Introduction to the Philosophy of Science

Concepts, Practice, and Case Studies

Dr. Yves J. Hilpisch[1] and GPT-5.1

December 2, 2025 (work in progress)

[1]Get in touch: `https://linktr.ee/dyjh`. Web page `https://hilpisch.com`.

Preface

Science is one of humanity's most ambitious storytelling projects. It tries to tell stories about the world that are reliable, revisable, and practically useful. This book is an invitation to step back from particular theories and experiments and look at how those stories are built, tested, and used.

Who This Book Is For

Primary readers:

- Practitioners in mathematics, physics, AI, quantitative finance, and neighbouring fields who want sharper language for thinking about models, evidence, and risk.
- Students in these areas who want a friendly but rigorous companion to more technical courses.

Secondary readers:

- Curious readers with a quantitative background who enjoy seeing how different sciences hang together.
- Mentors, team leads, and educators looking for ways to explain modelling choices and uncertainty to others.

You do not need prior philosophy training; familiarity with basic probability and calculus will help but is not essential.

Motivation and Scope

This book centres on the philosophy of science as it appears in day-to-day practice. Rather than surveying every historical position, we focus on questions that working scientists and engineers meet regularly, even if they do not call them "philosophy":

- What do we mean by "theory", "model", and "law" in concrete projects?
- How should we read evidence from experiments, observational studies, and simulations?
- When should we trust a black-box model, and what kind of explanation do we owe users and regulators?
- How do values and risks enter scientific decisions, especially in finance, AI, and policy-relevant science?

The emphasis is on concepts that travel well across domains, illustrated with examples from mathematics, physics, artificial intelligence, and quantitative finance.

Style and Approach

We adopt a light-hearted tone while treating arguments and notation with care. Figures and callouts carry much of the intuition; main text provides narrative and connections; short technical notes at the ends of some chapters supply formal details when they clarify the discussion.

How to Read This Book

Skim-first, deepen-later works well here:

- **First pass:** read Learning Objectives and At a Glance callouts, browse Everyday Analogy and From Lab to Life boxes, and glance at figures.
- **Second pass:** read the surrounding prose, paying attention to how examples and definitions are woven together.
- **Third pass (as needed):** work through Technical Notes and reflection questions when you want more formal or conceptual depth.

Different callout types serve different roles:

- *Everyday Analogy* boxes connect abstract ideas to familiar situations.
- *From Lab to Life* boxes show how concepts play out in mathematical work, physics labs, AI projects, and financial practice.
- *Common Pitfall* boxes warn about recurrent misunderstandings.
- *Try in 60 Seconds* prompts offer quick self-checks.
- *Technical Notes* provide compact formal clarifications you can revisit later.

How to Navigate

You can read the book linearly or along themed paths:

- **Conceptual walk:** read Parts I and II in order, then sample case studies in Part III.
- **Practitioner's shortcut:** start with measurement, causality, and statistics (Chapters 5–7), then dip into the case studies most relevant to your domain.

Chapters are largely self-contained; cross-references point you to earlier material when a concept reappears. If you are reading with a team or class, the summary sections and reflection questions at the end of each chapter make good starting points for discussion.

Disclaimers

Nothing in this book is financial, medical, or legal advice. Examples in finance, health-related contexts, or policy are simplified for illustration; real-world decisions in these areas should always involve appropriate professional expertise, domain-specific data, and regulatory guidance. The aim here is to strengthen your conceptual toolkit, not to replace specialised training.

Notation and Conventions

We keep notation light and reuse it consistently across chapters. This section is a quick reference; each chapter reintroduces specialised symbols as needed.

General Mathematical Notation

- Sets: capital letters such as A, B; real numbers $\mathbb{R}$; expectations $\mathbb{E}[\cdot]$.
- Functions: $f\colon X \to Y$; derivatives $f'(x)$ and partials $\partial f/\partial x$ where useful.
- Vectors: bold lower-case $\boldsymbol{x}$; matrices: bold upper-case $\boldsymbol{A}$.

Probability and Statistics

- Random variables: capital letters X, Y; realisations: lower-case x, y.
- Probabilities: $P(E)$ for an event E; conditional probability $P(E \mid H)$.
- Expectations: $\mathbb{E}[X]$; variance: $\mathrm{Var}(X)$; standard deviation: σ_X.
- Densities or mass functions: $p(x)$ or $p_X(x)$ when we need them.
- p-values: see Section 2.7 for a formal definition and discussion.

Causal Diagrams and Models

When discussing causality:

- Variables appear as nodes (for example treatment D, outcome Y, confounder C).
- Directed arrows (e.g. $C \to D$, $D \to Y$) indicate hypothesised direct causal influence.
- Graphs are assumed to be directed acyclic graphs (DAGs) when we use them as causal diagrams (see Section 6.6).

Domains and Examples

We refer to specific domains for illustrations:

- **Physics:** Newtonian mechanics, relativity, and quantum theory appear as running examples of theories, models, and laws.
- **AI:** supervised and reinforcement learning, benchmarks, and large models illustrate induction, overfitting, and explanation.
- **Finance:** asset-pricing models, trading strategies, and risk measures ground discussions of uncertainty, reflexivity, and model risk.

When domain-specific notation is introduced (for example risk-neutral probabilities, loss functions, or specific dynamical equations), it is defined locally in the relevant chapter.

Conventions

- Acronyms are spelled out on first use (for example efficient market hypothesis (EMH)).
- We use en-dashes for ranges (“5–10”) and proper LaTeX quotation marks (“like this”).
- References to chapters use Chapters 1 and 2 style labels; technical notes carry section labels such as Section 5.6.

Acknowledgments

The text was developed collaboratively by human and AI authors. Where the book speaks in a single voice, it reflects an iterative process: prompts, drafts, checks against examples and counterexamples, and refinement toward clear, practice-oriented explanations. Treat the AI contributions as part of an extended scientific conversation rather than as an oracle.

With gratitude, I acknowledge the artificial intelligence research community and OpenAI for advancing tools such as GPT-style models. Their work has made projects like this book both feasible and enjoyable to pursue on an ordinary laptop.

Although I am not a professional philosopher of science by formal training, I have been fascinated by these questions for decades while working in quantitative disciplines. Modern AI systems let me explore and write the kind of book I wish I had earlier in my own career, combining physics, mathematics, AI, and finance under one conceptual roof.

Any remaining oversimplifications or errors are mine. I would be grateful for feedback, suggestions, and pointers to improvements in examples or arguments; you can find ways to get in touch at `https://linktr.ee/dyjh`.

Contents

List of Figures

Part I

Theoretical Foundations

Part I Overview

This opening part clarifies what "science" is really about in practice. Chapter 1 starts from everyday images of science and refines them into a working concept built around explanation, prediction, and understanding. Chapter 2 then develops the logic of scientific reasoning—deduction, induction, and abduction; hypotheses, laws, and theories; and core ideas about evidence and falsification. Chapter 3 turns to the internal structure and dynamics of scientific theories as networks and model families, while Chapter 4 explores what theories are about and how scientific explanation works in realist and instrumentalist keys.

Chapter 1

What Is Science, Really?

This chapter is about getting a clean, honest picture of what science is and is not. Instead of starting with slogans ("the scientific method") or idealised diagrams, we will look at how science feels from the inside: as a mix of curiosity, messy practice, social negotiation, and surprisingly elegant structure.

Learning Objectives

After working through this chapter you should be able to:

- name and contrast several everyday images of science and say what each one captures and misses,
- distinguish four central aims of scientific work: knowledge, explanation, prediction, and understanding,
- state the core idea behind Popper-style falsifiability and why it does not settle every "science vs non-science" dispute,
- describe a three-layer picture of reality, theories, and models, using concrete examples from physics, AI, and finance,
- explain "what science is" in a short paragraph aimed at a curious non-expert.

At a Glance

Science is not just a bag of facts or a sacred method. It is a long-term, community effort to build reliable, revisable maps of the world—maps that help you explain what happened, predict what comes next, and understand why patterns hold. Those maps live in theories and models, and they are always drawn with human hands under uncertainty and constraint.

Everyday Analogy

Think of science like running a good kitchen. Recipes (theories) and tools (models) only matter because hungry people, real ingredients, and time pressure exist. You adjust the heat, taste as you go, throw out what fails, and keep what works—all while negotiating with other cooks sharing the same stove.

1.1 Everyday Images of Science

Most readers arrive with vivid pictures of what science is and how scientists work. It helps to lay these images on the table before refining them, because each one hides a different trap for learners and practitioners.

- Science as a neatly organised collection of facts, formulas, and definitions that appear in schoolbooks and exam sheets.

- Science as a step-by-step method from posters and infographics: ask a question, form a hypothesis, run an experiment, analyse data, draw conclusions.
- Science as a sprawling social institution with labs, journals, conferences, grant panels, and peer review shaping what gets studied and published.
- Analogy: following a cooking recipe versus standing in a real, noisy kitchen where time pressure, missing ingredients, and improvisation test the recipe.
- First contrast: a curious, self-correcting "scientific attitude" aimed at truth-seeking versus a rigid bureaucracy of checklists, forms, and prestige markers.

The first image—science as a tidy collection of facts—matches how many students first meet physics, chemistry, or biology. On the page everything is frozen and polished: Newton's laws, Ohm's law, definitions of energy or entropy. This view is not wrong, but it hides the work needed to produce those formulas and the uncertainty that accompanies them at the frontiers of knowledge.

The second image—science as a step-by-step method—comes from classroom posters. It emphasises procedure over judgment: you are told to "follow the method" almost like following assembly instructions from a furniture store. This framing is comforting because it suggests that if you follow the script, you will necessarily get the right answer. In practice, scientists improvise constantly: they switch hypotheses midstream, repair broken instruments, and reframe questions entirely.

The third image—science as an institution—highlights labs, large collaborations, journals, conferences, and funding agencies. This view is essential if you want to understand why some topics are studied intensely while others are ignored, why some methods become fashionable, and how social incentives can both support and distort the search for truth.

Common Pitfall

Equating "science" only with published papers or famous experiments quietly erases the background work: debugging code, calibrating sensors, cleaning data, double-checking units, and arguing about interpretations. When you later do physics, AI, or quantitative finance, remember that this invisible labour is part of science, not a distraction from it.

In the *Newtonian Physics* book, for example, we treat a flying ball as a point mass and happily ignore air resistance on a first pass. That decision is not written on a classroom poster; it is a modelling judgment informed by experience. Philosophy of science gives you names and tools for making such judgments explicit and discussable.

1.2 Knowledge, Explanation, Prediction, Understanding

The word "knowledge" is often used loosely. In this book we will work with a more structured picture: scientific practice weaves together at least four aims. Depending on the field and question, one aim may dominate, but none can be ignored for long.

- To know: building reliable belief about the world so that claims are more than opinion, gossip, or wishful thinking.
- To explain: answering "why?" in a way that connects particular events to general patterns, mechanisms, or principles.
- To predict: saying "what next?" with enough precision that decisions, designs, and safeguards can depend on the forecast.

- To understand: seeing structure, patterns, and unifying ideas well enough that new cases feel less like surprises and more like variations on a theme.
- Everyday anchor: a weather app on your phone that reports current conditions, offers explanations for changes, gives short- and long-term forecasts, and gradually teaches you to read the maps yourself.
- Tension: a model that predicts well but is opaque inside (a black box) versus a transparent explanation that feels clear but performs poorly on fresh data.

Knowledge is about reliability. When a civil engineer designs a bridge, they rely on knowledge about materials and forces that has survived many tests. We do not demand certainty—almost nothing in empirical science is certain—but we do demand that beliefs be robust against error-checking and alternative explanations.

Explanation is about satisfying the "why?" itch. A simple catalogue of correlations ("when A happens, B tends to follow") can be useful, but an explanation goes further: it connects A and B through a mechanism, a symmetry, or a deeper law. In Newtonian mechanics, for example, the explanation for a falling apple is not just that apples regularly fall but that the same gravitational law governs apples, the Moon, and the tides.

Prediction is about acting in time. A climate model, a volatility model in finance, or a disease-spread model earns its keep when it allows you to take action before events unfold. Sometimes prediction is local and short-term (tomorrow's rainfall), sometimes global and long-term (century-scale climate trends), but in all cases a scientific model must face the test of "what actually happened?" sooner or later.

Understanding is more than having a formula. It is the sense that you can see why a result must hold, how different pieces fit together, and what would happen if you tweak the setup. Many physicists feel they *understand* Newtonian gravitation only after they can navigate between energy diagrams, force diagrams, and actual orbits without getting lost.

Analogy: Weather on Your Phone

A weather app illustrates the four aims in your pocket:

- Knowledge: reporting current temperature and rainfall using calibrated instruments.
- Explanation: attaching labels such as "cold front" or "high-pressure system" to make sense of changes.
- Prediction: showing hourly and weekly forecasts with uncertainty bands.
- Understanding: interactive maps and animations that help you see larger patterns, so you start guessing tomorrow's weather before opening the app.

Modern machine learning models dramatise tensions between these aims. A deep neural network trained on market data can sometimes predict price movements more accurately than a simple economic model, yet offer almost no explanation a human can inspect. An interpretable but weaker model may explain more and predict less. Deciding which trade-off is acceptable in a given context is a philosophical question with real financial and social consequences.

From Lab to Life

Concrete echoes of these four aims show up in everyday scientific and engineering workflows:

- Physics lab: calibrating sensors and checking units to secure knowledge, using force or energy stories to explain results, predicting new runs before pressing the start button, and building intuition that links equations to traces on the screen.
- Machine learning project: estimating model performance curves for knowledge, using feature attributions or causal diagrams to explain behaviour, forecasting production metrics under drift, and developing an understanding of when a model will fail gracefully versus catastrophically.
- Quantitative finance desk: backtesting trading signals to establish knowledge, tying them to economic or behavioural mechanisms for explanation, generating forward risk and return scenarios for prediction, and cultivating an understanding of market regimes so that numbers are read in context.

1.3 The Demarcation Problem: Science vs Non-Science

Philosophers of science have long asked where, if anywhere, a clean boundary between science and non-science can be drawn. This is the *demarcation problem*. It matters whenever we decide which claims deserve extra trust, resources, or legal authority.

- Framing question: what separates scientific inquiry from pseudo-science, ideology, or mere storytelling, especially when the surface language sounds similar.
- Popper-inspired idea: scientific theories should make bold, risky predictions that could in principle be shown false by observation.
- Edge case: astrology versus astronomy, where both talk about planets and stars but only one anchors its claims in precise, testable predictions.
- Edge case: serious economics versus "financial astrology" newsletters that fit narratives to charts without exposing clear failure conditions.
- Analogy: shop return policies, where a claim that cannot be "returned" or revised in the face of failure behaves differently from one tied to observable outcomes.

Karl Popper famously proposed falsifiability as a key marker: a scientific theory should rule out possible observations. If nothing that could happen in a lab, at a telescope, or in a data stream would count as evidence against your view, Popper argued, you are not doing science but unfalsifiable storytelling.

This criterion works well for textbook contrasts like astronomy versus astrology. Astronomers publish precise predictions about eclipses and planetary positions and are visibly wrong when events do not match. Astrological forecasts are often vague and flexible. When they miss, they are retrofitted with new interpretations rather than forcing a revision of the core system.

However, real research is messier. Most theories come bundled with auxiliary assumptions about instruments, background conditions, and data processing. When a particle-physics experiment fails to find an expected signal, researchers can reasonably ask whether the theory is wrong, the detector is miscalibrated, the data cleaning pipeline is flawed, or the statistics were misapplied.

Analogy: Return Policies and Fine Print

Think of scientific theories like products with return policies:

- A strong policy: "If this fails under normal use, we replace or refund." This is like a theory that stakes its reputation on clear, risky predictions.
- A weak policy: "Returns only under conditions we define later." This resembles a system that always finds a way to excuse failure.
- The fine print: real return policies include clauses about misuse, damage, or unusual conditions—just as real experiments involve caveats about calibration, background noise, and model misspecification.

In quantitative finance, demarcation shows up when we compare statistically tested trading strategies with story-driven "market wisdom." A rule that has survived out-of-sample testing with clear criteria for failure belongs closer to the scientific side. A newsletter that always "would have been right" if only you had timed entries slightly differently leans toward pseudo-science.

Common Pitfall

It is tempting to label everything you dislike as "unscientific" and everything you admire as "scientific." Demarcation is not a moral ranking but an analysis of how claims are tied to evidence, error correction, and revision. An honest non-scientific practice (say, writing novels) can be more intellectually responsible than a bad "scientific" study.

1.4 Reality, Theories, and Models

To navigate talk about "truth" in science, it helps to distinguish the world, our theories about it, and the models we actually use in practice. Without this three-layer picture, arguments about realism, simulation, and approximation quickly turn into confusion.

- Layer 1: a world that exists independently of our descriptions, with many details we do not and cannot measure directly.
- Layer 2: theories as broad conceptual frameworks that aim to describe how parts of that world hang together.
- Layer 3: models as concrete, simplified tools—often mathematical or computational—that implement pieces of a theory for specific purposes.
- Map–territory analogy: a tourist city map, a live GPS map with traffic data, and a hand-drawn sketch all represent the same city with different compromises.
- Preview tension: scientific realism, which treats successful theories as approximately true descriptions of the world, versus instrumentalism, which treats them as instruments for organising experience and prediction.

At the bottom is the territory: the world "out there," including stars, cells, traders, and neural tissue. Whether you are a strict realist or a more cautious agnostic, you typically behave in the lab and on the trading floor as if this world does not bend to your preferences.

Theories live one level up. Newtonian mechanics, quantum field theory, evolutionary theory, and the efficient markets hypothesis are all examples of theories: they offer concepts, equations, and principles that claim to describe patterns in the world across many situations.

Models are more local, task-specific constructions that implement parts of a theory. A two-body gravitational model for Earth and the Moon, a Black–Scholes option pricing model, or a particular neural network architecture for image recognition are models: concrete mathematical objects or algorithms you can compute with.

Analogy: Three Kinds of City Map

Picture one city and three maps:

- A tourist paper map with major streets and landmarks—good for orientation, weak for traffic planning.
- A live GPS map on your phone with one-way streets and traffic overlays—excellent for routing, sometimes confusing at pedestrian scale.
- A hand-drawn sketch from a friend showing only the route from the train station to a favourite café.

All three are "about" the same city. None is the city itself. Each omits and distorts in different ways to serve a purpose.

The Newtonian physics book in this repository offers many examples of this layering:

- World: the actual motion of balls, planets, and vehicles, with air resistance, friction, imperfections, and human errors.
- Theory: Newton's laws, energy and momentum conservation, and the calculus-based framework for motion.
- Models: point particles on frictionless planes, ideal springs, uniform gravitational fields, and numerically simulated trajectories.

From Lab to Life

When you start a new modelling project—in physics, AI, or finance—it helps to check each layer explicitly:

- World: write down what you think actually exists and matters in the situation (objects, agents, signals, constraints).
- Theory: list the principles or frameworks you are implicitly relying on (conservation laws, no-arbitrage assumptions, learning theory, domain heuristics).
- Model: specify exactly which equations, algorithms, or code artefacts you will run and how they simplify the world and theory.

This quick audit reduces the risk of blaming "the world" for what is really a modelling shortcut or a fragile theoretical assumption.

Realism and instrumentalism are two contrasting attitudes toward the theory layer. A realist says roughly: "Our best theories are approximately true descriptions of the world, at least in their core claims." An instrumentalist says instead: "Theories are tools for organising experience and making predictions; asking whether electrons or fields "really exist" is often a meaningless or badly posed question." Both views will appear throughout the book, and part of your task is to see how they influence practice.

1.5 Summary and Short Reflection

To close the opening chapter, we pull together the most important threads and invite a quick self-check. Use the following points as a checklist rather than a script.

- Recap: science as a mix of facts, methods, and institutions that can be viewed through multiple everyday images.
- Recap: four central aims of scientific work—knowledge, explanation, prediction, and understanding—that sometimes trade off against each other.
- Recap: the demarcation problem and the role of falsifiability, case comparisons, and institutional safeguards in separating science from its look-alikes.
- Recap: the layered picture of reality, theories, and models, with map–territory analogies highlighting what is kept and what is left out.
- Reflection prompt: try to explain "what science is" to a curious twelve-year-old in three sentences that touch on practice, purpose, and limits.

Try the reflection prompt out loud. Imagine a twelve-year-old who has seen science mostly as colourful pictures in a textbook and dramatic scenes in films. Your three sentences might mention curiosity, evidence, fallibility, and the idea that good scientific stories can, in principle, be corrected when they go wrong.

Try in 60 Seconds

Quick exercises to cement the chapter:

- Point to one belief you hold about the world (for example about nutrition, markets, or climate) and ask: which of the four aims does it mostly serve—knowledge, explanation, prediction, or understanding?
- Take a claim you have seen online that presents itself as "scientific." Sketch a possible observation that, if it happened, would clearly count against it.
- Pick a favourite model from physics, AI, or finance and label which parts belong to the world, to the theory, and to the model in our three-layer picture.

Chapter 2

The Logic of Scientific Reasoning

Science is not just a pile of results; it is also a set of reasoning habits. This chapter introduces three basic forms of reasoning, the roles of hypotheses, laws, and theories, and some core ideas about evidence, falsification, and the ways our theories shape what we see.

Learning Objectives

After working through this chapter you should be able to:

- distinguish deduction, induction, and abduction using clear everyday examples,
- explain how hypotheses, laws, and theories relate using the falling-apple → law → Newtonian mechanics trio,
- describe what counts as confirmation or disconfirmation for a simple hypothesis,
- state Popper's falsification idea and name at least two practical limits,
- explain underdetermination and theory-ladenness with simple analogies and cases from physics, AI, and finance.

At a Glance

Scientific reasoning is a toolbox, not a single "scientific method." Deduction carries you from general rules to specific consequences; induction extracts patterns from repeated cases; abduction guesses the best explanation for puzzling data. Hypotheses, laws, and theories knit these moves together, while evidence, falsification, and theory-ladenness set the rules of the game.

Everyday Analogy

Imagine reading a detective novel. You know general rules about human behaviour (induction from many stories), you deduce what must follow if a suspect's story were true, and you abduce—you infer to the best explanation—when you decide who most likely did it. Science is a more disciplined, instrument-heavy version of this everyday reasoning.

2.1 Deduction, Induction, Abduction

We start with three classic patterns of reasoning. In real research they rarely appear in pure form, but naming them helps you see what you and others are doing.

Deduction: From Rules to Cases

Deduction moves from general premises to specific conclusions in a way that preserves truth: if the premises are true and the reasoning is valid, the conclusion must be true.

- Geometry example: "All interior angles of a Euclidean triangle sum to 180°. This figure is a Euclidean triangle. Therefore its interior angles sum to 180°." The move from rules to this particular triangle is deductive.
- Newtonian physics example: "If no net force acts on a body, its velocity remains constant. This cart experiences no net horizontal force. Therefore its horizontal velocity remains constant." The conclusion follows given the law and the description of forces.
- Everyday example: "All faculty receive a login for the new system. You are faculty. Therefore you receive a login." If the policy is correctly stated and applied, the conclusion is locked in.

In practice, the vulnerable point in a deductive argument is not the logical step itself but the premises: was there really no net force, is the policy really universal, are we really in Euclidean space?

Induction: From Cases to Patterns

Induction moves from repeated observations to a more general claim. It is never logically guaranteed, but it is indispensable for learning from experience.

- Sunrise example: you have seen the Sun rise every day of your life. From this you infer that the Sun rises every day and will probably rise tomorrow. The inference is strong but not infallible.
- Market example: a trader notices that a particular stock often bounces upwards after touching a certain price level. After dozens of such bounces, they infer a "support" level and consider trading on it.
- Lab example: repeated measurements of a pendulum's period at different amplitudes suggest that, within a certain range, the period is almost independent of amplitude. This pattern motivates a law-like statement.

Induction underlies much of statistics and machine learning: we fit models to past data and hope that learned patterns generalise. The logic is always: "This has worked in many cases under similar conditions; therefore it is likely to work again."

Abduction: Inference to the Best Explanation

Abduction, or inference to the best explanation, starts from puzzling data and searches for the most plausible story that would make those data unsurprising.

- Car example: you hear a grinding noise when braking. You consider explanations: a stone caught in the brake, worn brake pads, something loose in the trunk. You prefer the hypothesis that best fits the pattern of noise and context.
- Data example: a model that used to perform well suddenly degrades across regions. Possible explanations include data drift, deployment bugs, and a structural market change. The team weighs which explanation best accounts for all symptoms.
- Medical example: a doctor sees a combination of fever, rash, and recent travel history and infers the most likely disease given statistical frequencies and mechanisms.

