STUDY METHODOLOGIES

Scientific Methodology

The philosophy and practice underneath real science — demarcation, hypotheses and theories, controlled experiments, blinding, paradigm shifts, reproducibility, pseudoscience, peer review and preprints, and self-correction — grounded throughout in real, documented cases: Popper and Eddington's 1919 eclipse test, James Lind's 1747 scurvy trial, Wegener's continental drift, the 2015 Open Science Collaboration study, the NHMRC's homeopathy review, the string theory falsifiability debate, the Santa Clara antibody study, and the 2020 Surgisphere retraction — closing with a capstone that designs a falsifiable experiment from scratch and critiques a real one after the fact.

10 Chapters | Philip Osztromok | Generated with Claude

CHAPTER 1: WHAT MAKES SOMETHING "SCIENTIFIC"? THE DEMARCATION PROBLEM

Scientific Methodology

Chapter 1 · What Makes Something "Scientific"? The Demarcation Problem

Not every claim that sounds authoritative, uses technical language, or comes from a genuine expert is actually scientific in a meaningful sense. This course covers the real methodology underneath science itself — starting with the real observation that first made philosopher Karl Popper ask what separates science from everything that merely resembles it.

A Real Comparison That Sparked a Question

In autumn 1919, Popper — then a young man in Vienna, and briefly a former Marxist — noticed something genuinely striking: Albert Einstein's general theory of relativity and the psychological theories of Sigmund Freud and Alfred Adler were both widely treated as serious, authoritative bodies of knowledge. But they behaved completely differently when tested against the real world.

> **A REAL, RISKY PREDICTION — EDDINGTON'S 1919 ECLIPSE EXPEDITION**
>
> Einstein's theory made a genuinely risky, specific prediction: light from distant stars should bend by a precise, calculable amount as it passed near the sun's gravity. A British expedition led by Arthur Eddington tested this directly during the solar eclipse of 29 May 1919 — and the theory could have failed. It didn't; the measured deflection matched Einstein's prediction.

Freud's and Adler's theories, by contrast, seemed able to explain literally any human behavior after the fact — Popper later wrote there was "no conceivable human behaviour which could contradict them." Whatever someone did could always be reinterpreted as confirming the theory. Marxism, in Popper's own real assessment, had drifted into the same pattern.

The Falsifiability Criterion

Popper formalized this real distinction in his 1934 book *Logik der Forschung* (translated into English in 1959 as *The Logic of Scientific Discovery*): a genuinely scientific theory must make predictions specific enough that they could, in principle, be proven wrong by a real

observation. It's not a requirement that a theory actually be false — general relativity passed its real test — only that a real test capable of failing it must exist at all.

A REAL, CONCRETE EXAMPLE: ASTROLOGY

Popper himself pointed to astrology as a real case of the problem: horoscope predictions are typically vague enough to fit almost any real outcome after the fact, with no clear, specific claim that a contradicting event could ever actually falsify.

Falsifiable vs. Unfalsifiable, Compared

GENERAL RELATIVITY (1919)	PSYCHOANALYSIS (POPPER'S REAL ASSESSMENT)
Made a specific, risky, real prediction	Could accommodate any real behavior after the fact
A real test (the 1919 eclipse) could have failed it	No real, conceivable observation could contradict it
Passed a genuine, dangerous test	Never faced one

THE REAL, PRACTICAL LESSON

Falsifiability is a test of a theory's own structure, not a verdict on whether it happens to be true. Asking "what real observation would prove this wrong?" is often more revealing than asking "what evidence supports this?" — since a theory built to accommodate any possible evidence supports itself automatically, which is precisely the real problem.

What This Course Covers

2 **Theories & Laws**
Real definitional distinctions

3 **Experiment Design**
Variables, control groups, randomization

4 **Blinding**
Why the placebo effect demands it

5 **Paradigm Shifts**

6 **Reproducibility**

7 **Pseudoscience**

How consensus actually changes

The real, ongoing crisis

Demarcation applied to real cases

8

Open Science

Preprints and publishing reform

9

Self-Correction

Retraction and real fixes

10

Capstone

Design and critique a real experiment

Hands-On Exercises

EXERCISE 1

Explain, in your own words, exactly what real observation from the 1919 eclipse expedition would have falsified Einstein's theory — what specifically did Eddington's team measure, and what result would have proven the prediction wrong?

