source deck uses an airport construction problem. Sand must be transported from several locations in a nearby bay to several locations at the airport site. Each bay location has a limited amount of sand available. Each airport location requires a certain amount of sand. The goal is to minimize the total transportation distance or cost.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Sand Transportation in Airport Construction											
2												
3	Decision Variabl	Localiser	Extensio	Low Area	Roads	Carpark	Localiser	Fire Statio	Oil Industr	Perimiter Roa	Total Usag	Available
4	Apron	0	143	444	216	0	7	0	0	150	960	960
5	Terminal	62	49	0	0	0	0	0	90	0	201	201
6	Airline Area	0	21	0	0	50	0	0	0	0	71	71
7	Maint. Comp	0	4	0	0	0	0	20	0	0	24	24
8	Access Road	0	0	0	99	0	0	0	0	0	99	99
9	Total Received	62	217	444	315	50	7	20	90	150		
10	Total Demand	62	217	444	315	50	7	20	90	150		
11												
12												
13	Objective - Cost	Localiser	Extensio	Low Area	Roads	Carpark	Localiser	Fire Statio	Oil Industr	Perimiter Road		
14	Apron	22	26	12	10	18	18	11	8.5	20		
15	Terminal	20	28	14	12	20	20	13	10	22		
16	Airline Area	16	20	26	20	1.5	28	6	22	18		
17	Maint. Comp	20	22	26	22	6	999	2	21	18		
18	Access Road	22	26	10	4	16	999	24	14	21		
19	Total Cost										18863	
20												
21												
22												
23												
24												

Figure 4.1: Airport construction transportation table from the source materials. The top block contains decision quantities, row totals, source availability, and destination demand. The lower block contains unit transportation costs.

Let:

- $i = 1, \ldots, m$ index sources,
- $j = 1, \ldots, n$ index destinations,
- s_i be supply available at source i,
- d_j be demand required at destination j,
- c_{ij} be the unit transportation cost from source i to destination j,
- x_{ij} be the quantity transported from source i to destination j.

The cell x_{ij} now has a different meaning from the assignment problem. In assignment, x_{ij} is 0 or 1. In transportation, x_{ij} is a quantity such as tons of sand, kilograms of food, or pallets of goods.

4.3.3 Balanced Transportation Model

A transportation problem is **balanced** when total supply equals total demand:

$$\sum_i s_i = \sum_j d_j.$$

For a balanced transportation problem, the LP is

$$\min \sum_i \sum_j c_{ij} x_{ij}$$

subject to source constraints

$$\sum_j x_{ij} = s_i \quad \text{for each source } i,$$

destination constraints

$$\sum_i x_{ij} = d_j \quad \text{for each destination } j,$$

and nonnegativity

$$x_{ij} \geq 0 \quad \text{for all } i, j.$$

The source constraints say that every unit available at each source is shipped. The destination constraints say that every demand requirement is satisfied. Because total supply equals total demand, both sets of equalities can hold at the same time.

4.3.4 Unbalanced Transportation

Transportation problems are often unbalanced. Total supply may be less than total demand, or total supply may be greater than total demand. The source slides highlight two approaches.

Approach 1: write an unbalanced formulation directly. If total supply is less than total demand, all available supply can be used, but not every destination can be fully satisfied. We keep source equalities and change destination equalities to upper bounds:

$$\sum_j x_{ij} = s_i \quad \text{for each source } i,$$

$$\sum_i x_{ij} \leq d_j \quad \text{for each destination } j.$$

If total supply is greater than total demand, every destination can be satisfied, but some supply may remain unused. We keep destination equalities and change source equalities to upper bounds:

$$\sum_j x_{ij} \leq s_i \quad \text{for each source } i,$$

$$\sum_i x_{ij} = d_j \quad \text{for each destination } j.$$

This direct approach is honest about the business situation. It says exactly which side of the model has unused capacity or unmet demand.

Approach 2: add a dummy source or dummy destination. The second approach converts the problem into a balanced transportation table.

If total supply is less than total demand, add a **dummy source** with supply

$$\sum_j d_j - \sum_i s_i.$$

Shipments from the dummy source represent unmet demand, emergency supply, or another business interpretation chosen by the modeler. The dummy source usually has zero transportation cost unless unmet demand should be penalized.

If total supply is greater than total demand, add a **dummy destination** with demand

$$\sum_i s_i - \sum_j d_j.$$

Shipments to the dummy destination represent unused supply or unused capacity. The dummy destination is not a real customer.

Why make the model balanced? Balanced transportation tables have special structure. A basic solution uses only $m+n-1$ basic cells for m sources and n destinations, and transportation-simplex methods can work directly on the transportation table rather than on a full general simplex tableau. In practice, methods such as the transportation simplex, stepping-stone method, MODI method, and network simplex exploit this structure and are typically faster and easier to audit than treating the model as an arbitrary LP.

4.3.5 Prohibited Routes And Big M

Sometimes a route is not allowed. For example, source 1 may not be allowed to satisfy destination 2 because of regulations, timing, quality, or physical access.

There are two clean modeling choices.

1. Omit the decision variable for that route.
2. Keep the table cell but assign a very large unit cost M.

The second method is called the **Big M method**. The cost M should be large enough that the route will not be chosen in an optimal solution, unless there is no feasible alternative. It should not be unnecessarily huge, because extremely large coefficients can make numerical optimization less stable.

4.3.6 Assignment As A Special Transportation Problem

The assignment problem is a transportation problem with a very special structure:

- each resource is a source with supply 1,
- each task is a destination with demand 1,
- the decision x_{ij} indicates whether one unit of assignment goes from resource i to task j.

This connection explains why both models use row and column constraints. The difference is the business meaning of the cell:

Model	Cell meaning	Row/column meaning
Assignment	Match resource i to task j	Each row and column usually sums to 1
Transportation	Ship quantity from source i to destination j	Rows satisfy supply; columns satisfy demand

4.4 Managerial Takeaways

- Assignment and transportation models are both network models for allocation decisions.
- Assignment models match resources to tasks; each decision variable is normally a 0–1 selection.
- Transportation models move divisible quantities from sources to destinations; rows represent supply and columns represent demand.
- A balanced model has total supply equal to total demand. Dummy rows or columns can preserve a balanced table when the business situation is unbalanced.
- A dummy node is a modeling device, not a real resource or destination, so its flow must be interpreted carefully.
- Prohibited routes can be omitted or assigned a sufficiently large penalty cost M.
- Chapter 5 extends the transportation structure by allowing flow to pass through intermediate transshipment nodes.

Chapter 5

The Transshipment Problem

Source week: Week 5 lecture: Transshipment and Sustainability in Transportation

Transportation models assume that products move directly from supply locations to demand locations. Real distribution systems often include ports, cross-docks, warehouses, consolidation centers, and transfer cities. These intermediate facilities do not create the product or consume final demand; they redirect flow through the network.

A **transshipment problem** extends the transportation model by allowing a shipment to visit one or more intermediate nodes before reaching its final destination. The main modeling challenge is no longer only a row total or a column total. We must account for both inflow and outflow at every node.

This chapter develops the model in five stages:

1. identify supply, transshipment, and demand nodes;
2. formulate the allowed-route model for Flair Wheat;
3. convert the network into a spreadsheet-friendly complete table;
4. add route emissions and a carbon price to the objective; and
5. apply the structure to Tiger Beer's Singapore supply network.

5.1 From Direct Shipping to Transshipment

5.1.1 Three node roles

The meaning of a node determines its constraint.

Node type	Business role	Flow implication
Supply node	A brewery, factory, farm, or supplier where flow originates	Net outflow cannot exceed available capacity
Transshipment node	A port, cross-dock, hub, or warehouse that redirects flow	Inflow equals outflow when there is no storage, loss, or local demand
Demand node	A customer, store, market, or project site where flow is consumed	Net inflow must satisfy required demand

For a pure transshipment node k,

$$\sum_i x_{ik} = \sum_j x_{kj}.$$

The left side is total inflow and the right side is total outflow. For example, if 60 tons enter New York and New York has no supply, demand, storage, or loss, then 60 tons must leave New York.

5.1.2 Why use hubs?

A hub can consolidate shipments from several suppliers and then redistribute them to several destinations. Suppose a direct network has n suppliers and m destinations. If every supplier can serve every destination, the network contains

$$nm$$

possible direct arcs.

Now suppose the network uses k shared hubs, every supplier can reach every hub, and every hub can reach every destination. The two-stage network contains

$$nk + km = k(n + m)$$

arcs.

For $n = 8$, $m = 8$, and $k = 2$,

$$nm = 64, \qquad k(n + m) = 2(8 + 8) = 32.$$

The hub network uses 32 fewer arcs, a 50% reduction. Fewer links can create scale economies in transportation, coordination, and infrastructure. However, a hub also adds handling, delay, detour, congestion, and disruption exposure. The arc count is therefore a useful structural comparison, not proof that the hub design has the lowest total cost.

5.2 Flair Wheat: An Allowed-Route Model

5.2.1 Business setting and network balance

Flair Wheat supplies customers from Memphis and Denver. Wheat may move directly to customers or pass through New York and Chicago.

- Memphis has capacity of 150 tons.
- Denver has capacity of 200 tons.
- Los Angeles requires 130 tons.
- Boston requires 130 tons.
- New York and Chicago are transshipment nodes.

Maximum supply is

$$150 + 200 = 350 \text{ tons},$$

while required demand is

$$130 + 130 = 260 \text{ tons}.$$

The network has 90 tons of excess capacity. Therefore, the supply constraints are upper bounds, but both customer demands must be satisfied exactly.

The allowed unit shipping costs are:

From / To	Memphis	Denver	NY	Chicago	LA	Boston
Memphis	0	–	8	13	25	28
Denver	–	0	15	12	26	25
New York	–	–	0	6	16	17
Chicago	–	–	6	0	4	16
Los Angeles	–	–	–	–	0	–
Boston	–	–	–	–	–	0

A dash means the route is unavailable. Diagonal zeros in this first table are bookkeeping entries, not real shipping routes.

5.2.2 Decision variables and objective

Number the cities as

$$1 = \text{Memphis}, \quad 2 = \text{Denver}, \quad 3 = \text{New York},$$
$$4 = \text{Chicago}, \quad 5 = \text{Los Angeles}, \quad 6 = \text{Boston}.$$

For every illustrated route, define

$$x_{ij} = \text{tons of wheat shipped from city } i \text{ to city } j.$$

The first index identifies the origin and the second identifies the destination. Thus, x_{34} is the flow from New York to Chicago, not from Chicago to New York.

Using only allowed routes, the objective is

$$\begin{aligned} \min z = {} & 8x_{13} + 13x_{14} + 25x_{15} + 28x_{16} \\ & + 15x_{23} + 12x_{24} + 26x_{25} + 25x_{26} \\ & + 6x_{34} + 16x_{35} + 17x_{36} + 6x_{43} + 4x_{45} + 16x_{46}. \end{aligned}$$

5.2.3 Write one constraint for each node

The complete allowed-route formulation is

$$\begin{aligned} x_{13} + x_{14} + x_{15} + x_{16} &\le 150 && \text{Memphis capacity,} \\ x_{23} + x_{24} + x_{25} + x_{26} &\le 200 && \text{Denver capacity,} \\ x_{34} + x_{35} + x_{36} - x_{13} - x_{23} - x_{43} &= 0 && \text{New York balance,} \\ x_{43} + x_{45} + x_{46} - x_{14} - x_{24} - x_{34} &= 0 && \text{Chicago balance,} \\ x_{15} + x_{25} + x_{35} + x_{45} &= 130 && \text{Los Angeles demand,} \\ x_{16} + x_{26} + x_{36} + x_{46} &= 130 && \text{Boston demand,} \\ x_{ij} &\ge 0 && \text{for every allowed route.} \end{aligned}$$

Read the equations by node:

- Memphis and Denver can create net flow, but not beyond their capacities.
- New York and Chicago only redirect flow, so their net flow is zero.
- Los Angeles and Boston consume flow, so their net inflow is 130 tons each.

5.2.4 Interpret a feasible solution

A feasible plan from the Flair Wheat workbook is:

Route	Tons shipped
Memphis → New York	60
New York → Boston	60
Denver → Chicago	130
Chicago → Los Angeles	130
Denver → Boston	70

Los Angeles receives 130 tons through Chicago. Boston receives 60 tons through New York and 70 tons directly from Denver. The two transshipment balances are satisfied:

$$60 \text{ into New York} = 60 \text{ out},$$

$$130 \text{ into Chicago} = 130 \text{ out}.$$

The total cost is

$$60(8) + 60(17) + 130(12) + 130(4) + 70(25) = 5330.$$

The plan uses all 200 tons of Denver capacity and 60 of the 150 tons available at Memphis, leaving 90 tons of Memphis capacity unused.

5.3 A Complete-Table Formulation

5.3.1 Why keep every node-to-node cell?

The allowed-route model is intuitive, but its list of variables and formulas changes whenever the network changes. A spreadsheet can instead use one $n \times n$ decision table containing a cell x_{ij} for every ordered pair of nodes.

The cost table then encodes the network:

- a possible route receives its real unit cost;
- an impossible route receives a large positive penalty M; and
- a self-route also receives M.

For Flair Wheat, the extended cost matrix is:

From \ To	Mem	Den	NY	Chi	LA	Bos
Memphis	M	M	8	13	25	28
Denver	M	M	15	12	26	25
New York	M	M	M	6	16	17
Chicago	M	M	6	M	4	16
Los Angeles	M	M	M	M	M	M
Boston	M	M	M	M	M	M

Let $\tilde{c}_{ij}$ denote the corresponding extended cost. The objective becomes

$$\min z = \sum_{i=1}^{6} \sum_{j=1}^{6} \tilde{c}_{ij} x_{ij}.$$

If a prohibited or self-route carries positive flow, the objective pays the large penalty Mx_{ij}. When a feasible real-route solution exists and M is sufficiently large, a cost-minimizing optimizer avoids those cells. M is a modeling penalty, not an actual freight rate, and it should not be made needlessly enormous.

5.3.2 Row sums are outflow; column sums are inflow

In the complete table,

$$\text{Outflow}_k = \sum_{j=1}^{6} x_{kj}, \qquad \text{Inflow}_k = \sum_{i=1}^{6} x_{ik}.$$

The node constraints are

$$\begin{aligned}
\sum_{j=1}^{6} x_{1j} - \sum_{i=1}^{6} x_{i1} &\le 150 \quad \text{Memphis,} \\
\sum_{j=1}^{6} x_{2j} - \sum_{i=1}^{6} x_{i2} &\le 200 \quad \text{Denver,} \\
\sum_{j=1}^{6} x_{3j} - \sum_{i=1}^{6} x_{i3} &= 0 \quad \text{New York,} \\
\sum_{j=1}^{6} x_{4j} - \sum_{i=1}^{6} x_{i4} &= 0 \quad \text{Chicago,} \\
\sum_{i=1}^{6} x_{i5} - \sum_{j=1}^{6} x_{5j} &= 130 \quad \text{Los Angeles,} \\
\sum_{i=1}^{6} x_{i6} - \sum_{j=1}^{6} x_{6j} &= 130 \quad \text{Boston.}
\end{aligned}$$

A diagonal variable x_{kk} appears once in outflow and once in inflow, so it cancels from the balance equation. Assigning M to the diagonal ensures that positive self-flow is also unattractive in the objective.

5.3.3 General formulations

Let $N = \{1, \ldots, n\}$ be the node set and define

$$x_{ij} = \text{flow from node } i \text{ to node } j.$$

For a balanced network with exact net supply b_k at node k,

$$\begin{aligned} \min \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{c}_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{n} x_{kj} - \sum_{i=1}^{n} x_{ik} = b_k, \quad k = 1, \ldots, n, \\ & x_{ij} \geq 0, \qquad\qquad\qquad\qquad i, j = 1, \ldots, n. \end{aligned}$$

The sign of b_k identifies the node:

$$b_k > 0 \text{ for supply}, \qquad b_k = 0 \text{ for transshipment}, \qquad b_k < 0 \text{ for demand}.$$

The Week 5 lecture also uses a capacity-and-demand form. Let s_k be available supply and d_k be required demand at node k. Then

$$s_k + \sum_{i=1}^{n} x_{ik} \geq d_k + \sum_{j=1}^{n} x_{kj}, \qquad k = 1, \ldots, n.$$

This reads as

$$\text{available supply} + \text{inflow} \geq \text{required demand} + \text{outflow}.$$

Use $s_k = 0$ when node k is not a supply node and $d_k = 0$ when it is not a demand node. In a positive-cost model, unnecessary flow and disposal are unattractive, so the inequality is normally tight wherever flow is economically useful. When conservation must hold exactly—for example, at a cross-dock with no disposal or inventory—write the transshipment balance as an equality.

5.3.4 Spreadsheet implementation

A readable Solver workbook should separate the model into five blocks:

1. **Inputs:** the extended cost matrix, node capacities, and node demands.
2. **Decision table:** the complete $n \times n$ range of x_{ij} changing cells.
3. **Objective:** a `SUMPRODUCT` of the cost and flow tables.
4. **Flow audit:** a row for each node showing supply, inflow, demand, outflow, and slack.
5. **Solver setup:** minimize total cost, change the decision table, enforce the node constraints, and require nonnegative variables.

The complete-table design is reusable because every node uses the same row-and-column formulas. Changing the network usually requires editing costs and capacities rather than rebuilding lists of route variables.

5.4 Carbon-Aware Transshipment

5.4.1 Sustainability has three dimensions

A transportation network affects more than private freight cost. The Week 5 lecture organizes sustainability into three connected dimensions:

- **economic**: cost, productivity, service, and resilience;
- **social**: access, equity, safety, and health; and
- **environmental**: carbon emissions, air quality, and resource use.

Transport generated about 24% of global energy-related CO_2 emissions in 2018, and road passenger plus road freight activity accounted for about 74.5% of transport emissions.[1] This does not imply that mobility should be eliminated. Mobility connects people to jobs, education, healthcare, and markets. The design objective is accessible service with lower social and environmental cost.

5.4.2 Convert route movement into emissions

For the Flair Wheat classroom example, use the illustrative trucking factor

$$1.617 \text{ kg } CO_2 \text{ per mile-ton.}$$

If d_{ij} is route distance in miles, emissions per ton shipped are

$$e_{ij} = d_{ij}(1.617)(10^{-3}) \quad \text{tonnes of } CO_2 \text{ per ton shipped.}$$

Therefore, total route emissions are

$$x_{ij}e_{ij}.$$

For a 1,000-mile route,

$$e_{ij} = 1000(1.617)(10^{-3}) = 1.617$$

tonnes of CO_2 per ton shipped. A real study should replace the classroom factor with mode-, vehicle-, load-, and route-specific data.