Abduction is not guaranteed correct, but it is often the most creative and productive part of science. Successful theories are full of abductions that later survive further testing.

Common Pitfall

Two recurring confusions:

- Treating inductive patterns ("this has always worked so far") as if they were deductively certain laws, and then being surprised when a single counterexample arrives.
- Treating abduction as "just guessing" instead of disciplined hypothesis generation that must face further deductive and inductive tests.

Being explicit about which mode you are using prevents overconfidence and makes it easier to communicate your reasoning to collaborators.

Analogy: Detective Work

In a detective novel:

- Deduction checks consistency: "If this alibi were true, the suspect could not have been at the crime scene."
- Induction spots patterns: "In similar stories, the least obvious suspect often turns out guilty."
- Abduction chooses the best story: "Given all the clues, who most plausibly did it and why?"

Scientists play a similar game, but with telescopes, code, and equations instead of trench coats.

2.2 Hypotheses, Laws, and Theories

With the reasoning patterns on the table, we can clarify the units they operate on: hypotheses, laws, and theories. These terms are often used loosely in everyday speech; here we give them more precise roles.

Hypotheses: Focused, Testable Claims

Hypotheses are relatively specific, testable claims about the world.

- Everyday sense: "If I water my plants twice a week instead of once, they will grow faster."
- Scientific sense: "For small angles, the period of a simple pendulum is independent of its amplitude." Or: "Increasing this hyperparameter will reduce overfitting on the validation set."
- Finance example: "Stocks with lower price-to-book ratios will, on average, produce higher risk-adjusted returns over the next decade."

Hypotheses often emerge abductively: we see a pattern and propose a claim that would explain it. We then design inductive tests and deductive checks of what else must follow if the hypothesis holds.

Laws: Stable, Wide-Scope Regularities

Laws are more general statements that describe stable, widely applicable regularities.

- Physics example: the constant-acceleration kinematics used in the *Newtonian Physics* book, such as $s = \frac{1}{2}at^2$ for uniform acceleration from rest.

- Economics and finance example: no-arbitrage conditions that rule out risk-free profit loops under idealised trading assumptions.
- Everyday example: "Under similar traffic and weather, this commute usually takes about 30 minutes." This is a loose, probabilistic "law" of your personal life.

Laws summarise many successful hypotheses and experiments. They are not immune to revision, but they are treated as stable scaffolding until strong reasons arise to change them.

Theories: Frameworks That Generate Hypotheses

Theories are broader conceptual and mathematical frameworks that generate hypotheses and explain why certain laws hold.

- Classic trio from mechanics: falling apples as observations, constant-acceleration laws as regularities, and Newtonian mechanics as the theory that unifies and explains them.
- AI example: learning theory and optimisation frameworks that explain why certain architectures generalise and others overfit on particular data regimes.
- Finance example: asset-pricing theories such as the Capital Asset Pricing Model (CAPM) and its extensions, which generate hypotheses about expected returns and risk premia.

Theories knit together many laws and hypotheses into a network. They say what kinds of entities exist, which quantities matter, and which moves are legitimate when constructing models.

From Lab to Life

In real projects, keeping the three levels apart avoids confusion:

- Hypothesis: "This new data-cleaning step will reduce model error by 5%." Clear, local, testable.
- Law-like claim: "In this domain, adding more diverse training data tends to improve robustness." Broader, but still empirically grounded.
- Theory: "Generalisation in deep networks emerges from implicit regularisation in gradient-based optimisation." A wide-scope framework that connects many phenomena.

When a project stalls, ask yourself: am I adjusting hypotheses, questioning law-like regularities, or challenging the underlying theory?

Common Pitfall

It is easy to slide between levels:

- Dismissing a well-supported theory as "just a theory," as if that meant "mere guesswork," when in science theories are the most integrated, battle-tested frameworks we have.
- Announcing that a single successful experiment has "proven the theory," when in reality it has confirmed one hypothesis under specific conditions.
- Blaming a whole theory for local modelling mistakes, or, conversely, defending a shaky theory by hiding problems in vague hypotheses.

Keeping hypotheses, laws, and theories conceptually separate helps you debug ideas without overreacting in either direction.

2.3 Confirmation and Evidence

We now turn to evidence: what it means for data to support or undermine a hypothesis.

Confirming and Disconfirming Instances

For a simple universal claim like "All swans are white," we can distinguish:

- Confirming instances: observing many white swans supports the hypothesis but never proves it with absolute certainty.
- Disconfirming instances: a single black swan—literally or metaphorically—is enough, in ideal logic, to refute the universal claim.
- Fragile generalisations: hypotheses that survive only because we ignore counterexamples, redefine terms, or narrow the scope after the fact.

In practice, evidence is less clean. Measurements are noisy; categories are fuzzy; and we rarely have universal hypotheses with sharp boundaries. Still, the contrast between weak confirmation and strong disconfirmation is a useful compass.

Bayesian and Frequentist Flavours

Two broad statistical traditions interpret evidence differently.

- Bayesian flavour: treat probabilities as degrees of belief. You start with a prior probability for a hypothesis, observe data, and update to a posterior probability via Bayes' rule.
- Frequentist flavour: treat probabilities as long-run frequencies. You imagine repeating an experiment many times and ask how often a certain outcome would occur if the hypothesis were true.
- Common ground: both approaches formalise how strongly data speak for or against a hypothesis, but they answer slightly different questions.

Analogy: Restaurant Reviews

Imagine deciding whether a restaurant is "good." Each visit is a data point.

- Bayesian view: you start with a prior opinion (maybe based on a friend's recommendation), then adjust your belief after each meal. A surprisingly bad dinner sharply lowers your confidence.
- Frequentist view: you imagine a long run of visits and ask what fraction would be rated at least, say, four stars if the restaurant were truly good. You use observed visits as a sample to estimate that fraction.

Either way, you are turning repeated experience into a reasoned judgment rather than a single snap impression.

In physics and finance, both flavours appear. Particle physicists often talk frequentist language about p-values and "5σ" results, while Bayesian approaches are common in inference and model comparison. Quantitative traders update beliefs about strategies as new market data arrive, explicitly or implicitly performing Bayesian-style learning.

In much applied work, the meeting point of these traditions is the humble p-value. Informally, a p-value measures how surprising your data would be if a baseline "null" story were true: small values flag results that would be rare under that story, large values signal that the data are quite compatible with it. That idea belongs in the main text because it shapes how scientists talk about "significant" and "non-significant" findings; the later technical note in Section 2.7 simply spells out the formal definition and a worked example for readers who want the details.

Empiricism in One Page

Two broad attitudes about knowledge sit in the background of many scientific debates:

- *Empiricism*: experience and observation are the ultimate court of appeal. On this view, even the most elegant theory must answer to data, and concepts without possible empirical contact are suspect.
- *Rationalist strands*: give more weight to a priori reasoning or mathematical structure, sometimes treating certain principles as knowable independently of experience.

Most working scientists mix both: they rely on mathematics and modelling to extend ideas, but insist that measurements, experiments, and well-designed studies have the final say. In this book we mostly speak the language of evidence, experiment, and measurement rather than saying "empiricism" every time; this callout simply names the tradition that puts experience at the centre.

Common Pitfall

When reading statistical reports, watch for (and, if needed, revisit the technical note in Section 2.7):

- Interpreting a small p-value as "the probability the null hypothesis is false," which it is not; it is the probability of seeing data at least this extreme *if* the null were true.
- Ignoring base rates: even a highly accurate test can produce many false positives when the underlying event is rare.
- Treating any single significant result as decisive confirmation rather than as one noisy data point in a longer process of accumulation and replication.

Good scientific reasoning involves asking what else would need to be true for the reported numbers to be informative.

2.4 Falsification and Its Limits

In Chapter 1 we met Popper's falsification idea. We now look more closely at how it fits into real scientific practice.

Bold Hypotheses and Risky Predictions

Popper's central thought is simple:

- Theories should make bold, risky predictions—claims that could easily be shown false by observation.
- The more a theory forbids, the more it says about the world and the more meaningful it becomes.
- Surviving many serious tests without falsification raises a theory's status, but no finite number of tests ever proves it true once and for all.

In Newtonian mechanics, a bold prediction might be the precise path of a planet or the free-fall time of a dropped object under specified conditions. In finance, a bold prediction might be a clear arbitrage that should not persist if a given asset-pricing theory is right.

Ceteris Paribus and Auxiliary Assumptions

Real tests are rarely clean because hypotheses come wrapped in supporting assumptions.

- Ceteris paribus ("all else equal") clauses: predictions often hold only if background conditions remain within a certain range.
- Auxiliary assumptions: beliefs about instruments, calibration, background theories, and data pipelines that must hold for a test to be informative.
- Ambiguous failures: when a prediction fails, it is often unclear whether the core theory, a law, or an auxiliary assumption is at fault.

Everyday Analogy: Baking with a Recipe

Suppose a cake recipe promises a certain result if you follow it exactly.

- Bold prediction: "Bake at 180°C for 30 minutes and you will get a moist cake."
- Auxiliary assumptions: your oven temperature is accurate, your flour is fresh, and you measured ingredients correctly.
- When the cake fails, you must decide: is the recipe wrong, or did one of the background assumptions fail?

Scientific experiments are similar: a failed prediction sends you hunting through both the core theory and the "kitchen" of instruments and data handling.

In a failed physics experiment, for example, you might suspect the theory, but you also check for misaligned detectors, coding bugs in the analysis, or mislabelled cables. In quantitative finance, an apparent arbitrage that refuses to disappear might signal a bug in your backtest rather than a genuine market anomaly.

Common Pitfall

Falsification is powerful but can be misused:

- Declaring a mature, well-tested theory "disproven" on the basis of a single messy result without checking instruments, data quality, and auxiliary assumptions.
- Building theories that are so cushioned by vague clauses and exceptions that no conceivable observation could ever count against them.
- Forgetting that, in many fields, we learn more from patterns of partial successes and failures than from clean yes/no verdicts.

Aim for tests that are sharp enough to hurt when wrong, but honest enough about their background conditions to be interpretable.

2.5 Underdetermination and Theory-Ladenness

Finally, we examine two important ideas: underdetermination (different theories fitting the same data) and theory-ladenness (the way theories shape what we see).

Underdetermination: Many Stories, Same Data

Underdetermination arises when multiple theories or models can explain the same body of observations.

- Cosmology example: historically, different cosmological models could fit early astronomical data with small adjustments, even though they told very different stories about the universe.
- Volatility example in finance: several distinct stochastic volatility models can fit the same option price surface equally well, but they imply different stories about the underlying dynamics.
- Machine learning example: many different neural network architectures can reach similar accuracy on a benchmark dataset, yet they differ in robustness, interpretability, and inductive biases.

Data alone often cannot decide between such rivals. We then appeal to additional criteria: simplicity, coherence with other theories, computational tractability, or explanatory power in other domains.

Theory-Ladenness: Seeing Through Lenses

Theory-ladenness is the idea that what we "observe" is influenced by the concepts and expectations we bring with us.

- Chart example: two traders look at the same price chart. One sees "support" and "resistance" levels; the other, trained in different models, sees volatility regimes and order-flow imbalances.
- Physics example: an untrained observer sees blurred dots on a detector screen; a physicist sees traces of specific particles, shaped by years of learning what to look for.
- Everyday example: once you have learned some basic Newtonian mechanics, you cannot unsee parabolas and forces when you watch someone throw a ball.

Analogy: Camera Lenses and Filters

Imagine photographing the same scene with different lenses and filters:

- A wide-angle lens captures the whole scene but distorts edges.
- A telephoto lens zooms in, revealing details while hiding context.
- A coloured filter emphasises some features and mutes others.

The underlying scene is the same, but what stands out in the photograph depends on the optics. Theories and training are the "optics" of scientific observation.

Theory-ladenness does not mean we can never reach objective knowledge. It does mean we must be aware that observation, training, and theory are entangled. Recognising this helps explain why disputes in complex domains (macroeconomics, climate modelling, AI safety, financial regulation) can persist even with abundant data.

Common Pitfall

Underdetermination and theory-ladenness can be overextended:

- Sliding from "data underdetermine theory in some cases" to "any story is as good as any other," which ignores the hard work of cross-checks, independent measurements, and predictive tests.
- Using "everyone sees through lenses" as an excuse to stop arguing with evidence rather than as a prompt to make assumptions explicit and seek converging lines of support.
- Forgetting that, in many practical contexts (engineering tolerances, safety limits, risk management), different reasonable lenses still converge on similar numerical recommendations.

Philosophical subtlety is useful when it sharpens judgment, not when it paralyses it.

2.6 Summary and Where We Are Heading Next

We close this chapter by collecting the main threads and pointing to their uses later in the book.

- Recap: three basic reasoning patterns—deduction, induction, and abduction—and their detective-story analogies.
- Recap: the distinction between hypotheses, laws, and theories, illustrated with falling apples, constant acceleration, and Newtonian mechanics.
- Recap: the idea of confirmation and disconfirmation, with Bayesian and frequentist flavours as complementary ways to encode evidence.
- Recap: Popperian falsification, together with the practical messiness introduced by ceteris paribus clauses and auxiliary assumptions.
- Recap: underdetermination and theory-ladenness as reminders that data alone rarely pick a single theory and that observation is filtered through conceptual lenses.

These ideas will reappear in later parts of the book. When we study experiments and causal inference, we will lean on the distinctions between induction and abduction. When we analyse case studies in physics, AI, and finance, we will revisit underdetermination, theory-ladenness, and the trade-offs between explanation, prediction, and control.

Try in 60 Seconds

Short checks to test your grasp:

- Take a recent decision you made (about health, money, or study). Identify one deductive, one inductive, and one abductive step you used, even informally.
- Think of a favourite "rule of thumb" (for example about markets or coding). State it as a hypothesis, ask what evidence would disconfirm it, and note any auxiliary assumptions you normally ignore.
- Look at a familiar plot or dashboard from your work and ask: what theory or training shapes what my eyes pick out first?

2.7 Technical Note: Probability, Hypotheses, and p-Values

Because ideas like evidence, significance, and p-values are central to scientific reasoning, we end with a short, more formal clarification. You can skim this now and return to it whenever a later chapter uses these terms.

Basic Probability Language

We think in terms of experiments, outcomes, and events.

- An *experiment* is a repeatable setup that can produce different outcomes (tossing a coin ten times; drawing a sample of returns; running a clinical trial).
- An *outcome* is one possible result of the experiment (a particular sequence of heads and tails; a specific dataset).
- An *event* is a set of outcomes we care about ("at least 8 heads," "a daily loss worse than -5%").
- A *probability* $P(\text{event})$ is a number between 0 and 1 that encodes how often we expect that event in the long run under specified conditions.

Null and Alternative Hypotheses

In classical significance testing we compare two competing claims.

- The *null hypothesis* H_0 is a baseline story (for example "the coin is fair," "the treatment has no effect," "this trading rule has zero excess return").
- The *alternative hypothesis* H_1 is a competing story (for example "the coin is biased toward heads," "the treatment changes the mean outcome," "the rule has positive excess return").
- We choose a *test statistic* T (a function of the data) that should, under H_0, have a known probability distribution.

Definition of a p-Value

Given observed data and a chosen test statistic T, the p-value is:

- formally: $p = P\big(T \text{ is at least as extreme as } T_{\text{obs}} \,\big|\, H_0 \text{ is true}\big)$,
- in words: the probability, *assuming the null hypothesis is true*, of seeing a test statistic as extreme as (or more extreme than) the one we actually observed, purely by chance.

Small p-values mean that the observed data would be unusual if H_0 were true. They do *not* directly tell you the probability that H_0 itself is false; for that you would need a Bayesian calculation that combines a prior belief about H_0 with a likelihood for the data.

A Concrete Example

Consider a coin you suspect might be biased.

- Experiment: toss the coin $n = 100$ times.
- Null hypothesis: H_0: the coin is fair, so the number of heads X follows a binomial distribution with parameters $n = 100$ and $p = 0.5$.
- Test statistic: $T = X$, the number of heads.
- Suppose you observe $X_{\text{obs}} = 70$ heads.
- The p-value is the probability, under H_0, of seeing 70 or more heads in 100 tosses: $p = P(X \geq 70 \mid H_0)$.

If this probability is very small (for example less than 0.01), you say that the observation would be rare if the coin were fair. That is a reason to doubt H_0, but not a guarantee that it is false: unlikely events do sometimes happen.

Significance Thresholds and Caution

Practitioners often compare the p-value to a chosen threshold (such as 0.05 or 0.01) and say a result is "statistically significant" if p falls below that threshold. This convention:

- helps coordinate decisions across studies, but
- should never replace judgment about design quality, effect size, prior plausibility, and potential biases.

Later chapters on statistics and the replication crisis will revisit these points in more depth. For now, the key takeaway is: a p-value quantifies how surprising your data would be *if* a baseline story were true; it is not, by itself, the probability that the story is wrong.

To make this concrete, Figure 2.1 schematically shows a distribution centred near 50 heads with the tail region $X \geq 70$ highlighted; the exact probabilities would follow a binomial law, but the picture emphasises the idea of a small tail area under a baseline story.

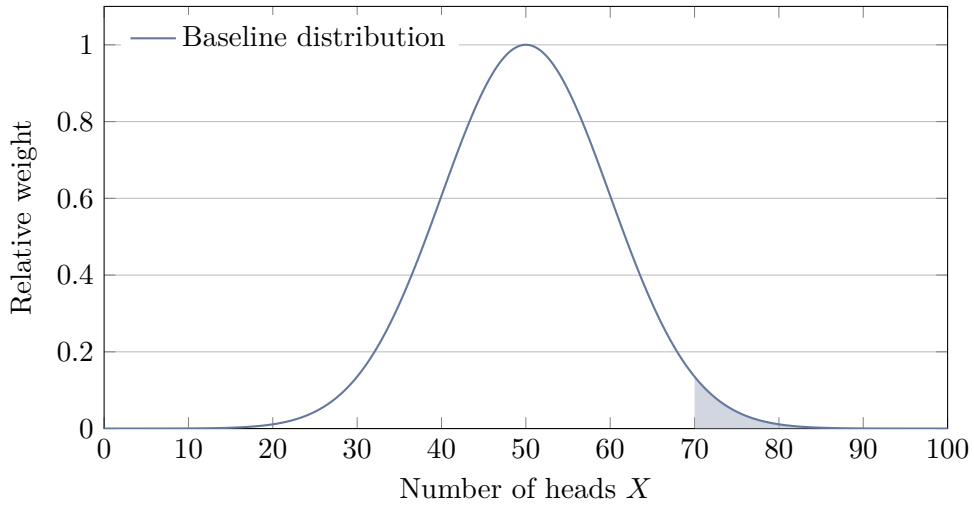

Figure 2.1: Schematic distribution of the number of heads in 100 tosses of a fair coin, with a right-hand tail region (here $X \geq 70$) highlighted to represent a small p-value event.

Chapter 3

The Structure and Dynamics of Scientific Theories

So far we have talked about what science is aiming at and how scientists reason. In this chapter we zoom in on theories themselves: not just as elegant equations on a page, but as living networks of principles, models, measurements, and practices that change over time.

Learning Objectives

After working through this chapter you should be able to:

- describe the "web" view of theories with core principles, auxiliary hypotheses, and measurement links,
- explain the semantic view that treats theories as families of models, with concrete examples from physics, economics, and finance,
- articulate why idealisations and approximations are used, and what can go wrong when they are taken too literally,
- summarise Kuhn's picture of paradigms, normal science, crisis, and revolution, and connect it to examples in physics and finance,
- compare different views of scientific progress: cumulative improvement, revolutions, and pluralist toolboxes.

At a Glance

Scientific theories are not single equations floating in a vacuum. They are structured networks of ideas, models, and measurements that evolve within communities. This chapter shows how those networks are built, how idealisation and approximation keep them manageable, and how large-scale shifts—paradigm changes—can reorganise the whole web.

Everyday Analogy

Think of a modern city's transportation system. There are core lines (subway or major roads), feeder routes (buses, trams), stations and stops (measurement points), and operating rules. The city as a whole is navigable because this network hangs together. Scientific theories work similarly: core principles and laws, auxiliary hypotheses, measurement procedures, and worked examples form a web you can travel through.

3.1 Theories as Networks, Not Just Equations

On a blackboard, a theory can look like a compact set of equations or postulates. In practice, a working theory is more like a web.

- Core principles: central laws or postulates (for example Newton's laws, conservation principles, or no-arbitrage conditions).
- Auxiliary hypotheses: domain-specific assumptions that connect the core to particular situations (for example "friction is negligible," "the market has no transaction costs," "measurement error is small and symmetric").
- Measurement links: operational procedures that tie abstract quantities to instruments and data streams (for example how we measure temperature, volatility, or neural activity).

Viewed this way, a theory is not a single statement but a network in which changing one piece can tug on many others. When an experiment fails, the resulting tension can propagate through core principles, auxiliaries, and measurement assumptions.

Analogy: City Transport Network

In a city's transport system:

- Core lines: underground or main train routes that carry the bulk of commuters.
- Feeder routes: buses and trams that connect neighbourhoods to the core.
- Stations: places where the abstract network meets concrete geography—boarding points, ticket machines, timetables.

A delay on one core line can ripple through the whole system. Similarly, a problem with a single auxiliary hypothesis or measurement link can propagate tension through a scientific theory.

In Newtonian mechanics, for example, the core includes Newton's three laws and conservation principles. Auxiliary hypotheses include idealisations such as "rigid body," "point mass," or "uniform gravitational field." Measurement links include how we define and read off positions, velocities, and forces in experiments and simulations.

Common Pitfall

It is tempting to identify a theory only with its prettiest equations. This can hide:

- the role of auxiliary assumptions (which often do the heavy lifting in applications),
- the dependence on measurement practices (which can change over time),
- the fact that many apparent "tests of a theory" are really tests of small subnetworks within the larger web.

Keeping the network picture in mind makes it easier to see where revisions are needed when something breaks.

3.2 The Semantic View: Models at the Centre

Older "syntactic" pictures of theories emphasised axioms and deductions: a theory was a set of sentences in a formal language, and models were secondary. The semantic view in philosophy of science flips this emphasis.

Theories as Families of Models

On the semantic view:

- A theory is associated with a family of models: mathematical or computational structures that satisfy its core principles and laws.
- Each model represents a possible way the world could be, according to the theory.
- Real systems are compared to models by checking how well particular models capture selected aspects of the system.

Instead of asking "Is the theory true?" in one leap, we often ask: "Which model from this family best fits the situation, and where do its simplifications matter?"

Concrete Examples

Some familiar examples make this view more tangible.

- Ideal gas model: within thermodynamics and statistical mechanics, the ideal gas model is one member of a family of models that share core equations of state but differ in parameters and boundary conditions.
- Supply-and-demand curves: in basic economics, simple supply-and-demand diagrams are models that instantiate more general equilibrium principles, with specific functional forms and exogenous shocks.
- Black–Scholes model: in finance, Black–Scholes is one model in a broader family of option pricing models, each embodying versions of no-arbitrage and stochastic process assumptions.

Analogy: Furniture Catalogue vs. Actual Room

A furniture catalogue shows many possible sofas, tables, and lamps that all respect a brand's design language. Your actual living room contains one specific configuration of these pieces, plus personal clutter. Similarly:

- The theory plays the role of the brand's design language.
- Individual models are catalogue entries: concrete, idealised possibilities.
- Real systems are rooms furnished with a particular combination, plus all the messy details the catalogue ignores.

From Lab to Life

In real projects, thinking "theory → model family → specific model" helps you:

- avoid over-committing to one favourite model when several candidates fit the data comparably well,
- see where disagreements are really about model choice versus about the underlying theory,
- design robustness checks by trying multiple models within the same theoretical family.

3.3 Idealisation and Approximation

Theories and models survive only if they are tractable and illuminating. Idealisation and approximation are the main tools to achieve this.

Why Idealise?

Inevitably, we strip away some complexity:

- Tractability: simplifying a system can turn an intractable problem into one with analytic solutions or feasible simulations.
- Conceptual clarity: idealised scenarios highlight key mechanisms without distraction from every real-world wrinkle.
- Communication: clean models are easier to teach, publish, and critique than tangled, fully realistic ones.

Examples abound:

- "Frictionless planes" in introductory mechanics.
- "Spherical cows" in physics jokes about over-idealised models.
- "Perfectly rational agents" in early economic and financial models.

Dangers of Forgetting the Idealisation

Idealisation becomes dangerous when:

- we confuse the model with the world and silently assume that neglected factors never matter,
- we push conclusions far outside the regime where the simplifications are reasonable,
- we forget which idealisations were made and cannot reconstruct the path from reality to the model.