View solution

EXERCISE 2

Take a real claim you've heard recently (not necessarily scientific) and ask Popper's own question of it: what specific, real observation would need to occur to prove it false? If you genuinely can't think of one, explain what that tells you about the claim.

View solution

EXERCISE 3

Explain, in your own words, why Popper's criterion evaluates a theory's own falsifiability rather than whether it happens to be true — using general relativity's own real 1919 test (which the theory passed) as the example.

View solution

CHAPTER 1 QUICK REFERENCE

- Karl Popper's real 1919 observation contrasted Einstein's risky, testable general relativity against psychoanalysis and Marxism's inability to be contradicted by any real observation
- Arthur Eddington's real 1919 solar eclipse expedition tested Einstein's theory directly — and it could have failed
- Popper formalized falsifiability in his real 1934 book, Logik der Forschung
- A genuinely scientific theory must permit a real observation capable of proving it wrong, even if that observation never actually occurs
- Popper himself cited astrology as a real example of claims too vague to be falsifiable

CHAPTER 2: HYPOTHESES, THEORIES & LAWS

Scientific Methodology

Chapter 2 · Hypotheses, Theories & Laws

"It's just a theory" is one of the most common real misunderstandings in public science discussion — and a real, landmark US federal court case ended up directly addressing why the argument doesn't hold up.

Three Real, Distinct Categories

Hypothesis

A specific, testable, tentative explanation for one observation — the real starting point of an investigation, not yet broadly tested.

Theory

A broad, well-substantiated explanation for a wide range of real observations, backed by extensive, repeatedly tested evidence.

Law

A concise, often mathematical description of what happens under given conditions — real, predictable regularity, without necessarily explaining why.

A REAL, COMMON MISCONCEPTION

Theories do not "graduate" into laws once enough evidence accumulates — they're genuinely different kinds of statements, not different levels of certainty. Newton's law of gravitation is a real mathematical description of *what* gravity does; it doesn't explain *why*. A theory of gravity (like general relativity) explains the underlying mechanism. Calling evolution "just a theory," as though a hypothesis were one step away from becoming a real law, misunderstands that a scientific theory already represents one of the strongest, most evidence-backed categories of scientific knowledge that exists.

A Real, High-Stakes Case

In 2005, the Dover, Pennsylvania school board required students be told evolution was "a theory ... not a fact," and directed them toward "intelligent design" as an alternative. Eleven parents sued in *Kitzmiller v. Dover Area School District*.

A REAL, DETAILED FEDERAL RULING — 20 DECEMBER 2005

Judge John E. Jones III issued a real 139-page ruling finding intelligent design was not science — concluding it "cannot uncouple itself from its creationist, and thus religious, antecedents," and that it rested on assumptions that were, in the real language of the ruling, "neither testable nor falsifiable."

A DIRECT, REAL CONNECTION TO CHAPTER 1

The ruling's own real language — "neither testable nor falsifiable" — is Popper's own demarcation criterion applied directly in a federal courtroom. Whatever position anyone takes on the underlying subject, the ruling's own real reasoning shows falsifiability isn't an abstract philosophical idea; it's an actively used, real, working standard for evaluating whether something belongs in a science classroom at all.

Theory vs. Law, Compared

LAW	THEORY
Describes what happens	Explains why it happens
Often a concise mathematical statement	A broad explanatory framework, often covering many phenomena
Example: Boyle's Law (pressure/volume relationship)	Example: the kinetic theory of gases (why the relationship exists)

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why "theories eventually become laws once proven" is a real, specific misunderstanding — using the real distinction between Newton's law of gravitation and a theory of gravity to illustrate your answer.

View solution

EXERCISE 2

Explain, in your own words, why the Dover ruling's real finding — that intelligent design was "neither testable nor falsifiable" — connects directly to Chapter 1's own falsifiability criterion, rather than being a separate, unrelated legal argument.

View solution

EXERCISE 3

A friend says, "Calling it a scientific theory means scientists aren't fully sure it's true." Using the real definitions from this chapter, explain what's misleading about that statement.