5.4.3 Add carbon cost to the objective

Let

- t_{ij} be the direct transportation and handling cost per unit,
- e_{ij} be tonnes of CO_2 emitted per unit shipped, and
- p be the carbon price per tonne of CO_2.

After expressing all monetary terms in the same currency, the carbon-aware objective is

$$\min \sum_i \sum_j (t_{ij} + pe_{ij})\, x_{ij}.$$

Equivalently,

$$\min \quad \underbrace{\sum_i \sum_j t_{ij}x_{ij}}_{\text{direct transportation cost}} \quad + \underbrace{p \sum_i \sum_j e_{ij}x_{ij}}_{\text{priced carbon cost}}.$$

Only the objective coefficients change. The physical supply, demand, capacity, and flow-balance constraints remain the same. As p rises, a cleaner route can become economically preferable even when its direct transportation cost is higher.

[1] https://ourworldindata.org/co2-emissions-from-transport

Singapore's carbon tax is S$45 per tonne of CO_2-equivalent for 2026–2027, with a stated 2030 range of S$50–80.[2] In a classroom scenario using $p = 45$, the route coefficient becomes

$$t_{ij} + 45e_{ij}.$$

The managerial question is not simply whether total cost increases. It is whether the relative effective costs change enough to alter routes, hub use, or sourcing decisions.

5.5 Tiger Beer's Singapore Supply Network

5.5.1 Classify the seven nodes

The Tiger Beer exercise applies the same logic to a regional-to-local network.

Node type	Locations
Supply	Malaysia brewery; Vietnam brewery
Transshipment	Port Cross-Dock; Tuas Logistics Hub
Demand	Retail DC; F&B customers; Hotels & Events

The breweries provide pallets. The Port Cross-Dock and Tuas Logistics Hub receive and redirect pallets. The three customer groups consume the delivered flow.

5.5.2 Build the extended cost matrix

The Week 5 group exercise uses the following feasible costs in S$ per pallet:

- Malaysia brewery → Tuas Logistics Hub: 25;
- Vietnam brewery → Port Cross-Dock: 20;
- Port Cross-Dock → Tuas Logistics Hub: 4;
- Port Cross-Dock → Retail DC, F&B, Hotels & Events: 12, 12, 24;
- Tuas Logistics Hub → Retail DC, F&B, Hotels & Events: 9, 12, 14.

Every other entry is M, including every self-to-self cell.

From \ To	Malaysia	Vietnam	Port	Tuas	Retail	F&B	Hotels
Malaysia brewery	M	M	M	25	M	M	M
Vietnam brewery	M	M	20	M	M	M	M
Port Cross-Dock	M	M	M	4	12	12	24
Tuas Logistics Hub	M	M	M	M	9	12	14
Retail DC	M	M	M	M	M	M	M
F&B	M	M	M	M	M	M	M
Hotels & Events	M	M	M	M	M	M	M

[2] https://www.nccs.gov.sg/singapores-climate-action/mitigation-efforts/carbontax/

5.5.3 Formulate and interpret the model

Let

$$x_{ij} = \text{pallets shipped from Tiger Beer node } i \text{ to node } j.$$

The transportation-cost objective is

$$\min \sum_i \sum_j \tilde{c}_{ij} x_{ij}.$$

Then write:

- one capacity constraint for each brewery;
- one inflow-equals-outflow constraint for the Port Cross-Dock;
- one inflow-equals-outflow constraint for the Tuas Logistics Hub;
- one demand constraint for each customer group; and
- $x_{ij} \geq 0$ for every table cell.

If a transshipment facility has a throughput limit U_k, add

$$\sum_i x_{ik} \leq U_k$$

or, equivalently under exact flow balance,

$$\sum_j x_{kj} \leq U_k.$$

This is separate from flow conservation: the balance equation says flow cannot disappear at the hub, while the throughput constraint limits how much the hub can handle.

A carbon extension assigns each feasible route an emissions coefficient e_{ij} and replaces $\tilde{c}_{ij}$ with $\tilde{c}_{ij} + pe_{ij}$ in the objective. A resilience extension can reduce a brewery or hub capacity and then re-optimize the full network. The new solution may reroute flow through another facility, change sourcing, or reveal that the demand plan is infeasible.

5.6 Common Modeling Mistakes

- Treating a transshipment node as a final customer and forgetting its outbound flow.
- Writing separate route constraints instead of one balance constraint for each node.
- Reversing the indices in x_{ij}; i is the origin and j is the destination.
- Forcing all capacity to be used when maximum supply exceeds demand.
- Giving an impossible route a normal cost instead of omitting it or assigning M.
- Choosing M so small that a prohibited route becomes attractive, or needlessly large and creating poor numerical scaling.
- Adding carbon cost without converting kilograms to tonnes or without using a common currency.
- Confusing flow conservation with throughput capacity; a hub often needs both constraints.

5.7 Managerial Takeaways

- Transshipment extends transportation by allowing flow through intermediate hubs.
- Supply nodes create net flow, demand nodes consume net flow, and pure transshipment nodes have inflow equal to outflow.
- Hubs can reduce the number of links and support consolidation, but they may add handling, delay, and disruption exposure.
- The allowed-route formulation mirrors the physical network; the complete-table formulation is easier to reuse in a spreadsheet.
- In the complete table, row sums measure outflow, column sums measure inflow, and the extended cost matrix distinguishes real routes from M-penalty routes.
- A carbon price changes effective route costs but does not change the network's physical flow constraints.
- Brewery capacities and hub throughput limits are managerial levers. When they change, the entire network should be re-optimized.

5.8 Self-Check

Before solving any transshipment model, verify that you can answer these questions:

1. What physical quantity and unit does x_{ij} represent?
2. Which nodes create flow, redirect flow, and consume flow?
3. Does each transshipment node have a complete inflow–outflow balance?
4. Is supply an exact quantity or a maximum capacity?
5. Are all impossible and self-routes excluded or assigned the intended penalty?
6. Are transportation cost, emissions, and carbon price expressed in compatible units?
7. Which brewery or hub capacity is likely to constrain the network, and what would an extra unit of that capacity be worth?

Chapter 6

Integer Linear Programming

Source week: Week 6

Many business decisions are not divisible. A company cannot open 2.4 warehouses, approve 3.6 projects, buy half of a standard machine, or sign 0.7 of a supplier contract. Other decisions are explicitly binary: launch or do not launch, locate or do not locate, accept or reject.

Linear programming is still the foundation for these decisions, but it needs an additional ingredient: selected decision variables must be integer. This chapter develops three connected ideas:

1. how an **LP relaxation** helps us understand and solve an integer linear program;
2. how **binary variables** model selection and logical relationships; and
3. how binary variables link a yes/no commitment to a quantity in fixed-cost and conditional-choice models.

6.1 When Divisibility Fails

An **integer linear program (ILP)** has a linear objective and linear constraints, just like an LP, but it also requires one or more decision variables to take integer values.

Three common model classes are:

Model	Integrality requirement	Business example
Pure integer program	Every decision variable is integer	numbers of products, trucks, or employees
Mixed-integer program	Some variables are integer and others continuous	open facilities and ship continuous quantities
Binary integer program	Selected variables satisfy $y_i \in \{0, 1\}$	fund projects, activate routes, or sign contracts

The integer restriction usually makes a model more realistic, but it also makes the problem harder to solve. The feasible region is no longer one continuous polygon or polyhedron. It becomes a collection of separated lattice points or mixed discrete–continuous pieces.

6.2 Harrison Electric: An Integer Product-Mix Model

Harrison Electric produces chandeliers and ceiling fans. A chandelier requires two wiring hours and six assembly hours. A ceiling fan requires three wiring hours and five assembly hours. The

company has 12 wiring hours and 30 assembly hours. Unit contributions are 7 and 6, respectively.

Define

$$x_1 = \text{number of chandeliers}, \qquad x_2 = \text{number of ceiling fans}.$$

The integer model is

$$\begin{aligned} \max \quad & 7x_1 + 6x_2 \\ \text{s.t.} \quad & 2x_1 + 3x_2 \le 12 \quad \text{(wiring)} \\ & 6x_1 + 5x_2 \le 30 \quad \text{(assembly)} \\ & x_1, x_2 \ge 0, \\ & x_1, x_2 \in \mathbb{Z}. \end{aligned}$$

The first three lines are an ordinary LP. The last line is what changes the problem: only whole-number production plans are allowed.

6.3 LP Relaxation

6.3.1 Definition

The **LP relaxation** of an ILP is created by removing the integer restrictions while keeping the objective and all other constraints unchanged. Harrison's relaxation is therefore

$$\begin{aligned} \max \quad & 7x_1 + 6x_2 \\ \text{s.t.} \quad & 2x_1 + 3x_2 \le 12, \\ & 6x_1 + 5x_2 \le 30, \\ & x_1, x_2 \ge 0. \end{aligned}$$

Every integer-feasible solution is feasible for the relaxation, but the relaxation also contains fractional solutions. Therefore,

$$\mathcal{F}_{\text{ILP}} \subseteq \mathcal{F}_{\text{LP}}.$$

This set relationship immediately gives a bound:

- for a maximization problem, $z^*_{\text{LP}} \ge z^*_{\text{ILP}}$;
- for a minimization problem, $z^*_{\text{LP}} \le z^*_{\text{ILP}}$.

The relaxation is useful because an LP can be solved efficiently and its objective value tells us how good any integer solution could possibly become.

6.3.2 Harrison's relaxation and integer optimum

The two resource boundaries intersect at

$$2x_1 + 3x_2 = 12, \qquad 6x_1 + 5x_2 = 30.$$

Solving gives

$$(x_1, x_2) = (3.75, 1.5), \qquad z_{\text{LP}} = 7(3.75) + 6(1.5) = 35.25.$$

This is not implementable because both product quantities are fractional. The best integer solution is

$$(x_1, x_2) = (5, 0), \qquad z_{\text{ILP}} = 35.$$

The relaxation therefore proves that no integer solution can exceed 35.25. Once we find an integer solution worth 35, only a very small gap remains.

6.3.3 Why simple rounding is unreliable

Rounding $(3.75, 1.5)$ up to $(4, 2)$ violates assembly capacity:

$$6(4) + 5(2) = 34 > 30.$$

Rounding down to $(3, 1)$ is feasible but gives only 27. Mixed rounding produces $(4, 1)$ with value 34 or $(3, 2)$ with value 33, both worse than the optimum $(5, 0)$.

The lesson is important:

> The LP relaxation provides a bound and a search direction. It does not provide an integer solution that can safely be rounded.

6.4 Branch-and-Bound

Branch-and-bound repeatedly solves LP relaxations over smaller regions.

1. **Solve a relaxation.** Its value is a bound for that subproblem.
2. **Branch.** If the solution is fractional, choose a fractional integer variable. For $x_k = 3.75$, create $x_k \leq 3$ and $x_k \geq 4$.
3. **Keep an incumbent.** The best integer-feasible solution found so far is the incumbent.
4. **Fathom or prune.** Stop exploring a node if it is infeasible, its solution is already integer, or its bound cannot beat the incumbent.

6.4.1 Harrison branch-and-bound calculation

At the root node, the relaxation gives $(3.75, 1.5)$ with upper bound 35.25. Branch on x_1:

$$x_1 \leq 3 \qquad \text{or} \qquad x_1 \geq 4.$$

Node B: $x_1 \leq 3$. The LP optimum is $(3, 2)$ with value 33. This solution is integer, so it becomes the first incumbent and the node is fathomed.

Node A: $x_1 \geq 4$. The LP optimum is $(4, 1.2)$ with upper bound 35.2. It is fractional, so branch on x_2:

$$x_2 \leq 1 \qquad \text{or} \qquad x_2 \geq 2.$$

The branch $x_2 \geq 2$ is infeasible. Under $x_2 \leq 1$, the LP optimum is

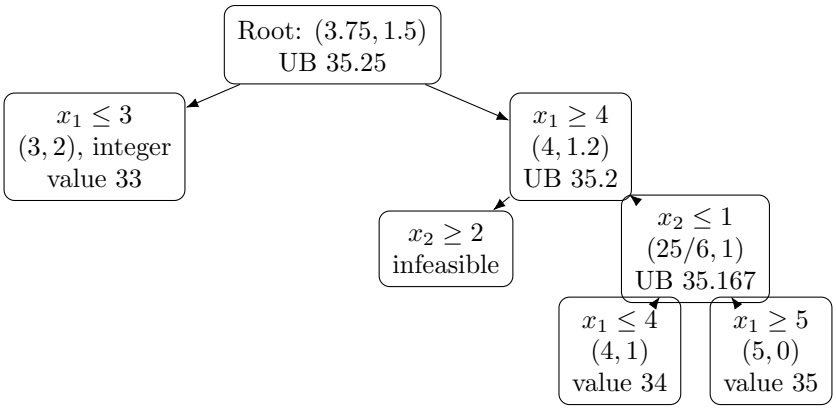
$$(x_1, x_2) = \left(\frac{25}{6}, 1\right), \qquad z = \frac{211}{6} \approx 35.167.$$

Branch on x_1 again:

$$x_1 \leq 4 \qquad \text{or} \qquad x_1 \geq 5.$$

The first branch gives integer solution $(4, 1)$ with value 34. The second gives integer solution $(5, 0)$ with value 35, which becomes the new incumbent.

All remaining active nodes are now infeasible, already integer, or have upper bounds at most 35. Therefore, $(5, 0)$ is proven optimal.



The standalone Week 06 animation reproduces this search interactively and labels every branch, bound, incumbent, and pruning decision.

6.5 Binary Variables and Project Selection

A binary variable records a yes/no commitment:

$$y_i = \begin{cases} 1, & \text{if option } i \text{ is selected,} \\ 0, & \text{otherwise.} \end{cases}$$

Binary variables are useful because many managerial actions are commitments: fund a project, open a facility, use a supplier, advertise in a market, or activate a route.

6.5.1 The knapsack structure

Suppose project i requires resource w_i and creates value u_i. With resource capacity W, the project-selection model is

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} u_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} w_i y_i \le W, \\ & y_i \in \{0,1\}, \qquad i = 1, \dots, n. \end{aligned}$$

This is the **0–1 knapsack problem**. The same structure appears when selecting projects under a capital budget, choosing features under a development-time limit, or loading indivisible items under a weight limit.

6.5.2 Eight-item example

The original Week 06 example has capacity 35:

Item i	1	2	3	4	5	6	7	8
Weight w_i	1	4	7	6	7	12	5	14
Utility u_i	20	50	80	60	50	75	30	70

The model is

$$\begin{aligned} \max \quad & 20y_1 + 50y_2 + 80y_3 + 60y_4 + 50y_5 + 75y_6 + 30y_7 + 70y_8 \\ \text{s.t.} \quad & y_1 + 4y_2 + 7y_3 + 6y_4 + 7y_5 + 12y_6 + 5y_7 + 14y_8 \le 35, \\ & y_i \in \{0,1\}, \qquad i = 1, \dots, 8. \end{aligned}$$

The optimal selection is items $1, 2, 3, 4, 6, 7$. It uses exactly 35 units of capacity and produces utility 315.

6.6 Logical Constraints with Binary Variables

The following patterns should be read in words before they are memorized as formulas.

6.6.1 At most, at least, and exactly

For a set S of binary decisions:

$$\sum_{i \in S} y_i \le k \quad \text{means at most } k \text{ are selected,}$$

$$\sum_{i \in S} y_i \ge k \quad \text{means at least } k \text{ are selected,}$$

$$\sum_{i \in S} y_i = k \quad \text{means exactly } k \text{ are selected.}$$

Thus, "exactly one of items 1, 2, and 3" is

$$y_1 + y_2 + y_3 = 1.$$

6.6.2 Mutual exclusion

"Items 6 and 7 cannot both be chosen" is

$$y_6 + y_7 \le 1.$$

The pair $(1, 1)$ is excluded, while $(0, 0)$, $(1, 0)$, and $(0, 1)$ remain feasible.

6.6.3 Same choice

"Items 6 and 7 are either both chosen or both rejected" is

$$y_6 = y_7.$$

6.6.4 Implication

"If item 1 is chosen, item 5 must be chosen" is

$$y_1 \le y_5.$$

The only forbidden pair is $(y_1, y_5) = (1, 0)$. The reverse implication is not imposed: item 5 may be selected without item 1.

6.6.5 Two antecedents imply a third

"If items 1 and 2 are both chosen, item 3 must be chosen" is

$$y_3 \geq y_1 + y_2 - 1.$$

The right-hand side becomes 1 only when both antecedents equal 1.

6.6.6 Either–or

For two selectable alternatives, "choose exactly one" is simply

$$y_A + y_B = 1.$$

For either–or constraints involving continuous expressions, use one binary switch. For example, at least one of

$$a^T x \leq b \qquad \text{or} \qquad c^T x \leq d$$

can be modeled as

$$\begin{aligned} a^T x &\leq b + My, \\ c^T x &\leq d + M(1-y), \\ y &\in \{0,1\}. \end{aligned}$$

When $y = 0$, the first constraint is active; when $y = 1$, the second is active. M must be a valid bound, not an arbitrary astronomical number.

6.7 Panther Manufacturing: Fixed Costs and Conditional Production

Panther manufactures compact, midsize, and large cars. Let x_i be the number of cars of type i.

	Compact	Midsize	Large	Available
Steel (tons)	1.5	3	5	6,000
Labor (hours)	30	25	40	60,000
Unit profit	2,000	3,000	4,000	

Without fixed costs, the production constraints are

$$1.5x_1 + 3x_2 + 5x_3 \leq 6000,$$

$$30x_1 + 25x_2 + 40x_3 \leq 60000, \qquad x_i \in \mathbb{Z}_+.$$

6.7.1 Fixed setup cost

Suppose each car type needs a dedicated facility with a one-time setup cost of 30,000. Define

$$y_i = 1 \text{ if facility } i \text{ is set up, and 0 otherwise.}$$

The objective becomes

$$\max \quad 2000x_1 + 3000x_2 + 4000x_3 - 30000(y_1 + y_2 + y_3).$$

The model must prevent positive production when a facility is closed. Use

$$x_i \le M_i y_i.$$

If $y_i = 0$, then $x_i \le 0$ and nonnegativity forces $x_i = 0$. If $y_i = 1$, production may rise as high as the valid bound M_i.