Analogy: Caricatures vs. Portraits

A caricature exaggerates a few features of a face (nose, hairstyle, posture) to capture personality quickly. It is powerful for:

- highlighting what stands out at a glance,
- communicating character in a crowded newspaper.

But you would not use a caricature for a passport photo. Scientific models are often caricatures: excellent for certain questions, misleading for others. Knowing which is which is part of expertise.

Common Pitfall

Three recurring mistakes with idealisation:

- Treating an ideal model as a universal truth rather than a controlled exaggeration.
- Forgetting to "come back down" to reality by checking whether ignored effects (friction, liquidity, bounded rationality) could change key conclusions.
- Comparing two models as if they were about the same idealisation level when one is deliberately coarse and the other fine-grained.

When in doubt, ask: what did we throw away to make this model simple, and could that matter here?

3.4 Paradigms, Normal Science, and Revolutions

Thomas Kuhn popularised a picture of scientific change that emphasises long periods of "normal science" punctuated by crises and revolutions.

Paradigms as Shared Packages

For Kuhn, a paradigm is more than a theory in isolation. It includes:

- Exemplary problems: standard textbook and lab problems that train newcomers in how to "see" and solve tasks.
- Shared standards: what counts as a good explanation, an acceptable approximation, or a decisive experiment.
- Training and textbooks: the narratives, diagrams, and worked examples that shape how students learn to think.
- Instrumentation and tools: standard apparatus, software, and data sources that define what is easily measurable.

Under a stable paradigm, most researchers work on puzzles whose solutions are expected to exist within the existing framework. This is normal science.

Crisis and Revolution

Crises emerge when:

- anomalies accumulate—results that systematically resist explanation within the current paradigm,
- rival approaches start to solve important problems more naturally,
- confidence in core standards erodes.

Revolutions occur when a new paradigm reorganises the field:

- Classical to quantum physics: moving from deterministic orbits to probabilistic wavefunctions and operator algebra.
- Deterministic to probabilistic risk models in finance: shifting from point predictions to distributions, value-at-risk, and scenario analysis.

From Lab to Life

You can often spot paradigm boundaries by looking at:

- which equations appear on lecture hall walls,
- which datasets and benchmarks "everyone" uses,
- which kinds of questions are considered respectable for PhD theses.

If you find yourself asking questions that feel important but "out of scope," you may be nudging at the edges of a paradigm.

3.5 Scientific Progress: Truth, Tools, or Both?

How should we describe long-term scientific change? Several pictures compete.

Cumulative vs. Revolutionary Views

A cumulative view emphasises:

- gradual accumulation of results,
- extension and refinement of existing theories,
- incremental improvement in predictive accuracy and explanatory scope.

A revolutionary view emphasises:

- discontinuous conceptual breaks (Newton → Einstein; classical → quantum),
- incommensurability between some old and new concepts,
- periods where scientists across paradigms "talk past" each other.

Real history usually contains both: long stretches of cumulative work punctuated by occasional conceptual overhauls.

Realist, Relativist, and Pluralist Flavours

Beyond the pace of change, philosophers disagree about what progress is *towards*.

- Realist flavour: science progresses by getting closer to the truth about how the world is, even if our best theories are still approximations.
- Relativist flavour: what counts as progress is heavily shaped by shifting values, interests, and power structures; talk of "closer to the truth" is suspect.
- Pluralist flavour: progress can mean building a richer toolbox of models and methods that serve different purposes well, without assuming there is a single final theory.

Analogy: Updating Software vs. Switching Operating Systems

Think of your scientific toolkit as software:

- Cumulative updates: minor version bumps that fix bugs and add features without changing the basic interface.
- Revolutionary changes: switching operating systems, which may require relearning commands and abandoning some old apps.
- Pluralist setups: running multiple operating systems or virtual machines side by side because different tasks are best served by different environments.

Scientific fields often show all three behaviours across different subdomains.

3.6 Summary and Short Discussion Prompts

We close by highlighting key takeaways and a few prompts for your own case studies.

- Recap: theories as networks of core principles, auxiliary hypotheses, and measurement links, not just isolated equations.
- Recap: the semantic view that treats theories as families of models, with concrete catalogue-like examples.

- Recap: idealisation and approximation as essential tools that require careful boundary awareness.
- Recap: Kuhn's paradigms, normal science, crises, and revolutions as one lens on scientific change.
- Recap: contrasting pictures of progress—toward truth, shaped by values, or expanding toolboxes—that we will revisit in later case studies.

Short discussion prompts:

- Is current AI research in a stable paradigm, in pre-paradigmatic exploration, or in a transition phase between paradigms?
- What would count as a genuine "revolution" in quantitative finance: a new asset-pricing theory, a regulatory regime shift, a radically different market structure, or something else?
- In your own field, can you sketch the core principles, auxiliaries, and measurement links that make up its main theory web?

Try in 60 Seconds

Quick reflections:

- Sketch, in three bullet points, the core principles, auxiliaries, and measurement links of one theory you use regularly.
- Name one idealisation in a model you like (for example "frictionless" or "rational agents") and one way it might fail.
- Decide whether you feel more drawn to a cumulative or revolutionary picture of progress in your own field, and why.

3.7 Technical Note: Theories as Graphs and Model Families

For readers who like a slightly more formal picture, this note sketches one way to represent theories as networks and model families. The main text does not require these details, but they can help organise complex examples.

A Simple Graph View

We can treat a theory as a typed graph:

- Nodes represent elements such as core principles (P_i), auxiliary hypotheses (A_j), and measurement rules (M_k).
- Directed edges represent relations such as "supports," "depends on," or "is operationalisation of."
- Subgraphs correspond to local modelling setups used in specific experiments or applications.

Failures and revisions can then be described as:

- deleting or weakening nodes (abandoning an auxiliary hypothesis),
- adding new nodes (introducing a new measurement rule or correction term),

- rewiring edges (changing which core principle an auxiliary depends on).

This graph view is deliberately coarse, but it captures the intuition that theories are structured objects, not flat lists of sentences.

Figure 3.1 sketches a toy theory network with core principles, auxiliaries, and measurement links represented as nodes and arrows.

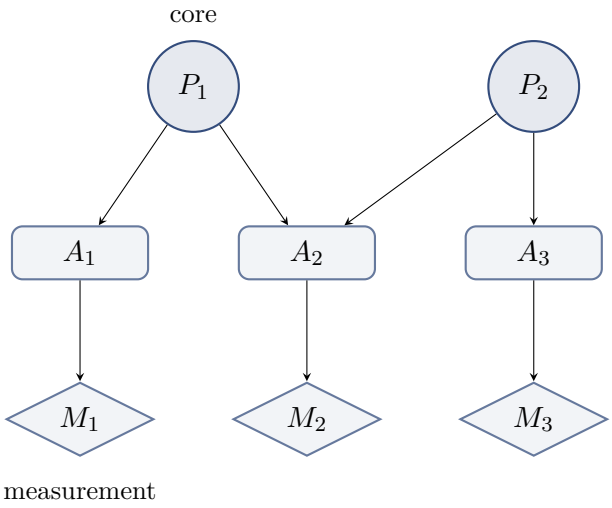

Figure 3.1: Toy "theory as graph": core principles P_i, auxiliary hypotheses A_j, and measurement links M_k connected by dependency arrows.

Theories and Model Classes

In the semantic view:

- A model is a mathematical or computational structure $\mathcal{M}$ that makes all the theory's core sentences come out true (or approximately so) when interpreted within $\mathcal{M}$.
- The theory corresponds to the class $\{\mathcal{M}_\alpha\}$ of all such models, together with mappings that link elements of each $\mathcal{M}_\alpha$ to features of real systems.

You need not master model theory to use this idea. It is enough to remember:

- "Theory" $\approx$ structured family of models plus rules for mapping them to reality.
- "Applying a theory" $\approx$ choosing, calibrating, and sometimes combining models from that family for a given task.

Later, when we talk about model pluralism and complex systems, this picture will help explain why using several models from different families can be a strength rather than a sign of confusion.

Chapter 4

Realism, Instrumentalism, and the Nature of Explanation

We now turn to two questions that sit just beneath the surface of scientific practice: what are theories about, and what does it mean to explain something scientifically? The answers matter for how we treat unobservable entities, choose models, and balance prediction, understanding, and control.

Learning Objectives

After working through this chapter you should be able to:

- distinguish scientific realism and instrumentalism and illustrate each with examples from physics, AI, and finance,
- describe and contrast causal, statistical, and unifying explanations using everyday analogies and scientific cases,
- explain the tension between mechanistic explanations and black-box models that predict well but are hard to interpret,
- analyse trade-offs between explanation, prediction, and control in domains such as medicine, regulation, and risk management,
- reflect on your own preferences for transparent versus opaque but accurate models in high-stakes decisions.

At a Glance

Realists treat successful theories as telling us, more or less, how the unobservable world is. Instrumentalists treat theories as tools for organising experience, prediction, and control without demanding that their entities "really exist." Explanations themselves come in different flavours—causal, statistical, and unifying—and real practice often trades off depth of understanding against raw predictive power.

Everyday Analogy

Think of a navigation app:

- A realist attitude says: the map is useful because it roughly matches real streets, junctions, and traffic flows.
- An instrumentalist attitude says: I do not care whether the app's internal representation matches the world, as long as it gets me to my destination reliably.
- Different explanations for a delay ("accident ahead," "rush hour pattern," "roadworks on a key artery") mirror causal, statistical, and unifying stories.

We apply similar attitudes and explanatory styles, often implicitly, when we use scientific theories.

4.1 What Are Scientific Theories About?

Scientific theories talk about electrons, fields, genes, spacetime, latent factors, and risk-neutral probabilities—entities we do not observe directly. How seriously should we take this talk?

Scientific Realism

Scientific realism, in a first approximation, says:

- Our best theories aim to describe both observable and unobservable aspects of the world.
- When a theory is mature and successful across many tests, we have reason to believe that at least its central claims are approximately true.
- Entities posited by such theories (electrons, fields, quarks, underlying risk factors) probably exist in something like the way the theory describes.

Realists often appeal to a "no miracles" intuition: it would be miraculous if a theory made consistently accurate, wide-ranging predictions while being completely wrong about what the world is like.

Instrumentalism

Instrumentalism offers a different attitude:

- Theories are instruments for organising experience, making predictions, and guiding action.
- Questioning whether unobservables "really exist" is often seen as metaphysical excess or simply not well posed.
- What matters is whether a theory works—predictively, computationally, or practically—for our aims.

Instrumentalists are not anti-science; they simply suspend or downplay ontological commitments. For them, the Black–Scholes model in finance or a deep neural network in AI can be extremely valuable even if we hesitate to say that *volatility surfaces* or *latent features* correspond to specific things out there in the world.

Do Risk-Neutral Probabilities "Exist"?

Risk-neutral probabilities in finance make the contrast vivid.

- In pricing theory, we often work under a "risk-neutral measure": a probability distribution under which all appropriately discounted asset prices are martingales.
- This measure is extremely convenient for deriving prices of derivatives and for proving neat theorems.
- But no real investor is truly risk-neutral, and observed empirical frequencies do not match the risk-neutral distribution.

A realist about risk-neutral probabilities might say: there is an underlying structure in markets that our risk-neutral measure is capturing, even if imperfectly. An instrumentalist might say: the measure is a clever re-parameterisation that simplifies calculations and enforces no-arbitrage, but asking whether it "really exists" is like asking whether isobars on a weather map "really exist" in the air.

From Lab to Life

Your stance matters:

- In physics, a realist attitude may push you toward interpretations of quantum mechanics that treat the wavefunction as a real physical field, while an instrumentalist stance may keep you content with a rulebook for calculating measurement outcomes.
- In AI, realists about learned features may look for mechanistic circuits inside networks; instrumentalists may treat networks as opaque but useful gadgets.
- In finance, realists about factors may search for underlying economic drivers; instrumentalists may focus on predictive factors whether or not they map neatly to stories.

There is an extra twist when models are used for control rather than mere description. In AI alignment debates (Chapter 12) and in reflexive markets (Chapter 13), objectives, loss functions, and pricing rules do not just represent the world; they help shape it. Agreement on a particular risk-neutral valuation practice, for example, can stabilise derivatives markets around that convention. The book will not force you into one camp, but it will keep asking you to notice which hat you are wearing—and how your models act back on the systems they describe.

4.2 Causal, Statistical, and Unifying Explanations

Explanations come in different shapes. Three especially important ones are causal, statistical, and unifying explanations.

Causal Explanations

In a causal explanation, we say roughly:

- "Event C caused event E," or
- "This mechanism produced that outcome."

Everyday and scientific examples include:

- Coffee machine: "The coffee tastes burnt because the heating element is stuck on and overheats the water."
- Medicine: "The patient's symptoms improved because the drug blocked a specific receptor, interrupting a disease pathway."
- Engineering: "The bridge collapsed because of resonance with wind gusts at a particular frequency."

Good causal explanations usually identify a mechanism and show how intervening on the cause would change the effect.

Statistical Explanations

Statistical explanations appeal to probabilities and regularities rather than single, sharp causal chains.

- Weather: "It rained today because this region is in a season where such fronts bring rain 70% of the time; the patterns matched a typical rainy configuration."
- Credit risk: "This portfolio experienced a default because, given its risk profile, there was a small but significant probability of such an event, and such events occur at roughly the expected rate over many portfolios."
- Epidemiology: "Non-smokers sometimes develop lung cancer because, even though smoking greatly increases risk, there is still a baseline probability of disease due to other factors."

Here, the explanation lies in showing how the observed event fits into a broader pattern of frequencies or distributions.

Unifying Explanations

Unifying explanations reduce the number of independent stories we need to tell.

- Maxwell's electromagnetism: unifies electricity, magnetism, and light under a single field theory.
- No-arbitrage in finance: unifies diverse pricing relations under a common constraint that rules out free money loops.
- Symmetry principles: explain multiple conservation laws (energy, momentum, charge) as consequences of underlying symmetries.

A unifying explanation is powerful when it lets us see many seemingly different phenomena as instances of the same underlying structure.

Analogy: Fixing Machines, Pricing Insurance, and Master Recipes

- Causal: a technician explaining a broken coffee machine points to a specific clogged valve or broken pump.
- Statistical: an actuary explaining life insurance premiums points to mortality tables and risk classes rather than a mechanism for each individual's lifespan.
- Unifying: a chef explaining a restaurant's menu points to a handful of master recipes that, with variations, generate most dishes.

Scientific explanations mix these flavours in different proportions depending on the question and domain.

Common Pitfall

Two frequent confusions:

- Treating any statistical regularity as if it were automatically causal, without checking for confounders or alternative mechanisms.
- Assuming that unification is always better, even when a more modest, domain-specific explanation may be more reliable for decisions.

In later chapters on causality and statistics we will return to these distinctions with more technical tools.

4.3 Mechanisms, Models, and "Black Boxes"

Modern practice, especially in AI and quantitative finance, makes the tension between mechanistic and black-box explanations concrete.

Mechanistic Explanations

A mechanistic explanation opens the box and shows you how the parts interact.

- Biochemistry: tracing a signalling pathway from receptor activation through intermediate molecules to gene expression changes.
- Engineering: decomposing a control system into sensors, controllers, and actuators and showing how signals move through the loop.
- Neuroscience: linking behaviour to activity in specific circuits or cell types, with a story about how information flows.

Mechanistic stories are often satisfying because they support counterfactuals: "If we block this step, the effect should disappear."

Black-Box Models

Black-box models prioritise predictive performance over interpretability.

- Deep learning: large networks that classify images or predict language with high accuracy but resist simple mechanistic interpretation.
- Quant strategies: complex, proprietary algorithms that trade based on a combination of signals whose individual roles are opaque even to their designers.
- Ensemble methods: random forests or boosted trees that combine many weak learners into a strong predictor that is hard to summarise succinctly.

In many competitive environments, black-box models win on performance metrics while raising concerns about transparency, bias, and robustness.

Analogy: Driving a Car vs. Knowing the Engine

Most people can drive a car confidently without understanding internal combustion, fuel injection, or motor control software. They rely on:

- an interface (steering wheel, pedals, dashboard),
- rules of thumb (how quickly the car accelerates, how it handles in rain),
- trust in engineers and regulations.

Similarly, you can use a powerful model via its interface (API, dashboard, high-level documentation) without fully understanding its internal workings. The question is when that is acceptable and when deeper mechanistic understanding is required.

From Lab to Life

In practice, teams often blend approaches:

- Use interpretable models or mechanistic reasoning in safety-critical components where understanding is non-negotiable.
- Allow more black-box components in peripheral or exploratory parts of a system where failure is tolerable and performance is key.
- Layer explanation tools (feature attributions, sensitivity analyses, scenario tests) on top of black-box models to recover partial understanding.

Philosophy of science provides language for these design decisions and for communicating them to stakeholders.

4.4 Explanation, Prediction, and Control

We now connect explanation to two other central aims of science: prediction and control.

When Prediction Outruns Explanation

There are domains where models predict extremely well but offer limited explanatory insight.

- Complex ML systems that beat human benchmarks in games or pattern recognition tasks without yielding simple, human-scale stories about how they succeed.
- Short-term weather nowcasting using high-resolution data and machine learning, which can outperform simpler dynamical models in accuracy while being harder to interpret mechanistically.
- High-frequency trading algorithms that exploit fleeting statistical patterns without clear economic narratives.

Here, the temptation is to say: "As long as it works, we do not care why." But regulators, clinicians, and risk managers often *must* care why.

When Explanation Outruns Prediction

We also find elegant explanations that do not translate directly into strong predictive performance.

- Toy models in economics or ecology that clarify mechanisms but omit enough realism that their numerical forecasts are rough at best.
- Conceptual models in climate science or epidemiology that capture core feedbacks and qualitative behaviour but require heavier numerical machinery for precise forecasts.
- Educational models in physics (for example simple harmonic oscillators) that are crucial for understanding but need corrections for real-world systems.

These models earn their keep by structuring thinking, organising questions, and guiding the design of more detailed predictive tools.

Balancing Aims in Practice

Different roles weight explanation, prediction, and control differently.

- Regulators and risk managers: often require a minimum level of explanation for models that affect capital requirements, safety margins, or systemic risk assessments.
- Doctors and patients: may accept some black-box elements in diagnostic tools, but want mechanistic or at least semi-interpretable stories when deciding on major treatments.
- Engineers and traders: may lean more on prediction and control in fast-moving contexts, while still needing enough understanding to diagnose failures.

Common Pitfall

It is easy to:

- romanticise explanation and dismiss high-performing but opaque models out of hand, or
- celebrate prediction so much that any call for understanding or accountability is labelled "anti-innovation."

Good practice asks: for this decision, in this domain, with these stakes, what mix of explanation, prediction, and control is responsible?

4.5 Summary and Reflection Questions

We close by recapping the main ideas and posing a few questions for reflection.

- Recap: scientific realism and instrumentalism as two attitudes toward what theories say about unobservables.
- Recap: causal, statistical, and unifying explanations as complementary ways of making sense of phenomena.
- Recap: the tension between mechanistic explanations and black-box models, especially in AI and finance.
- Recap: the need to balance explanation, prediction, and control differently across domains and decision types.

Reflection questions:

- In a medical context, would you prefer a perfectly transparent model that is slightly less accurate, or a black-box model that is more accurate but only partially explainable? Why?

- In algorithmic trading or credit scoring, where do you think the line should be drawn between raw predictive power and explainability for regulators and clients?
- For one theory you regularly use (in physics, AI, finance, or another field), are you mostly a realist or an instrumentalist—and does your behaviour match your answer?

Try in 60 Seconds

Short exercises:

- Pick one theory you rely on and write a single sentence from a realist stance and a single sentence from an instrumentalist stance about it.
- Classify an explanation you saw recently (in news, work, or class) as mainly causal, statistical, or unifying.
- Think of one model you use and say whether you treat it more like a mechanism you understand or a black box you monitor by performance.

4.6 Technical Note: Sketching Causal and Statistical Explanation

This brief technical note provides a minimal formal vocabulary for later chapters on causality and statistics. The main text of this chapter can be read without it, but returning here will help when we start drawing causal diagrams and talking about interventions.

Causal Claims and Interventions

A simple way to sharpen causal talk is to imagine interventions.

- We say that C is a cause of E (in a given model) if, roughly, intervening to change C while holding relevant background conditions fixed would change the probability or occurrence of E.
- Symbolically, we compare $P(E \mid \mathrm{do}(C = c_1))$ and $P(E \mid \mathrm{do}(C = c_0))$ in a causal model, where $\mathrm{do}(\cdot)$ denotes an intervention.

The technical machinery of causal graphs and structural equations will appear later; for now, the key idea is that causal explanations support "what if we changed this?" questions, not just "what tends to go with what?" descriptions.

Statistical Regularities

In a purely statistical explanation, we highlight patterns such as:

- $P(E \mid C) > P(E)$: the event E is more likely when C occurs than in general.
- Long-run frequencies: in repeated trials, events of type E happen in a certain fraction of cases that matches or clarifies observed data.

Such patterns can be informative without yet specifying which arrows in a causal diagram are present. Later, we will see how to combine statistical regularities with causal assumptions to move from mere association to intervention-ready knowledge.

To fix ideas, Figure 4.1 contrasts a simple causal diagram, where changing C would change E, with a purely statistical link that encodes correlation without committed arrows.

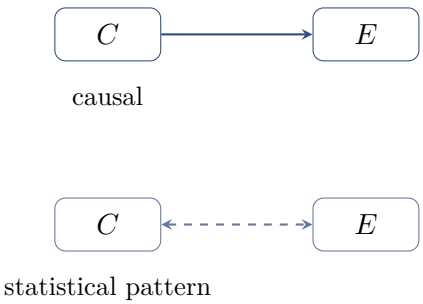

Figure 4.1: Simple contrast between a causal link (top), where intervening on C would change E, and a purely statistical association (bottom), which encodes correlation between C and E without yet fixing the direction of causation.

Unification in Simple Terms

We can also sketch unification informally:

- A theory provides a small set of principles or equations.
- Many different models instantiate those principles in different settings.
- A unifying explanation shows how seemingly unrelated phenomena correspond to different models within the same theoretical framework.

When you see the same mathematical structure (for example a harmonic oscillator or a diffusion process) cropping up in physics, biology, and finance, you are witnessing unification at work.

Part II
Practical Considerations

Part II Overview

This part moves from abstract structure to hands-on practice. Chapter 5 examines how concepts become numbers, how measurement error and latent constructs shape data, and why standards and units matter for cumulative work. Chapter 6 surveys experimental, quasi-experimental, and observational designs, highlighting randomisation, confounding, and core identification strategies that support causal claims in physics, AI, and finance. Chapter 7 looks at statistics as uncertainty bookkeeping, the dangers of overfitting and p-hacking, and lessons from the replication crisis. Chapter 8 analyses how values, ethics, and risk attitudes shape scientific practice and its applications. Chapter 9 then sketches complexity, interdisciplinarity, and the limits of models as a bridge to the case studies.

Chapter 5

Measurement, Operationalisation, and Error

Theories and models only touch the world through measurements. This chapter focuses on how vague concepts become numbers, how errors creep in, and how to think responsibly about the data you feed into models in physics, AI, and finance.

Learning Objectives

After working through this chapter you should be able to:

- explain what it means to operationalise a concept and give examples from different domains,
- distinguish systematic from random error and describe their practical consequences,
- describe latent constructs and factors and how they are inferred indirectly,
- discuss why standards, units, and comparability matter for cumulative science,
- apply a short checklist before trusting a number in a model or decision.

At a Glance

Measurement links the abstract language of theories to concrete instruments, records, and datasets. Operationalisation turns ideas like "intelligence," "volatility," or "temperature" into agreed procedures. Every such link carries assumptions and errors. Learning to see and question these links is as much a philosophical skill as it is a technical one.

Everyday Analogy

Imagine buying running shoes labelled with a "comfort score" from 1 to 10. Somewhere behind that label lies an operationalisation: perhaps a lab protocol, perhaps a survey, perhaps a single designer's hunch. The number only means what the underlying procedure makes it mean. Scientific measurements work similarly: the operational story matters.

5.1 From Concepts to Measurements

Scientific work begins with concepts: temperature, stress, learning progress, market risk. To bring them into models, we must operationalise them.

Operationalisation: Making Concepts Measurable

Operationalisation means specifying how a concept is to be measured or observed in practice.

- Temperature: defined via thermodynamic principles but in practice measured by thermometers calibrated against fixed points and standard scales.