View solution

CHAPTER 2 QUICK REFERENCE

- A hypothesis is a specific, testable, tentative explanation; a theory is a broad, well-substantiated explanation; a law describes what happens, often mathematically
- Theories don't graduate into laws — they're genuinely different kinds of statements, not different confidence levels
- The real 2005 Kitzmiller v. Dover case addressed a school board requirement to call evolution "a theory, not a fact"
- Judge John E. Jones III's real 20 December 2005 ruling found intelligent design "neither testable nor falsifiable" — directly applying Popper's own criterion

CHAPTER 3: DESIGNING A CONTROLLED EXPERIMENT

Scientific Methodology

Chapter 3 · Designing a Controlled Experiment

A hypothesis stays untested until it's put through a real, structured comparison. This chapter covers the real building blocks of a controlled experiment through one of history's earliest and most famous examples — including its own honest limitations.

The Real, Foundational Case

JAMES LIND, HMS SALISBURY, 20 MAY 1747

Ship's surgeon James Lind selected twelve sailors suffering from scurvy, at similarly early stages of the disease, and divided them into six pairs. Each pair received a different traditional remedy — cider, elixir of vitriol, vinegar, seawater, a purgative mixture, or two oranges and one lemon daily — while everything else about their care stayed the same.

After six days, only the citrus pair showed real, dramatic improvement; the cider pair improved slightly; the other four pairs showed no real improvement at all. It's now considered one of the first controlled clinical trials in the history of medicine.

The Real Components of a Controlled Experiment

Independent Variable

What's deliberately changed between groups — Lind's own six different remedies.

Dependent Variable

What's measured as the outcome — how much each pair's real symptoms improved.

Controlled Variables

What's deliberately kept the same — disease stage, basic diet, and shipboard conditions across all six pairs.

A REAL, HONEST LIMITATION

Lind's own trial is genuinely celebrated, but it wasn't perfect: two sailors per group is a very small sample, and Lind didn't randomly assign sailors to their pairs — he simply grouped them.

Randomization, covered next, protects against unknown factors (a sailor's own prior health, diet before boarding) unevenly influencing one group over another purely by chance. Even history's most famous early controlled trial had real, genuine room for improvement by later methodological standards.

Why Randomization Matters

Randomly assigning participants to groups — rather than choosing by hand — spreads out real, unknown differences between individuals (age, health, unmeasured habits) evenly across groups by pure chance, rather than letting a researcher's own unconscious choices skew who ends up where. It's a real, structural safeguard against a confounding variable quietly explaining a result that looks like it came from the actual treatment.

THE REAL, SOBERING POSTSCRIPT

Despite Lind's own clear real result, the Royal Navy didn't formally introduce citrus rations until 1795 — 48 years later. Good experimental design doesn't automatically translate into real-world action; that gap between evidence and adoption is its own genuine, separate problem.

Hands-On Exercises

EXERCISE 1

Identify the independent variable, dependent variable, and at least two controlled variables in Lind's real trial, in your own words.

View solution

EXERCISE 2

Explain, in your own words, a real, specific way the lack of randomization could have affected Lind's own results — what unmeasured factor might have differed between the pairs purely by how he happened to group them?

View solution

EXERCISE 3

Design a simple, real, controlled experiment (any topic you like) explicitly naming its independent variable, dependent variable, and at least two controlled variables.

View solution

CHAPTER 3 QUICK REFERENCE

- James Lind's real 1747 scurvy trial compared six treatments across twelve sailors in matched pairs — one of medicine's first controlled clinical trials
- Only the citrus pair showed dramatic real improvement
- Independent variable = what's changed; dependent variable = what's measured; controlled variables = what's kept the same
- Lind's trial genuinely lacked randomization and used a tiny sample — real limitations even in a celebrated case
- Randomization spreads unknown confounding factors evenly across groups by chance
- The Royal Navy didn't act on Lind's findings for a real 48 years — good evidence doesn't guarantee real-world adoption

CHAPTER 4: THE PLACEBO EFFECT & WHY BLINDING MATTERS

Scientific Methodology

Chapter 4 · The Placebo Effect & Why Blinding Matters

Belief alone can measurably change how a person feels — which means an experiment that doesn't account for that fact can mistake a participant's own expectation for a real treatment effect. This chapter covers the real, surprisingly old origin of blinding as a fix, and the real, complicated modern history of the placebo effect itself.