6.7.2 Deriving tight values of M_i

The resource constraints imply product-specific bounds:

$$M_1 = \min\left\{\frac{6000}{1.5}, \frac{60000}{30}\right\} = 2000,$$

$$M_2 = \min\left\{\frac{6000}{3}, \frac{60000}{25}\right\} = 2000,$$

$$M_3 = \min\left\{\frac{6000}{5}, \frac{60000}{40}\right\} = 1200.$$

These are the smallest obvious valid bounds from the existing capacities. Tight bounds generally make the LP relaxation stronger and branch-and-bound faster.

The complete fixed-cost formulation is

$$\begin{aligned}
\max \quad & 2000x_1 + 3000x_2 + 4000x_3 - 30000(y_1 + y_2 + y_3) \\
\text{s.t.} \quad & 1.5x_1 + 3x_2 + 5x_3 \le 6000, \\
& 30x_1 + 25x_2 + 40x_3 \le 60000, \\
& x_1 \le 2000y_1, \\
& x_2 \le 2000y_2, \\
& x_3 \le 1200y_3, \\
& x_i \in \mathbb{Z}_+, \quad y_i \in \{0,1\}, \qquad i = 1,2,3.
\end{aligned}$$

The validated solution is

$$(x_1, x_2, x_3) = (570, 1715, 0), \qquad (y_1, y_2, y_3) = (1, 1, 0),$$

with net objective 6,225,000.

6.7.3 Conditional minimum: zero or at least 1,000

Suppose a car type is economically feasible only if Panther produces at least 1,000 units. The pair of links is

$$1000y_i \le x_i \le M_i y_i.$$

When $y_i = 0$, both bounds force $x_i = 0$. When $y_i = 1$, they give

$$1000 \le x_i \le M_i.$$

The validated optimum is to produce 2,000 midsize cars only:

$$(x_1, x_2, x_3) = (0, 2000, 0), \qquad (y_1, y_2, y_3) = (0, 1, 0),$$

with objective 6,000,000.

6.8 Solver Workflow and Model Audit

The Week 06 workbook contains four models. Build the paper formulation before using Solver, then check:

1. the objective formula uses the decision cells;
2. every capacity formula is a `SUMPRODUCT`;
3. integer or binary restrictions are applied to the correct cells;
4. linking constraints are entered explicitly;
5. "Ignore Integer Constraints" is not selected; and
6. reported decisions satisfy both the numerical constraints and the business logic.

Unlike an ordinary LP, a solved ILP does not have the standard LP sensitivity report. Managerial analysis therefore relies more heavily on scenario changes, re-solving, bounds, and optimality gaps.

6.9 Common Modeling Errors

- **Rounding an LP solution.** Rounding can violate constraints or miss the best integer plan.
- **Reversing an implication.** "If A, then B" is $y_A \leq y_B$, not necessarily $y_A = y_B$.
- **Charging a fixed cost without a link.** A binary cost in the objective is ineffective unless positive activity requires $y = 1$.
- **Using an invalid M.** If M is too small, good solutions are removed. If it is needlessly huge, the relaxation becomes weak and numerical performance may suffer.
- **Forgetting the lower link.** "Zero or at least L" needs both $x \leq My$ and $x \geq Ly$.

6.10 Chapter Summary

1. An ILP adds integer restrictions to a linear model; a binary model uses variables in $\{0, 1\}$.
2. The LP relaxation enlarges the feasible set and supplies an objective bound.
3. Branch-and-bound partitions the feasible region, maintains an incumbent, and prunes nodes by infeasibility, integrality, or bound.
4. Knapsack models select indivisible options under limited capacity.
5. Binary constraints express at-most, at-least, exactly-one, mutual exclusion, prerequisites, and implications.
6. Linking constraints connect quantities to commitments: $Ly \leq x \leq My$.

6.11 Self-Check Questions

1. In a maximization ILP, why is the LP-relaxation objective an upper bound?
2. For the Harrison model, why is $(4, 2)$ not a valid rounded solution?

3. Translate "select at least two of projects 1–4" into one constraint.
4. Translate "project 2 can be selected only if project 5 is selected."
5. Explain what happens to x in $1000y \leq x \leq 3000y$ when $y = 0$ and when $y = 1$.
6. Give two independent reasons a branch-and-bound node can be pruned.

Chapter 7

Probability, Expected Value, and Bayes' Theorem

Source week: Week 9

7.1 Chapter overview

Decisions must be made before uncertain outcomes are known. Probability gives us a precise language for describing that uncertainty, expected value summarizes the long-run value of a risky prospect, and Bayes' theorem explains how beliefs should change when new evidence arrives.

This chapter develops these ideas in the following sequence:

1. describe experiments, outcomes, sample spaces, and events;
2. assign and interpret probabilities;
3. combine events using set operations and probability rules;
4. distinguish joint, marginal, and conditional probabilities;
5. diagnose independence and dependence;
6. use probability trees for sequential decisions;
7. compute expected value, variance, and standard deviation; and
8. update prior beliefs with Bayes' theorem.

7.2 Why probability matters

7.2.1 From uncertain possibilities to measurable risk

Every important decision begins with possibilities. A student may say, "I am likely to receive an offer after the interview." A manager may believe that demand will be high, or an insurer may expect a policyholder to survive the next year. Words such as *likely*, *unlikely*, and *possible* communicate direction, but not precision. Probability converts these vague beliefs into numbers between 0 and 1:

$$0 \leq P(E) \leq 1.$$

A probability near 0 means that an event is very unlikely; a probability near 1 means that it is very likely. A probability is not a promise about one case. It is a disciplined statement about uncertainty.

7.2.2 A motivating application: life insurance

A life insurer must estimate the probability that a person of a given age will survive or die during a specified period. That probability affects the premium, the reserve the company must hold, and whether the insurance pool will remain solvent. An estimate that is too optimistic can bankrupt the fund; one that is too pessimistic can make coverage unaffordable.

This problem helped turn insurance from an informal wager into a quantitative institution. In 1744, Scottish ministers Alexander Webster and Robert Wallace created a fund to support the widows and orphans of deceased Church of Scotland clergymen. When a minister died, his family could lose both income and the manse, the church-provided home. Each minister would contribute to a common pool, but the organizers first had to estimate how many claims the fund was likely to face and how much each member should pay. Mortality probabilities made collective financial protection possible.

7.2.3 A good decision can still produce a bad outcome

Probability separates the quality of a decision from the outcome of a single trial. A carefully maintained helicopter can still suffer a rare mechanical failure, while a poorly maintained helicopter may complete one flight safely. The outcome is evidence, but it is not by itself a complete evaluation of the decision. A sound decision uses the best available information and weighs both probabilities and consequences.

7.3 Experiments, outcomes, sample spaces, and events

7.3.1 Random experiments

A **random experiment** is a repeatable process whose exact outcome is not known in advance.

Experiment	One outcome	Sample space S
Toss a coin	Head	$\{H, T\}$
Roll a six-sided die	4	$\{1, 2, 3, 4, 5, 6\}$
Observe an applicant	Hired	$\{\text{hired}, \text{not hired}\}$

An **outcome** is one possible result. The **sample space** S is the set of all possible outcomes. If the outcomes are $s_1, s_2, \ldots, s_n$, a valid probability assignment satisfies

$$0 \le P(s_i) \le 1 \quad \text{and} \quad \sum_{i=1}^{n} P(s_i) = 1.$$

7.3.2 Events

An **event** is a set of outcomes. For a die roll, define

$$A = \{1, 2, 3\} \quad \text{(the result is less than 4)}$$

and

$$B = \{2, 4, 6\} \quad \text{(the result is even)}.$$

For a discrete sample space, the probability of an event is the sum of the probabilities of the outcomes it contains:

$$P(A) = \sum_{s_i \in A} P(s_i).$$

For a fair die, $P(A) = 3/6 = 1/2$ and $P(B) = 3/6 = 1/2$.

7.4 Where probabilities come from

7.4.1 Equally likely outcomes

When a finite experiment has equally likely outcomes,

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S}.$$

For a fair die,

$$P(\text{roll at least 5}) = \frac{|\{5, 6\}|}{6} = \frac{1}{3}.$$

7.4.2 The frequency approach

Sometimes we do not know whether a physical device is fair. A coin may be biased because of its shape, weight distribution, or manufacturing. We can estimate $P(H)$ by repeatedly tossing the coin:

$$\widehat{P}(H) = \frac{\text{number of observed heads}}{\text{number of tosses}}.$$

The **law of large numbers** says that, under stable repeated conditions, the cumulative relative frequency tends to settle near the true probability as the number of trials grows. Early observations can fluctuate sharply; larger samples are more stable.

This is why a coin toss can be used before an American football game to determine which team receives the ball first: the procedure is transparent and, with a fair coin, neither team has an advantage.

7.4.3 Subjective probability

Not every uncertainty can be repeated many times. The probability of receiving a particular job offer or launching a successful product is a belief based on the information currently available. Such a probability should still be coherent, clearly communicated, and revised when new evidence appears.

7.5 Operations on events

Events are sets, so event language corresponds directly to set operations.

Everyday statement	Set notation	Meaning
A or B occurs	$A \cup B$	at least one occurs
A and B occur	$A \cap B$	both occur
A does not occur	A^c	the complement of A

Using $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$,

$$A \cup B = \{1, 2, 3, 4, 6\}, \qquad A \cap B = \{2\}, \qquad A^c = \{4, 5, 6\}.$$

The complement rule is

$$P(A^c) = 1 - P(A).$$

7.5.1 Mutually exclusive and collectively exhaustive events

Two events are **mutually exclusive** if they cannot occur on the same trial:

$$A \cap B = \varnothing \quad \Longrightarrow \quad P(A \cap B) = 0.$$

Events are **collectively exhaustive** when their union covers the entire sample space. If $E_1, \ldots, E_k$ are both mutually exclusive and collectively exhaustive, then

$$\bigcup_{i=1}^{k} E_i = S, \qquad E_i \cap E_j = \varnothing \ (i \neq j), \qquad \sum_{i=1}^{k} P(E_i) = 1.$$

7.5.2 The addition rule

Adding $P(A)$ and $P(B)$ counts the overlap $A \cap B$ twice. Therefore,

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}.$$

If A and B are mutually exclusive, the rule simplifies to

$$P(A \cup B) = P(A) + P(B).$$

7.6 Conditional probability

7.6.1 Information changes the reference group

Before rolling a fair die, let

$$B = \{\text{even}\} = \{2, 4, 6\}.$$

Then $P(B) = 3/6 = 1/2$. Now suppose we learn

$$A = \{\text{less than } 4\} = \{1, 2, 3\}.$$

Within this smaller reference group, only one of the three possibilities is even:

$$P(B \mid A) = \frac{1}{3}.$$

In general,

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}} \qquad \text{when } P(A) > 0.$$

Likewise,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

These two conditional probabilities usually differ because they use different denominators.

7.6.2 The multiplication rule

Rearranging the conditional-probability formula gives

$$\boxed{P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)}.$$

This identity is especially useful for sequential experiments. We multiply along a path because the second probability is conditional on what happened first.

7.7 Independence, dependence, and mutual exclusion

Events A and B are **independent** if learning that one occurred does not change the probability of the other:

$$P(B \mid A) = P(B).$$

Equivalently,

$$P(A \cap B) = P(A)P(B).$$

For the fair-die events $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$,

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{2}, \quad P(A \cap B) = \frac{1}{6}.$$

Because $1/6 \neq (1/2)(1/2) = 1/4$, the events are dependent.

Do not confuse independence with mutual exclusion. Mutually exclusive events with positive probabilities are dependent: observing one tells us that the other did not occur. Independent events may occur together.

7.8 Joint, marginal, and conditional probabilities

Suppose an organization reviewed 1,000 transactions:

Transaction location	Fraud F	Not fraud F^c	Total
Far from usual location D	200	50	250
Not far D^c	100	650	750
Total	300	700	1,000

A **joint probability** refers to a cell:

$$P(F \cap D) = \frac{200}{1000} = 0.20.$$

A **marginal probability** refers to a row or column total:

$$P(F) = 0.30, \qquad P(D) = 0.25.$$

A **conditional probability** filters the table and changes the denominator:

$$P(F \mid D) = \frac{200}{250} = 0.80, \qquad P(F^c \mid D) = 0.20,$$

$$P(F \mid D^c) = \frac{100}{750} \approx 0.133, \qquad P(F^c \mid D^c) \approx 0.867.$$

The location signal is informative because the conditional fraud rates differ.

7.9 Sequential experiments and probability trees

7.9.1 Sampling with and without replacement

An urn contains 7 green balls and 3 blue balls. If a selected ball is returned before the next draw,

$$P(G_1 \cap G_2) = \frac{7}{10} \cdot \frac{7}{10} = \frac{49}{100}.$$

The draws are independent.

Without replacement, the second-draw probabilities change:

$$P(G_1 \cap G_2) = \frac{7}{10} \cdot \frac{6}{9} = \frac{7}{15}, \qquad P(B_2 \mid G_1) = \frac{3}{9} = \frac{1}{3}.$$

The draws are dependent because the first draw changes the urn.

A probability tree records each conditional branch. Multiply along a complete path. Add the probabilities of mutually exclusive paths that lead to the same event.

7.9.2 Practice: color and label

An urn contains 10 balls: 4 white lettered, 2 white numbered, 3 yellow lettered, and 1 yellow numbered. If L means lettered and Y means yellow,

$$P(L \mid Y) = \frac{3}{4} = 0.75.$$

But $P(L) = 7/10 = 0.70$, so L and Y are not independent.

7.9.3 Does drawing first create an advantage?

An urn contains 5 black and 5 white balls. Bianca wins if she draws black; Eric wins if he draws white. Bianca draws first.

With replacement, both have an unconditional winning probability of 1/2. Without replacement, Eric's chance depends on Bianca's result:

$$P(E \mid B) = \frac{5}{9}, \qquad P(E \mid B^c) = \frac{4}{9}.$$

The law of total probability gives

$$P(E) = \frac{5}{9}\frac{1}{2} + \frac{4}{9}\frac{1}{2} = \frac{1}{2}.$$

Order changes the conditional probabilities, but neither player has a higher unconditional probability of winning.

7.10 The birthday puzzle

For a class of n people, the complement is easier to compute:

$$P(\text{at least one shared birthday}) = 1 - P(\text{no shared birthday}).$$

Ignoring leap days and assuming birthdays are equally likely,

$$P(\text{no shared birthday}) = \prod_{k=0}^{n-1} \frac{365-k}{365}.$$

The first person can have any birthday. The second must avoid one occupied day, the third must avoid two, and so on. For $n = 46$,

$$P(\text{no shared birthday}) \approx 0.0517,$$

so

$$P(\text{at least one shared birthday}) \approx 0.9483.$$

The result feels surprising because the number of pairs, $n(n-1)/2$, grows quickly.

7.11 Prediction markets and expected profit

A prediction market lets participants put money behind probability judgments. A YES event contract pays \$1 if the event occurs and \$0 otherwise. If a YES contract costs q, the trader's probability assessment is p, and the fee is f, then

$$\boxed{\text{EV(YES)} = p - q - f}.$$

For example, if $p = 0.65$, $q = 0.55$, and $f = 0.02$,

$$\text{EV(YES)} = 0.65 - 0.55 - 0.02 = \$0.08.$$

A positive expected profit does not guarantee profit on one contract. It is an average over repeated comparable opportunities.

Prediction markets are controversial. Platforms such as Kalshi describe event contracts as federally regulated financial products. Critics, including some states, tribal gaming leaders, and casino representatives, argue that sports and event contracts can function like gambling and may bypass state gaming rules. The classroom lesson is analytical rather than advisory: a market price can be a probability-like signal, but incentives, fees, liquidity, and regulation matter.

7.12 Bayes' theorem

7.12.1 Prior, evidence, and posterior

Bayesian reasoning is a disciplined learning cycle:

$$\text{prior belief} + \text{new evidence} \longrightarrow \text{posterior belief}.$$

The posterior is not a permanent truth. It becomes the prior for the next update.

Suppose A is a hypothesis and B is observed evidence. Then

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

If A and A^c form a mutually exclusive, collectively exhaustive partition, the total probability of the evidence is

$$P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c).$$

Therefore,

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}}.$$

7.12.2 Reading Bayes through a probability tree

Start with two branches, A and A^c, weighted by their priors. Each branch then divides into B and B^c using conditional probabilities. The two paths ending in B have probabilities

$$P(A)P(B \mid A) \quad \text{and} \quad P(A^c)P(B \mid A^c).$$

Together they represent all ways B can occur. Among those B cases, the fraction that came through A is $P(A \mid B)$.

7.13 Case study: interpreting an HIV screening result

Let

$$H = \{\text{person has HIV}\}, \quad N = H^c, \quad + = \{\text{test is positive}\}.$$

For this classroom case, use

$$P(H) = 0.01228\% = 0.0001228,$$
$$P(+ \mid H) = 0.999 \quad \text{(sensitivity)},$$

and

$$P(- \mid N) = 0.995 \quad \text{(specificity)}.$$

Thus $P(+ \mid N) = 0.005$. The positive result can arise through two paths:

$$P(H \cap +) = 0.0001228(0.999) = 0.0001226772,$$
$$P(N \cap +) = 0.9998772(0.005) = 0.004999386.$$

Therefore,

$$P(H \mid +) = \frac{0.999(0.0001228)}{0.999(0.0001228) + 0.005(0.9998772)}$$
$$\approx 0.02395 = 2.40\%.$$

The test has high sensitivity and specificity, yet one positive result does not imply a 99.9% probability of HIV. The low base rate matters. In a hypothetical population of 100,000 people, the model produces about 12.27 true positives and 499.94 false positives. Noise from the much larger non-HIV population dilutes the positive signal.

7.13.1 A second independent test

Use the posterior after the first positive result as the new prior:

$$P(H \mid +_1) = 0.02395.$$

Suppose the second test has

$$P(+_2 \mid H) = 0.99, \qquad P(+_2 \mid N) = 0.005.$$

If the second test is also positive,

$$P(H \mid +_1, +_2) = \frac{0.99(0.02395)}{0.99(0.02395) + 0.005(1 - 0.02395)}$$
$$\approx 0.8293 = 82.93\%.$$

Independent evidence compounds. Repeating a test with correlated errors may add less information, which is why confirmatory testing should use an appropriate independent method.

7.14 Bayes as a discipline of continuous learning

The HIV example illustrates a general sequence:

$$\begin{array}{ccccc} \text{prevalence prior} & \xrightarrow{\text{first positive}} & \text{updated belief} & \xrightarrow{\text{second positive}} & \text{updated belief} \\ 0.01228\% & & 2.40\% & & 82.93\%. \end{array}$$

Belief should move in proportion to the evidence: do not overreact to one signal, but do not protect an old belief when strong, independent evidence accumulates. Bayes gives a rational way to respond to evidence.