- Learning progress: might be measured via test scores, time to mastery on tasks, error rates of a model over epochs, or self-report questionnaires.
- Market stress: could be proxied by volatility indices, credit spreads, liquidity measures, or combinations thereof.

Each choice embeds assumptions about which aspects of the underlying concept matter for the questions at hand.

Analogy: Measuring Comfort in Shoes

There are many ways to "measure" comfort:

- subjective ratings from wearers after a week,
- lab tests of pressure distribution on the foot,
- durability tests on cushioning materials.

Each method captures a different slice of "comfort." No single operationalisation is the concept itself, but some will be more appropriate than others for specific uses (long-distance running vs. office wear).

Common Pitfall

It is easy to:

- forget that a measurement is an operationalised slice of a concept, not the concept in full,
- treat whatever is easy to measure as "what matters," simply because it appears in dashboards and models,
- compare numbers across studies or products without checking whether they use the same operationalisation.

Whenever you see a number, ask: what procedure, instrument, or convention produced it?

5.2 Measurement Error and Uncertainty

No measurement is perfect. Understanding error types helps you judge reliability and design better studies and systems.

Systematic vs. Random Error

Two broad categories are especially important:

- Systematic error (bias): measurements are consistently off in one direction (for example a bathroom scale that always reads 1 kg too high; a miscalibrated volatility estimator).
- Random error (noise): measurements scatter around a true value in an unpredictable way (for example small timing variations in a reaction time experiment; tick-level price noise).

Systematic error shifts averages; random error widens variability. Both can mislead, but they demand different countermeasures.

Everyday and Scientific Examples

Examples make the distinction concrete:

- Bathroom scale: standing on the scale repeatedly shows small variation (random error) around a value that might itself be shifted up or down (systematic error).
- Fitness tracker: step counts fluctuate because the movement classifier sometimes misfires (random error) and because some activities are systematically undercounted (systematic error).
- Options data: stale or misaligned feeds can introduce systematic mispricing in a dataset, while bid–ask bounce and microstructure noise add random variation.

From Lab to Life

In practice:

- random error can often be reduced by averaging over repeated measurements or larger samples,
- systematic error requires calibration, redesign, or explicit correction—averaging does not fix a biased instrument,
- model uncertainty (for example about which volatility estimator to use) adds a third layer beyond pure measurement error.

Good experimental and data-engineering practice always asks: are we mostly facing noise, bias, or model misspecification?

Common Pitfall

Do not:

- assume that high precision (many decimal places) implies low error,
- treat all uncertainty as random noise and hope it "cancels out",
- forget that aggregation (summing across assets, averaging across patients) can hide large systematic errors.

Numbers can look crisp while resting on shaky measurement foundations.

5.3 Constructs and Latent Variables

Some of the most important scientific concepts are not directly measurable at all. Instead, we infer them indirectly as latent variables or constructs.

Latent Constructs Across Domains

Latent constructs appear in many fields:

- Psychology: intelligence, anxiety, personality traits.
- Economics and finance: risk appetite, expected inflation, latent factors in factor models.
- Machine learning: representation quality, cluster structure, latent topics in text models.

We never observe these constructs directly. Instead we observe indicators (test items, price series, behaviour) and fit models that infer latent structure.

Factor Models as Hidden-Cause Stories

Factor models in finance and other domains provide a clean example.

- Observed returns are expressed as combinations of a small number of latent factors plus noise.
- Each factor is interpreted as capturing some underlying driver (market, value, size, momentum, liquidity).
- The strength of the loadings tells us how sensitive each asset is to each factor.

Even when the math is solid, interpretation requires care: several different factor structures can fit the same data about equally well.

Analogy: Shadows and Hidden Shapes

Imagine watching shadows cast on a wall by moving objects behind a screen:

- You only see the 2D silhouettes (observable indicators).
- You infer the 3D shapes and motions (latent causes) that could have produced them.
- Different hidden setups can, in principle, generate very similar shadow patterns.

Latent variable models are like systematic ways of guessing the hidden shapes from their shadows.

Common Pitfall

With latent constructs it is tempting to:

- reify the construct as if it were a directly measurable object (treating "intelligence" or "risk appetite" as if you could pour it into a beaker),
- ignore model dependence and act as if one factor decomposition were uniquely correct,
- forget that cultural, institutional, and market contexts shape which indicators are even considered.

Treat latent constructs as carefully crafted stories that earn trust by their track record, not as direct readings from nature.

5.4 Standards, Units, and Comparability

Beyond concepts and error, shared standards and units are what make science cumulative.

Why Standards and Units Matter

Common standards let different labs, markets, and eras talk to each other.

- Physics: SI units for length, mass, time, current, temperature, amount of substance, and luminous intensity.
- Finance: conventions for day counts, compounding, volatility scaling (for example annualising daily volatility).
- Data science: agreed preprocessing pipelines and benchmark datasets.

Without clear units and conventions, numerical comparisons quickly become meaningless.

Cross-Study and Cross-Market Comparisons

Comparability issues arise when:

- one study uses Celsius, another Kelvin; one uses metres per second, another kilometres per hour,
- one volatility series is computed from close-to-close returns, another from high-frequency data with a different sampling scheme,
- two AI models report accuracy on nominally the same task but with different data cleaning or class distributions.

From Lab to Life

Before combining or comparing numbers across sources:

- check units and scaling conventions carefully,
- examine definitions of key quantities (“default,” “event,” “episode”),
- document any transformations you apply (normalisations, rescalings).

This small discipline often prevents large interpretive errors.

5.5 Summary and Practical Checklist

Key takeaways from this chapter:

- Measurement is about operationalising concepts; the story behind a number matters.
- Errors can be systematic or random, and they demand different responses.
- Latent constructs and factors are inferred from indicators; they are powerful but model-dependent.
- Standards and units make results comparable across labs, markets, and time.

Before trusting a number enough to base a decision on it, ask:

- What exactly does this number measure, operationally?
- What are the likely sources of systematic and random error?
- Is the underlying construct latent, and if so, which model defines it?
- Are units and conventions compatible with other numbers I am comparing it to?

Try in 60 Seconds

Quick checks:

- Take one metric you use at work or in study and write down, in one sentence, how it is actually measured.
- Decide whether a recent surprising number you saw is more likely affected by bias, noise, or both.
- Open a report or dashboard and verify units and definitions for one key quantity.

5.6 Technical Note: A Simple Error Model

To anchor talk of measurement error, we sketch a minimal mathematical picture. This is not a full statistics course, but it gives a language for later chapters.

Additive Error Decomposition

Suppose we want to measure a quantity X (for example true temperature or true volatility). Our measurement instrument produces a reading

$$Y = X + B + \varepsilon,$$

where:

- B is a systematic error term (bias), which may be constant or depend on conditions,
- ε is a random error term with mean zero and some variance.

Then:

- The expected value of the measurement is $\mathbb{E}[Y] = X + B$.
- The variance of the measurement is $\mathrm{Var}(Y) = \mathrm{Var}(\varepsilon)$ if B is treated as fixed.

Calibration aims to estimate and correct B; repeated measurements can reduce the impact of ε via averaging.

Figure 5.1 visualises these components: the true value, the biased mean measurement, and the spread due to random error.

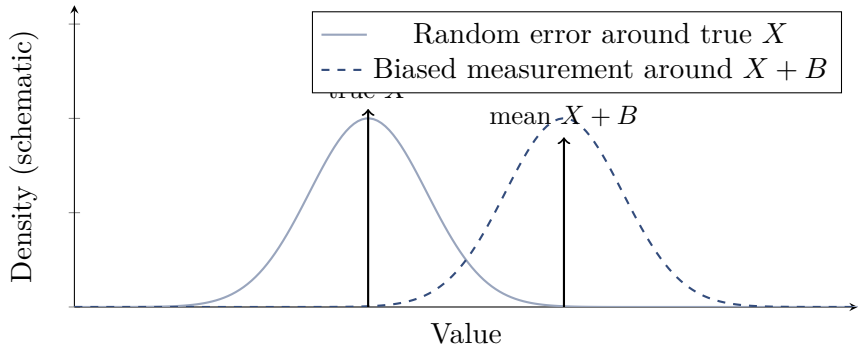

Figure 5.1: Schematic view of measurement error: random error produces spread around a true value X, while systematic error (bias) shifts the mean away from X.

Repeated Measurements

If we take n independent measurements $Y_1, \ldots, Y_n$ of the same quantity under similar conditions, each obeying $Y_i = X + B + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i] = 0$, then the sample mean

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

has:

- expected value $\mathbb{E}[\bar{Y}] = X + B$,
- variance $\mathrm{Var}(\bar{Y}) = \mathrm{Var}(\varepsilon)/n$.

So averaging reduces random error but leaves bias untouched. This formalises the earlier intuition that noise can be averaged down, but systematic shifts require calibration or design changes.

Latent Variables and Indicators

For a simple latent-factor picture, suppose we observe indicators $Y_1, \ldots, Y_k$ that depend on one latent variable Z and noise:

$$Y_j = \lambda_j Z + \varepsilon_j, \quad j = 1, \ldots, k,$$

where λ_j are loadings and ε_j are error terms. Fitting such a model lets us infer Z up to scaling, but:

- the choice of indicators Y_j and model structure (how many factors, which loadings are allowed) shapes what "Z" becomes,
- different, equally good-fitting models can yield different interpretations of Z.

This is one reason to treat latent constructs as model-dependent tools rather than as directly observed quantities.

Chapter 6

Experiments, Observations, and Causality

Science often wants more than patterns; it wants to know what causes what. This chapter introduces the practical and conceptual distinctions between experiments and observational studies, and sketches core ideas about randomisation, confounding, and common identification strategies.

Learning Objectives

After working through this chapter you should be able to:

- distinguish experimental, quasi-experimental, and purely observational studies with concrete examples,
- explain why randomisation helps for causal inference and how confounding can mislead us,
- describe, at a high level, identification strategies such as regression with controls, instrumental variables, and difference-in-differences,
- articulate limits of causal discovery from data alone in complex domains like macroeconomics and finance,
- apply a short checklist when evaluating causal claims in research papers, reports, or products.

At a Glance

Experiments let us intervene and see what changes; observational studies watch what happens without direct control. Randomisation, control groups, and careful design make causal stories more credible, but no method is magic. Many real-world questions require a mix of domain insight, quasi-experimental strategies, and humility about what data can show.

Everyday Analogy

Trying a new diet yourself is like a (very small) experiment: you change something deliberately and monitor effects. Watching friends who choose different diets is more like an observational study: you see patterns but cannot separate food choices from lifestyle, genetics, and motivation. Much of scientific causality lives between these two poles.

6.1 Experimental vs. Observational Studies

We begin by clarifying three broad types of study: randomised experiments, quasi-experiments, and observational studies.

Randomised Controlled Trials

In a randomised controlled trial (RCT):

- participants or units are randomly assigned to treatment and control groups,
- the treatment is applied to the treatment group; the control group serves as a baseline,
- outcomes are compared across groups to estimate causal effects.

Examples include:

- Medicine: testing a new drug versus placebo.
- Online product development: A/B tests where users are randomly shown version A or B of a feature.
- Operations research: randomising which branches of a retail chain adopt a new process first.

Randomisation aims to balance both observed and unobserved characteristics across groups, so that outcome differences can more credibly be attributed to the treatment.

Quasi-Experiments and Natural Experiments

Sometimes we do not randomise, but external events or institutional rules create something close.

- Policy change: a tax reform affects only firms above a certain size threshold.
- Regulatory rule: a new disclosure requirement applies only to banks with assets above a cutoff.
- Timing differences: some regions roll out a programme earlier than others for logistical reasons.

These situations can mimic experiments if we can argue that, around the cutoff or roll-out timing, assignment is "as good as random" for the purposes of the question.

Purely Observational Studies

In purely observational studies, we do not control who receives which "treatment"; we simply observe choices and outcomes.

- Epidemiology: observing lifestyles and health outcomes in a cohort without assigning behaviours.
- Finance: observing which investors adopt a new trading strategy and how their portfolios fare.
- Macroeconomics: studying economic growth and policy variables across countries.

These designs are often the only ethical or practical option, but causal inference becomes much more delicate.

Common Pitfall

It is tempting to:

- label any study with a control group as "experimental," even when assignment was not random or nearly random,
- read causal language into observational correlations without checking for alternative explanations.

Whenever you see a causal claim, ask: who or what chose who received the "treatment," and could that choice be related to the outcome?

6.2 Randomisation, Control, and Confounding

Randomisation and control groups are tools for dealing with confounding—situations where other variables influence both treatment and outcome.

Why Randomisation Helps

Random assignment:

- breaks systematic links between potential confounders and treatment,
- ensures (in expectation) that treated and control groups are comparable on both observed and unobserved characteristics,
- lets simple comparisons of group averages estimate causal effects under minimal assumptions.

In a well-run A/B test, for example, differences in click-through rates or conversion rates can often be interpreted as causal effects of the design change, because user attributes were balanced by randomisation.

Confounding and Selection Bias

Without randomisation, confounders can easily distort apparent effects.

- Health example: people who choose to exercise more may also differ in diet, income, and genetics—all of which affect health outcomes.
- Finance example: firms that adopt a new risk system early may also be more sophisticated or better resourced in ways that affect performance.
- Tech example: users who opt into a new feature may be more engaged or tech-savvy than those who do not.

Selection bias occurs when the process that determines who is in the sample or who receives treatment is related to the outcome in ways not fully controlled for.

Analogy: Recipe Testing with Self-Selected Tasters

Imagine comparing two cake recipes by:

- letting guests choose which slice to try,
- then asking them to rate satisfaction.

If food enthusiasts disproportionately pick the more exotic-looking cake, differences in ratings may reflect your guests, not the recipes. Randomly assigning slices would do a better job of isolating the recipe effect.

From Lab to Life

Even in observational settings, you can:

- design data collection to reduce obvious confounding (for example, gather rich covariates for later adjustment),
- use stratification or matching to compare more similar units,
- pilot randomised interventions where ethically and logistically feasible.

The goal is always the same: to approximate the balanced comparisons that randomisation would have given you.

6.3 Identification Strategies in Practice

When randomised experiments are impossible or limited, researchers use identification strategies to approximate causal effects.

Regression with Controls

Regression analysis with control variables is the workhorse.

- We model an outcome Y (for example test scores, returns) as a function of a treatment D and covariates X: $Y = \alpha + \beta D + \gamma^\top X + \varepsilon$.
- The coefficient β is interpreted as a treatment effect, conditional on X, under assumptions about how confounding works.
- The key assumption is that, after conditioning on X, treatment assignment is as good as random with respect to the error ε.

This strategy can fail if important confounders are unobserved or poorly measured.

Instrumental Variables (IV)

Instrumental variables aim to separate variation in treatment that is "as if random" from variation driven by confounders.

- An instrument Z affects the treatment D but affects the outcome Y only through D, not directly or through omitted confounders.
- Classic examples include policy rules, eligibility thresholds, or distance-based instruments.
- Under suitable assumptions, one can estimate causal effects using only the component of D that is driven by Z.

Intuitively, the instrument provides a built-in mini-experiment inside the observational data.

Difference-in-Differences (DiD)

Difference-in-differences compares changes over time in a treated group to changes over time in a control group.

- We observe outcomes before and after a policy or treatment in both groups.
- If, absent treatment, the groups would have followed parallel trends, then the difference in their changes can be interpreted as a treatment effect.
- This strategy is widely used in policy evaluation, labour economics, and some finance contexts.

Analogy: Which Ingredient Matters?

Think of testing which ingredient in a recipe makes the real difference:

- Regression with controls: hold as many other ingredients as constant as possible across batches.
- Instrumental variables: use a constraint (for example, which ingredients the shop had in stock) that changes only one component across friends trying the recipe.
- Difference-in-differences: compare how two versions of the recipe evolve over time under similar kitchen conditions, attributing persistent differences to the changed ingredient.

Each approach tries to isolate one influence in the midst of many.

Common Pitfall

Identification strategies are powerful but fragile:

- forgetting that regression with controls still relies on untestable assumptions about unobserved confounders,
- using weak or invalid instruments that in fact correlate with omitted variables,
- assuming parallel trends in DiD without checking pre-treatment patterns.

In applied work, much of the real skill lies in diagnosing when these assumptions are or are not plausible.

6.4 Causal Discovery and Its Limits

Can we learn causal structure directly from data? The short answer is "sometimes, up to a point, under strong assumptions."

Data Alone Are Rarely Enough

In many settings, observational data alone cannot distinguish between alternative causal graphs that fit the same statistical patterns.

- Time series: volatility and volume in financial markets may move together; many different causal stories can generate similar correlations and lead-lag structures.
- Macroeconomics: GDP, interest rates, and inflation are entangled in ways that make clean causal identification from historical data extremely hard.

- Complex networks: in gene regulation or social networks, many feedback loops and hidden variables can mimic each other statistically.

Domain knowledge, experimental interventions, and robust design remain essential.

AI and ML Approaches to Causal Discovery

Recent work in AI and statistics has pursued algorithms for causal discovery.

- Constraint-based methods: infer possible causal graphs from patterns of (conditional) independencies.
- Score-based methods: search for graph structures that best balance fit and simplicity under a chosen score.
- Hybrid and deep-learning approaches: use neural nets with built-in constraints that encourage certain causal structures.

These tools can suggest hypotheses and narrow down possibilities, but they cannot, by themselves, eliminate the need for judgement and intervention-based tests.

From Lab to Life

In practice:

- treat algorithmic causal discovery as a hypothesis generator, not a final oracle,
- cross-check algorithmic suggestions against domain expertise and feasible interventions,
- be explicit about which assumptions (for example, no hidden confounders, certain time-ordering) the algorithm relies on.

Transparent communication about these assumptions is as important as the algorithms themselves.

6.5 Summary and Short Cases

Key points from this chapter:

- Experiments, quasi-experiments, and observational studies differ in how much control we have over treatment assignment.
- Randomisation and control groups help, but real-world constraints introduce confounding and selection issues.
- Identification strategies such as regression with controls, instrumental variables, and difference-in-differences approximate experimental logic under assumptions.
- Purely data-driven causal discovery has promise but also real limits.

Short cases to think through:

- Policy intervention: a city introduces congestion pricing; some neighbourhoods are more affected than others. How might you design a study to estimate causal effects on traffic and air quality?
- A/B test in a fintech app: you test a new interface for a trading feature. What would you randomise, and which outcomes and covariates would you track?

- Market event: a central bank surprises markets with an announcement. What can and cannot be said about causality from price and volume reactions alone?

Try in 60 Seconds

Tiny tasks:

- Classify a study you recently encountered (paper, blog, internal analysis) as experimental, quasi-experimental, or observational.
- Draw a three-node DAG for a simple situation in your domain and mark at least one possible confounder.
- Name one assumption behind an identification strategy you have seen (for example "no hidden confounders after X") and ask yourself how plausible it really is.

6.6 Technical Note: Simple Causal Diagrams

This technical note introduces minimal diagrammatic language for causal reasoning that we will build on later.

Directed Acyclic Graphs (DAGs)

A basic causal diagram is a directed acyclic graph (DAG):

- Nodes represent variables (for example treatment D, outcome Y, confounder C).
- Directed edges $C \to D$ or $D \to Y$ represent hypothesised direct causal influence.
- Absence of an edge encodes a claim of no direct causal effect (given the rest of the graph).

Confounding corresponds to a common cause C that points to both treatment and outcome: $C \to D$ and $C \to Y$.

Back-Door Paths and Adjustment

In this language:

- A *back-door path* between D and Y is any path that starts with an arrow into D (for example $D \leftarrow C \to Y$).
- If back-door paths exist, naive associations between D and Y may be confounded.
- Adjusting for appropriate variables (conditioning on C, for example) can block such paths under certain conditions.

The "back-door criterion" in causal graph theory formalises when adjustment sets suffice to identify causal effects from observational data. We do not need full formalism here; it is enough to see that diagrams help organise which variables we must measure and control.

Randomisation in DAG Terms

Randomising treatment D can be pictured as:

- breaking all arrows from pre-treatment variables into D,
- ensuring that there are no back-door paths from D to Y through earlier variables.

This diagrammatic picture matches the earlier intuition: randomisation severs hidden links between who gets treated and what would have happened anyway.

Figure 6.1 shows a simple confounding structure and how randomisation conceptually removes back-door paths.

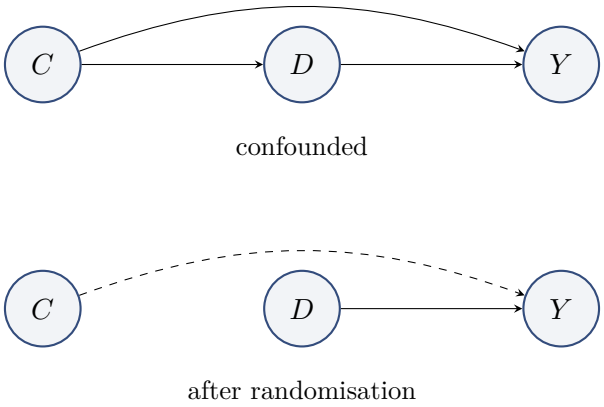

Figure 6.1: Simple causal diagrams: above, confounding with C affecting both treatment D and outcome Y; below, randomisation conceptually breaks the arrow $C \to D$, leaving only the direct effect $D \to Y$ and any residual influence of C on Y.

Chapter 7

Statistics, Models, and the Replication Crisis

Statistics is the nervous system of modern science: it carries signals, filters noise, and can misfire in subtle ways. This chapter looks at what statistics is for, how models can be overfitted or abused, and what the "replication crisis" has taught us about incentives and safeguards.

Learning Objectives

After working through this chapter you should be able to:

- describe statistics as uncertainty bookkeeping and signal-from-noise extraction,
- explain overfitting and p-hacking with simple examples from curve-fitting and trading strategies,
- summarise key features of the replication crisis in psychology, medicine, and economics, and their analogues in finance and AI,
- list concrete better practices such as pre-registration, open data, and robustness checks,
- use a checklist to critically read "flashy" results in papers, reports, or benchmark leaderboards.

At a Glance

Statistics is not a magic truth machine. It helps keep track of uncertainty and extract signals, but modelling choices, researcher freedom, and incentives can distort results. Understanding overfitting, multiple testing, and replication failures is essential for using statistical tools responsibly in physics, AI, and finance.

Everyday Analogy

Think of building a playlist. If you judge a song only by how well it fits the mood of the last hour of your life, you will overfit badly. If you test dozens of playlists but only show your friends the one that gets lucky applause, you are "p-hacking" social feedback. Statistics formalises these temptations—and offers tools to resist them.

7.1 The Role of Statistics in Science

Statistics helps scientists navigate uncertainty and noise.

Uncertainty Bookkeeping

Statistics is partly about keeping honest accounts of uncertainty.

- Estimation: quantifying how precisely we know a parameter (for example a mean, a slope, a treatment effect).
- Interval estimates: attaching intervals or regions (confidence intervals, credible intervals) rather than single best guesses.
- Error rates: specifying how often a procedure will mislead us under repeated use.

Used well, this bookkeeping prevents overconfident statements and highlights where data are thin.

Signal-from-Noise Extraction

Statistics is also about separating structure from randomness.

- Fitting models: finding relationships between variables (for example regression, time-series models, classification boundaries).
- Detecting patterns: identifying non-random structure in data while accounting for chance fluctuations.
- Quantifying evidence: assessing how strongly data support or conflict with a given model or hypothesis.

Intuitive and formal views of probability meet here: our informal sense of "rare vs frequent" is sharpened into explicit models and tests.

From Lab to Life

Across domains:

- In physics, statistics turns noisy detector counts into estimates of particle properties with quantified uncertainties.
- In AI, validation sets and cross-validation estimate how well models will generalise beyond their training data.
- In finance, risk metrics and stress tests summarise the distribution of possible portfolio outcomes.

In all cases, the value lies in honest uncertainty bookkeeping and careful pattern detection, not in forcing certainty from thin data.

7.2 Overfitting, p-Hacking, and Researcher Degrees of Freedom

Two central ways statistics can go wrong in practice are overfitting and p-hacking. Both exploit researcher degrees of freedom: the many choices you can make about data, models, and analyses.

Overfitting: Too Close to the Sample

Overfitting occurs when a model captures noise in the training data as if it were signal.

- Curve-fitting example: with enough polynomial terms, you can fit almost any wiggly line through a small dataset, but such a curve will typically extrapolate very poorly.
- Trading strategy example: by testing many parameter combinations on historical price data, you can usually find a rule that would have performed spectacularly in the past purely by chance.

- Machine learning example: a neural network with many parameters can achieve near-perfect accuracy on a training set while performing poorly on new data.