A Real, 1784 Origin of Blinding

> **THE ROYAL COMMISSION ON ANIMAL MAGNETISM**
>
> In 1784, King Louis XVI commissioned a real investigation into Franz Mesmer's claimed "animal magnetism" — a supposed invisible fluid practitioners could channel to heal patients. The commission included Benjamin Franklin, chemist Antoine Lavoisier, and astronomer Jean-Sylvain Bailly. Its real method was genuinely innovative: blindfolding both subjects and, in some tests, the investigators themselves, and applying "sham" and "genuine" procedures to patients with both real and fabricated conditions — the first documented use of blindfolding in a controlled trial.

The commission's real conclusion, after five months of testing: the mesmeric "fluid" didn't exist, and the convulsions and other effects attributed to it were products of the subjects' own imagination and imitation — a striking, direct demonstration that belief and expectation alone could produce real, physically observable effects.

The Placebo Effect: A Real, Complicated Modern History

Physician Henry Beecher, running a hospital for wounded American soldiers during World War II, ran out of morphine and gave saline solution instead, letting soldiers believe it was the real painkiller. Almost half reported their pain had lessened or disappeared. He later formalized this into a landmark 1955 JAMA paper, "The Powerful Placebo," estimating that up to 35% of therapeutic effects across the trials he reviewed could be attributed to placebo response alone.

A REAL, HONEST COMPLICATION

Beecher's own famous paper hasn't held up perfectly under later scrutiny. A real reanalysis of the specific studies he cited found no clear evidence of an actual placebo effect in them, and his own methodology has since been described as genuinely sloppy, even by 1955's standards. The placebo effect itself remains real and measurable in modern, better-controlled research — but the specific famous evidence that first popularized its size is shakier than its own reputation suggests, the same real pattern this course has already seen with other landmark studies.

Why Double-Blind Design Exists

A single-blind study keeps the participant unaware of which treatment they're receiving. A **double**-blind study keeps the researcher unaware too — because a researcher who knows which patient received the real treatment can unconsciously treat that patient differently, ask leading questions, or interpret ambiguous results more favorably, introducing a real bias entirely separate from the participant's own expectations.

DESIGN	WHAT IT PROTECTS AGAINST
Single-blind	The participant's own expectation influencing their reported outcome
Double-blind	Both the participant's expectation and the researcher's own unconscious bias

THE REAL, PRACTICAL LESSON

Blinding isn't a formality — it's a direct, structural response to a real, demonstrated fact: belief changes outcomes, on both sides of an experiment. The 1784 commission proved this about patients nearly 170 years before Beecher's paper; modern double-blind design closes the same gap for researchers too.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why blindfolding both the subjects AND, in some tests, the investigators was necessary for the 1784 commission's own real conclusion to be trustworthy.

View solution

EXERCISE 2

Explain, in your own words, why the real reanalysis finding "no clear evidence of a placebo effect" in Beecher's own cited studies doesn't mean the placebo effect itself is fake — what's the real difference between those two claims?

View solution

EXERCISE 3

Describe a real, specific scenario where a single-blind design (participant doesn't know, researcher does) could still produce a biased result — what would the researcher's own knowledge affect?

View solution

CHAPTER 4 QUICK REFERENCE

- The 1784 Royal Commission on Animal Magnetism (Franklin, Lavoisier, Bailly) used history's first documented blindfolded controlled trial
- It concluded mesmeric effects came from imagination and imitation, not a real physical "fluid"
- Henry Beecher's real 1955 "The Powerful Placebo" (JAMA), grounded in his own WWII saline-for-morphine experience, estimated placebo effects at up to 35%
- A real later reanalysis found no clear evidence of a placebo effect in Beecher's own specific cited studies, and criticized his methodology — though the broader effect remains real in better modern research
- Double-blind design protects against both the participant's expectation and the researcher's own unconscious bias

CHAPTER 5: PARADIGM SHIFTS: HOW SCIENCE ACTUALLY CHANGES

Scientific Methodology

Chapter 5 · Paradigm Shifts: How Science Actually Changes

Science doesn't always advance through smooth, gradual accumulation of evidence. Sometimes a real, correct idea sits rejected for decades — not because the evidence was wrong, but because the field lacked the mechanism needed to accept it. This chapter covers the real framework for understanding that pattern, and the real case that best illustrates it.