7.14.1 When a prior becomes doctrine

Consider a stylized account of the smartphone transition. A hardware-first manufacturer begins with the belief that customers mainly value durable devices, physical keyboards, battery life, and manufacturing quality. The iPhone and the growth of app ecosystems provide new evidence that software experience and developer participation are becoming central.

A learning organization updates its belief, invests in its platform and app ecosystem, and adapts its product strategy. An organization that treats the old belief as identity can dismiss the evidence, protect the prior, and fall behind. The point is not that every new signal should trigger a radical change. The point is that a belief should remain testable and revisable.

Max Planck's observation that scientific progress can occur "one funeral at a time" captures the cost of refusing to update: new ideas sometimes prevail only when a new generation replaces defenders of the old doctrine. A stronger learning habit is to *learn, unlearn, and relearn* while evidence is still useful.

7.15 Strategic evidence: the Monty Hall problem

There are three doors: one hides a car and two hide goats. A contestant chooses Door 1. The host, who knows the locations, opens Door 3 to reveal a goat. Should the contestant switch to Door 2?

Let C_i mean that the car is behind Door i, and let H_3 mean that the host opens Door 3. Initially,

$$P(C_1) = P(C_2) = P(C_3) = \frac{1}{3}.$$

Under the standard rules, the host must reveal a goat and offer a switch:

$$P(H_3 \mid C_1) = \frac{1}{2}, \qquad P(H_3 \mid C_2) = 1, \qquad P(H_3 \mid C_3) = 0.$$

Thus,

$$P(C_2 \mid H_3) = \frac{1(1/3)}{(1/2)(1/3) + 1(1/3) + 0(1/3)} = \frac{2}{3}.$$

The probability that Door 1 contains the car remains 1/3, so switching wins with probability 2/3.

The host's action is informative because it is constrained and strategic, not a random door opening. In business, a rival's bid, a supplier's concession, or a negotiating partner's refusal may likewise reveal private information. Bayes provides a way to update beliefs using the likelihood of the observed action under different hidden states.

7.16 Random variables and discrete probability distributions

A **random variable** maps each outcome of a random experiment to a number. For a discrete random variable X, its probability distribution is a table of possible values and their probabilities:

Possible value	x_1	x_2	$\cdots$	x_n
Probability	p_1	p_2	$\cdots$	p_n

A valid distribution satisfies

$$p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1.$$

7.16.1 Mean, variance, and standard deviation

The mean, or expected value, is the probability-weighted average outcome:

$$\boxed{\mu = E[X] = \sum_{i=1}^{n} x_i p_i}.$$

For a class GPA distribution, $E[X]$ is the average GPA. For an investment, $E[X]$ is the long-run average return.

Variance measures the probability-weighted squared distance from the mean:

$$\boxed{\mathrm{Var}(X) = \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 p_i}.$$

Because variance is measured in squared units, standard deviation converts the spread back to the same unit as the original variable:

$$\boxed{\sigma = \sqrt{\mathrm{Var}(X)}}.$$

Greater standard deviation means greater dispersion around the mean.

7.17 Worked comparison: two investments

Return x	$P(X = x)$	$P(Y = x)$
3	0.05	0.15
4	0.20	0.20
5	0.35	0.25
6	0.30	0.15
7	0.10	0.15
8	0.00	0.10

For Investment X,

$$E[X] = 3(0.05) + 4(0.20) + 5(0.35) + 6(0.30) + 7(0.10) = 5.20,$$

$$\mathrm{Var}(X) = 1.06, \qquad \mathrm{SD}(X) = \sqrt{1.06} \approx 1.03.$$

For Investment Y,

$$E[Y] = 3(0.15) + 4(0.20) + 5(0.25) + 6(0.15) + 7(0.15) + 8(0.10) = 5.25,$$

$$\mathrm{Var}(Y) = 2.3875 \approx 2.39, \qquad \mathrm{SD}(Y) = \sqrt{2.3875} \approx 1.55.$$

Investment Y has a slightly higher mean but substantially greater spread. The decision depends on how the decision maker values return relative to risk.

The companion workbook contains a completed solution sheet and a student formula template:

Open the Week 9 investment probability-distribution workbook

7.18 Common mistakes

- **Treating an outcome as a probability.** Probability describes uncertainty before the outcome is known.
- **Confusing $P(A \mid B)$ with $P(B \mid A)$.** The conditioning event determines the denominator.
- **Adding without subtracting overlap.** Use the general addition rule unless events are mutually exclusive.
- **Multiplying marginal probabilities automatically.** The shortcut $P(A \cap B) = P(A)P(B)$ requires independence.
- **Treating mutual exclusion as independence.** Positive-probability mutually exclusive events are dependent.
- **Ignoring base rates.** A highly accurate test can still yield many false positives when the condition is rare.
- **Interpreting expected value as a guaranteed payoff.** Expected value is an average, not a promise for one trial.
- **Comparing means without spread.** Similar expected values can conceal different risk profiles.

7.19 Practice questions

1. A fair die is rolled. Let $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$. Find $P(A \cup B)$, $P(A \cap B)$, and $P(B \mid A)$.
2. In the fraud table, compute $P(D \mid F)$. Explain why this is not the same as $P(F \mid D)$.
3. An urn contains 7 green and 3 blue balls. Two balls are drawn without replacement. Find $P(G_1 \cap B_2)$.
4. For a class of 23 students, write the expression for the probability of at least one shared birthday.
5. A YES contract costs \$0.62, your probability assessment is 0.70, and the fee is \$0.02. What is the expected profit per contract?
6. Explain why the first positive HIV test moves the probability from 0.01228% to only 2.40%, while a second independent positive test moves it to 82.93%.

7. In Monty Hall, which assumption makes the revealed goat informative?
8. Which investment has the larger expected return? Which has more spread?

Selected answers

1. $P(A \cup B) = 5/6$, $P(A \cap B) = 1/6$, and $P(B \mid A) = 1/3$.
2. $P(D \mid F) = 200/300 = 2/3$. It conditions on fraud, whereas $P(F \mid D) = 200/250 = 0.80$ conditions on distance.
3. $P(G_1 \cap B_2) = (7/10)(3/9) = 7/30$.
4. $1 - \prod_{k=0}^{22}(365 - k)/365$.
5. $0.70 - 0.62 - 0.02 = \$0.06$.
6. The first result contains many false positives because HIV is rare. After the first positive, HIV is less rare in the filtered group, so an independent second positive has much stronger diagnostic value.
7. The host knows the car location, must open a goat door, and must offer the switch.
8. Y has the larger mean, 5.25 versus 5.20, and the larger standard deviation, 1.55 versus 1.03.

7.20 Chapter summary

- Probability measures uncertainty on a scale from 0 to 1.
- Events are sets; unions, intersections, and complements translate everyday logic into mathematics.
- Conditional probability changes the reference group when information becomes known.
- Independence means that learning one event occurred does not change the probability of the other.
- Probability trees organize conditional and joint probabilities.
- Expected value measures the probability-weighted average; variance and standard deviation measure spread.
- Bayes' theorem combines a prior, likelihood, and total evidence probability to produce a posterior.
- Rational learning means updating in proportion to evidence: neither overreacting to one signal nor clinging to an obsolete prior.

Chapter 8

Decision Making under Risk and Uncertainty

Source week: Week 10

8.1 Chapter overview

An attractive outcome is not necessarily an attractive decision. A new product may earn millions in a good market and lose money in a bad one. The decision maker chooses an action *before* learning which market will occur. This chapter develops a systematic way to describe that choice, compare actions, account for risk preferences, and value information that arrives before the decision.

After studying this chapter, you should be able to construct a payoff table; distinguish certainty, risk, and uncertainty; apply five rules for decisions without known probabilities; compute expected monetary value and switching probabilities when probabilities are available; use utility and certainty equivalents to express risk preference; and calculate the expected value of perfect information (EVPI).

8.2 Describe the decision before choosing

8.2.1 Three decision environments

Environment	Information available before the choice
Certainty	The consequence of each alternative is known.
Uncertainty	Several states can arise, but their probabilities are not known.
Risk	Several states can arise, and their probabilities are known or supplied for the analysis.

Here, *risk* is a technical label for a decision with a probability model; it does not mean that every outcome is bad. Known probabilities do not make the eventual outcome certain. In practice, the probabilities may be estimates, so their quality also deserves scrutiny.

An **alternative** is an action the decision maker controls. A **state of nature** is an outcome, such as a favorable or unfavorable market, that the decision maker does not control. A **payoff** is the consequence of an alternative in a particular state, measured here in money. A **payoff table** places alternatives in rows and states in columns. The timing matters: unless information is explicitly available beforehand, one must select a single row before learning which column will occur.

A useful decision-analysis process is to (1) define the problem and objective, (2) list feasible alternatives, (3) identify states of nature, (4) calculate the payoff of every alternative–state combination, (5) select a model or decision rule, and (6) apply it and test the result's sensitivity. A decision tree can show the same alternatives and outcomes in chronological order; the payoff table is more compact for the examples in this chapter.

8.2.2 Running example: Stephen Marts's product idea

Stephen Marts may develop a new product for \$7 million. It would yield \$10 million in a good market or \$5 million in a bad market. Alternatively, he may sell the idea to Company A for a fixed \$0.75 million, sell to Company B under a profit-sharing arrangement, or do nothing. All payoffs in Table 8.1 are in *millions of dollars*. Development payoffs are $10 - 7 = 3$ and $5 - 7 = -2$. The Company B values are the payoffs supplied in the classroom case; its precise loss-sharing contract is not specified, so we should not infer its bad-market payoff from the profit-sharing percentage alone.

Table 8.1: Entrepreneur payoff table (\$ million)

Alternative	Good market	Bad market
Develop the product	3	−2
Sell to Company A	0.75	0.75
Sell to Company B	1.2	0
Do nothing	0	0

Under *certainty*, if Stephen knows the market will be good before choosing, he develops and earns \$3 million. If he knows it will be bad, he sells to Company A and earns \$0.75 million. These are separate information conditions; they do not permit him to wait and choose retrospectively when the market was unknown at the time of the decision.

8.3 When probabilities are known: expected monetary value

For alternative j, the **expected monetary value** is the probability-weighted average of its payoffs:

$$\mathrm{EV}(j) = \sum_i p_i V_i(j), \qquad \sum_i p_i = 1, \tag{8.1}$$

where p_i is the probability of state i and $V_i(j)$ is the payoff of alternative j in that state. If monetary expected value is the decision objective, choose the alternative with the largest EV. EV is an average across possible outcomes, not the guaranteed payoff of one attempt.

Let p be the probability of a good market for Stephen, so the bad-market probability is $1 - p$. The four alternatives have expected payoffs

$$\begin{aligned} \mathrm{EV}(\text{Develop}) &= 3p - 2(1-p) = 5p - 2, \\ \mathrm{EV}(\text{Company A}) &= 0.75p + 0.75(1-p) = 0.75, \\ \mathrm{EV}(\text{Company B}) &= 1.2p + 0(1-p) = 1.2p, \\ \mathrm{EV}(\text{Nothing}) &= 0. \end{aligned}$$

At $p = 0.6$, these values are 1, 0.75, 0.72, and 0 million, respectively. Developing has the largest EV, although it still has a \$2 million loss if the market is bad.

8.3.1 Sensitivity to the probability

Probability estimates can change. To see when the recommendation changes, equate the expected payoffs of the two alternatives that could be best:

$$5p - 2 = 0.75 \quad \Longrightarrow \quad p = 0.55.$$

For $p < 0.55$, Company A is best; for $p > 0.55$, Develop is best; at $p = 0.55$, they tie at \$0.75 million. Company B is never the highest-EV choice on $0 \le p \le 1$: Company A beats it at lower probabilities, and Develop beats it at higher probabilities. Notice the precise claim: Company B is not *strictly dominated by one fixed alternative in every state*. Rather, it lies below the *upper envelope* of the EV lines for every probability.

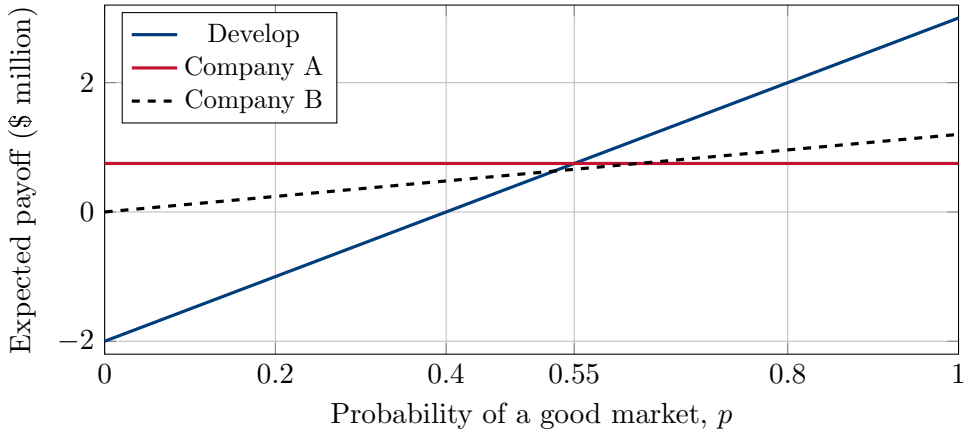


Figure 8.1: The highest expected-payoff line changes at $p = 0.55$. Company B never reaches the upper envelope.

8.4 Risk preference, utility, and certainty equivalents

EV describes monetary outcomes but does not fully describe how everyone values risk. Consider a sure \$50 (Option A) and a lottery paying \$100 or \$0 with equal probability (Option B). Both have EV \$50. A **risk-averse** person prefers the sure \$50; a **risk-neutral** person is indifferent; a **risk-seeking** person prefers the lottery. The spread of outcomes, which can be measured by variance, differs even though the means match.

8.4.1 Why a concave utility curve represents risk aversion

Utility represents the decision maker's value for an outcome rather than its dollar amount. For an increasing, concave utility function, each additional dollar adds less utility than the preceding one. Starting from a given wealth point, a loss of a given size reduces utility more than an equal-sized gain raises it. For the \$100/\$0 lottery, concavity gives

$$U(50) > \tfrac{1}{2}U(100) + \tfrac{1}{2}U(0).$$

Thus the utility of receiving the sure EV is greater than the lottery's *expected utility*. This is the geometric reason risk aversion is linked to concavity; it is not merely a dislike of negative outcomes.

The **certainty equivalent (CE)** is the guaranteed amount that gives the same utility as a risky prospect:

$$U(\text{CE}) = \mathbb{E}[U(W)]. \tag{8.2}$$

For a risk-averse person, the CE of a nonconstant lottery is below its EV. The difference $\text{EV}(W) - \text{CE}$ is the **risk premium**: the most that person would pay to replace the lottery with its expected amount for certain, within this model.

8.4.2 Constant absolute risk aversion (CARA)

The lecture uses the negative-exponential utility function

$$U(w) = -\exp(-aw), \qquad a > 0, \tag{8.3}$$

where w is wealth or payoff measured in a consistent unit and a is the absolute risk-aversion coefficient. A larger positive a produces greater risk aversion. "Constant absolute risk aversion" means that $-U''(w)/U'(w) = a$, independent of w. Units matter: if w is measured in millions of dollars, a is per million dollars. At $a = 0$, the displayed formula becomes the constant -1, so risk neutrality must be handled as a *limit* of its CE formula, rather than by directly substituting $a = 0$ into (8.3). Likewise, the unchanged formula with $a < 0$ decreases with wealth, so a risk-seeking utility representation needs a separately specified increasing, convex function. This distinction avoids assigning a preference for less wealth by accident.

For a lottery paying \$100 with probability p and \$0 otherwise, set the utility of a sure CE equal to expected utility:

$$\begin{aligned} -\exp(-a\,\text{CE}) &= p[-\exp(-100a)] + (1-p)[-\exp(0)], \\ \text{CE} &= -\frac{1}{a}\ln[p\exp(-100a) + (1-p)\exp(0)]\,. \end{aligned}$$

With $p = 0.5$, $a = 0.01$ gives CE = \$37.99 and $a = 0.15$ gives CE = \$4.62 (rounded). The sure \$50 is therefore preferred at both coefficients. As a approaches zero from above, the CE approaches the monetary EV, \$50. Do not confuse the p in this \$100/\$0 illustration with the good-market probability in the entrepreneur example; the symbol has the same probability meaning but a different event.

8.5 When probabilities are unknown: five decision rules

The entrepreneur table also illustrates decisions under *uncertainty*, when no credible probability for the market states is available. No single rule can be derived from absent probabilities. Each rule expresses a different attitude or convention, so it is useful to report which rule produced the recommendation.

8.5.1 Maximax and maximin

Maximax chooses the alternative with the largest *best-case* payoff. The best cases for Develop, Company A, Company B, and Nothing are $3, 0.75, 1.2, 0$; maximax chooses Develop. It highlights upside but can overlook severe losses.

Maximin chooses the alternative with the largest *worst-case* payoff. The worst cases are $-2, 0.75, 0, 0$; maximin chooses Company A. It protects the worst case but can miss an attractive opportunity.

8.5.2 Criterion of realism and equally likely

The **criterion of realism** (Hurwicz rule) places a weight $\alpha \in [0, 1]$ on each alternative's best payoff and $1 - \alpha$ on its worst payoff:

$$H_j(\alpha) = \alpha \max_i V_i(j) + (1 - \alpha) \min_i V_i(j).$$

The parameter α is an optimism *weight*, not necessarily the probability of a good market. For the four alternatives, the scores are $5\alpha - 2,\ 0.75,\ 1.2\alpha,\ 0$. At $\alpha = 0.4$, these become $0, 0.75, 0.48, 0$, so Company A wins. The weight balances extremes but can be difficult to justify empirically.

The **equally likely** (Laplace) rule treats the two states as having equal weights when no probability estimate is available. The average payoffs are $0.5, 0.75, 0.6, 0$, so it also chooses Company A. This is an *assumption for a decision rule*, not evidence that the markets truly have probability 0.5 each.

8.5.3 Minimax regret

Regret or **opportunity loss** in a state is the difference between the best payoff achievable in that state and the payoff of the chosen alternative:

$$R_i(j) = \max_k V_i(k) - V_i(j).$$

The state-best payoffs are 3 in a good market and 0.75 in a bad market. Subtracting each cell from its column best gives Table 8.2.