The hallmark of overfitting is a large gap between training performance and out-of-sample or test performance.

p-Hacking and Researcher Freedom

p-hacking refers to trying many analyses and selectively reporting only those that yield "significant" results.

- Choices include: which variables to include, which outcome to focus on, how to transform data, when to stop collecting more observations.
- If each decision is motivated by looking at the results, the nominal p-values no longer reflect true error rates.
- Even without bad intent, a researcher can unintentionally search until something "works," then forget the many paths not taken.

Researcher degrees of freedom are the accumulated choices that, if used opportunistically, inflate false positive rates.

Analogy: Tossing Coins Until You Like the Run

Imagine:

- tossing a fair coin many times,
- only showing your friends the one sequence where you happened to get 8 heads in a row,
- claiming this run as mysterious evidence of bias.

If you do not reveal how many sequences you tried before picking this one, your friends cannot properly judge how surprising the run really is. p-hacking is this logic in statistical clothing.

Common Pitfall

Signals of trouble:

- very flexible models justified only by training performance, with little attention to validation or out-of-sample testing,
- many exploratory analyses presented as if they were a single pre-planned test,
- complicated data pipelines with undocumented choices that are tuned to a desired conclusion.

None of these guarantees dishonesty, but all warrant extra scepticism and a search for robust checks.

7.3 Replication Crisis and Its Lessons

The "replication crisis" refers to large-scale efforts in several fields that failed to reproduce many published findings.

What Happened?

In psychology, medicine, and parts of economics, organised replication projects:

- selected published studies and repeated them with similar or improved designs,
- found that a substantial fraction yielded weaker or non-significant effects,
- highlighted that headline results are often less stable than initial reports suggest.

Similar worries arise in:

- quantitative finance, where promising backtested strategies sometimes fail in live trading,
- AI, where benchmark results can hinge on subtle choices in data processing, hyperparameters, or compute budgets.

Why Do Results Fail to Replicate?

Many factors interact:

- Low power: small sample sizes and noisy measurements make genuine effects hard to detect reliably.
- Publication bias: journals and conferences prefer positive, surprising results over null or mundane findings.
- Researcher degrees of freedom: flexible analysis pipelines increase false positive rates.
- Context sensitivity: effects may depend on cultural, institutional, or technological contexts that change over time.

From Lab to Life

Replication crises are not just embarrassments; they are learning opportunities:

- They push fields toward larger, better-powered studies and multi-site collaborations.
- They encourage sharing code and data so others can inspect pipelines.
- They motivate reforms in incentives, such as valuing robustness and transparency alongside novelty.

Similar reforms are underway in parts of finance and AI, where backtests and benchmarks are being scrutinised more critically.

7.4 Better Practices: Pre-Registration, Open Data, Robustness

Several concrete practices can mitigate overfitting and fragile findings.

Pre-Registration and Registered Reports

Pre-registration means specifying your research plan in advance.

- You record hypotheses, primary outcomes, and analysis strategies before seeing the final data.
- Deviations from the plan can still be explored but are clearly labelled as exploratory.
- Registered reports go further: journals review and conditionally accept a paper based on the proposed methods before results are known.

This separates confirmatory tests from exploratory fishing.

Open Data, Code, and Robustness Checks

Openness allows others to stress-test your findings.

- Sharing data (where ethical and legal) lets others reanalyse and combine results.
- Sharing code reveals analysis pipelines, making hidden researcher choices visible.
- Robustness checks and sensitivity analyses show how conclusions change when assumptions or specifications are varied.

Analogy: Building a Stable Chair

Think of a study as a chair:

- A single, unexamined specification is like a chair balanced on one thin leg.
- Robustness checks add extra legs and cross-braces: small wobbles do not topple the structure.
- Open materials are like inviting others to sit, wiggle, and inspect the joints.

You would not trust a chair that collapses the moment someone shifts their weight; treat empirical findings similarly.

7.5 Summary and Practical Checklist

Key points:

- Statistics helps with uncertainty bookkeeping and signal extraction, but modelling choices matter.
- Overfitting and p-hacking exploit flexibility to produce fragile or illusory results.
- The replication crisis revealed systemic issues in incentives and practices across multiple fields.
- Better practices include pre-registration, open materials, and systematic robustness checks.

When you encounter a flashy empirical result, ask:

- How big is the effect, and how precisely is it estimated?
- How complex is the model relative to the amount of data?
- Were analyses pre-specified, or does the story arise from extensive exploration?
- Are code, data, and robustness checks available for scrutiny?

Try in 60 Seconds

Quick sanity checks:

- Take a headline result (from a paper, report, or blog) and list two ways it could be sensitive to researcher degrees of freedom.
- Think of one model you use and say how you would detect if it is overfitting (which extra data or split would you use?).
- Write down one robustness check you could add to a current or past project.

7.6 Technical Note: Overfitting and Generalisation

We close with a minimal formal sketch of overfitting and generalisation.

Training and Test Error

Consider:

- a data-generating process producing input–output pairs (X, Y),
- a model f_θ with parameters θ,
- a loss function $L(f_\theta(X), Y)$ measuring prediction error.

We distinguish:

- Training error: the average loss on the data used to fit θ.
- Test (or generalisation) error: the expected loss on new data from the same process, not used during training.

Overfitting corresponds to a situation where training error is low but test error is substantially higher than necessary.

Model Complexity and Capacity

Schema:

- Increasing model flexibility (more parameters, richer function classes) typically reduces training error.
- Beyond a certain point, additional flexibility can increase test error because the model begins to fit noise.
- Conceptually, there is often a U-shaped relationship between model complexity and test error: underfitting at low complexity, overfitting at high complexity.

Formal learning theory (for example bounds involving VC dimension or Rademacher complexity) quantifies how model capacity, sample size, and generalisation error relate. For this book it suffices to remember that complexity must be matched to data and task.

High Dimensions and the Curse of Dimensionality

In high-dimensional spaces, two further constraints appear:

- Data become sparse: as the number of features grows, the volume of the space explodes, and any fixed sample occupies an ever-thinner slice. Intuitions from low dimensions ("nearby points are easy to find") break down.
- Computation becomes expensive: searching over many features, parameters, or model classes can require time that grows faster than linearly with dimension, even before training a single large model.

This "curse of dimensionality" helps explain why very flexible models are not automatically better, even with large datasets. It also connects statistical worries about overfitting with computational ones: in practice we must choose architectures and hypothesis classes that are not only expressive enough but also learnable in a reasonable amount of time. Later chapters on complexity (Chapter 9) and AI (Chapter 12) return to these limits from dynamical and engineering perspectives.

Figure 7.1 offers a classical schematic view: as complexity increases, training error falls monotonically while test error first falls then rises when overfitting dominates.[1]

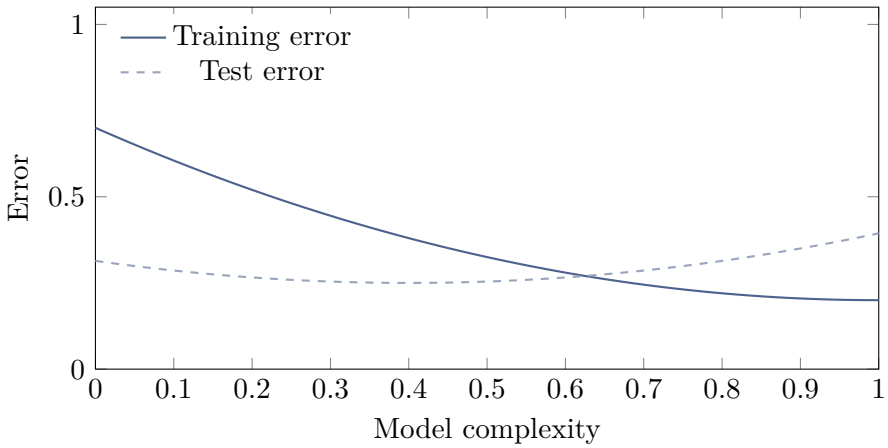

Figure 7.1: Schematic training and test error as a function of model complexity: training error decreases monotonically; test error achieves a minimum at intermediate complexity and rises again when the model overfits.

[1] Recent work on so-called "double-descent" behaviour shows that, in some heavily overparameterised regimes, test error can dip again after an initial increase. The qualitative message of the sketch still holds: pushing complexity far beyond what the data support risks unstable generalisation unless carefully controlled.

Chapter 8

Values, Ethics, and Risk in Scientific Practice

Science does not float above human concerns. Choices about what to study, how to study it, and how to use the results all reflect values. This chapter explores how values and ethics enter scientific practice, and how risk and uncertainty shape responsible decisions in domains like climate science, public health, AI, and finance.

Learning Objectives

After working through this chapter you should be able to:

- explain how values influence research questions, hypothesis framing, and interpretation,
- describe key ethical principles for experimentation and data use, including privacy and consent,
- distinguish risk from Knightian uncertainty and discuss different attitudes toward precaution and innovation,
- analyse how scientific expertise interacts with policy and public communication,
- reflect on concrete value-laden choices in your own domain.

At a Glance

Values are not optional extras in science; they shape questions, methods, and uses. Ethics constrains what we may do in experiments and with data. Risk and uncertainty force us to make decisions under incomplete knowledge. Recognising and articulating these elements is part of good scientific practice, not a departure from it.

Everyday Analogy

Think of planning a family road trip:

- Values show up in destination choices (quiet nature vs. busy city), budgets, and what counts as "fun."
- Ethics appears in how you drive, how tired you allow yourself to get, and how you treat other road users.
- Risk and uncertainty arise when you decide whether to drive through the night, how fast to go in bad weather, and how to react to conflicting GPS advice.

Scientific projects operate on similar axes, just with higher stakes and more complex institutions.

8.1 Are Values Inevitable in Science?

There is a long tradition of talking about science as "value-free." In practice, values enter at many stages.

Choosing Questions and Framing Hypotheses

Values guide:

- which phenomena are considered worthy of study (for example diseases affecting wealthy vs. poor populations),
- how questions are framed (for example focusing on individual behaviour vs. structural constraints),
- what counts as a "success" (for example GDP growth vs. distributional fairness vs. environmental impact).

Two different framing choices can lead to very different research programmes, even with the same data and methods.

Interpreting and Communicating Results

Values also appear when interpreting findings and deciding how to present them.

- Emphasis: highlighting certain outcomes (for example short-term gains) over others (for example long-term risks).
- Language: describing trade-offs as "costs" or "investments," "side effects" or "harms."
- Uncertainty: choosing whether to stress consensus, disagreement, or lack of knowledge.

Climate science and public health offer clear examples: how to communicate uncertain but potentially severe risks involves both scientific judgement and value choices about caution and equity.

From Lab to Life

You can ask, for any project:

- Who benefits directly from this work, and who bears the risks?
- Which alternative questions or designs were possible but not chosen?
- How would someone with different priorities describe the same findings?

These questions do not make science "subjective" in a trivial sense; they make explicit the value-laden decisions already present.

8.2 Ethics in Experimentation and Data Use

Ethical norms and regulations govern what we may do with human and animal subjects and with data about people.

Human and Animal Subjects

Core principles include:

- Respect for persons: informed consent, voluntariness, and the right to withdraw.
- Beneficence: minimising harm and maximising potential benefits.
- Justice: fair distribution of burdens and benefits across groups.

Institutional review boards (IRBs) or ethics committees review proposed studies to ensure these principles are honoured.

Data Privacy and Consent

In the age of digital traces, data ethics is central.

- Transaction data, clickstreams, and sensor logs can reveal sensitive patterns about individuals and groups.
- Consent obtained once for one purpose does not automatically justify all later uses.
- Anonymisation is difficult; re-identification can be surprisingly easy with enough auxiliary data.

Analogy: Borrowed Keys

Having keys to a friend's house:

- does not entitle you to enter whenever you like,
- does not let you invite strangers into their living room,
- comes with an implicit expectation of restraint and respect.

Access to detailed data about people's lives is similar: possession of data is not a blank cheque for all possible analyses.

Common Pitfall

Watch for:

- "We have the data, so we might as well use it" reasoning without revisiting consent and risk,
- assuming that technical anonymisation alone solves all ethical issues,
- treating ethics review as a box-ticking exercise instead of a real design constraint.

Ethical reflection should shape study design from the beginning, not patch it at the end.

8.3 Risk, Uncertainty, and Decision-Making

Scientific advice often guides decisions under uncertainty. It helps to separate risk from deeper forms of uncertainty.

Risk vs. Knightian Uncertainty

Following a classic distinction:

- Risk refers to situations where we can reasonably assign probabilities to outcomes (for example well-understood casino games, many insurance products, some financial instruments).
- Knightian uncertainty refers to situations where we cannot reliably specify probabilities (for example unprecedented technologies, long-term climate tipping points, novel financial products with limited history).

Different attitudes—from aggressive innovation to strong precaution—make sense in different parts of this spectrum.

Precautionary Principle vs. Innovation

Two poles:

- Precautionary emphasis: when potential harms are large, irreversible, or poorly understood, we should err on the side of caution even if probabilities are uncertain.
- Innovation emphasis: delaying beneficial technologies or policies also has costs; overcaution can entrench harms that better tools might reduce.

Seatbelts and vaccines are cases where initial hesitations gave way to strong adoption as evidence accumulated. New trading algorithms or AI systems may demand a more cautious approach when systemic or opaque risks are involved.

Analogy: New Routes in a Car

- Taking a slightly faster unfamiliar route on a sunny afternoon with time to spare: low stakes, innovation-friendly.
- Choosing an untested mountain shortcut at night in a storm with tired passengers: high stakes, precaution-friendly.

Decisions about new drugs, AI models, or financial products are often closer to the second scenario than the first.

8.4 Science, Policy, and Expertise

"Follow the science" is a common slogan, but its meaning is subtle.

What It Means to "Follow the Science"

Science can:

- describe what is happening (for example infection rates, emission trajectories),
- estimate likely consequences of different actions,
- clarify trade-offs and uncertainties.

Science cannot, on its own:

- decide which trade-offs are acceptable,
- set values such as how to weigh health vs. economic output vs. privacy.

Policy decisions inevitably mix scientific inputs with value judgements.

Expertise, Overconfidence, and Humility

Scientific advisory boards and expert panels sit at the science–policy interface.

- They must communicate clearly what is known, what is uncertain, and what assumptions underlie models.
- They must manage conflicts of interest and avoid overstating confidence for political or media simplicity.
- They should avoid both technocratic arrogance (pretending to dictate values) and false modesty (withholding clear warnings).

From Lab to Life

Good expert communication often:

- separates descriptive statements (what data show) from evaluative ones (what outcomes we prefer),
- uses scenarios rather than single-point forecasts,
- explains how new data might update recommendations.

This style helps non-experts see where science ends and value choices begin.

8.5 Summary and Reflection

Key points:

- Values influence scientific practice at many stages, from question choice to communication.
- Ethics in experimentation and data use constrains what we may do, not just what we can do.
- Distinguishing risk from deep uncertainty helps calibrate precaution vs. innovation.
- Scientific expertise informs policy but does not replace value-based decision-making.

Thought experiments:

- You design an AI system for credit scoring. How do you balance predictive performance against fairness, transparency, and privacy?
- You evaluate a proposed geoengineering experiment with uncertain side effects. Which values and risks would you weigh, and how would you communicate your stance?
- You advise a regulator about a new high-frequency trading technology. What ethical and systemic-risk questions would you raise beyond raw profitability?

Try in 60 Seconds

Short reflections:

- Identify one value-laden choice in a current project (for example choice of metric, target population, or outcome to emphasise).
- Ask whether the main uncertainty you face in that project feels more like quantifiable risk or deeper uncertainty.
- Check one dataset you use and note what, if anything, you know about consent and intended use.

8.6 Technical Note: Simple Risk Calculus and Uncertainty

We finish with a minimal vocabulary for risk and decision under uncertainty.

Expected Value Under Risk

When probabilities are reasonably specified:

- The expected value of an outcome X is $\mathbb{E}[X] = \sum_i p_i x_i$ in the discrete case, where x_i are possible values and p_i their probabilities.
- Decisions can be compared by their expected values or by expected utility if we account for risk preferences.

In finance, expected return and risk measures (such as variance or Value-at-Risk) summarise aspects of this distribution.

Risk Aversion and Utility

Risk-averse decision-makers often prefer:

- a sure outcome to a risky one with the same expected value,
- distributions with lower variance when downside outcomes are severe.

Formally, a concave utility function $u(X)$ captures risk aversion through $\mathbb{E}[u(X)]$ rather than $\mathbb{E}[X]$ alone. We will not lean heavily on this machinery, but it underlies many models in economics and finance.

To make this concrete, Figure 8.1 compares a linear "risk-neutral" utility curve with a concave risk-averse curve as functions of wealth, showing how the same spread of outcomes can have the same expected value but different expected utilities.

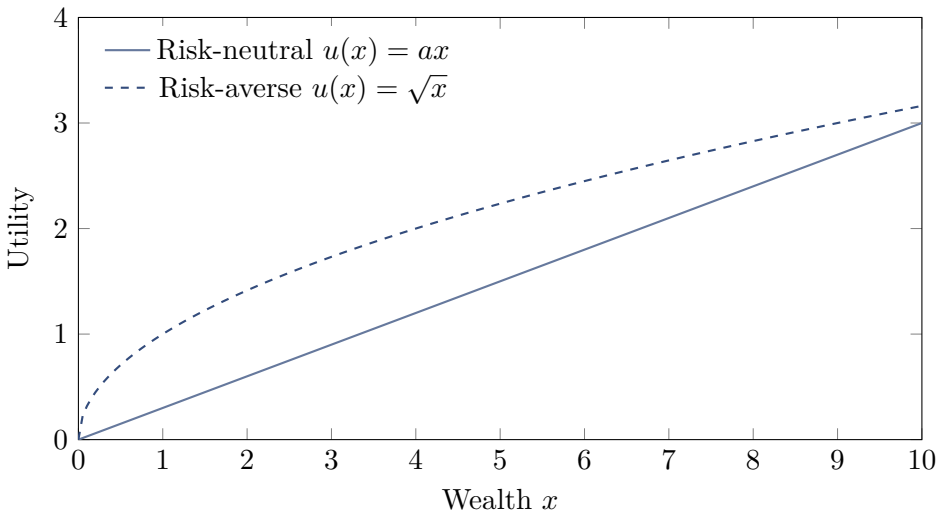

Figure 8.1: Illustration of risk-neutral (approximately linear) versus risk-averse (concave) utility as functions of wealth. The same spread of outcomes can have the same expected value but lower expected utility for a risk-averse decision-maker.

Knightian Uncertainty and Robustness

Under deep uncertainty:

- probability estimates are themselves fragile or contested,
- scenario analysis and robustness considerations become central,
- decision-makers may adopt maximin, minimax-regret, or safety-margin strategies.

In such cases, philosophy of science meets decision theory: we must decide how to act when our best models are known to be incomplete or unstable. Later chapters on complex systems and case studies will return to this theme.

Chapter 9

Complexity, Interdisciplinarity, and the Limits of Models

Many of the most important systems we care about—ecosystems, economies, markets, climate, social networks—are complex. This chapter sketches what "complex" means in a scientific sense, how disciplines talk to each other about such systems, and why we often need plural, modest modelling strategies.

Learning Objectives

After working through this chapter you should be able to:

- distinguish simple from complex systems with reference to linearity, feedback, and emergence,
- describe examples of complex behaviour in traffic, finance, and ecology,
- explain why interdisciplinary science is often necessary and what makes it hard,
- discuss limits of prediction and control in chaotic or regime-shifting systems,
- articulate the case for model pluralism and modesty in modelling practice.

At a Glance

Complex systems are those where many interacting parts, feedback loops, and nonlinearities make behaviour hard to predict from simple sums of components. In such systems, small changes can have large effects, and different models may capture different aspects of reality. This chapter argues for interdisciplinary dialogue and for treating models as tools in a portfolio rather than as single, final pictures.

Everyday Analogy

Think of city traffic at rush hour. No single driver's plan explains the jam. Small events (a minor lane change, a brief hesitation) can ripple backward and create large waves of slowdowns. Understanding and managing such a system requires more than just knowing the rules of the road; you need ideas from physics, psychology, engineering, and even economics.

9.1 Simple Systems vs Complex Systems

We start with a contrast between simple and complex systems.

Simple, Mostly Linear Systems

Simple systems:

- often have a small number of important variables,

- behave approximately linearly in the range of interest (inputs and outputs scale proportionally),
- can be decomposed into parts whose effects add up straightforwardly.

Examples include:

- a mass on a spring under small oscillations,
- a low-dimensional electrical circuit with linear components,
- a simple queue with roughly constant arrival and service rates.

These systems often admit clean analytic solutions and reliable control strategies.

Complex, Nonlinear, Feedback-Rich Systems

Complex systems typically:

- involve many interacting components (agents, species, devices),
- contain nonlinear interactions (thresholds, saturations, multiplicative effects),
- exhibit feedback loops (reinforcing or balancing),
- show emergent behaviour—patterns at the system level that are not obvious from individual rules alone.

Examples include:

- Traffic networks where local driving decisions and infrastructure interact.
- Financial markets where heterogeneous traders, institutions, and regulations coevolve.
- Ecosystems where species compete, cooperate, and modify their environment.

Analogy: One Guitar vs. a Jam Session

A single, well-tuned guitar playing a simple melody is like a linear system: you can often predict the next note from the score alone. A live jam session with several musicians improvising is more like a complex system:

- each musician responds to the others,
- small variations can change the whole vibe,
- the "song" is an emergent property of the interaction, not just a fixed script.

Common Pitfall

Two tempting oversimplifications:

- treating genuinely complex, feedback-rich systems as if they were simple and linear, then being surprised by sudden failures,
- labelling any messy situation "complex" without specifying which features (nonlinearity, feedback, heterogeneity) actually matter.

Using the word "complex" should invite sharper questions, not end the conversation.

9.2 Interdisciplinary Science

Complex systems rarely respect disciplinary boundaries. Interdisciplinary work brings tools and metaphors across fields.

Cross-Fertilisation Examples

Some classic flows of ideas:

- Physics to finance: stochastic processes, diffusion models, and option-pricing techniques.
- Computer science to biology: algorithms for sequence alignment, network analysis, and simulation.
- Statistics and AI to many domains: hierarchical models, reinforcement learning, causal inference.

These transfers are rarely one-way; receiving fields push back with new constraints and questions.

Challenges of Interdisciplinary Work

Interdisciplinary projects must manage:

- different vocabularies and notations for similar ideas,
- different standards of evidence and acceptable simplification,
- differing incentives (what counts as a contribution) in each field.

Analogy: Multilingual Conversation

Imagine a group conversation in which each person speaks a different language but shares some phrases:

- words that sound similar may have different meanings,
- some concepts exist in one language but have no direct translation in another,
- good communication requires patience, paraphrasing, and verification.

Interdisciplinary work feels like this: misunderstandings are easy, but new ideas emerge precisely because not everyone thinks in the same way.

From Lab to Life

In your own work:

- notice when you import a method or metaphor from another field (for example physics-style modelling in finance, or ML metrics in healthcare),
- ask which assumptions travel safely and which need adaptation,
- consider co-authoring or consulting across disciplines when stakes or complexity are high.

Interdisciplinarity is not just about sharing tools; it is about negotiating standards and expectations.

9.3 Limits of Prediction and Control

Even with powerful models, complex systems impose fundamental limits on what we can predict and control.

Chaos and Sensitivity to Initial Conditions

Chaotic systems are deterministic but highly sensitive to initial conditions.

- Tiny differences in starting states can grow exponentially, making long-term prediction practically impossible.
- Classic examples include certain weather models and simple nonlinear maps.
- Short-term forecasts may still be accurate, but horizons for reliable prediction are limited.

This sensitivity is not mere noise; it is a structural feature of the dynamics.

Structural Breaks and Regime Changes

In many real-world systems, the rules themselves change.

- Markets can shift regimes when regulation, technology, or dominant strategies change.
- Climate systems may cross tipping points that alter feedback structures.
- Social systems can reconfigure after policy shifts or cultural changes.

Models estimated in one regime may perform poorly in another; past data do not always speak clearly about unprecedented futures.

Simulation vs. Understanding

Computers let us simulate highly detailed models of complex systems.

- Simulations can reproduce realistic behaviours and explore "what if" scenarios.
- However, reproducing behaviour does not guarantee deep understanding of the mechanisms or tipping points involved.
- We can easily be dazzled by visual similarity while missing structural mismatches.

Analogy: Flight Simulators

Flight simulators can mimic the experience of flying, including turbulence and emergencies. They are invaluable for training. Yet:

- they may not capture every rare combination of failures,
- they rely on models of aircraft and weather that can be wrong in edge cases.