Table 8.2: Entrepreneur opportunity loss ($ million)

Alternative	Good market	Bad market	Largest regret
Develop the product	0	2.75	2.75
Sell to Company A	2.25	0	2.25
Sell to Company B	1.8	0.75	1.8
Do nothing	3	0.75	3

Minimax regret selects the alternative with the smallest largest regret, Company B in this case. Compare this with Figure 8.1: Company B is never the highest-EV choice for any specified good-market probability. That is not an arithmetic contradiction. The two rules optimize different quantities. It is a warning that a decision motivated chiefly by anticipated regret can forgo expected payoff if probabilities become available. The lecture describes this emotion-driven possibility as an "animal spirit" in decision making.

8.6 Thompson Lumber: expected value and sensitivity

John Thompson's lumber company is considering a line of backyard storage sheds. It can construct a large plant, construct a small plant, or construct no plant. The favorable and unfavorable markets initially have probabilities 0.50 each.

For example, the small plant's EV is $100{,}000(0.5) - 20{,}000(0.5) = \$40{,}000$. The large plant's EV is $200{,}000(0.5) - 180{,}000(0.5) = \$10{,}000$. With no plant, the EV is zero. Hence the small plant is best when $p = 0.5$.

Table 8.3: Thompson Lumber payoffs and expected values ($)

Alternative	Favorable	Unfavorable	EV at $p = 0.5$
Large plant	200,000	$-180,000$	10,000
Small plant	100,000	$-20,000$	40,000
No plant	0	0	0

Now let p denote the favorable-market probability. The EV functions are

$$\begin{aligned} \text{EV(Large)} &= 200{,}000p - 180{,}000(1-p) = 380{,}000p - 180{,}000, \\ \text{EV(Small)} &= 100{,}000p - 20{,}000(1-p) = 120{,}000p - 20{,}000, \\ \text{EV(None)} &= 0. \end{aligned}$$

The first relevant switch solves $120{,}000p - 20{,}000 = 0$, giving $p = 1/6 \approx 0.167$. The second solves $120{,}000p - 20{,}000 = 380{,}000p - 180{,}000$, giving $p = 8/13 \approx 0.615$. Therefore no plant has the highest EV below 1/6, a small plant between 1/6 and 8/13, and a large plant above 8/13. Adjacent alternatives tie at each exact boundary. When a probability estimate is near a boundary, its uncertainty matters more to the choice than when it is far from any boundary.

8.7 The newsvendor: construct the payoff, then choose

A newsvendor must order $Q = 300, 400$, or 500 newspapers before knowing daily demand D. Demand is $300, 400, 500$ with probabilities $0.5, 0.3, 0.2$. Each paper costs \$2, a sold paper earns \$4, and an unsold paper can be returned for \$1. Unmet demand is lost; there is no backorder payoff. The profit formula is

$$\Pi(Q, D) = 4\min(Q, D) + 1\max(Q - D, 0) - 2Q. \tag{8.4}$$

The first term is sales revenue, the second is return value for unsold papers, and the last is the cost of *all papers ordered*. For example, ordering 400 when demand is 300 gives $4(300) + 1(100) - 2(400) =$ \$500; ordering 500 when demand is 400 gives $4(400) + 1(100) - 2(500) =$ \$700.

Table 8.4: Newsvendor profits and expected values ($)

Order quantity	$D = 300$	$D = 400$	$D = 500$	EV
300	600	600	600	600
400	500	800	800	650
500	400	700	1,000	610
Probability	0.5	0.3	0.2	

The EV of ordering 400 is $500(0.5) + 800(0.3) + 800(0.2) =$ \$650. Since the two higher-demand payoffs are identical, this can also be written $500(0.5) + 800(0.5) =$ \$650. The EV of ordering 500 is $400(0.5) + 700(0.3) + 1000(0.2) =$ \$610; ordering 300 gives \$600. Therefore order 400 newspapers to maximize expected profit.

8.8 What is perfect information worth?

Perfect information reveals the state *before* the decision must be made and does so without error. The **expected value with perfect information** (EVwPI) first selects the best action separately for each state, then averages those state-best payoffs:

$$\text{EVwPI} = \sum_i p_i \max_j V_i(j). \tag{8.5}$$

Without that information, the best available action must be selected before the state is observed:

$$\text{EVwoPI} = \max_j \sum_i p_i V_i(j), \qquad \text{EVPI} = \text{EVwPI} - \text{EVwoPI}\,. \tag{8.6}$$

The position of the maximum relative to the probability-weighted sum is the key distinction. EVPI is nonnegative and is an upper bound on what a risk-neutral decision maker should pay for truly perfect information under the stated model. An imperfect forecast generally has less value.

For Thompson Lumber, the best favorable-market payoff is \$200,000 from a large plant; the best unfavorable-market payoff is \$0 from no plant. Hence EVwPI $= 200{,}000(0.5) + 0(0.5) =$ \$100,000. The best fixed decision without information is the small plant, worth \$40,000 in EV. The additional value is EVPI $= 100{,}000 - 40{,}000 =$ $\boxed{\$60{,}000}$.

For the newsvendor, knowing demand before ordering permits $Q = 300$ when $D = 300$, $Q = 400$ when $D = 400$, and $Q = 500$ when $D = 500$. The state-best profits are \$600, \$800, and \$1,000. Thus

$$\text{EVwPI} = 600(0.5) + 800(0.3) + 1000(0.2) = \$740, \qquad \text{EVPI} = 740 - 650 = \boxed{\$90}.$$

This is the gain above ordering 400 every day, not the total profit from perfect information. The calculation assumes demand can be learned before the order is placed; knowing it only after ordering has no such decision value.

8.9 Common mistakes and practical checks

- **Subtracting development cost twice.** Stephen's 3 and -2 million table entries are already net of the \$7 million cost.
- **Inventing a market probability under uncertainty.** An equal weight in the Laplace rule or an optimism weight in the realism rule is a chosen convention, not an observed probability.
- **Confusing EV with a certain payoff.** At $p = 0.6$, Stephen's Develop EV is \$1 million, but the realized payoff is either \$3 million or a \$2 million loss.
- **Using the wrong comparison for regret.** Regret is computed within a *column*: state-best payoff minus that cell's payoff.
- **Treating a low-regret choice as automatically high-EV.** Company B illustrates why the objective must be named.
- **Mixing money units in a utility calculation.** The coefficient a must use the inverse of the unit used for w.
- **Choosing after the state without paying for information.** Only EVwPI allows state-contingent actions; EVwoPI commits to one action.
- **Omitting salvage value or charging only sold units.** In the newsvendor case, returned papers earn \$1 each, but all ordered papers cost \$2 each.

8.10 Practice questions

1. For Stephen's product, compute the four expected payoffs when the good-market probability is 0.4. Which alternative has the largest EV?
2. At $p = 0.55$, which entrepreneur alternatives tie? Why is Company B not selected by the EV rule at that point?
3. Starting from Table 8.1, calculate the four minimax-regret row maxima without looking at Table 8.2.
4. For Thompson Lumber, which alternative is best by EV when the favorable-market probability is 0.7? Compute its EV.
5. Explain why a concave utility function gives the \$100/\$0 gamble a certainty equivalent below \$50 for a risk-averse person.
6. Use Table 8.4 to calculate EVwPI and EVPI for the newsvendor. State when the demand information must arrive.

Selected answers

1. At $p = 0.4$, EVs (\$ million) are Develop 0, Company A 0.75, Company B 0.48, and Nothing 0. Choose Company A.
2. Develop and Company A each yield 0.75 million; Company B yields $1.2(0.55) = 0.66$ million.
3. The largest regrets (\$ million) are $2.75, 2.25, 1.8, 3$, so minimax regret selects Company B.
4. Large plant: $380{,}000(0.7) - 180{,}000 = \$86{,}000$. Small plant gives \$64,000; no plant gives zero.
5. Concavity implies $U(50) > [U(0) + U(100)]/2$, so the sure amount that equals the lottery's expected utility must be less than \$50.
6. EVwPI is \$740 and EVPI is \$90. Demand must be revealed before choosing the order quantity.

8.11 Chapter summary

- A payoff table separates alternatives under the decision maker's control from uncertain states outside that control.
- With probabilities, maximize expected monetary value when money EV is the objective; sensitivity analysis identifies where the preferred choice changes.
- Without probabilities, maximax, maximin, realism, equally likely, and minimax regret can recommend different actions because they embody different criteria.
- Concave utility captures risk aversion; a certainty equivalent turns a risky prospect into a sure amount with equal utility.
- EVPI equals the best state-contingent expected payoff minus the best fixed-action expected payoff. It measures the maximum expected improvement from knowing the state before deciding.

Chapter 9

Decision Trees and the Value of Information

Source week: Week 11

9.1 Chapter overview

Many decisions are not made just once. A student may decline a job offer today, learn whether a better offer arrives next month, and then choose again. A firm may buy a market survey, observe its result, and only then decide how much capacity to build. A **decision tree** makes this sequence explicit. It distinguishes choices from events, records what is known at each choice, and works backward from eventual outcomes to find a strategy.

This chapter develops trees through Bill Sampra's summer-job decision and Thompson Lumber's plant and survey decision. You will learn to draw a valid tree, compute expected monetary values (EMVs), roll it back, check a recommendation's sensitivity, update probabilities with Bayes' theorem, and distinguish the value of imperfect information from the value of perfect information. A behavioral example also shows why a plan based on future self-discipline may fail.

9.2 Structure a sequential decision

A **decision node**, drawn as a square, is a point where the decision maker chooses among feasible actions. An **event node**, drawn as a circle, is a point where an uncertain outcome occurs. Branches from an event node have probabilities; branches from a decision node are actions, not probabilities. Terminal branches end in payoffs or utilities.

For a tree to represent the actual problem, check five things:

1. Time flows in chronological order, conventionally from left to right.
2. Each decision node lists the alternatives available *at that time*.
3. Event branches are mutually exclusive and collectively exhaustive.
4. Probabilities leaving each event node sum to one.
5. Every terminal branch has a numerical consequence in a consistent unit.

The order of nodes prevents a common error: choosing an action as though the decision maker already knew an outcome that had not occurred. A tree should also identify a **strategy**, not only a first action. A strategy states what to do after each possible later observation.

9.2.1 Bill Sampra's job offers

Bill, an SMU undergraduate, must decide at the end of October whether to accept J&J's $12,000 offer for a 12-week summer job. The offer expires. If he rejects it, Maersk may offer $14,000 around mid-November with probability 0.60. If no Maersk offer arrives, Bill can seek a job at the January–February campus job fairs. If Maersk does offer a job, Bill may accept it or try the fairs instead. The fair salaries and Bill's probabilities are shown in Table 9.1.

Table 9.1: Bill's possible 12-week job-fair payoffs

Salary ($)	Probability	Contribution to EMV ($)
21,600	0.05	1,080
16,800	0.25	4,200
12,000	0.40	4,800
6,000	0.25	1,500
0	0.05	0
Total	1.00	11,580

The job-fair event node has value

$$J = 0.05(21{,}600) + 0.25(16{,}800) + 0.40(12{,}000) + 0.25(6{,}000) + 0.05(0) = \$11{,}580. \tag{9.1}$$

This number is an expected payoff, not a guaranteed fair salary. Once Bill has a Maersk offer, the later decision node chooses the larger of $14,000 and $11,580, namely Maersk. Figure 9.1 shows the compact tree after the job-fair distributions are rolled back to their EMV.

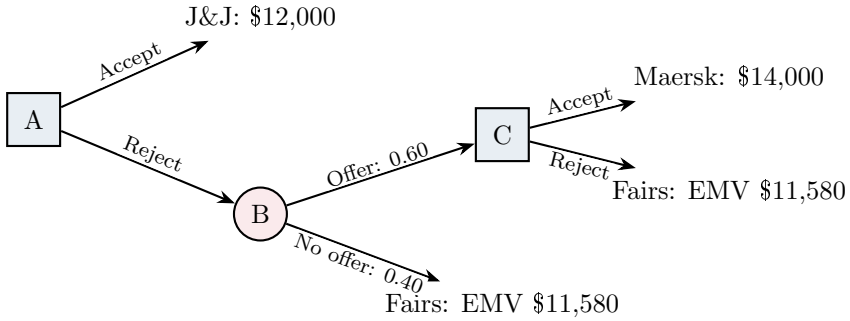


Figure 9.1: Bill's simplified tree. Each "Fairs" terminal stands for the five-outcome event in Table 9.1.

9.3 Solve a tree by rolling it backward

Rollback starts at terminal payoffs and proceeds toward the root. At an event node, calculate a probability-weighted average. At a decision node, choose the branch with the highest value when

the objective is to maximize EMV:

$$\text{Event node: } \sum_i p_i v_i, \qquad \text{Decision node: } \max_j v_j. \tag{9.2}$$

The branch chosen at a decision node also records one component of the optimal contingent strategy. The value at the first, or *root*, decision node is the EMV of that strategy.

For Bill, the Maersk decision node is worth $C = \max(14{,}000, 11{,}580) = \$14{,}000$. If he rejects J&J, the mid-November event node is

$$B = 0.60(14{,}000) + 0.40(11{,}580) = \$13{,}032. \tag{9.3}$$

At the October root, compare \$13,032 with J&J's sure \$12,000. Bill's EMV-maximizing strategy is to **reject J&J now, accept Maersk if offered, and attend the fairs if it is not offered**. Describing only "reject J&J" would leave the later decisions unspecified.

9.3.1 Sensitivity and a reservation value

Let q be Bill's assessed probability of a Maersk offer. The value of rejecting J&J is

$$R(q) = q(14{,}000) + (1-q)(11{,}580) = 11{,}580 + 2{,}420q.$$

If $q = 0.20$, then $R(0.20) = \$12{,}064$, which still beats J&J by \$64. The probability at which Bill is indifferent solves

$$11{,}580 + 2{,}420q = 12{,}000 \quad \Longrightarrow \quad q = \frac{420}{2{,}420} \approx 0.174.$$

At the original $q = 0.60$, \$13,032 is Bill's **reservation value** for accepting an immediate J&J offer under this EMV model. A revised J&J offer of at least that amount would match or exceed the best alternative, assuming the other inputs stay fixed. This is a useful way to turn a tree into a negotiation benchmark.

9.4 When the future self may change the plan

Standard rollback assumes future decisions will be made according to the preferences and payoffs used when the tree was drawn. That assumption can fail. The lecture's ice-cream-and-gym example illustrates **time inconsistency**: the person planning today expects to exercise later, but the future self faces a stronger immediate aversion to going.

Suppose not eating ice cream gives current utility -1. Eating gives $+1$ now and creates a later gym choice. Today's self predicts that the future gym choice will be "Go" with future utility -1, rather than "Not go" with utility -3. On that prediction, eating yields $+1 - 1 = 0$, above -1, so today's self plans to eat.

At night, however, suppose the future self assesses going as -6 and not going as -3. The future self will *not* go. A **sophisticated** current self anticipates this actual future choice. Eating would then total $+1 - 3 = -2$, so not eating (-1) is the better decision now. The calculation illustrates why a current plan can be overoptimistic about future discipline.

A **commitment device** changes the future choice environment before the temptation arrives: schedule exercise, lock in a deadline, or ask someone to hold you accountable. If that arrangement makes going to the gym the future choice as modeled (utility -1 rather than a less attractive alternative), eating again totals 0. A commitment device is effective only if it actually changes later incentives or feasible actions; writing "I will go" on the tree does not by itself make the future self follow through. Similar problems arise when a business keeps postponing maintenance, cybersecurity, or technical-debt work.

9.5 Thompson Lumber: a decision before the market is known

Thompson Lumber must decide whether to build a large plant, build a small plant, or do nothing before knowing whether demand for backyard storage sheds will be favorable. The two market states initially each have probability 0.50. Table 9.2 restates the payoffs from Chapter 8 so the information problem can be solved on its own.

Table 9.2: Thompson Lumber without a survey ($)

Alternative	Favorable	Unfavorable	EMV
Large plant	200,000	−180,000	10,000
Small plant	100,000	−20,000	40,000
No plant	0	0	0

Without further information, the small plant is best: its EMV is $0.5(100{,}000) + 0.5(-20{,}000) =$ \$40,000. This is the **no-survey baseline**. An information source has value only if it can improve on this optimized baseline, not on an arbitrarily selected plant.

9.6 A fallible market survey and Bayesian updating

Thompson can buy a survey for \$10,000. Let F and U denote the actual favorable and unfavorable markets; let SF and SU denote a survey that predicts favorable and unfavorable. The survey is accurate 70% of the time when the market is favorable and 80% of the time when it is unfavorable:

$$P(SF \mid F) = 0.70, \quad P(SU \mid F) = 0.30, \qquad P(SF \mid U) = 0.20, \quad P(SU \mid U) = 0.80.$$

These are *likelihoods*: probabilities of a signal conditional on the actual state. They are not the probabilities of the actual state after seeing a signal.

9.6.1 From likelihoods to signal probabilities

Begin with $P(F) = P(U) = 0.50$. Multiplying each prior by its conditional survey probability produces the four joint probabilities in Table 9.3.

Table 9.3: Market–survey joint probabilities

Actual market	Survey SF	Survey SU	Row total
F	$0.50(0.70) = 0.35$	$0.50(0.30) = 0.15$	0.50
U	$0.50(0.20) = 0.10$	$0.50(0.80) = 0.40$	0.50
Column total	0.45	0.55	1.00

Thus $P(SF) = 0.45$ and $P(SU) = 0.55$. The result is intuitive: a favorable signal may be a true positive from a favorable market or a false positive from an unfavorable one.

9.6.2 Posterior probabilities after each signal

Bayes' theorem divides the joint probability of a state and signal by the total probability of that signal:

$$P(F \mid SF) = \frac{0.35}{0.45} = \frac{7}{9} \approx 0.78, \qquad P(U \mid SF) = \frac{0.10}{0.45} = \frac{2}{9} \approx 0.22, \tag{9.4}$$

$$P(F \mid SU) = \frac{0.15}{0.55} = \frac{3}{11} \approx 0.27, \qquad P(U \mid SU) = \frac{0.40}{0.55} = \frac{8}{11} \approx 0.73. \tag{9.5}$$

After a favorable signal, Thompson's assessed chance of a favorable market rises from 0.50 to about 0.78. After an unfavorable signal, it falls to about 0.27. A survey result changes beliefs; it does not reveal the state with certainty. Because the result is observed before the plant choice, the plant decision can depend on the signal.

9.7 Roll back the imperfect-information tree

The timeline is: choose whether to survey, observe SF or SU if surveying, choose a plant, then experience the actual market. If Thompson surveys, subtract the \$10,000 cost from *every* terminal payoff on that branch. Table 9.4 shows the resulting payoffs. The no-survey branch continues to use Table 9.2 and the 0.50/0.50 priors.