Similarly, complex-system simulations are powerful training grounds for intuition, but they do not replace empirical checks and theoretical analysis.

Common Pitfall

Beware:

- treating model output as data rather than as conditional "what if" answers,
- assuming that because a model has many parameters and realistic visuals, it must be accurate,
- extrapolating far beyond the validated range of a complex model.

The more complex the model, the more discipline is needed in how we interpret its outputs.

Computational Limits and Intractability

There is another layer of limitation that comes not from physics or social structure but from computation itself.

- Some problems are *intractable* in the sense of computational complexity theory: even with perfect models and data, any exact algorithm would require time that grows explosively with system size.
- The famous distinction between problems in class P (solvable in roughly polynomial time) and hard problems like those in NP is a way of drawing a theoretical line between what is tractable in principle and what is likely out of reach for large instances.
- Many realistically modelled tasks—optimal routing on large networks, combinatorial portfolio selection, certain scheduling and design problems—sit on the hard side of this line.

Practically, this means that for complex systems we often settle for approximations, heuristics, or simulations that run within available time and energy budgets. Even if the world obeys simple equations at a micro level, our limited computational resources ensure that *effective* theories and coarse-grained models will remain central. Later chapters on AI and finance make this constraint concrete when discussing large-scale learning systems and high-frequency markets.

9.4 Modesty and Pluralism in Modelling

Given these limits, a modest, pluralist attitude toward models is often wise.

Model Pluralism

Model pluralism is the practice of using several models for different questions or as a robustness check.

- Different models may emphasise different mechanisms or scales.
- Disagreements between models can illuminate where knowledge is thin.
- Agreement across diverse models can build confidence.

In finance, for example, practitioners often combine multiple risk models and scenario analyses rather than relying on a single number.

"All Models Are Wrong, Some Are Useful"

George Box's aphorism is often quoted; unpacking it helps.

- All models simplify and idealise; in that sense they are "wrong" as literal descriptions of reality.
- "Useful" means fit for a purpose: clarifying mechanisms, making short-term predictions, exploring policy trade-offs.
- Utility is context-dependent: a model that is helpful for teaching may be too crude for engineering design.

Analogy: Portfolio of Tools

Think of a toolkit:

- You would not expect a single tool (say, a hammer) to solve all problems.
- You choose between tools based on the task (screwdriver, wrench, saw).
- You may use several tools in sequence for one project.

Model pluralism treats models like tools: you build and maintain a portfolio rather than searching for one universal instrument.

9.5 Summary and Transition to Case Studies

Key points from this chapter:

- Complex systems feature many interacting parts, feedback, nonlinearity, and emergence.
- Interdisciplinary work is often necessary to study such systems and comes with both opportunities and challenges.
- Prediction and control have intrinsic limits in chaotic, regime-shifting, or poorly understood systems.
- Model pluralism and modesty are strategies for navigating these limits responsibly.

Looking ahead:

- Part III applies these ideas in specific case domains: mathematics, physics, AI, and finance.
- As you read the case studies, keep asking: what makes each domain simple or complex, and which modelling attitudes are appropriate where?

Try in 60 Seconds

Small diagnostics:

- List three systems you know well and classify each as mostly simple or mostly complex in the sense of this chapter.
- For one complex system, name a feedback loop that could amplify or dampen changes.
- Write down two different models you might reasonably use for the same complex system and what each is best at capturing.

9.6 Technical Note: A Toy Dynamical System

To anchor the idea of complexity and sensitivity, consider a simple discrete-time dynamical system.

Linear vs. Nonlinear Updates

Compare:

- Linear system: $x_{t+1} = ax_t$. Behaviour is easy to classify: solutions grow, decay, or stay constant depending on $|a|$.
- Nonlinear logistic map: $x_{t+1} = rx_t(1 - x_t)$ for $0 < x_t < 1$, $r > 0$. For certain values of r, small differences in x_0 can lead to very different long-term behaviour.

The logistic map, despite its simple formula, can produce fixed points, cycles, or chaotic motion depending on r. This illustrates how nonlinearity and feedback can generate rich behaviour.

Sensitivity to Initial Conditions

In chaotic regimes of the logistic map:

- two trajectories starting at x_0 and $x_0 + \delta$ with tiny δ can diverge rapidly,
- long-term prediction becomes practically impossible beyond a limited horizon,
- qualitative structures (for example invariant sets) may still be studied mathematically.

You do not need full chaos theory for this book; the key message is that simple equations can generate complex behaviour and that in such systems we may need to shift from precise long-term prediction to probabilistic or scenario-based reasoning.

Figure 9.1 illustrates how linear and logistic updates differ for nearby initial conditions.

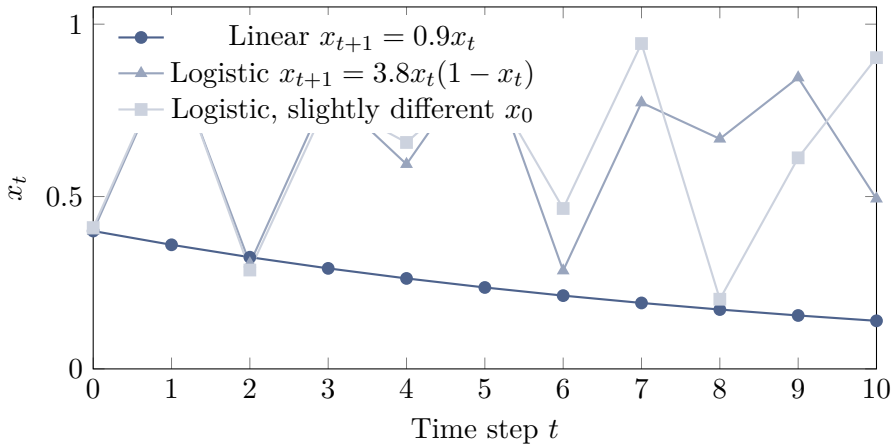

Figure 9.1: Example trajectories: a linear decay (solid, circles) and two logistic-map trajectories with nearby initial conditions (triangles and squares). The linear system forgets its starting point smoothly; the nonlinear system shows sensitive dependence on initial conditions.

Part III

Case Studies

Part III Overview

This part applies the earlier tools to concrete domains and, in several places, turns the lens back on systems that respond to being modelled. Chapter 10 examines mathematics itself as a case study: how proof, axioms, and structures relate to truth and applicability, and how abstract mathematics interacts with physics and finance. Chapter 11 then looks at physics as the archetypal hard science, highlighting laws, symmetry, and debates over determinism and chance. Chapter 12 treats artificial intelligence as a laboratory for large-scale induction, black-box modelling, and questions of understanding, alignment, and computational limits. Chapter 13 uses quantitative finance to explore uncertainty, reflexivity, and model risk in human-designed, model-aware systems, while Chapter 14 closes by synthesising themes across domains.

Chapter 10

Mathematics: Proof, Truth, and Applicability

Mathematics feels both central to science and strangely different from it. This chapter explores what kind of knowledge mathematics offers, how proof differs from empirical evidence, and why abstract mathematics turns out to be so effective in domains like physics and finance.

Learning Objectives

After working through this chapter you should be able to:

- explain in what sense mathematics is and is not a "science",
- distinguish mathematical proof from empirical evidence and from legal or everyday "proof",
- describe how axioms, structures, and models fit together in modern views of mathematics,
- summarise key ideas about why mathematics applies so well to physics and finance,
- reflect on whether mathematics is better thought of as discovered or invented, and what that means for its role in science.

At a Glance

Mathematics offers a different kind of certainty than empirical sciences: once a result is rigorously proved from given axioms, it stays proved. Yet the choice of axioms and structures is guided by curiosity, aesthetics, and applications. The "unreasonable effectiveness" of mathematics in physics and finance is not a miracle; it reflects a long co-evolution between abstract theory and the kinds of systems we choose to model.

Everyday Analogy

Think of designing board games:

- You set up a small world with rules (axioms) and pieces (objects).
- Within that world, some outcomes follow inevitably from the rules—you can analyse them without playing every game.
- Players then use the game to explore strategies and patterns.

Mathematicians design and explore such rule-based worlds at a much higher level of abstraction. Scientists borrow these worlds when they fit real phenomena well.

10.1 Is Mathematics a Science?

Opinions differ on how to classify mathematics.

A Priori vs. Empirical

In contrast to empirical sciences:

- Mathematics deals primarily with consequences of definitions, axioms, and formal rules.
- Its theorems are justified by proof, not by experiments or observations.
- Once a proof is correct, new data cannot overturn it (though they can inspire revision of axioms or definitions).

In this sense, mathematics is often seen as an *a priori* discipline: it does not depend on empirical measurement for its core claims.

Mathematics as the Language of Science

At the same time:

- Physical theories are written in mathematical language: differential equations, vector spaces, manifolds.
- Finance models use stochastic calculus, linear algebra, and optimisation.
- Many scientific advances are inseparable from advances in the mathematics used to express them.

So while mathematics has its own internal life, it also functions as the shared language and toolkit for many sciences.

From Lab to Life

For practitioners:

- mathematics provides symbolic compression: a short equation can summarise many empirical relationships,
- proofs give confidence that certain patterns will hold wherever the assumptions are met,
- new mathematical structures can suggest new hypotheses about what to look for in data.

Thinking of mathematics only as "calculations" misses its role in shaping the questions we ask.

10.2 Proof and Certainty

Proof is central to mathematics, but the word "proof" means different things in different contexts.

Mathematical vs. Legal or Everyday Proof

In everyday and legal contexts:

- "proof" is often a matter of high probability based on evidence and argument,
- new evidence can overturn a verdict, even if it was previously "proved" in court.

In mathematics:

- a proof is a chain of reasoning from axioms and previously proved results that can, in principle, be checked line by line,
- once accepted as correct, the theorem is considered established within the chosen formal system,
- errors sometimes slip into published proofs, but they are corrected by the community; the standard remains deductive certainty.

Informal and Formal Proofs

Working mathematicians:

- usually communicate in compressed, informal proofs that rely on shared background and omitted routine steps,
- sometimes use computers to verify large or intricate arguments,
- view fully formalised proofs (down to the level of logic symbols) as possible in principle, but often unnecessary in practice.

This mirrors science more broadly: we rely on trusted chains of reasoning and communities of checking, not on constant re-derivation from first principles.

Analogy: Legal vs. Mathematical Proof

In a trial:

- the standard might be "beyond reasonable doubt", not absolute certainty,
- evidence is weighed by human judgement under time and resource constraints.

In mathematics:

- the standard is logical necessity, given the axioms,
- the main resource constraints are human attention and clarity, not new data.

Both notions of proof are rigorous in their own settings but aim at different kinds of assurance.

Common Pitfall

Two opposite mistakes:

- assuming that because mathematical results are certain, any mathematically expressed scientific model is equally certain about nature,
- dismissing mathematical proofs as irrelevant to practice, ignoring their role in ruling out entire classes of failure modes.

The right balance recognises that proofs certify consequences *if* assumptions hold; empirical work must still test those assumptions.

10.3 Structures, Axioms, and Models

Modern perspectives often describe mathematics as the study of structures.

Axiomatic Systems

An axiomatic system consists of:

- a language (symbols and formation rules),
- axioms: basic statements taken as starting points,
- inference rules: how to derive new statements from old ones.

Examples include:

- Euclidean vs. non-Euclidean geometries (changing the parallel postulate),
- various set theories (with or without certain large-cardinal axioms),
- algebraic structures such as groups, rings, and fields.

Models of Axioms

In logic, a *model* of an axiomatic theory is a structure in which all the axioms are true.

- A Euclidean plane can be modelled by ordinary $\mathbb{R}^2$ with the standard distance.
- Hyperbolic geometry can be modelled in different ways (for example the Poincaré disk model).
- Probability theory can be modelled as measures on sigma-algebras over sample spaces.

Mathematicians often study properties that hold across all models of a given theory, or across large classes of structures.

Discovered or Invented?

Views differ:

- Platonist flavour: mathematical objects and truths exist independently; mathematicians discover them.
- Formalist flavour: mathematics is about manipulating symbols according to rules; structures are invented and explored.
- Structuralist flavour: focus on patterns of relations rather than on objects themselves.

You need not pick a camp to use mathematics well, but awareness of these views clarifies debates about truth and existence in math.

10.4 Why Does Mathematics Work So Well in Physics and Finance?

Eugene Wigner called it "the unreasonable effectiveness of mathematics"; here we sketch some partial explanations.

Selection Effects and Model Choice

One line of thought:

- We notice and celebrate domains where mathematics works well and quietly ignore those where it fails.
- We choose to model systems that already exhibit regularity and symmetry amenable to mathematical description.
- When a model fits poorly, we change the model or the question, not the underlying mathematics.

In this view, the match is less miraculous and more the result of co-selection.

Co-Evolution of Math and the World We Model

Another angle emphasises co-evolution:

- Scientists and engineers develop mathematical tools precisely in response to features of physical and economic systems (calculus for mechanics, stochastic calculus for finance).
- Successful models feed back into practice, changing the systems we build (for example engineered control systems, algorithmic trading).
- Over time, we inhabit a world increasingly shaped by devices and institutions designed with mathematical tools.

From this perspective, mathematics is effective partly because we have engineered parts of the world to behave in mathematically tractable ways.

Examples Across Domains

- Mechanics: differential equations and energy methods describe planetary motion and bridges.
- Finance: martingale measures and partial differential equations underpin option pricing.
- Data science: linear algebra and optimisation shape everything from recommendation systems to risk models.

In each case, the mathematical formalism and the domain practice grew together.

10.5 Case Vignettes

A few short vignettes illustrate interactions between mathematics and science.

Symmetry in Math and Physics

Symmetry groups:

- organise patterns in geometry, algebra, and combinatorics,
- underlie conservation laws in physics via Noether-type connections,
- appear in finance when pricing problems reduce under symmetry or invariance assumptions.

Mathematical classification of symmetries has enabled predictions about particles and interactions before direct observation.

Probability as a Bridge

Probability theory links:

- mathematical structures (measure spaces, random variables),
- empirical frequencies and uncertainty in science,
- risk and pricing in finance.

The same binomial and normal distributions appear in fields as diverse as genetics, queuing theory, and option pricing, with different interpretations.

A Simple Binomial Model

As a toy example, consider a one-period binomial model for an asset:

- Current price S_0.
- Next period, the price is either $S_0 u$ (up) or $S_0 d$ (down), with $u > d > 0$.
- A risk-free rate r allows borrowing or lending at factor $1 + r$.

Under mild conditions, there exists a risk-neutral probability q such that the arbitrage-free price of a derivative with payoffs V_u and V_d is the discounted expected value under q. The mathematics of this model is exact; whether the assumptions match a real market is an empirical question.

10.6 Summary and Questions

Key points:

- Mathematics differs from empirical sciences in its methods of justification, but it is deeply entwined with them.
- Proof provides conditional certainty: if the axioms and definitions hold, then the theorem follows.
- Axioms, structures, and models offer multiple lenses on mathematical practice and its connection to reality.
- The effectiveness of mathematics in science reflects selection, co-evolution, and careful choice of what to model.

Reflection questions:

- In your own work, when do you rely on mathematical proof, and when on empirical evidence? How do you combine them?
- Can you think of a time when a mathematical model shaped the way you or your field saw a problem, for better or worse?
- Do you lean more toward thinking of mathematics as discovered or invented, and does that affect how you trust it in applications?

Try in 60 Seconds

Quick prompts:

- Pick one result you use often and state, in one sentence, which assumptions it depends on.
- Note one place in your work where you could, in principle, rely on a proof but currently rely on empirical checking (or vice versa).
- Identify a mathematical structure (for example vector space, graph, Markov chain) that appears in your domain and name two different roles it plays.

10.7 Technical Note: Axioms, Models, and a Binomial Market

We finish with a compact, slightly more formal summary connecting axioms, models, and a simple finance example.

Abstract Structures and Models

In logic:

- A theory T is a set of sentences in a formal language.
- A model $\mathcal{M}$ of T is a structure in which all sentences of T are true.
- Mathematical "truth" in this sense is always truth *in* a model relative to axioms.

In applications:

- We map elements of a mathematical model to features of the world (for example states to physical configurations, paths to price trajectories).
- Whether the mapping is adequate is an empirical and pragmatic question, not a purely mathematical one.

Binomial Asset-Pricing Model

In the one-period binomial setting:

- Assume no-arbitrage and the ability to borrow and lend at rate r.
- Under these axioms, one can derive a unique risk-neutral probability

$$q = \frac{(1+r) - d}{u - d},$$

provided $d < 1 + r < u$.

- The arbitrage-free price of a derivative with payoffs V_u (up state) and V_d (down state) is

$$V_0 = \frac{1}{1+r}\big(qV_u + (1-q)V_d\big).$$

Within the mathematical model, these results are theorems. In reality:

- the no-arbitrage assumption may hold only approximately,
- markets have more than two states and frictions,
- calibration and stress testing are needed to see how robust conclusions are,
- widespread use of risk-neutral valuation itself can help coordinate prices and expectations, turning a convenient mathematical device into part of the market's microstructure (a theme revisited in Chapter 13).

This toy case illustrates the general pattern: clear axioms → precise mathematical consequences → empirical work to test whether the axioms are good enough for the task.

Figure 10.1 shows the simple one-period binomial tree underlying this model.

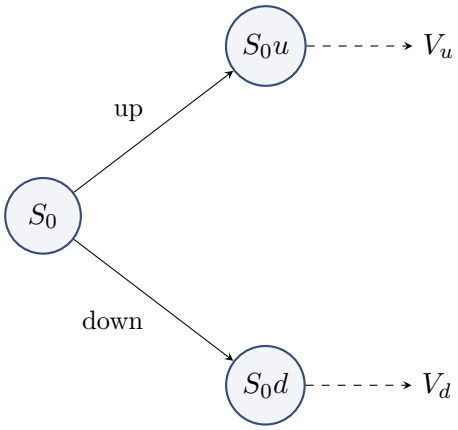

Figure 10.1: One-period binomial asset-pricing model: from initial price S_0 to either S_0u (up) or S_0d (down), with derivative payoffs V_u and V_d at the end of the period.

Chapter 11

Physics: Laws, Symmetry, and Reality

Physics has often been the poster child for "real" science: precise laws, accurate predictions, and deep connections between mathematics and the world. This chapter uses physics as a case study for how laws, symmetry, determinism, and chance shape our image of scientific theories and reality.

Learning Objectives

After working through this chapter you should be able to:

- describe how physics came to be seen as the archetypal hard science and why that image is both helpful and misleading,
- sketch the shift from Newtonian mechanics to relativity and what it implies for theory change,
- summarise key features of quantum theory and the measurement problem at a conceptual level,
- explain the role of symmetry and conservation laws in modern physics and why they matter philosophically,
- distinguish different senses of determinism and chance and relate them to prediction and risk.

At a Glance

Physics offers some of our clearest examples of laws, explanation, and theory change. Newtonian mechanics once seemed to describe the world completely; relativity and quantum theory showed both its power and its limits. Symmetry principles and conservation laws reveal deep structure, while debates about determinism and chance show how even precise theories leave room for philosophical disagreement.

Everyday Analogy

Think of the human body:

- the skeleton provides structure and constraints—analogous to physical laws,
- organs and tissues add function and complexity—like the variety of specific systems studied in biology or economics,
- behaviour emerges from interactions among all parts.

Physics often plays the role of the "backbone" in our picture of science: it sets constraints within which other fields operate, but it does not exhaust all the richness of the scientific enterprise.

11.1 Physics as the Archetypal Hard Science

Popular images of physics emphasise precision and universality.

Why Physics Looks Special

Several features make physics a natural anchor for philosophy of science:

- High-precision tests: from planetary orbits to particle scattering, physics often achieves extremely accurate quantitative predictions.
- Clear mathematical structure: theories like classical mechanics, electromagnetism, and quantum field theory have elegant, well-developed formalisms.
- Stable cores: many physical laws have remained robust across vast domains, even as theories expanded or were reinterpreted.

As a result, many early philosophical accounts of science implicitly took physics as the model and tried to fit other sciences into that template.

Why the Image Can Mislead

However:

- Not all areas of physics resemble idealised textbook cases; frontiers like turbulence or complex materials are messy and model-rich.
- Other sciences (biology, psychology, social sciences, AI) face different kinds of constraints and uncertainties that do not map neatly onto the physics template.
- Over-generalising from physics can obscure the diversity of legitimate scientific methods.

Common Pitfall

It is tempting to:

- treat fields as "less scientific" whenever they depart from the physics pattern of simple laws with clean predictions,
- ignore the role of idealisation and modelling choices even in physics itself.

Using physics as a reference point is helpful; using it as a rigid standard for all science is not.

11.2 From Newton to Einstein (Very Compressed)

The transition from Newtonian mechanics to relativity is a classic example of theory change.

Newtonian Mechanics in Brief

Newtonian mechanics:

- models bodies as moving in an absolute space and time,
- uses forces and masses to relate acceleration to applied forces via $F = ma$,
- treats time as a universal parameter that flows identically for all observers.

Within its domain (moderate speeds, weak gravitational fields), Newtonian mechanics is extraordinarily successful.

Relativity and Conceptual Shift

Relativity introduces several conceptual surprises:

- Special relativity: the speed of light is constant for all inertial observers; simultaneity becomes relative; time and space merge into spacetime.
- General relativity: gravity is not a force in space but curvature of spacetime itself; mass–energy tells spacetime how to curve, spacetime curvature tells matter how to move.

For intuition, Figure 11.1 sketches a spacetime diagram with light cones and different slices of "now" for different observers.

From a philosophical angle:

- the meaning of "law" changes (from forces in a fixed background to geometric relations in a dynamic spacetime),
- what counts as a "framework" versus a "model" shifts accordingly,
- older theories remain as approximations within a broader picture.

From Lab to Life

The Newton-to-Einstein story shows that:

- theories can be deeply successful yet still incomplete,
- new frameworks may re-interpret old entities (for example gravitational "force" vs. spacetime curvature),
- approximations (like Newtonian mechanics) remain valuable even after more general theories appear.

This pattern recurs in other fields: new paradigms reinterpret, rather than simply discard, earlier models.

11.3 Quantum Theory and the Measurement Problem

Quantum theory adds another layer of conceptual challenge.

Key Features of Quantum Theory

At a minimalist level:

- States are represented by vectors (or wavefunctions) in a Hilbert space.
- Observables correspond to operators; measurement outcomes are eigenvalues with probabilities given by the Born rule.
- Evolution between measurements is given by a deterministic equation (for example Schrödinger's equation).

So we have a mix of deterministic evolution (between measurements) and probabilistic outcomes (at measurement).

The Measurement Problem, Sketch

The measurement problem arises because:

- the theory seems to treat measuring devices and observers as quantum systems like any other,
- yet in practice we talk as if measurements have definite outcomes ("the pointer points here"),
- standard formulations often invoke a special "collapse" process that is not described by the same evolution equations.

Interpretations vary:

- Copenhagen-flavoured views emphasise classical descriptions of measurement devices and accept collapse as an effective rule.
- Many-worlds-flavoured views treat all evolution as unitary and interpret probabilities as branching weights in a multiverse picture.
- Other approaches (pilot waves, objective collapse models) add additional structure.

Analogy: Fuzzy vs. Crisp Photos

Imagine:

- a camera that stores many slightly different exposures of the same scene (a "fuzzy" superposition),
- a viewing process that always shows you just one sharp frame (a "collapse").

Quantum theory, in many standard presentations, feels like this: the underlying description is spread out, but our experience is of crisp, definite outcomes. The measurement problem is about making sense of that bridge.

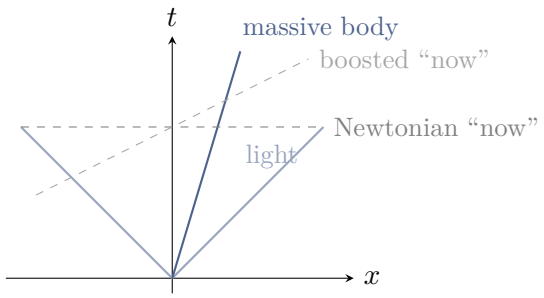

Figure 11.1: Schematic spacetime diagram: time t runs vertically and space x horizontally. Light rays (thin diagonal lines) define a light cone; worldlines of slower bodies must stay within it. Horizontal and slanted dashed lines indicate different “surfaces of simultaneity” for different observers in relativity.

11.4 Symmetry, Conservation, and Unification

Much of modern physics is organised around symmetry principles.

Symmetries and Conservation Laws

Roughly:

- A symmetry of a system is a transformation (for example a rotation, translation, or gauge transformation) that leaves key equations or structures invariant.
- Noether-type results connect continuous symmetries with conservation laws (for example time-translation symmetry with energy conservation).

Philosophically, this shifts emphasis from particular forces to invariant structures that organise many phenomena at once. For a concrete picture, Figure 11.2 shows rotational symmetry in a simple central-force orbit and its link to conserved angular momentum.

Unification as an Ideal

Unification seeks to:

- describe apparently different interactions (for example electricity and magnetism) within a single framework,
- reduce the number of independent postulates needed to account for observed regularities,
- sometimes predict new phenomena (for example new particles) by exploring the consequences of symmetry.

From Lab to Life

Symmetry-based thinking has echoes beyond physics:

- In statistics and AI, invariance to certain transformations (for example translations in images) motivates convolutional architectures.
- In finance, no-arbitrage conditions can be viewed as invariance under certain portfolio transformations.
- In mathematics, group theory and representation theory formalise symmetry far beyond spatial intuition.

Seeing symmetries helps you recognise when different-looking problems are instances of the same underlying structure.

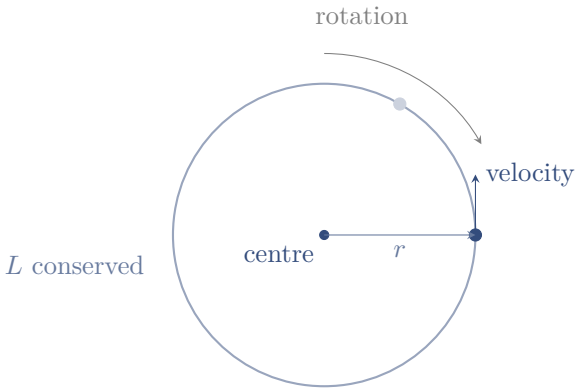

Figure 11.2: Rotational symmetry in a central-force system: rotating the whole configuration leaves the dynamics unchanged. This invariance under rotations is tied to conservation of angular momentum L.

11.5 Determinism, Chance, and Initial Conditions

Physics has long been a testing ground for ideas about determinism and chance.

Deterministic Laws and Unknown Initial Conditions

Classical mechanics is often described as deterministic:

- given exact laws and precise initial conditions, the future (and past) evolution is fixed,
- uncertainty stems from our ignorance of initial states or parameters, not from inherent randomness.

In practice:

- we never know initial conditions perfectly,
- small uncertainties can grow rapidly in chaotic systems (as discussed in Chapter 9),
- probabilistic descriptions become practically necessary even when underlying laws are deterministic.

Intrinsic Randomness

Quantum theory appears to introduce intrinsic randomness:

- even with a fully specified wavefunction, measurement outcomes are probabilistic,
- no deeper hidden variables (within standard formulations) determine which individual decay or detection event will occur.

Whether this randomness is truly fundamental or reflects incomplete description depends on one's preferred interpretation, but at least operationally it plays a different role than ignorance-based randomness in classical settings.

Implications for Prediction and Risk

For prediction and risk:

- deterministic chaos and unknown initial conditions limit forecast horizons even when laws are simple,
- intrinsic randomness imposes irreducible variability on certain outcomes,
- both phenomena motivate probabilistic and scenario-based planning rather than exact long-term forecasts.

Analogy: Weather, Coins, and Quantum Decay

- Weather: governed by deterministic equations but chaotic; we predict probabilities over days, not exact states months ahead.
- Coin toss: often modelled as random because tiny, uncontrolled differences in initial conditions and air currents dominate outcomes.
- Quantum decay: even in carefully controlled conditions, only decay probabilities and distributions are predicted.

These different sources of uncertainty shape how we use physical theories in risk assessment and decision-making.

11.6 Summary and Cross-Domain Links

Physics serves as a rich testbed for many themes from earlier chapters.

- Laws and frameworks: Newtonian mechanics, relativity, and quantum theory illustrate how theories structure reality and how they can change.
- Symmetry and unification: show how deep structures can organise many phenomena and guide new discoveries.
- Determinism and chance: highlight different sources of uncertainty and their implications for prediction and risk.

Cross-domain links:

- To mathematics: physics motivates and is organised by mathematical structures, especially those built from symmetry.
- To AI: statistical models and learning algorithms often borrow probabilistic and dynamical ideas from physics.

- To finance: stochastic processes and risk measures echo physical thinking about randomness and dynamics.

As you move into later case studies, keep physics in mind as both inspiration and caution: it showcases what clear laws can achieve and what conceptual upheavals may still arise.

Try in 60 Seconds

Short checks:

- Think of three phenomena and decide for each whether a Newtonian, relativistic, or quantum description is most appropriate.
- Name one symmetry (spatial, temporal, or internal) that appears in a system you know, and what is conserved because of it.
- Classify a source of uncertainty you face (for example weather at a launch, device noise, or quantum decay) as mainly initial-condition sensitivity, chaos, or intrinsic randomness.

Chapter 12

Artificial Intelligence: Data, Models, and Understanding

Artificial intelligence (AI) and machine learning (ML) sit at the intersection of engineering, empirical science, and mathematical theory. This chapter treats AI as a case study in large-scale induction, black-box modelling, and the challenges of understanding and governing powerful predictive systems.

Learning Objectives

After working through this chapter you should be able to:

- describe different facets of AI/ML as engineering practice, empirical science, and mathematical/statistical theory,
- explain in high-level terms how supervised, unsupervised, and reinforcement learning relate to induction and generalisation,
- discuss what it might mean to "understand" a complex model and how interpretability tools fit in,
- evaluate claims based on benchmarks and leaderboards with attention to robustness and replication,
- analyse questions of agency and responsibility for AI-assisted or AI-driven decisions.

At a Glance

AI systems learn patterns from data and deploy them in the world. Their successes raise classic questions about induction, explanation, and scientific method in a new key: what counts as a good model when we cannot fully interpret it, and how do we align such systems with human values and goals?

Everyday Analogy

Think of a very experienced colleague:

- they cannot always explain every judgement they make,
- their track record matters, but so do your questions about where they might be biased or brittle,
- you might use their advice as input, not as unquestionable authority.

AI systems are like such colleagues built from data and code: powerful in some domains, limited in others, and always embedded in human decision processes.

12.1 AI as Engineering, Science, or Something Else?

AI/ML research and practice blend several modes of work.

Engineering Discipline

As engineering:

- AI focuses on building systems that achieve specific performance goals (accuracy, latency, revenue, safety margins),
- success is judged largely by whether systems work reliably in deployment,
- methods may be chosen pragmatically rather than for theoretical elegance.

Empirical Science

As empirical science:

- AI research designs experiments (benchmarks, ablation studies, controlled comparisons) to test hypotheses about architectures, training regimes, and data,
- communities track results via leaderboards and shared datasets,
- replication, robustness, and cross-lab validation become central concerns.

Mathematical and Statistical Theory

On the theoretical side:

- learning theory studies conditions under which models generalise from finite samples,
- optimisation theory analyses training dynamics,
- probabilistic modelling links AI back to statistics and information theory.

Analogy: Hybrid Lab

An AI lab often looks like:

- a physics lab (experiments, careful measurement, error analysis),
- a software startup (iterative product building, deployment pipelines),
- a psychology department (studying behaviour of systems under varied conditions).

Recognising these multiple roles helps when you read AI papers or evaluate AI products.

12.2 Learning from Data: Induction on Steroids

AI systems operationalise induction at scale.

Supervised, Unsupervised, and Reinforcement Learning

At a high level:

- Supervised learning: learn a mapping from inputs to labelled outputs (for example images to classes, text to sentiment, market states to returns).

- Unsupervised learning: find structure in unlabelled data (for example clusters, latent factors, compressed representations).
- Reinforcement learning: learn policies for acting in environments to maximise cumulative reward (for example games, robotics, algorithmic trading).

In each case, models generalise from observed data to new situations, revisiting themes of overfitting, inductive bias, and evaluation from earlier chapters.

Examples Across Domains

Concrete instances:

- Image classification: convolutional networks trained on millions of labelled images to recognise objects.
- Language models: large neural networks trained on text corpora to predict the next token, enabling downstream tasks.
- Trading bots: RL or supervised systems trained on historical market data to propose trades, subject to risk constraints.

Scale, Cost, and Computational Limits

Modern AI systems also make brute-force limits visible.

- Training state-of-the-art models can cost tens or hundreds of millions of dollars in compute and energy alone; re-running a full training run is often economically out of the question.
- Even when algorithms are, in principle, efficient, high-dimensional inputs and parameter spaces mean that exploration, data movement, and experimentation are constrained by wall-clock time.
- Engineers therefore treat compute as a scarce resource on par with data and talent, designing architectures, curricula, and evaluation pipelines with budgeted time and energy in mind.

From a philosophy-of-science angle, this makes *computational time* a first-class constraint on discovery, not just a technical detail. In complex domains we may know how to pose an ideal inference or control problem but still be unable to solve it at scale within realistic budgets. This is another face of the curse of dimensionality from Chapter 7 and the complexity limits discussed in Chapter 9.

From Lab to Life

Key parallels with earlier discussions:

- training vs test splits mirror the training vs generalisation distinction in Chapter 7,
- inductive biases (architecture choices, regularisation) echo model assumptions in statistics and physics,
- deployment under drift resembles causal and structural-break issues in Chapters 6 and 9.

AI systems make these issues highly visible because their failures can be abrupt and public.

12.3 Black-Box Models and the Question of Understanding

Deep networks and other flexible models challenge traditional notions of understanding.

What Does It Mean to Understand a Model?

Possible senses of understanding:

- Predictive: being able to anticipate outputs in new situations.
- Mechanistic: knowing how internal components contribute to behaviour.
- Counterfactual: being able to say what would change if certain parts or inputs were altered.
- Explanatory: having a concise, human-usable story that links inputs, internals, and outputs.

Many modern models deliver strong predictive understanding but weaker mechanistic and explanatory understanding.

Interpretability and Explainability Techniques

Interpretability work tries to bridge the gap:

- Feature attributions (for example saliency maps, SHAP values) highlight which inputs contributed to a prediction.
- Probing and representation analysis study internal activations or neurons.
- Simplified surrogate models approximate complex models locally for explanation.

These tools provide partial insight, but they do not always yield a simple "theory" of how a model works.

Analogy: Understanding a Friend

You can:

- predict many of a friend's reactions without knowing all the neural details of their brain,
- explain some behaviours with high-level traits or histories,
- still find them surprising in new contexts.

Understanding complex models may be more like understanding people than like understanding simple mechanical devices: layered, approximate, and never fully complete.

12.4 AI Benchmarks, Leaderboards, and Scientific Claims

Benchmarks organise AI research, but they bring their own philosophical questions.

Benchmarks as Experimental Theories

Benchmarks:

- define specific tasks and datasets that function as shared experimental environments,
- implicitly encode what the community considers important or representative,
- act as provisional "theories" of what constitutes progress on a problem.

Leaderboards add a competitive layer, ranking models by performance metrics.

Replication, Robustness, and Scaling Laws

Concerns include:

- Replicability: whether reported results reproduce under reimplementation, different random seeds, or slightly changed data.
- Robustness: whether performance holds under distribution shifts, adversarial perturbations, or minor dataset changes.
- Scaling laws: empirical regularities relating model size, data size, and performance, which some treat as quasi-laws.

Common Pitfall

It is easy to:

- over-interpret small leaderboard differences as meaningful scientific progress,
- conflate performance on narrow benchmarks with general intelligence or understanding,
- neglect failure modes that benchmarks do not probe.

Critical reading of AI results combines earlier lessons about p-values, overfitting, and replication with domain-specific awareness of benchmark design.

12.5 AI, Agency, and Responsibility

AI systems increasingly participate in decisions with real consequences.

Decision-Support vs. Decision-Making

We can locate systems along a spectrum:

- Decision-support tools: provide recommendations, risk scores, or summaries while humans retain final authority.
- Semi-autonomous systems: act within constraints but can be overridden or audited (for example trading bots with risk limits).
- Fully autonomous systems: act with minimal human oversight in real time (for example some industrial control systems, future autonomous vehicles).

Responsibility attribution differs across these modes.

Alignment and Value-Loading

Alignment questions ask:

- What objectives are systems optimising (explicitly or implicitly)?
- How do those objectives relate to human values, institutional goals, and legal norms?
- How do we handle trade-offs between accuracy, fairness, transparency, and other desiderata?

From a philosophy-of-science angle, alignment is partly about what "targets" we choose for learning: loss functions, datasets, and evaluation metrics embody value judgements.

From Lab to Life

In concrete projects:

- document intended use cases and out-of-scope uses,
- specify who is accountable for decisions and how errors will be detected and corrected,
- revisit objectives and metrics as contexts and values evolve.

These steps turn abstract alignment talk into day-to-day engineering and governance practice.

12.6 Summary and Open Questions

Key points:

- AI/ML blends engineering, empirical, and theoretical modes of science.
- Learning from data at scale revisits classical inductive questions under modern constraints.
- Understanding powerful models involves multiple senses beyond pure interpretability.
- Benchmarks and leaderboards shape research, but must be read with attention to robustness and replication.
- Agency and responsibility for AI systems centre on how we encode and monitor objectives and values.

Open questions:

- Can AI systems participate meaningfully in scientific discovery, and what would "understanding" look like for such systems?
- How should institutions allocate responsibility when AI tools contribute to failures or harms?
- What new philosophical tools are needed as AI becomes more entangled with all stages of scientific practice?

Try in 60 Seconds

Quick applications:

- Map one AI system you use or know to supervised, unsupervised, or reinforcement learning, and state its main loss or reward.
- List two ways that system might fail or be brittle that are *not* obvious from its benchmark score.
- Write down who, in practice, is responsible for decisions the system currently influences.

12.7 Technical Note: A Minimal Learning Abstraction

We close with a very compact abstraction that ties AI learning problems to earlier statistics and decision-theory ideas.

Supervised Learning as Risk Minimisation

In supervised learning:

- data are drawn from an unknown distribution over input–output pairs (X, Y),
- we choose a model class $\{f_\theta\}$ and a loss function $L(f_\theta(X), Y)$,
- the goal is to find θ that (approximately) minimises the expected loss $\mathbb{E}[L(f_\theta(X), Y)]$.

In practice, we approximate this expectation with empirical averages over a finite dataset and worry about overfitting, as in Section 7.6.

Reinforcement Learning as Sequential Decision-Making

In reinforcement learning:

- an agent interacts with an environment over time, receiving states S_t, choosing actions A_t, and obtaining rewards R_t,
- the goal is to learn a policy π that maximises expected cumulative reward (for example $\mathbb{E}[\sum_t \gamma^t R_t]$).

This links AI back to decision theory: we face trade-offs between exploration and exploitation, short-term vs. long-term gains, and risk in uncertain environments.

Connection to Earlier Themes

This abstraction highlights:

- the role of loss functions and rewards as carriers of value judgements,
- the importance of generalisation beyond training data,
- the need to think carefully about causality and distribution shift when deploying learned policies.

These themes tie AI back to the broader philosophy-of-science questions that run through the book. To make the loop structure concrete, Figure 12.1 expands the usual agent–environment sketch into policy, tools, memory, and environment blocks so you can literally trace where information and feedback circulate.

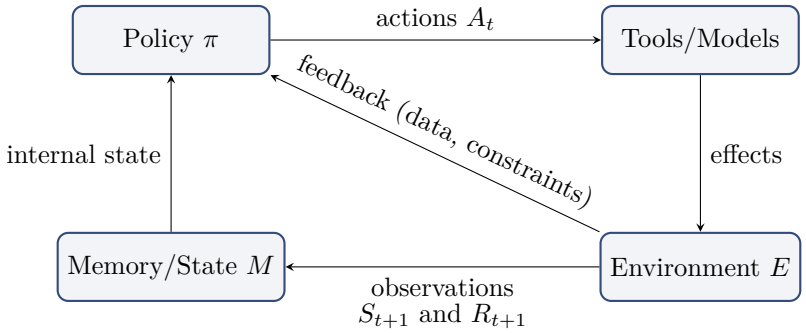

Figure 12.1: Expanded agent–environment loop: a policy chooses actions, tools and models implement them, the environment responds, memory updates an internal state, and feedback closes the learning and governance loop.

Chapter 13

Quantitative Finance: Markets, Models, and Uncertainty

Financial markets are real-world laboratories for uncertainty. Prices move quickly, data are plentiful, and models directly influence behaviour. This chapter uses quantitative finance to explore idealisation, efficient markets, model risk, reflexivity, and structural uncertainty.

Learning Objectives

After working through this chapter you should be able to:

- describe how financial markets serve as laboratories for uncertainty with rich but unstable data,
- summarise the historical arc from Bachelier through Black–Scholes and what it teaches about idealisation,
- explain the efficient market hypothesis (EMH), evidence for and against it, and behavioural critiques,
- analyse overfitting, backtest risk, and model risk in trading and risk management,
- discuss reflexivity, regime shifts, and structural uncertainty in markets.

At a Glance

Quantitative finance compresses many themes of this book: strong mathematical models, noisy and shifting data, agents reacting to models, and high-stakes decisions under uncertainty. It shows both the power of clean probabilistic theories and the limits imposed by structural breaks, feedback, and human behaviour.

Everyday Analogy

Imagine a casino where:

- the rules of the games slowly change,
- some players can influence the rules by how they play,
- new games appear and old ones vanish over time.

Quantitative finance is like doing probability theory and risk management in such a casino: there is structure, but it shifts, and your models are part of the story.

13.1 Finance as a Laboratory for Uncertainty

Financial markets are stochastic systems with unusual characteristics.

Data Richness vs. Structural Instability

On the one hand:

- high-frequency trading generates enormous amounts of tick data,
- exchanges and vendors record time-stamped prices, volumes, and order-book states,
- backtesting infrastructure makes it easy to simulate many strategies.

On the other hand:

- the structure of markets changes with regulation, technology, and strategy adoption,
- historical data reflect past market microstructures, not necessarily current or future ones,
- rare events (crashes, liquidity freezes) are underrepresented but crucial.

From Lab to Life

Practitioners must constantly balance:

- exploiting apparent regularities in historical data,
- guarding against regime shifts and model breakdowns,
- communicating uncertainty about both probabilities and structures to stakeholders.

Finance thus makes very concrete the difference between risk (known distributions) and deeper uncertainty.

13.2 From Bachelier to Black–Scholes and Beyond

The development of quantitative finance illustrates how idealised models can shape entire industries.

Random Walks and Brownian Motion

Key milestones:

- Bachelier (1900) modelled asset prices as random walks in continuous time, anticipating Brownian motion, though his work was largely overlooked for decades.
- Later, Brownian motion became central in both physics (diffusion) and finance (lognormal price models).

These models treat price changes as increments of a stochastic process with specified distributional properties.

Black–Scholes–Merton and Idealisation

The Black–Scholes–Merton framework:

- assumes frictionless markets (no transaction costs, continuous trading),
- models asset prices as geometric Brownian motion (continuous paths, lognormal distribution),
- derives a partial differential equation whose solution yields option prices.

Philosophically:

- the model embodies strong idealisations (continuous trading, constant volatility),
- yet it captures key qualitative features and provides a benchmark for more realistic extensions,
- risk-neutral valuation appears as a change of measure, echoing model-theoretic perspectives from Chapter 10, and, once widely adopted, helps coordinate how market participants quote and hedge derivatives.

Common Pitfall

Two opposite misreadings:

- treating Black–Scholes as a literal description of markets, ignoring its idealisations and domain limits,
- dismissing it entirely when deviations appear, overlooking its role as a baseline for thinking about arbitrage and risk.

Good practice treats such models as starting points, not as final truths.

13.3 Efficient Markets, Rational Expectations, and Their Critics

The efficient market hypothesis (EMH) has been central in finance and philosophy debates.

Efficient Market Hypothesis as a Scientific Claim

In one classic form:

- EMH states that asset prices "fully reflect" available information, making systematic excess returns hard to achieve after adjusting for risk.
- Variants differ in what counts as "available information" (weak, semi-strong, strong forms).

Viewed as a scientific claim:

- EMH proposes a link between information, expectations, and price dynamics,
- it suggests empirical tests based on predictability of returns and profitability of strategies.

Evidence and Behavioural Critiques

Empirically:

- some patterns (for example simple linear predictability of returns from past prices) are weak, consistent with certain EMH formulations,
- other anomalies (momentum, value, size effects, calendar effects) have persisted, at least for some periods,
- behavioural finance documents systematic deviations from rational expectations at the individual and market level.

A canonical example of elegant but empirically strained simplicity is the Capital Asset Pricing Model (CAPM): in its pure form, expected excess returns line up on a single "security market line" as a function of beta. Figure 13.1 sketches this theoretical line together with stylised data points that wander around it, hinting at the many factors real-world returns depend on. For decades, such one-factor linear models were almost the only tractable option; with modern computing power and powerful AI methods, finance research increasingly embraces high-dimensional and nonlinear structures instead of forcing everything onto a single line.

Philosophically, EMH illustrates:

- underdetermination: different models can explain overlapping sets of facts,
- demarcation challenges: is EMH falsifiable in practice, or can it be rescued with auxiliary assumptions?

From Lab to Life

For practitioners and regulators:

- taking EMH too literally may encourage complacency about bubbles and mispricings,
- ignoring EMH entirely may lead to overconfidence in supposed "edges" that are actually noise.

Philosophy of science encourages a nuanced stance: treat EMH as a useful limiting case and a discipline on modelling, not as a universal law.

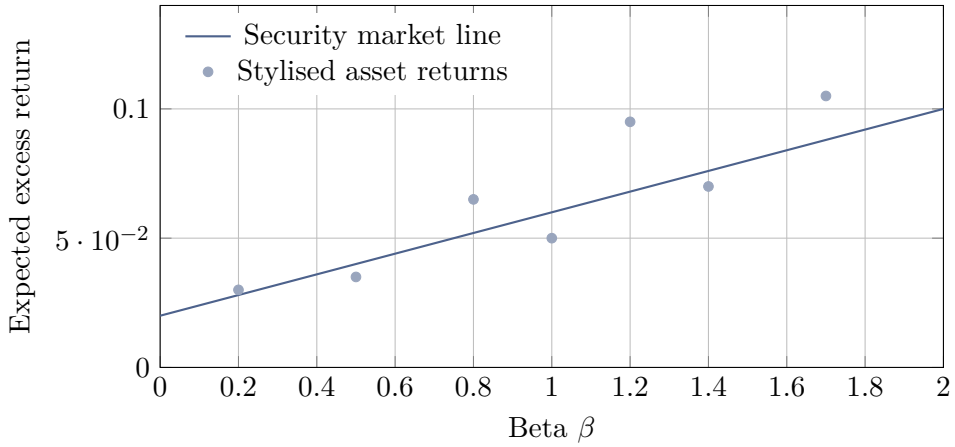

Figure 13.1: Stylised CAPM picture: the security market line (solid) expresses a one-factor linear view in which expected excess returns depend only on beta. Realised average returns for individual assets or portfolios (points) often scatter noticeably around this line, reflecting additional risk sources and model misspecification.

13.4 Backtests, Overfitting, and Model Risk

Backtesting is indispensable in finance and also a prime arena for overfitting.

The Garden of Forking Paths in Strategy Design

In strategy research:

- researchers test many signals, parameter choices, and portfolio constructions,

- only the best-performing backtests are typically reported or pursued,
- this search process inflates the apparent performance of selected strategies.

This mirrors p-hacking and researcher degrees of freedom in other sciences.

Model Risk and Stress Testing

Model risk arises when:

- models misrepresent key dynamics (for example assuming normal returns in fat-tailed markets),
- parameters are estimated from unrepresentative periods,
- users deploy models outside their domain of validity.

Stress testing:

- explores extreme but plausible scenarios,
- asks how portfolios or institutions would fare under shocks,
- helps identify vulnerabilities not obvious from day-to-day volatility.

Analogy: Trying Many Keys on One Lock

If you:

- try a bucket of random keys on a lock,
- eventually find one that happens to open it,
- later claim you "predicted" the right key from insight,

you are doing the backtest version of overfitting. Real prediction would require your choice to work on many different locks and days, not just the one you trained on.

Common Pitfall

Beware:

- reading too much into single impressive backtests without out-of-sample evidence,
- ignoring model risk because a framework is industry-standard,
- treating stress tests as box-ticking exercises rather than as probes of genuine fragility.

Robust finance practice aligns closely with robust scientific practice.

13.5 Reflexivity and the Observer Effect

Markets are populated by agents who read and act on models; this reflexivity changes the game.