Table 9.4: Terminal payoffs after paying for the survey (\$)

Plant decision after either signal	Actual F	Actual U
Large plant	190,000	$-190,000$
Small plant	90,000	$-30,000$
No plant	$-10,000$	$-10,000$

Using the lecture slides' two-decimal *rounded* posteriors, the decision-node EMVs are

Alternative	After SF (0.78/0.22)	After SU (0.27/0.73)
Large plant	106,400	$-87,400$
Small plant	63,600	2,400
No plant	$-10,000$	$-10,000$

For example, after SF, the large plant has rounded-posterior EMV $190{,}000(0.78-0.22) = \$106{,}400$. After SU, the small plant has $0.27(90{,}000)-0.73(30{,}000) = \$2{,}400$. Therefore the *contingent* policy is to survey, build a large plant after SF, and build a small plant after SU. With those rounded node values, the survey branch's net EMV is

$$0.45(106{,}400) + 0.55(2{,}400) = \$49{,}200, \tag{9.6}$$

above the no-survey EMV of \$40,000. This reproduces the classroom slide. The \$10,000 fee has *already* been subtracted in the terminal payoffs; subtracting it again here would double-count it.

9.7.1 Why carrying exact probabilities gives a slightly different value

The slide first rounds 7/9 to 0.78 and 3/11 to 0.27, then uses those rounded values in the payoff calculation. For an exact result, retain the joint probabilities from Table 9.3. The selected policy's

pre-fee expected payoff is

$$\underbrace{0.35(200{,}000) + 0.10(-180{,}000)}_{\text{favorable signal, large plant}} + \underbrace{0.15(100{,}000) + 0.40(-20{,}000)}_{\text{unfavorable signal, small plant}} = \$59{,}000.$$

Subtracting the survey fee once yields \$49,000, rather than the slide's \$49,200. The \$200 difference is caused by rounding intermediate posterior probabilities; it does *not* change the optimal survey policy or the conclusion that surveying is worthwhile at \$10,000. For reproducible monetary calculations, keep full precision until the final answer.

9.8 The value of imperfect and perfect information

9.8.1 Value of the fallible survey

The **expected value of imperfect information** (EVII) is the improvement in the best expected payoff that a fallible signal makes possible *before paying for the information*. The lecture slides use their rounded rollback value of \$49,200, which is net of the \$10,000 survey fee. The gain *at that fee* is $49{,}200 - 40{,}000 = \$9{,}200$. Adding the fee back gives the slide's gross information value and maximum willingness to pay:

$$\text{EVII}_{\text{slide}} = 49{,}200 - 40{,}000 + 10{,}000 = \$19{,}200. \tag{9.7}$$

If full precision is retained instead, the pre-fee survey EMV is \$59,000 and $\text{EVII} = 59{,}000 - 40{,}000 = \$19{,}000$. At a \$10,000 survey price the corresponding exact net gain is \$9,000. The difference from the slide numbers is only the rounding described above, not a different definition of EVII.

More generally, with signal S, market state M, and action j,

$$\text{EVII} = \sum_s P(S = s) \max_j \mathbb{E}[V(j, M) \mid S = s] - \max_j \mathbb{E}[V(j, M)].$$

The decision maker should not pay more than this gross expected gain for the survey under the monetary-EMV objective.

9.8.2 Perfect information as an upper bound

A **perfect survey** would identify the actual market before the plant decision. After SF, $P(F \mid SF) = 1$; after SU, $P(F \mid SU) = 0$. Thompson would then build a large plant in a favorable market for \$200,000 and no plant in an unfavorable market for \$0. His expected payoff *before any survey fee* would be

$$\text{EVwPI} = 0.50(200{,}000) + 0.50(0) = \$100{,}000. \tag{9.8}$$

The **expected value of perfect information** is

$$\text{EVPI} = 100{,}000 - 40{,}000 = \boxed{\$60{,}000}. \tag{9.9}$$

Perfect information is an upper bound on an imperfect survey's value, provided both are evaluated with the same payoffs, timing, and objective. Here both the slide's \$19,200 and the exact \$19,000 EVII fall below \$60,000. An information source can always be ignored if it is unhelpful; additional information cannot reduce the best achievable expected payoff before its cost.

9.9 Common mistakes

- **Putting information after the decision.** A survey helps only if its result arrives before the plant choice it could change.
- **Averaging at a decision node.** Choose the best available branch at a square; average probabilistic branches at a circle.
- **Using an incomplete strategy.** "Reject J&J" does not specify what Bill does if Maersk calls or does not call.
- **Reversing conditional probabilities.** Survey accuracy $P(SF \mid F)$ is not the posterior $P(F \mid SF)$.
- **Using priors after a survey signal.** The plant decision after SF or SU should use the corresponding posterior.
- **Counting a survey fee twice.** Terminal payoffs on the survey branch already include the \$10,000 fee in this case.
- **Confusing net gain, EVII, and EVPI.** The slide's \$9,200 net gain is at the stated survey price; \$19,200 EVII is the rounded pre-fee information value; \$60,000 EVPI assumes a perfect signal.
- **Rounding too early.** The lecture's rounded posterior probabilities shift the survey result by \$200; exact joint probabilities give the reproducible \$49,000 net EMV.
- **Assuming future follow-through.** If preferences change, rollback should use the future self's likely action or model a real commitment device.

9.10 Practice questions

1. Use Table 9.1 to verify the job-fair EMV. Then calculate Bill's root EMV if the Maersk-offer probability is 0.30. Would he accept J&J's \$12,000 offer?
2. At an event node, three exhaustive outcomes have probabilities $0.2, 0.5, 0.3$ and payoffs $10, 4, -2$ in thousands of dollars. What value should be rolled back? How would a decision node with a sure \$5,000 alternative use that value?
3. Explain why $P(SF \mid F) = 0.70$ does not mean $P(F \mid SF) = 0.70$ in Thompson's example. Calculate the latter.
4. Using the slide's two-decimal posteriors, calculate the small plant's EMV after SU with the survey cost included. Which plant choice is best after that signal?
5. Using the *exact* joint probabilities, calculate the gross payoff, net payoff at a \$10,000 survey fee, and EVII for Thompson's optimal survey policy.
6. Calculate Thompson's EVPI. Explain why the survey's price and information accuracy must be kept separate in the comparison.

Selected answers

1. The fair EMV is \$11,580. At $q = 0.30$, rejecting J&J gives $0.30(14{,}000) + 0.70(11{,}580) =$ \$12,306; reject J&J.

2. The event-node value is $0.2(10) + 0.5(4) + 0.3(-2) = 3.4$ thousand dollars. The decision node chooses the sure \$5,000 instead.

3. The first probability conditions on the actual market; the second conditions on the signal. $P(F \mid SF) = 0.35/0.45 = 7/9 \approx 0.78$.

4. $0.27(90{,}000) + 0.73(-30{,}000) = \$2{,}400$. It exceeds the large plant's $-\$87,400$ and no plant's $-\$10,000$ after the unfavorable signal.

5. Gross survey EMV is \$59,000; net survey EMV is \$49,000; exact EVII is $59{,}000 - 40{,}000 =$ \$19,000.

6. EVPI is $100{,}000 - 40{,}000 =$ \$60,000. The survey's accuracy determines its possible decision improvement; its price is then subtracted to judge the net benefit of buying it.

9.11 Chapter summary

- A decision tree separates choices (squares) from uncertain events (circles) and places them in the order they occur.
- Rollback averages at event nodes and chooses the best branch at decision nodes; the result is a contingent strategy and its root EMV.
- Sensitivity analysis tests whether a recommendation survives changes in uncertain inputs, while a reservation value helps compare or negotiate an immediate offer.
- Time-inconsistent preferences can invalidate a naive future plan; sophisticated rollback and real commitment devices address this in different ways.
- A fallible survey changes market beliefs through Bayes' theorem. Its value comes from choosing an action after the signal, using the appropriate posterior probabilities.
- EVII is the gross expected improvement from imperfect information; EVPI is the improvement from perfect information and provides an upper bound under the same decision model.

Chapter 10

Simulation for Decisions under Uncertainty

Source week: Week 12

10.1 Chapter overview

A business decision is often made before its most important inputs are known. An insurer sets a premium before learning how many claims will arrive. A baker chooses how many cakes to make before seeing the day's customers. A newspaper vendor orders copies before demand is known. We can calculate an expected value directly for a small model with a few outcomes. A realistic model may have many interacting uncertain inputs, business rules, and competing goals. In those settings, deriving a best decision directly can be difficult.

Simulation offers another route. Specify a candidate decision, generate possible values for the uncertain inputs, calculate the consequence of each set of inputs, and repeat. The resulting outcomes describe both what happens *on average* and how often an unfavorable event occurs. Comparing these summaries across candidate decisions helps a manager choose. Simulation evaluates the decisions included in the model; it does not make the input assumptions true or automatically search every possible decision.

This chapter begins with an illustrative laptop-sales distribution to show how a random draw becomes a business input. It then builds an annual model of the Scottish Ministers' Widows' Fund, followed by Maria's Bakery and a newsvendor. The fund's historical setting motivates the business question; its numerical probabilities are simplified classroom assumptions, not historical estimates.

After studying the chapter, you should be able to:

- map a uniform random number into an outcome from a discrete distribution;
- define a complete replication and estimate a mean from many replications;
- simulate age-group death counts with a binomial model;
- distinguish an average break-even policy from one with low deficit risk;
- write sales and profit formulas that respect available inventory; and
- compare candidate decisions using simulated outcomes while checking the quality of the model's inputs.

10.2 Build a simulation around a decision

The first step is to say what the manager controls. Call that decision q. Next identify uncertain inputs, represented by X, and a rule $g(q, X)$ that turns the decision and inputs into an outcome Y:

$$Y = g(q, X). \tag{10.1}$$

The rule is the **business logic**. It might calculate annual surplus, daily profit, unsold inventory, or shortages. A complete model specifies the distribution of X, the timing of q, and any dependence among uncertain inputs. Table 10.1 applies this structure to the cases in this chapter.

Table 10.1: Decision, uncertainty, and outcome in three simulation problems

Case	Decision q	Uncertain input X	Calculated outcome Y
Insurance fund	Annual premium per member	Death counts in age groups	Annual surplus and deficit indicator
Maria's Bakery	Cakes baked each morning	Regular demand; possibly bargain demand	Daily profit, discards, and shortages
Newsvendor	Papers ordered each morning	Daily newspaper demand	Sales, returns, and daily profit

A **replication** is one complete simulated instance of the system. For the fund it is one possible annual outcome across all age groups, not one person's death draw. For the bakery it is one possible day. If the decision is held fixed and the replications are independent, they provide a sample from that decision's outcome distribution. We can then change the decision and repeat the evaluation.

10.3 Turn random numbers into business outcomes

10.3.1 A discrete weekly-sales example

Suppose X is the number of laptops sold by a small store in one week. The classroom distribution in Table 10.2 is illustrative; assume enough laptops are in stock so potential demand becomes sales. It assigns a probability to each possible value of X.

Table 10.2: Illustrative distribution for weekly laptop sales

Laptops sold, x	Probability $P(X = x)$	Uniform interval	Two-digit bin
0	0.20	$[0, 0.20)$	00–19
1	0.40	$[0.20, 0.60)$	20–59
2	0.20	$[0.60, 0.80)$	60–79
3	0.10	$[0.80, 0.90)$	80–89
4	0.10	$[0.90, 1.00)$	90–99
Total	1.00		

Imagine a random number U drawn uniformly from $[0, 1)$: every subinterval of a given length is equally likely. **Partition** the range into intervals whose lengths match the five probabilities, then **draw** U and **read** the corresponding sales value. For example, $U = 0.39$ lies between 0.20 and

0.60, so that simulated week has one laptop sold. A draw of 0.73 lies between 0.60 and 0.80, giving two laptops. The 00–99 bins are the same construction on a 100-number scale.

The expected weekly sales from the stated distribution are

$$E[X] = 0(0.20) + 1(0.40) + 2(0.20) + 3(0.10) + 4(0.10) = 1.5. \tag{10.2}$$

No particular week can sell 1.5 laptops in this model. The number is an average over many possible weeks. A short simulated run can have a sample average above or below 1.5.

10.3.2 Use RAND and a cumulative lookup in Excel

Excel's `=RAND()` returns a decimal in $[0, 1)$. Put the cumulative *lower bounds* 0, 0.20, 0.60, 0.80, and 0.90 in an ascending column, with sales values 0, 1, 2, 3, and 4 beside them. If the draw is in B9 and the two-column lookup range is D2:E6, use

```
=VLOOKUP(B9,$D$2:$E$6,2,TRUE).
```

The `TRUE` argument requests an approximate match. Excel finds the largest lower bound that does not exceed B9 and returns the corresponding sales value. The bounds must be sorted upward. An exact-match lookup would usually fail because a RAND draw is unlikely to equal one of the five bounds. `RANDBETWEEN(0,4)` would also be wrong here: it assigns equal probability to all five integers rather than the stated 20/40/20/10/10 percent distribution.

Copy the RAND and lookup formulas down one row for each simulated week. A recalculation produces a different set of draws. Keep the original probability table visible so a reader can check that the generated outcomes match the intended distribution.

10.4 Replicate, average, and interpret

Let $Y_1, \ldots, Y_R$ be the outcomes from R independent replications of one fixed decision. Their **sample mean** is

$$\bar{Y}_R = \frac{1}{R} \sum_{r=1}^{R} Y_r. \tag{10.3}$$

For laptop sales, a running average after r weeks is $\bar{X}_r = (X_1 + \cdots + X_r)/r$. Plotting $\bar{X}_r$ as more weeks are simulated shows how the estimate changes. A one-week "average" is just one random outcome. Ten weeks remain noisy. With many independent draws from a stable distribution, the average tends to settle near the distribution's mean, although it does not move smoothly or become exact at a fixed replication count.

The approximate **Monte Carlo standard error** of a simulated mean is $s_R/\sqrt{R}$, where s_R is the sample standard deviation of the R outcomes. More replications reduce sampling noise; roughly four times as many replications are needed to halve this standard error. This formula describes uncertainty caused by using a finite simulation run. It does *not* repair a wrong demand distribution, a biased mortality estimate, or incorrect business logic. Those are modeling errors, not Monte Carlo noise.

An average alone can hide unfavorable outcomes. When a decision maker cares about a threshold, also calculate its frequency. For example, if $L_r = 1$ when replication r has a loss and $L_r = 0$ otherwise, then $R^{-1} \sum_r L_r$ estimates the probability of a loss. A histogram of outcomes can show both the common cases and the lower tail that a single mean conceals.

10.5 Insurance premiums: the Scottish Ministers' Widows' Fund

The historical fund provides an early business motivation for using a model to plan contributions and benefits under uncertainty. Setting a sustainable premium requires an estimate of how many claims may occur. Collecting relevant mortality records and calibrating the probabilities are substantial parts of the work. The following numerical model is a classroom simplification: it is not a reconstruction of the fund's historical premiums or benefit rules.

10.5.1 Model inputs and an annual death count

Suppose the fund has 100 ministers divided into four age groups, as in Table 10.3. Each person in a group has the listed annual death probability. These probabilities are illustrative inputs, not historical estimates.

Table 10.3: Illustrative annual mortality inputs for the fund

Age group	Ministers n_i	Death probability p_i	Expected deaths $n_i p_i$
30–39	20	0.005	0.10
40–49	30	0.010	0.30
50–59	30	0.020	0.60
60–69	20	0.040	0.80
Total	100		1.80

For age group i, let D_i be the number of deaths in one year. The classroom model treats each person's annual death event as an independent yes/no trial with the group's fixed probability:

$$D_i \sim \text{Binomial}(n_i, p_i), \qquad D = \sum_{i=1}^{4} D_i. \tag{10.4}$$

For the 50–59 group, $D_i \sim \text{Binomial}(30, 0.02)$. Excel can generate one possible count with `=BINOM.INV(30,0.02,RAND())`. The expected count is $30(0.02) = 0.6$, not a possible count in one year. Repeat the binomial draw for each age group using its own n_i, p_i, then add the four results. Because the groups have different probabilities, the total is not modeled as one Binomial$(100, p)$ count with a single shared p.

One possible year might produce group counts $0, 0, 1, 2$: three deaths in total. A different replication may produce none or more than three. The model assumes the group sizes and probabilities are reset for each independent replication. Thus "1,000 replications" means 1,000 possible annual outcomes under the same policy, not a forecast that follows one fund through 1,000 calendar years of aging members.

10.5.2 Claims, premium income, and the annual result

Suppose the fund pays £500 per death and charges each of the 100 ministers an annual premium P, measured in pounds. The annual claims and surplus are

$$C = 500D, \qquad I(P) = 100P, \qquad S(P) = 100P - 500D. \tag{10.5}$$

For the illustrative three-death year, claims are £1,500. At a £9 premium, income is £900 and the annual result is a £600 deficit. The premium needed to cover *that year's* claims is $1{,}500/100 = £15$

per member. This number changes with the death count, which is why one replication cannot set a reliable premium.

The expected death count from Table 10.3 is 1.8, so expected annual claims are $500(1.8) = £900$. Consequently,

$$E[S(P)] = 100P - 900, \qquad E[S(9)] = 0. \tag{10.6}$$

The £9 premium is **break-even on average** in this model. It does not guarantee a nonnegative surplus each year. Two deaths cause £1,000 of claims and a £100 deficit at this premium. Indeed, the deficit event is $D \geq 2$, which has probability about 54 percent under the stated group model. A simulation estimates that probability by counting the fraction of replications with $S(9) < 0$.

Table 10.4: Illustrative premium comparison under the classroom fund model

Premium P (£)	Expected surplus (£)	Approx. deficit chance	Deficit when
9	0	54.0%	$D \geq 2$
10	100	26.9%	$D \geq 3$
15	600	10.6%	$D \geq 4$
20	1,100	3.4%	$D \geq 5$

Higher premiums raise expected surplus and reduce the chance of an annual deficit in this fixed model, but affordability and fairness also matter. A manager must decide what risk is acceptable; simulation does not supply that preference. Moreover, an **annual deficit** means this year's premium income is below this year's claims. It is not automatically **insolvency**: reserves or investment income may cover a shortfall. A full solvency model would need to track reserves, timing, membership changes, and other benefits over multiple years.

10.5.3 Input quality is part of the decision

The historical problem of gathering mortality data illustrates a general rule: *garbage in, garbage out*. If probabilities are estimated from an irrelevant group or a small, biased sample, thousands of replications can produce a precise-looking answer to the wrong model. Before using the result, ask whether the age groups, probabilities, assumed independence, benefit amount, and fixed membership reflect the actual decision. Testing how recommendations change when uncertain inputs vary is often as valuable as increasing the number of replications.