Self-Fulfilling and Self-Defeating Prophecies

Examples:

- A widely-followed technical rule may become self-fulfilling: if enough traders buy when a certain pattern appears, prices may indeed rise.
- A highly profitable niche strategy may become self-defeating: once crowded, its edge disappears or reverses.

George Soros popularised "reflexivity" to describe such feedback loops between beliefs and market realities.

Difference from Planetary Motion

Unlike planets:

- market participants can learn about and adapt to models describing them,
- regulation and technology can change fundamental constraints,
- expectations can move prices in ways that then validate or invalidate those expectations.

This makes finance a particularly vivid example of systems that contain model-aware agents, complicating notions of law and prediction. Agent-based models (ABMs) and other simulation frameworks explicitly place many heterogeneous traders, strategies, and institutions inside the model, letting their rules update in response to prices and risk measures. This connects back to the complex-systems themes of Chapter 9: models here are not just about a system, they are also ingredients in the system's evolution.

13.6 Regimes, Crises, and Structural Uncertainty

Finance also showcases regime shifts and structural uncertainty.

Regime Shifts and Crises

Markets cycle through regimes:

- low-volatility, high-liquidity periods,
- stressed conditions with spikes in volatility and correlations,
- crisis episodes with liquidity evaporating and institutions failing.

Model parameters and even structures may differ across regimes.

To visualise this, Figure 13.2 shows a toy volatility time series with calm, stressed, and crisis phases. The picture is schematic, but it captures the idea that the same model may fit poorly across all segments without explicit regime or structural modelling.

Tails, Fat Tails, and Structural Breaks

Tail behaviour matters:

- return distributions often exhibit "fat tails" compared to Gaussians,
- structural breaks (for example regulatory shifts, technological disruptions) can create new regimes with different tail properties,
- Knightian uncertainty looms large: probabilities estimated from past calm periods may severely understate future risks.

From Lab to Life

For risk managers:

- combining probabilistic models with scenario analysis and expert judgement is standard,
- model governance frameworks explicitly track model risk, validation, and limitations,
- philosophy-of-science ideas about underdetermination and robustness are implicitly at work.

Finance thus illustrates how abstract epistemic worries turn into concrete governance challenges.

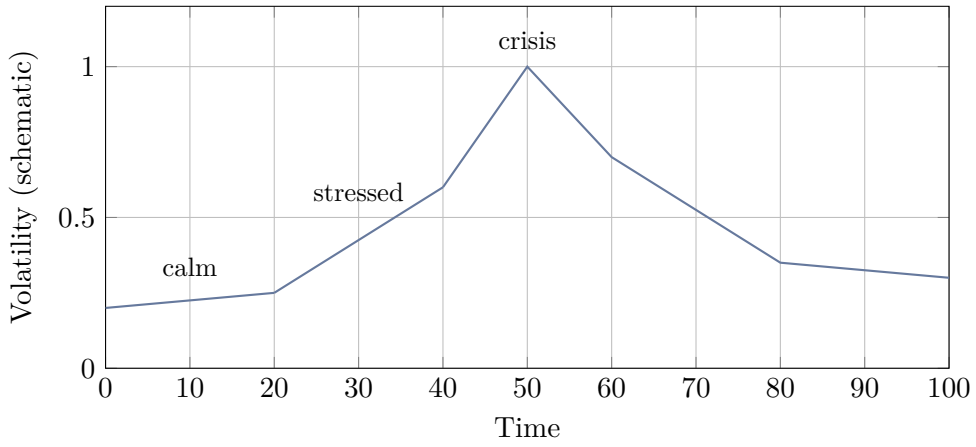

Figure 13.2: Toy volatility path illustrating regimes: an extended calm phase with low volatility, a stressed phase with rising volatility, and a crisis spike followed by partial normalisation. Real markets show richer behaviour, but regime structure like this often drives where simple models succeed or fail.

13.7 Summary and Synthesis

Key lessons from quantitative finance:

- Models can be mathematically elegant and commercially important while resting on strong idealisations.
- Overfitting and multiple-testing issues in backtests parallel replication concerns in other sciences.
- Markets are reflexive systems where models and agents co-evolve, challenging simple notions of law and prediction.
- Regime shifts, fat tails, and structural breaks push us from pure risk calculations toward scenario-based reasoning.

Taken together, these features make finance a powerful case study for thinking about model validity, uncertainty, and human behaviour in scientific systems.

Try in 60 Seconds

Rapid reflections:

- Imagine a simple backtest you might run and name one way it could be misleading because of overfitting or regime dependence.
- List two distinct market regimes you have observed (for example calm vs crisis) and one variable that behaves very differently between them.
- Note one way in which knowledge of a model or forecast could change behaviour in your domain and thereby change outcomes.

Chapter 14

Putting It All Together

We now step back from the details and look across domains. This closing chapter collects cross-cutting themes, reflects on scientific methods in the plural, highlights the human side of science, and sketches how philosophy of science can support practitioners in an AI-rich future.

Learning Objectives

After working through this chapter you should be able to:

- compare how mathematics, physics, AI, and finance use models, evidence, and idealisations,
- articulate why there is no single scientific method, but a family of related practices,
- describe the human elements that drive scientific progress and error,
- explain how philosophy of science can help practitioners think more clearly about models, uncertainty, and communication,
- reflect on open questions about the future of science with increasingly capable AI systems.

At a Glance

Across fields, science combines modelling, measurement, reasoning, and community norms—but in different proportions and with different constraints. Philosophy of science does not hand out rigid rules; it offers lenses and language to reason about methods, models, and values. In an era of powerful AI, these lenses become even more important.

Everyday Analogy

Think of a long-running TV series written by many authors:

- there is a shared world and continuity, but also shifts in tone and focus,
- some episodes are tightly plotted, others experimental or character-driven,
- fans and critics develop meta-knowledge about recurring patterns and tropes.

Science is like such a collaborative narrative: each field writes episodes with its own style, yet themes and characters cross over. Philosophy of science is partly the analysis of this evolving story.

14.1 Cross-Cutting Themes from the Four Domains

We briefly compare mathematics, physics, AI, and finance along a few axes.

Role of Mathematics

- Mathematics: mathematics is the subject; structures, axioms, and proofs are central.
- Physics: mathematics encodes laws and symmetries; experiments test where the structures apply.
- AI: mathematics underlies learning algorithms and optimisation, but large-scale computation and empirical performance drive much of the agenda.
- Finance: stochastic calculus, optimisation, and statistics shape models; institutional and behavioural realities push back.
- Philosophy theme: same mathematical structures (for example diffusions, Markov processes) appear across these fields with different interpretations.

Nature of Evidence

- Mathematics: proof is the primary form of evidence; examples and computations guide conjecture but do not replace proof.
- Physics: precise experiments and observations, often replicated across labs and scales, constrain theories.
- AI: benchmark performance, ablations, and robustness tests serve as evidence; replication and cross-lab validation are still maturing.
- Finance: historical data and backtests provide evidence, but structural breaks and reflexivity limit their stability.
- Philosophy theme: evidence always speaks through models and methods; we must ask "evidence for what, given which assumptions?"

Typical Idealisations and Uncertainties

- Mathematics: idealises by working in perfectly defined structures (no measurement error, no noise).
- Physics: uses idealised systems (frictionless planes, point particles) and then adds corrections.
- AI: idealises via training distributions and loss functions; real-world deployment introduces drift and adversarial effects.
- Finance: idealises away frictions and heterogeneous behaviour; real markets add microstructure, regulation, and feedback.
- Philosophy theme: idealisation is not a vice by itself; unexamined idealisations are.

14.2 Scientific Method(s), Plural

The idea of a single "scientific method" is attractive but misleading.

Family Resemblance

Across fields we see recurring elements:

- Systematic observation and measurement, often with increasing precision over time.
- Modelling: constructing abstract representations (mathematical, computational, conceptual) to capture patterns.
- Testing and revision: confronting models with data, revising or replacing them when they fail.
- Community norms: peer review, replication, credit assignment, and shared standards of evidence.

Different domains emphasise different pieces. Physics leans heavily on precise experiment and mathematical structure; AI on large-scale experiments and engineering iteration; finance on backtests and risk governance.

No One-Size-Fits-All Recipe

Therefore:

- insisting on a single stepwise "scientific method" obscures the diversity of successful practices,
- at the same time, calling anything "science" just because it involves data or equations dilutes the term,
- philosophy of science seeks a middle path: mapping patterns of practice without forcing uniformity.

From Lab to Life

For practitioners:

- it is more useful to ask "Which methods are appropriate here, and why?" than to ask whether a project fits a textbook method,
- explicit reflection on modelling choices, tests, and norms encourages better design and communication.

Thinking in terms of families of methods rather than a single recipe can be liberating and sharpening at once.

14.3 The Human Side of Science

Science is done by humans, with all our virtues and vices.

Curiosity, Creativity, Stubbornness, and Error

Human traits that drive science:

- Curiosity: motivates exploration of new questions and anomalies.
- Creativity: generates new models, analogies, and experimental designs.
- Stubbornness: sometimes needed to pursue risky ideas against initial scepticism.

- Error: inevitable; the key is building systems that detect and correct it.

Narratives, metaphors, and visualisations help us think and communicate, but they can also mislead if taken too literally.

Science as a Collaborative Narrative

Over time:

- theories accumulate, are revised, and sometimes replaced,
- new generations reinterpret past work through fresh lenses,
- institutions and technologies (journals, archives, code repositories) shape what is remembered and reused.

Analogy: A Long Novel with Many Authors

Science resembles a novel:

- early chapters define characters and settings (core concepts and theories),
- later chapters introduce twists (new data, new paradigms),
- different authors (disciplines) focus on different plot threads.

Philosophy of science is partly literary criticism of this evolving novel: it examines structure, themes, and tensions to help authors write better future chapters.

14.4 How Philosophy of Science Helps Practitioners

What does all this reflection buy a working physicist, AI engineer, quant, or data scientist?

Clearer Thinking About Models and Evidence

Philosophy of science offers:

- concepts for distinguishing hypotheses, laws, and theories,
- tools for analysing underdetermination, idealisation, and model pluralism,
- language for separating what data show from what models and values add.

This clarity helps when designing studies, interpreting results, and choosing among model families.

Better Communication of Uncertainty

Practitioners must explain uncertainty to colleagues, regulators, and the public.

- Knowing the difference between risk and deep uncertainty supports more honest risk communication.
- Understanding p-values, confidence intervals, and Bayesian updates avoids misleading language.
- Recognising the role of values in framing results helps manage expectations and trust.

More Robust Decision-Making

Philosophical tools support robustness:

- thinking about alternative models and scenarios reduces overreliance on any single view,
- being explicit about assumptions makes it easier to revisit them when conditions change,
- awareness of incentives and replication issues guards against overconfidence in early results.

From Lab to Life

You do not need to quote philosophers in your code or lab notes. But:

- when you ask "What would really count against this hypothesis?", you are doing philosophy of science,
- when you design a robustness check, you are implementing a philosophical concern in technical form,
- when you explain limits and trade-offs to non-experts, you are playing philosopher and teacher at once.

This book aims to make those activities more deliberate and enjoyable.

Try in 60 Seconds

Final prompts:

- Choose one concept from this book (for example model, risk, paradigm) and write two sentences about how it appears in your own field.
- Sketch, in a few words, your typical "method mix": how you combine modelling, data, and judgement in a representative project.
- Name one human factor in your own work (for example curiosity, time pressure, incentives) that helps or hinders good scientific practice.

14.5 Closing Thoughts and Outlook

Finally, we look ahead.

Science in an AI-Rich World

As AI systems become more capable:

- they will assist in data analysis, model discovery, and even hypothesis generation,
- they may propose models or explanations that are hard for humans to interpret directly,
- they will create new forms of collaboration between human and machine reasoning.

Questions follow:

- What counts as "understanding" when much of the computational work is done by systems we only partially grasp?
- How should credit, responsibility, and trust be allocated in human–AI scientific teams?
- Which parts of scientific practice should remain distinctly human, and why?

Can Scientific Discovery Be Automated?

Some tasks are already partly automated (symbolic regression, structure search, simulation-based inference). Others remain deeply human.

- Automation will likely expand the space of candidate models and explanations we can explore.
- Human judgement will remain central in choosing questions, interpreting results, and integrating knowledge across domains.
- Philosophy of science can help design interfaces and workflows where human and machine strengths are aligned.

Final Reflection

As you return to your own work:

- treat models as tools rather than tiny universes,
- view uncertainty as something to be mapped and communicated, not magically eliminated,
- remember that good science is both rigorous and humane: it cares about truth, usefulness, and the people affected by its applications.

The story of science is far from finished. You are, in a very real sense, one of its authors.

Part IV
Appendices

Appendix Overview

These appendices collect supporting material: core logical and probabilistic tools, a glossary of key terms, and pointers to further reading. They are designed as quick references you can dip into as needed.

Appendix A

Basic Logical and Probabilistic Notions

This appendix collects basic concepts in logic and probability that underpin many chapters. It is not a full course, but a compact reference you can consult when needed.

A.1 Propositions, Arguments, Validity, Soundness

Propositions and Arguments

- A *proposition* is a statement that is either true or false (for example "It is raining"; "This model overfits").
- An *argument* is a sequence of propositions where some (premises) are offered as reasons to accept another (the conclusion).

Validity and Soundness

- An argument is *valid* if, assuming the premises are true, the conclusion must be true; validity is about form, not actual truth.
- An argument is *sound* if it is valid and its premises are in fact true.

Many scientific arguments are not strictly deductive; they mix deductive, inductive, and abductive steps (see Chapter 2). Still, the ideas of validity and soundness help clarify when a step follows from assumptions and when we are making a probabilistic or explanatory leap.

A.2 Conditional Probability and Bayes' Rule

Conditional Probability

For events A and B with $P(B) > 0$, the conditional probability of A given B is

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

Informally: the probability that A occurs, restricting attention to cases where B occurs.

Bayes' Rule

Bayes' rule rearranges conditional probabilities:

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

This underlies Bayesian updating: starting from a prior $P(A)$, we update to a posterior $P(A \mid B)$ after observing B.

Toy Example: Medical Test

Suppose:

- A: a person has a disease; $P(A) = 0.01$ (1% prevalence).
- B: a test is positive; the test has 99% sensitivity ($P(B \mid A) = 0.99$) and 95% specificity ($P(\text{negative} \mid \neg A) = 0.95$).

Then:

$$P(B) = P(B \mid A)P(A) + P(B \mid \neg A)P(\neg A) = 0.99 \cdot 0.01 + 0.05 \cdot 0.99 = 0.0594,$$

and Bayes' rule yields

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \approx \frac{0.0099}{0.0594} \approx 0.167.$$

So even with a sensitive and fairly specific test, a positive result in a low-prevalence setting implies only about a 17% chance of disease. This illustrates why base rates matter in interpreting test results.

The same calculation can be visualised as a probability tree, shown in Figure A.1, which makes the flow from priors and likelihoods to the posterior $P(A \mid B)$ explicit.

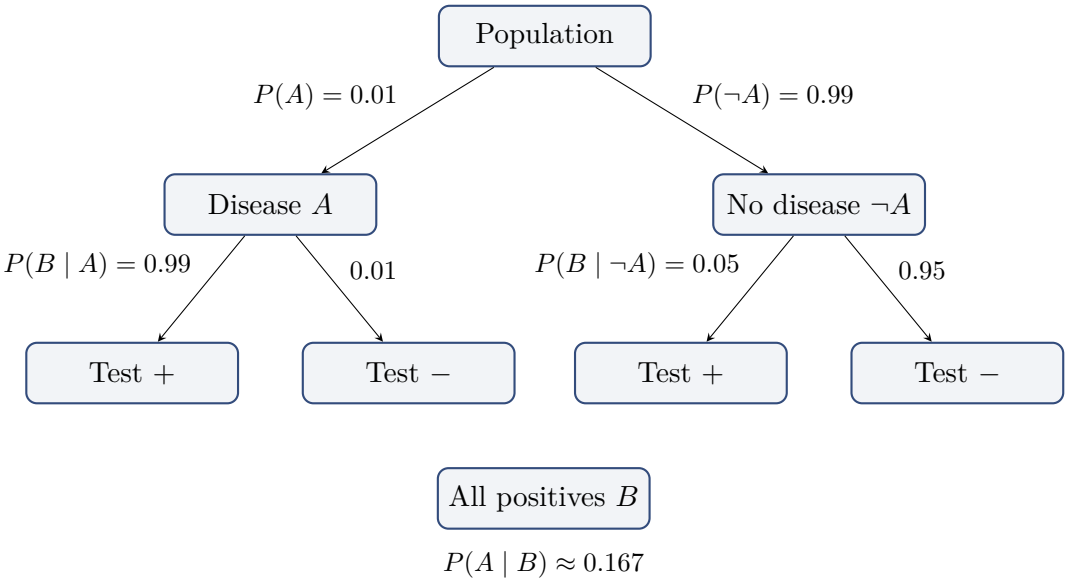

Figure A.1: Bayesian updating for the toy medical test example: starting from prevalence $P(A)$ and test characteristics $P(B \mid A)$ and $P(B \mid \neg A)$, the tree shows how the joint probabilities of each branch combine to give the marginal $P(B)$ and the posterior $P(A \mid B)$ for those who test positive.

A.3 Independence and Correlation

Independence

Events A and B are *independent* if $P(A \cap B) = P(A)P(B)$, equivalently $P(A \mid B) = P(A)$ (when $P(B) > 0$). Knowledge about one does not change probabilities for the other.

Correlation

- Correlation measures linear association between random variables; zero correlation does not imply independence in general.
- Correlation (or other dependence) alone does not establish causation; see Chapter 6.

A.4 Connecting Back to the Main Text

These notions appear throughout:

- conditional probabilities and Bayes' rule in discussions of evidence and updating,
- independence, correlation, and base rates in thinking about confounding and replication,
- logical structure (validity, soundness) when analysing arguments about theories and models.

Appendix B

Glossary of Key Terms

This glossary provides short, informal definitions of central terms as used in the book. Entries point back to chapters where concepts are developed in context.

Core Concepts

Abduction Inference to the best explanation: selecting the most plausible story that would make observed data less surprising (Chapter 2).

Anomaly A persistent result or pattern that resists accommodation within a current paradigm (Chapter 3).

Deduction Reasoning from general premises to specific conclusions in a way that preserves truth if the premises are true (Chapter 2).

Hypothesis A relatively specific, testable claim about aspects of the world (Chapter 2).

Induction Reasoning from repeated observations to more general claims, never logically guaranteed but central to learning from data (Chapter 2).

Law A stable, widely applicable regularity that summarises many observations and hypotheses (Chapters 2, 11).

Model A concrete representation (often mathematical or computational) of some part of a theory used for specific tasks (Chapters 3, 10).

Paradigm A Kuhnian package of exemplary problems, standards, tools, and training practices that structure normal science in a field (Chapter 3).

Theory A broader framework of concepts, principles, and models that generates hypotheses and explains laws (Chapters 3, 4).

Underdetermination The situation where multiple theories or models can explain the same body of data (Chapters 2, 9).

Philosophical Stances

Empiricism The attitude that experience and observation are the ultimate source of scientific knowledge and the final test of theories; elegant principles count only insofar as they survive contact with data (Chapter 2).

Constructivism A family of views that emphasise the role of social, historical, or conceptual construction in shaping scientific knowledge (appears implicitly throughout).

Instrumentalism The view that theories are tools for organising experience and making predictions, without strong commitments about the existence of their unobservable entities (Chapter 4).

Realism The view that successful scientific theories aim to describe both observable and unobservable aspects of reality, at least approximately (Chapter 4).

Statistical and Methodological Terms

Control Group A comparison group in experiments or quasi-experiments that does not receive the active treatment (Chapter 6).

Knightian Uncertainty Deeper uncertainty where probabilities themselves are poorly known or unstable (Chapters 8, 9, 13).

Latent Variable An unobserved quantity inferred indirectly from observed indicators via a model (Chapter 5).

Operationalisation The process of turning a concept (for example "intelligence" or "volatility") into a specific measurement procedure (Chapter 5).

Overfitting When a model captures noise in the training data as if it were signal, leading to poor generalisation (Chapters 7, 12).

***p*-hacking** Trying many analyses and only reporting those that achieve "significance," inflating false positive rates (Chapter 7).

Random Error Unpredictable measurement noise that scatters results around a value (Chapter 5).

Randomisation Assigning units to treatments by chance to break links with confounders (Chapter 6).

Replication Repeating a study or analysis to see whether results hold under similar or improved conditions (Chapter 7).

Risk Uncertainty that can be characterised by reasonably well-defined probability distributions (Chapters 8, 13).

Systematic Error Bias in measurement that consistently shifts results in one direction (Chapter 5).

Causality and Complexity

Causal Diagram A directed graph used to represent hypothesised causal relationships among variables (Chapter 6).

Confounder A variable that influences both a treatment and an outcome, potentially biasing naive comparisons (Chapter 6).

Complex System A system with many interacting components, nonlinearities, feedback, and emergence, making behaviour hard to infer from parts alone (Chapter 9).

Instrumental Variable A variable used in causal inference that affects treatment but not the outcome directly, helping isolate quasi-random variation (Chapter 6).

Difference-in-Differences A method comparing before–after changes in treated and control groups to estimate treatment effects under a parallel-trends assumption (Chapter 6).

Model Pluralism The practice of using multiple models for different questions or as robustness checks, rather than seeking a single universal model (Chapter 9).

Domain-Specific Terms

Backtest An evaluation of a trading or investment strategy using historical data (Chapters 7, 13).

Benchmark A standard dataset or task for comparing AI models or other systems (Chapter 12).

Alignment The problem of ensuring that AI systems' objectives and behaviours match human values and goals (Chapters 8, 12).

Black-Box Model A model whose internal workings are opaque or too complex to summarise, even if its input–output behaviour is known (Chapters 4, 12).

Interpretability The extent to which a model's behaviour or predictions can be understood and explained in human terms (Chapter 12).

Risk-Neutral Probability A probability measure under which discounted asset prices are martingales, used in pricing derivatives (Chapters 10, 13).

Symmetry A transformation that leaves key structures or laws invariant; in physics, linked to conservation laws (Chapter 11).

Unification Explaining diverse phenomena within a single theoretical framework (Chapters 11, 4).

Appendix C

Further Reading and Directions

This appendix lists pointers for deeper study. It is selective rather than exhaustive; many excellent sources are omitted.

General Philosophy of Science

- Samir Okasha, *Philosophy of Science: A Very Short Introduction*. Oxford University Press.
  A concise entry point covering core topics: explanation, realism, underdetermination, and scientific change.
- Peter Godfrey-Smith, *Theory and Reality*. University of Chicago Press.
  A clear, historically informed survey of twentieth-century philosophy of science, including Popper, Kuhn, and later developments.
- James Ladyman, *Understanding Philosophy of Science*. Routledge.
  A slightly more advanced treatment with careful attention to models, laws, and realism.

Philosophy of Physics

- Tim Maudlin, *Philosophy of Physics: Space and Time*. Princeton University Press.
  A focused introduction to conceptual issues in classical and relativistic spacetime theories.
- Tim Maudlin, *Philosophy of Physics: Quantum Theory*. Princeton University Press.
  A companion volume on quantum theory and the measurement problem.
- David Wallace, *The Emergent Multiverse*. Oxford University Press.
  A detailed defence of an Everettian (many-worlds) view of quantum mechanics with clear exposition of probabilities and branching.

Philosophy of AI and Cognitive Science

- Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*. Pearson.
  A broad AI textbook; useful background for connecting philosophical questions to concrete algorithms.
- Judea Pearl, *Causality*. Cambridge University Press.
  A foundational treatment of causal diagrams and interventions, highly relevant to experiments, observational studies, and AI.
- Brian Christian, *The Alignment Problem*. W. W. Norton.
  A readable survey of fairness, interpretability, and alignment issues in contemporary machine learning.

Philosophy of Economics and Finance

- Nancy Cartwright, *The Dappled World*. Cambridge University Press.
  A discussion of how laws and models work in messy, domain-specific settings, including economics.
- Donald MacKenzie, *An Engine, Not a Camera*. MIT Press.
  A sociological study of how financial models shape markets, with reflections on reflexivity and model risk.
- John H. Cochrane, *Asset Pricing*. Princeton University Press.
  A technical text on modern asset-pricing theory; useful for readers who want to see how finance models implement ideas about risk and expectation.

Suggested Reading Paths

- **Conceptual emphasis:** pair Part I of this book with general philosophy of science texts, then sample domain-specific readings.
- **Domain emphasis:** choose the case study (mathematics, physics, AI, or finance) closest to your interests and follow references outwards.

The literature is large and evolving; treat this appendix as a starting map rather than a finished atlas.