10.6 Maria's Bakery: convert demand into profit

Maria must choose how many cakes to bake before observing the day's regular demand R. Each cake costs \$3 to make and sells for \$10; unsold cakes are discarded in the base policy. Table 10.5 gives the classroom demand distribution.

Table 10.5: Maria's daily regular demand

Demand R	5	10	15	20	25
Probability	0.05	0.20	0.40	0.20	0.15

If Maria bakes Q cakes, the number sold cannot exceed either supply or demand:

$$\begin{aligned} \text{Regular sales} &= \min(Q, R), \\ \text{Discarded cakes} &= \max(Q - R, 0), \\ \text{Unmet demand} &= \max(R - Q, 0), \\ \pi_0(Q, R) &= 10 \min(Q, R) - 3Q. \end{aligned} \tag{10.7}$$

For the starting policy $Q = 20$ and a demand draw $R = 15$, Maria sells 15, discards five, and earns $15(10) - 20(3) = \$90$. For a large number of demand replications, average the daily profit cells to estimate expected profit at $Q = 20$, then change Q to compare other policies. As a check available in this small model, the expected number sold at $Q = 20$ is 15.25, giving expected daily profit $10(15.25) - 60 = \$92.50$. The simulated average will usually be near, but not exactly equal to, that value.

10.6.1 A discount market for leftovers

Now suppose Maria first sells to regular buyers at \$10, then offers any leftovers to bargain hunters at \$5 before discarding the remainder. The number of bargain hunters H is 5, 10, or 15 with probabilities 0.45, 0.30, and 0.25. The supplied distributions do not specify how regular demand and bargain demand move together, so the classroom spreadsheet *assumes independent draws*. That assumption should be reviewed if the model is used for a real shop.

The order of operations matters. Let

$$\begin{aligned} S_R &= \min(Q, R), & L &= Q - S_R, \\ S_H &= \min(L, H), & U &= L - S_H. \end{aligned} \tag{10.8}$$

Here S_R is regular sales, L is inventory left after regular sales, S_H is discount sales, and U is the final discarded amount. Profit becomes

$$\pi_1(Q, R, H) = 10S_R + 5S_H - 3Q. \tag{10.9}$$

If $Q = 20$, $R = 15$, and $H = 10$, then regular sales are 15 and only five cakes remain. Discount sales are therefore $\min(5, 10) = 5$, not 10. The day's profit is $15(10) + 5(5) - 20(3) = \$115$. Under the independent classroom distributions, the exact expected number of discount sales at $Q = 20$ is four, so expected profit is $10(15.25) + 5(4) - 60 = \$112.50$. A 1,000-replication workbook provides a simulated estimate and makes the sequence of sales visible row by row.

10.7 The newsvendor: compare order quantities

A street-corner vendor pays \$2 for each newspaper ordered in the morning. Each paper sold earns \$4, and each unsold paper can be returned for a \$1 credit. Daily demand D is 20, 22, 24, 26, 28, or 30 with probabilities 0.05, 0.10, 0.25, 0.30, 0.20, and 0.10, respectively. The decision Q is the morning order quantity. For one day,

$$\pi(Q, D) = 4 \min(Q, D) + \max(Q - D, 0) - 2Q. \tag{10.10}$$

If $Q = 26$ and $D = 22$, the vendor sells 22 and returns four. Profit is $22(4) + 4(1) - 26(2) = \$40$. Omitting the return credit would understate profit, while recording sales of 26 would sell four papers for which no customer appeared.

For each simulated demand draw, calculate profit under every candidate order quantity using the *same* demand. This makes the candidate comparison easy to inspect: the policies face identical simulated days. Average each policy's profit across replications. The exact averages in Table 10.6 are a check for this small six-outcome model; the workbook's finite-sample averages fluctuate.

Table 10.6: Expected daily profit for the listed newsvendor candidates

Order Q	20	22	24	26	28	30
Expected profit ($)	40.00	43.70	46.80	48.40	48.20	46.80

Among the listed candidates, ordering 26 maximizes expected daily profit under the supplied distribution. The advantage over ordering 28 is only $0.20 per day. A 1,000-day simulation can occasionally rank these two policies in the opposite order because of sampling noise. Repeating the simulation and increasing the replication count helps distinguish a small profit difference from random variation. Even then, the conclusion is conditional on the stated demand distribution, prices, and return policy.

10.8 A reliable spreadsheet workflow

The companion workbooks use one row per replication. A student can build a similar model with the following sequence:

1. **Separate inputs from calculations.** Put probabilities, prices, group sizes, benefit amounts, and candidate decisions in labeled cells. Check that each discrete distribution sums to one.
2. **Generate uncertain inputs.** Use `RAND()` with a sorted cumulative lookup for discrete demand. Use `BINOM.INV(n,p,RAND())` for an age group's binomial death count. Draw each input needed for a complete replication.
3. **Apply business logic in visible steps.** Calculate sales, leftovers, claims, revenue, cost, and profit or surplus in separate columns when doing so makes the rule auditable. Use `MIN` to cap sales at available supply and demand.
4. **Repeat rows and summarize.** Copy formulas through the chosen number of replications. Calculate average profit or surplus, as well as loss frequency, discards, or shortages when those outcomes matter.
5. **Compare policies and check the model.** Change the decision input or calculate several candidates on the same random draws. Hand-check a simple row, test boundary cases, and compare a small example with a direct calculation when one is available.

A spreadsheet that returns a number is not necessarily a correct model. Check, for example, that a lookup's lower bounds are sorted, that probabilities and units match the intended period, and that every cost or return credit enters profit with the correct sign. In the fund, inspect both the expected surplus and the share of deficit replications. In the bakery, verify that regular sales never exceed Q, discount sales never exceed leftovers, and discarded cakes never become negative.

10.9 Common mistakes

- **Treating one replication as an average.** One possible year or day can be unusually good or bad; estimate an average from many draws.

- **Calling 1,000 independent replications 1,000 calendar years.** The fund classroom model resets its age mix each time and does not age a single cohort through time.
- **Assuming more replications fix poor inputs.** Repetition reduces sampling noise, not errors in estimated probabilities or logic.
- **Using the wrong lookup rule.** Approximate-match VLOOKUP requires sorted cumulative lower bounds for the discrete demand model.
- **Using one death probability for every minister.** The four age groups have different probabilities and need separate draws.
- **Equating break-even on average with safety.** At a £9 premium, the fund's expected surplus is zero but many individual years have deficits.
- **Selling more units than are available.** Both regular and discount sales must be capped by remaining inventory and the relevant demand.
- **Confusing the sample winner with a guaranteed optimum.** A finite simulation can misrank close candidates, and the best tested candidate need not be best among all possible choices.

10.10 Practice questions

1. Use Table 10.2 to map uniform draws 0.39 and 0.73 to weekly laptop sales. Calculate the distribution's expected sales.
2. Explain why `RANDBETWEEN(0,4)` does not generate the laptop sales distribution in Table 10.2.
3. Find the expected annual deaths in the fund's 50–59 age group and across all four age groups. Why can neither expectation be read as the death count in one particular year?
4. A fund replication has three deaths. Calculate claims, the premium per member needed to cover those claims, and the annual result at a £9 premium.
5. Under a £10 premium, how many deaths are required for an annual deficit? Explain why two deaths do not count as a deficit.
6. Maria bakes 20 cakes and regular demand is 15. Calculate base-policy sales, discards, shortages, and profit.
7. With the same bake quantity and regular demand, suppose ten bargain hunters arrive. How many cakes can be sold at the discount price, and what is total profit?
8. A newsvendor orders 26 papers and demand is 22. Calculate sales, returns, and profit.
9. Why might a 1,000-replication newsvendor simulation occasionally prefer $Q = 28$ to $Q = 26$, even though Table 10.6 places $Q = 26$ first?

Selected answers

1. The draws map to one and two laptops, respectively. Expected sales are 1.5 laptops per week.
2. `RANDBETWEEN` makes 0, 1, 2, 3, and 4 equally likely; the stated probabilities differ.
3. The group mean is $30(0.02) = 0.6$; the four-group total is 1.8. A realized count must be a nonnegative whole number and varies by year.

4. Claims are £1,500. Covering them requires £15 per member. At £9, surplus is $900 - 1{,}500 = -600$ pounds, a £600 deficit.
5. Income is £1,000. Two deaths cost exactly £1,000, so they break even; three or more cause a deficit.
6. Maria sells 15, discards five, has no unmet demand, and earns $15(10) - 20(3) = \$90$.
7. Only five cakes remain after regular sales. Discount sales are five, and total profit is $15(10) + 5(5) - 20(3) = \$115$.
8. Sales are 22 and returns are four. Profit is $22(4) + 4(1) - 26(2) = \$40$.
9. The expected-profit gap is only \$0.20 per day. Finite random samples contain Monte Carlo noise, so the sample ranking can vary.

10.11 Chapter summary

- Simulation evaluates a decision by sampling uncertain inputs, applying business logic, and repeating the complete calculation.
- A cumulative probability lookup turns a uniform random draw into an outcome from an unequal-probability discrete distribution.
- Repeated independent replications estimate averages and risk frequencies. A running average can stabilize, but model assumptions still need scrutiny.
- The insurance fund's £9 premium breaks even on average under the classroom inputs; an annual deficit remains possible and is different from insolvency.
- Bakery and newsvendor profit models must cap sales at both demand and inventory. Candidate policies are compared using their distributions of profit, not one favorable day.
- Simulation supports a managerial choice, while input calibration, model checks, risk tolerance, and the range of tested decisions determine how trustworthy and useful the recommendation is.

Chapter 11

The Psychology of Human Decisions

Source week: Week 13

11.1 Chapter overview

Earlier chapters showed how to define alternatives, calculate expected values, express risk preferences, build decision trees, and simulate uncertain outcomes. Those methods remain useful. Yet a calculation cannot explain by itself which information a person notices, when a person stops searching, or why two equivalent descriptions can lead to different choices. This chapter examines the *decision process* as well as the decision model.

Three questions organize the chapter. First, how do people choose when time, information, and attention are limited? Second, how do they experience gains, losses, and probabilities? Third, how can the presentation of a choice steer behavior? The goal is to understand behavior and design better decisions, not to label every departure from an expected-value calculation a mistake.

After studying this chapter, you should be able to:

- explain satisficing, bounded rationality, rational inattention, and the useful roles of fast and deliberate thought;
- distinguish a monetary expected value from reference-dependent value and probability weighting;
- identify framing, sunk-cost influence, anchoring, decoys, and defaults in concrete decisions;
- separate an empirical finding from an illustrative classroom example;
- redesign a business choice while checking its effect on both the firm and the customer; and
- apply the same checks to your own spending and subscription choices.

11.2 Search for a workable choice

Imagine that you have interviewed with three companies for a three-month internship. One offers S$1,600 per month with a 35-minute commute each way. Its role and hours suit you, but the offer expires tomorrow. You have not heard from the other two companies. Would you accept?

The offer cannot be judged from pay alone. A student who needs at least S$1,500 per month and will accept a commute of up to 40 minutes might find it acceptable. A student whose minimum pay is S$1,800 would not. Even the first student might continue searching if a better offer is sufficiently

likely and losing the current offer is bearable. A complete calculation would require uncertain future offers, the value of each job attribute, and the cost of waiting. Those inputs may not be available before tomorrow.

11.2.1 Satisficing and bounded rationality

Satisficing is a search rule: set an *aspiration level*, look for an alternative that meets it, and stop when an acceptable alternative appears. The word combines "satisfy" and "suffice." In the internship example, minimum pay and maximum commute are two possible thresholds. The order in which offers arrive can matter: accepting the first offer that qualifies need not select the best offer that could ever arrive.

Herbert A. Simon developed this view of organizational choice. His research on decision making in economic organizations was recognized with the 1978 Nobel Memorial Prize in Economic Sciences. Simon's **bounded rationality** starts from a practical observation: people try to make sensible choices with limited information, time, and computational ability. They use search rules, routines, and **heuristics**—simple rules of thumb that reduce the work of a decision. "Reject internships with a commute above my limit" is one such rule. It can save effort, but it may also exclude an exceptional job with compensating benefits.

Bounded rationality does not mean that people choose arbitrarily, nor that every shortcut is a bias. A search process should be evaluated against the environment: how costly is further search, what may be missed, and how good are the chosen thresholds? A manager can improve the process by making important information easier to find, setting appropriate stopping rules, and revisiting rules when circumstances change.

11.2.2 Rational inattention: decide what to learn

Limited capacity also affects which facts we process. Christopher A. Sims's **rational inattention** model treats information processing as a scarce resource. A decision maker can choose which signals to examine and how precisely to understand them. Ignoring some information can be a rational allocation of capacity when its expected contribution to the decision is small. The theoretical constraint is expressed using information capacity; minutes of reading are only a classroom analogy.

Suppose you have 30 minutes to compare two laptops within a S$1,500 budget. Software compatibility and total price could immediately rule out a model, so they deserve a careful check. A review summary may be enough for battery life until it reveals a concern. A tiny benchmark difference may be worth setting aside if it cannot change your choice. A student with different needs could allocate attention differently. The decision rule is not "always read less," but *read further when the expected benefit of a better choice justifies the attention it consumes.*

The same issue appears in financial markets: an investor cannot process every price movement and report. Choosing which signals to follow is part of the decision, not merely a failure to notice everything. Rational inattention offers a model of purposeful selection; it does not prove that every omission observed in practice is optimal. Sims's 2011 economics prize recognized his macroeconomic research on cause and effect, not rational inattention specifically.

11.3 Fast responses, deliberate checks, and emotion

Consider a short production question. Five identical machines make five items in five minutes. How long do 100 identical machines, operating at the same time and at the same rate, need to make 100 items? A tempting response is 100 minutes. Checking one machine reveals the answer:

each machine makes one item in five minutes, so 100 machines make 100 items in *five minutes*. The check changes the number of machines and items together while keeping the time per item fixed.

Daniel Kahneman's *Thinking, Fast and Slow* popularized the labels **System 1** for fast, relatively automatic responses and **System 2** for effortful, deliberate reasoning. Rapid responses are useful. Moving aside when a vehicle approaches leaves little time for an explicit cost-benefit calculation; familiar routines and skilled recognition also save effort. Deliberate thought helps us compare offers, check an arithmetic result, question assumptions, and plan. A good decision process uses each where it helps. Fast thinking is not inherently wrong, and deliberate thinking can still begin from incorrect assumptions. The labels describe families of processes, not two separate anatomical brain locations or a diagnostic test of an individual student.

Neuroscience adds evidence about the feelings that accompany decisions under risk. In repeated investment decisions during fMRI, Kuhnen and Knutson (2005) found that activity associated with reward anticipation preceded risky choices, while activity associated with anticipatory negative feelings preceded safe choices. The authors also related these patterns to choices classified as excessive or insufficient risk taking *relative to their risk-neutral benchmark*. That benchmark matters: a cautious choice can be sensible for someone with a different legitimate risk preference. fMRI activity is an indirect measurement and an association; it is not a direct reading of a person's thoughts.

In another study, Shiv and coauthors (2005) compared people with lesions in emotion-related brain areas, people with lesions elsewhere, and people without lesions. In 20 repeated investment rounds, the first group invested more often in a favorable gamble and earned more. Members of the comparison groups became more reluctant after previous outcomes even though each new opportunity remained favorable. This finding shows how reactions to a past result can alter the next choice. It does *not* imply that brain damage generally improves decisions or that emotion should be removed. Emotion can warn us about danger and support goals; its value depends on the task. Neither study establishes fixed genetic "risk types." Experience, context, institutions, and learning also shape behavior.

11.4 Prospect theory: gains, losses, and chances

Would you prefer S$500 for certain or a 50% chance of S$1,000 and a 50% chance of nothing? Both have the same expected monetary value:

$$\text{EMV(certain)} = \text{S\$500}, \qquad \text{EMV(gamble)} = 0.5(\text{S\$1,000}) + 0.5(\text{S\$0}) = \text{S\$500}. \tag{11.1}$$

Equal expected money does not require equal preference. A person may value certainty; expected utility from earlier chapters can represent that preference. Daniel Kahneman and Amos Tversky's **prospect theory** (1979) adds a descriptive account of how people evaluate outcomes and probabilities in many risky choices.

11.4.1 Reference dependence and loss aversion

Imagine, explicitly as an *assumption*, that the sure S$500 becomes your reference point. The certain option now feels like no change. The gamble becomes an equal chance of ending S$500 *above* or S$500 *below* that reference. The low outcome is a shortfall from an expectation, not necessarily an out-of-pocket loss. A different reference point could change the interpretation without changing the money paid.

Reference dependence means that an outcome is evaluated as a gain or loss relative to a benchmark. The current situation, an expected bonus, a target, or a comparison with peers can

help form that benchmark. **Loss aversion** means that, at the same reference point, a loss of a given size may have more psychological impact than an equal-sized gain. This is a qualitative tendency, not a universal fixed "two-to-one" law. **Diminishing sensitivity** is a different property: moving another S$100 tends to matter less farther from the reference, on either side. The value curve is commonly drawn concave for gains and convex for losses, with a steeper loss side near the reference. Figure 11.1 is a schematic example, not a fitted curve for this class.

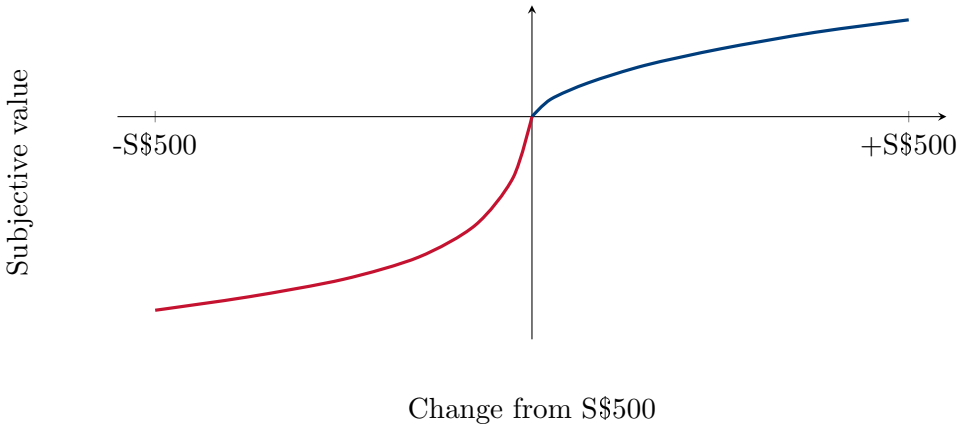


Figure 11.1: Illustrative reference-dependent value curve. The vertical scale is arbitrary; its shape conveys diminishing sensitivity and a steeper loss side, not a measured loss-aversion coefficient.

The same reference-point logic helps explain different responses to a bonus. Employee 1 expected no bonus and receives S$1,000: the payment may feel like a S$1,000 gain. Employee 2 expected S$2,000 and receives the same S$1,000: the payment may feel like a S$1,000 shortfall. Now offer each employee a sure S$1,000 or a 50–50 gamble between S$2,000 and zero. Employee 1 may secure a perceived gain, while Employee 2 may gamble for a chance to reach the expected S$2,000. These are possible reactions, not predictions that every person must follow. Curvature within the gain or loss domain helps describe risk attitude; loss aversion compares equal changes *across* the two domains.

11.4.2 Probability weighting

Prospect theory also distinguishes a stated probability (p) from its **decision weight** (w(p)), the influence a chance has in an evaluation. A common descriptive pattern gives small probabilities more than proportional weight and many moderate or high probabilities less than proportional weight. Figure 11.2 contrasts that pattern with the diagonal $w(p) = p$. It is illustrative, not an estimated curve. At the endpoints, an impossible event remains impossible and a certain event remains certain: $w(0) = 0$ and $w(1) = 1$.

A Singapore TOTO S$1 ordinary entry has a Group 1 jackpot probability of roughly one in 14 million. A buyer may know how tiny that chance is and still attend strongly to the possibility of becoming rich. That is one possible role for probability weighting; enjoyment, hope, and other preferences are also possible explanations. Ordinary lottery play has a negative monetary return on average, but a draw-specific calculation would require the prize structure, possible sharing, and any rollover. Buying a ticket by itself does not diagnose a person's probability weights. Likewise, a person who assigns importance to a 95% event while not treating it as certain need not have misread the number. Decision weights describe choice behavior, not necessarily a mistaken estimate of the stated odds.

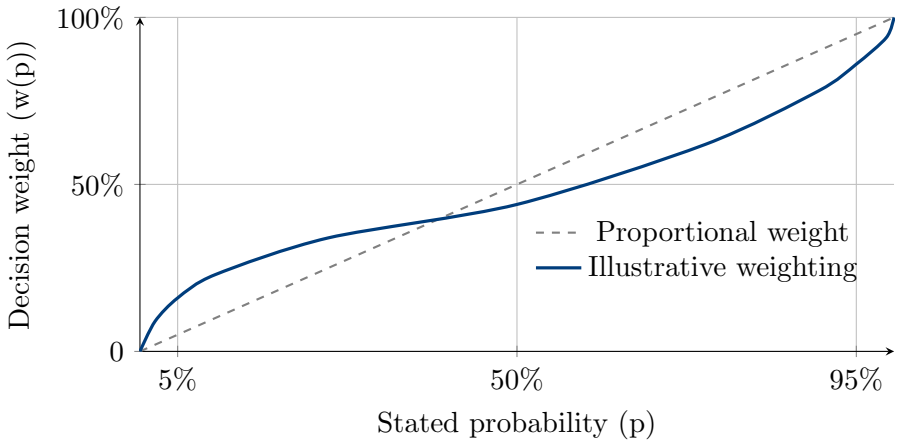


Figure 11.2: A schematic weighting curve. A 5% chance can receive more than 5% weight; a 95% chance can receive less than 95% weight. Certainty still has full weight.

11.5 Framing: equivalent facts, different focus

Framing changes the description of a choice while holding its underlying facts fixed. "90% of deliveries arrive on time" and "10% arrive late" describe the same set of 100 deliveries, assuming each is either on time or late. One wording emphasizes success, the other failure. That change in focus can alter how reliable the service feels, without changing its actual performance.

The well-known disease problem makes the effect more striking. A disease threatens 600 people. Table 11.1 describes the same hypothetical treatments in lives saved and lives lost.

Table 11.1: Equivalent gain and loss descriptions of two treatments

Treatment	Gain frame: lives saved	Loss frame: lives lost
A, certain	200 saved	400 die
B, gamble	(1/3) chance all 600 are saved; (2/3) chance none are saved	(1/3) chance nobody dies; (2/3) chance all 600 die

Treatment A has the same certain consequence in each column. Treatment B has the same lottery in each column. Its expected number saved is ((1/3)(600)+(2/3)(0)=200); this equality of expected lives does not make certain and risky outcomes identical. In the original Tversky and Kahneman (1981) experiment, *72% chose the certain option* under the gain frame, compared with *22% under the loss frame*. These were two respondent groups, not the same individuals changing their minds and not this course's class votes. The classroom version retains A for the certain treatment and B for the gamble in both frames; the original paper used different letters in the second version. The evidence concerns the effect of describing risky outcomes as gains or losses; it should not be reduced to a claim that loss aversion alone explains every response.

Businesses can also frame an unchanged contract differently. A one-year S$120 warranty covering eligible repairs up to S$1,000 could be described as protection from covered repair costs, or as avoiding the possibility of paying those costs yourself. The second wording highlights a potential loss, but neither wording tells a customer the chance of a covered failure, exclusions, or whether another warranty overlaps. A subscription can be promoted as "get an extra 10% discount" or "don't miss an extra 10% discount." The offer is identical, while the second message presents

inaction as a missed opportunity. It may increase urgency; it does not establish an actual deadline or guarantee a sale.

When assessing a frame, translate both versions back into the same outcomes, probabilities, time horizon, and total price. This reveals whether only the wording changes or whether the contract itself changes.

11.6 Sunk costs and decisions from today forward

A **sunk cost** has already been incurred and cannot be recovered by the current choice. Suppose you paid S$15 for a movie and, after 30 minutes, realize you dislike it. Whether you stay or leave, the S$15 remains spent. The question is how you prefer to use the next 90 minutes. Staying only because you "must get your money's worth" is the **sunk-cost trap**. Staying because the movie may improve or because leaving has a new cost could still be reasonable. What matters is the *future* difference between the alternatives.

The same reasoning applies to a medical startup that has invested heavily in a technology and then receives disappointing trial results. It must decide whether to stop or pay for another trial. The spending already made cannot be recovered by the new trial. However, the *information* learned from the earlier trial is relevant: it should update beliefs about whether another test is worth funding. A useful question is, "If we encountered this technology for the first time today, with the evidence and resources we now have, would we pay for another trial?" A simple forward-looking comparison is

$$\begin{aligned}\text{Incremental value of continuing} = {} & \text{expected future benefits, including useful learning} \\ & - \text{additional future costs.}\end{aligned} \tag{11.2}$$

No numbers in this hypothetical case determine a unique stop-or-continue answer. Past expenditure should not tip the comparison *by itself*; knowledge, assets, obligations, and future switching costs created along the way can still matter. **Path dependence** describes how a previous course of action can shape the current state. It is not always a fallacy. The error is treating irrecoverable expenditure as an independent reason for another expenditure.

11.7 Anchors, comparisons, and decoys in pricing

An **anchor** is a salient starting value that influences a later judgment. Consider identical headphones at a final price of S$120. "S$120" and "was S$200, now S$120" state the same current price, but S$200 may make the latter display feel like a bargain. The listed comparison price is not by itself evidence of product quality, market value, or a genuine discount. An independent benchmark is more useful. Anchoring is related to, but distinct from, evaluating gains or losses relative to a personal reference point: the former emphasizes the pull of a salient number in judgment, while the latter describes how outcomes are coded and valued.

A **decoy** changes the *set of comparisons*. In Dan Ariely's reported *Economist* subscription example, the historical annual menu included digital at US$59 and a print-plus-digital bundle at US$125. A second version also offered print only at US$125. The print-only option was never selected in the reported group, yet adding it made the equally priced bundle easier to favor: for the same price, the bundle included digital access as well. Table 11.2 reports Ariely's two groups of 100 MIT students. These are reported historical choices, not class responses or current publisher prices.

Table 11.2: Reported subscription choices in Ariely's classroom example

Annual option	Two-option menu	Three-option menu
Digital, US$59	68%	16%
Print only, US$125	Not offered	0%
Print + digital, US$125	32%	84%

The bundle's share differs by *52 percentage points* (84% minus 32%), not "52% more." Under the additional *illustrative* assumption that every stated choice becomes one purchase, revenue from 100 customers would be $68(59) + 32(125) = 8{,}012$ US dollars under the two-option menu and $16(59) + 84(125) = 11{,}444$ US dollars under the three-option menu, an increase of about 42.8% relative to the first calculation. These are not observed sales or profits; printing, distribution, and other costs are omitted. The groups were different, so we cannot say that 52 particular students switched. The observed pattern is consistent with a decoy effect, but one menu comparison does not reveal every individual's thought process.

For a new plan to be a dominated decoy, the target plan must be at least as good on all relevant attributes and better on one, including price, quality, obligations, and access. Two coffee sizes with different prices are not automatically a decoy; the larger size may represent a genuine trade-off. An annual subscription also changes commitment and upfront cost, not just the displayed monthly equivalent.

11.8 Choice architecture, nudges, and friction

Choice architecture is the way options are arranged and presented. Richard Thaler and Cass Sunstein's *Nudge* (2008) discusses ways to guide behavior while keeping options available and without materially changing financial incentives. Thaler received the 2017 economics Nobel Memorial Prize for contributions to behavioral economics; the prize was not awarded to Sunstein or solely for the book.

Placing fruit at eye level makes it easier to notice and reach while other foods remain available. An employer can make enrollment in a savings plan the default while offering an easy opt-out. A timely assignment reminder with a direct submission link makes an intended action easier to start. A checklist next to a schoolbag provides a cue at the moment items are packed. These are *mechanism examples*, not claims of measured effect sizes or universal success. A cash bonus, fine, or compulsory choice can influence behavior too, but is not the same kind of nudge.

Defaults deserve special attention because many people continue with what is preselected. For a study-app subscription, a screen might preselect Plus and monthly auto-renewal. If Basic is visible and switching is simple, the preselected plan is a transparent default. A cancellation route hidden under several layers of settings is different: it increases **friction** after purchase, sometimes called *sludge* or a dark pattern. It may keep payments flowing while preventing customers from carrying out a decision to leave. A design that supports informed choice shows renewal terms clearly and makes cancellation easy to find.

The practical lesson is not that all influence is forbidden. Every menu has an order, a description, and some default. Evaluate a design by asking whose objectives it serves, which alternatives remain genuinely accessible, whether fees and commitments are clear, and what happens to customers after they choose.

11.9 Apply the ideas: redesign a study-app offer

Consider a campus study app with two monthly plans. Basic costs S$6 and provides flashcards. Plus costs S$10 and provides flashcards plus practice quizzes. The business wants more Plus subscriptions. A useful exercise is to change *one feature of the sign-up screen at a time*, predict what will happen, and then test it. The first four ideas can keep the original two prices and features unchanged:

1. **Incremental-price frame.** Show Basic at S$6 and describe Plus as "add practice quizzes for S$4 more per month." Disclose the S$10 total before payment. The emphasis changes; the bill does not.
2. **Decoy.** Add quizzes-only at S$10 alongside Basic and Plus. If access and other terms are comparable, Plus at the same S$10 also includes flashcards and dominates the quizzes-only alternative.
3. **Transparent default.** Preselect Plus while keeping Basic visible and one click away. Explain the monthly charge and renewal.
4. **Lower sign-up friction.** Use one clear checkout screen and avoid repeated data entry. Do not confuse a smoother sign-up process with making cancellation difficult later.

Changing the *offer* creates additional strategies that should be analyzed separately. A 20% first-month Plus discount changes that month's price to S$8, with S$10 per month afterward. The loss-framed message "Don't miss 20% off your first month" emphasizes a saving that could be forgone; the financial incentive and the wording are different mechanisms. An annual Plus plan at S$108 upfront costs S$12 less than twelve S$10 monthly payments, a 10% saving relative to S$120. "Don't miss the annual option: save 10%" is a loss frame, but annual billing also changes upfront payment and commitment. No promotion deadline should be invented to add pressure.

11.9.1 Test business and customer outcomes

To estimate whether a screen works, randomly show comparable visitors the original or one redesigned version while holding other conditions fixed. To isolate *wording*, compare gain and loss descriptions of the *same* financial offer; comparing two different prices cannot identify a pure framing effect. Define a business outcome such as paid Plus sign-ups per visitor, not only Plus's share among those who buy. Measure net revenue after discounts and relevant costs before claiming a profit improvement. Define a customer outcome too: use of practice quizzes, satisfaction, downgrades, cancellations, or complaints. A menu that raises the Plus share while reducing total purchases or increasing regretted subscriptions has not necessarily improved the decision environment. An observed difference needs adequate data and a fair comparison; a predicted effect is not a measured effect.

11.10 Use the ideas for your own decisions

The same tools that can influence customers can help you examine your own choices. Before a purchase or subscription, ask what you will pay in total, including later renewals; whether an initial price or carefully placed alternative is anchoring your comparison; whether a message merely rephrases the same outcome; and whether the choice fits your goals and budget. If an interface preselects a plan, read the selection and the cancellation terms. For a past investment, use what you learned from it without treating irrecoverable spending as a reason to continue.

For further reading, the instructor recommends Morgan Housel's *The Psychology of Money*. The Week 13 classroom deck closes with a professor-supplied video linked from 3 minutes 11 seconds. The book and video are optional extensions; the decision checks above are the course's synthesis, not a quotation from either source.

11.11 Common mistakes

- **Calling every shortcut irrational.** Search and attention have costs. A heuristic can be sensible even if it is imperfect.
- **Calling System 1 a "bad brain region."** The labels describe kinds of processing. Fast responses can be useful; neither system is one anatomical location.
- **Inferring loss aversion from one risk choice.** Expected utility, reference points, curvature, and probability weighting can all matter. One answer does not identify a person's mechanism.
- **Confusing decision weights with beliefs.** A person can know the stated probability while giving it disproportionate impact.
- **Extending high-probability underweighting to 100%.** In standard weighting, certainty has full weight.
- **Treating two frames as two different outcomes.** Translate the village treatments into the same saved-or-dead consequences.
- **Ignoring everything learned from a past investment.** Money already spent is sunk; evidence from a trial is not.
- **Calling any middle price a decoy.** Dominance depends on all relevant attributes and obligations, not the label or position.
- **Calling more subscriptions more profit.** Discounts, delivery costs, retention, and customer harm must be considered.

11.12 Practice questions

1. In the internship example, explain how two reasonable students could make different choices without either miscalculating the pay.
2. Five machines make five items in five minutes. How long do 100 identical independent machines need to make 100 items simultaneously?
3. Calculate the expected money from the S$500 gamble. If S$500 is the reference, identify the relative gain and loss outcomes.
4. Two employees receive S$1,000 bonuses. One expected zero; the other expected S$2,000. Identify their gains or shortfalls relative to those expectations. Does this prove how each will feel?
5. For the village, calculate Treatment B's expected lives saved and show why its gain and loss descriptions are the same lottery. What do 72% and 22% refer to in the original framing study?

6. A startup has spent S$4 million on a product. A new trial would cost S$1 million. Which past and future information belongs in the decision about another trial? Is S$4 million an independent reason to continue?
7. In the subscription example, how much larger is the reported bundle-choice share when print-only is included? Why is that not a percentage increase in profit?
8. A Plus plan costs S$10 per month. Calculate the first month's price after a 20% discount and the saving from paying S$108 for a year rather than twelve monthly payments. Which comparisons would isolate the effect of loss-framed wording?
9. Explain why preselecting Plus and hiding cancellation under five settings screens are different kinds of choice architecture.

Selected answers

1. Students may have different minimum pay, commute limits, beliefs about later offers, or ability to risk losing this offer. An acceptable stopping rule depends on those facts and preferences.
2. Five minutes. Each machine makes one item in that interval.
3. The expected money is S$500. Relative to the assumed S$500 reference, the gamble produces +S$500 or -S$500, each with probability one half.
4. The first employee has a S$1,000 gain from the expected zero; the second is S$1,000 below the expected S$2,000. Their feelings cannot be proved from these amounts alone.
5. Expected lives saved are 200. A one-third chance all are saved means a one-third chance nobody dies; a two-thirds chance none are saved means a two-thirds chance all die. The 72% and 22% are the shares choosing the *certain* treatment in separate gain- and loss-frame groups.
6. The trial result updates beliefs about success and the value of learning from another test. Compare its expected future benefits with its S$1 million future cost and other alternatives. The S$4 million already spent is not a reason by itself to continue.
7. The difference is (84-32=52) percentage points. Profit requires costs and actual purchases; the reported figures are stated choices.
8. The first month costs S$8. Twelve monthly payments cost S$120, so annual billing saves S$12, or 10% of S$120. Compare gain and loss copy for the same first-month offer or the same annual offer, holding prices and terms fixed.
9. Preselection is a default that can leave alternatives readily available. Buried cancellation creates effort after purchase and can obstruct a customer's decision to leave.

11.13 Chapter summary

- People can choose sensibly under limits by setting search rules and allocating attention; a heuristic is not automatically a bias.
- Fast responses, deliberate checks, and emotion all contribute to decisions. Their usefulness depends on context.

- Prospect theory describes reference-dependent gains and losses, diminishing sensitivity, loss aversion, and probability weighting.
- Framing changes the focus of equivalent facts. Sunk-cost reasoning wrongly treats irrecoverable spending as a reason to continue.
- Anchors, decoys, and defaults can change comparisons or effort. Transparent design and accessible cancellation matter for customers.
- A behavioral redesign is a hypothesis. Test business performance and customer outcomes while holding the relevant terms constant.

Further reading and sources

- Simon, H. A. (1978). *Rational Decision-Making in Business Organizations*. https://www.nobelprize.org/uploads/2018/06/simon-lecture.pdf.
- Sims, C. A. (2003). "Implications of Rational Inattention." *Journal of Monetary Economics*. https://doi.org/10.1016/S0304-3932(03)00029-1.
- Kahneman, D., & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*. https://www.jstor.org/stable/1914185.
- Tversky, A., & Kahneman, D. (1981). "The Framing of Decisions and the Psychology of Choice." *Science*. Read the paper.
- Tversky, A., & Kahneman, D. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." https://doi.org/10.1007/BF00122574.
- Kuhnen, C. M., & Knutson, B. (2005). "The Neural Basis of Financial Risk Taking." *Neuron*. https://doi.org/10.1016/j.neuron.2005.08.008.
- Shiv, B., et al. (2005). "Investment Behavior and the Negative Side of Emotion." *Psychological Science*. https://doi.org/10.1111/j.0956-7976.2005.01553.x.
- Ariely, D. (2008). *Predictably Irrational*, chapter 1; author interview on the subscription example.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge*.
- Housel, M. (2020). *The Psychology of Money*. Publisher page.