Kuhn's Real Framework

Philosopher Thomas Kuhn's 1962 book *The Structure of Scientific Revolutions* — which sold over a million real copies and introduced the term "paradigm shift" into everyday language — described science as moving through a real, recurring cycle rather than steady linear progress.

Normal Science

Scientists work within an accepted paradigm, solving smaller puzzles without questioning its own basic assumptions.

Anomaly

Observations start appearing that the current paradigm can't easily explain.

Crisis

Anomalies accumulate until confidence in the existing paradigm genuinely breaks down.

Revolution

A new paradigm replaces the old one — not gradually, but as a real, comparatively sudden shift.

A Real, Textbook Case: Continental Drift

On 6 January 1912, meteorologist Alfred Wegener presented his hypothesis that Earth's continents had once been joined and had since drifted apart, publishing it fully in 1915. The idea explained real anomalies — matching coastlines, shared fossils on now-separated

continents — that the existing geological paradigm struggled with. It was met with real, documented ridicule; by 1930, most geologists had rejected it, and it faded into obscurity for decades.

THE REAL MISSING PIECE

Wegener's own theory had one serious, real weakness: he couldn't propose a credible mechanism for how entire continents could actually move. Without one, geologists had a real, legitimate reason for skepticism, not just institutional stubbornness.

THE REAL VINDICATION

Starting in the mid-1950s, real paleomagnetic evidence — the direction rocks were magnetized in when they formed — began supporting continental movement. In the 1960s, new instruments revealed real seafloor spreading and subduction zones, finally supplying the missing mechanism. Continental drift was resurrected as the foundation of modern plate tectonics.

Mapping the Story onto Kuhn's Framework

KUHN'S STAGE	WHAT HAPPENED, IN REAL TERMS
Normal science	Pre-1912 geology, working within a static-continents framework
Anomaly	Matching coastlines, shared fossils across oceans, unexplained by the old framework
Crisis	Decades of unresolved tension, with Wegener's own idea rejected for lack of mechanism
Revolution	1960s seafloor-spreading evidence supplies the mechanism — plate tectonics becomes the new paradigm

THE REAL, PRACTICAL LESSON

A correct idea being rejected for decades isn't automatically evidence of closed-mindedness — Wegener's own theory genuinely lacked a real, necessary piece (a mechanism) that later evidence supplied. The lesson isn't "trust every rejected idea," but that real scientific consensus can lag behind a correct conclusion until the supporting evidence a field actually needs becomes available.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why "lacking a mechanism" was a real, legitimate scientific objection to Wegener's theory, rather than simply institutional bias against a new idea.

View solution

EXERCISE 2

Using Kuhn's own four-stage framework, map a real or hypothetical example of your own choosing (scientific or otherwise) onto normal science, anomaly, crisis, and revolution.

View solution

EXERCISE 3

Explain, in your own words, why the chapter's own closing lesson is "trust the evidence a field actually needs" rather than "every rejected fringe idea eventually turns out to be right" — what's the real difference, and why does it matter?

View solution

CHAPTER 5 QUICK REFERENCE

- Thomas Kuhn's real 1962 book described science moving through normal science, anomaly, crisis, and revolution rather than steady linear progress
- Alfred Wegener's real 1912 continental drift hypothesis was ridiculed and largely rejected by 1930, mainly for lacking a credible mechanism
- Real paleomagnetic evidence (mid-1950s) and seafloor-spreading discoveries (1960s) supplied that missing mechanism
- Continental drift was resurrected as the foundation of modern plate tectonics

- A correct idea can be rejected for decades not from mere stubbornness, but a genuine, real missing piece of evidence

CHAPTER 6: REPRODUCIBILITY: CAN OTHER SCIENTISTS GET THE SAME RESULT?

Scientific Methodology

Chapter 6 · Reproducibility: Can Other Scientists Get the Same Result?

A single published result, however striking, isn't the end of the scientific process — it's a claim waiting to be checked. This chapter covers real, hard evidence of how often that check fails, and a real, influential theory of why.

A Real, Landmark Test

> **THE OPEN SCIENCE COLLABORATION, SCIENCE, 2015**
>
> A large team of researchers attempted to replicate 100 real psychology studies, originally published in 2008 across three prestigious journals. In the original studies, 97% reported a statistically significant effect. In the real replications — using the original materials wherever possible — only about **36%** produced a statistically significant effect, and average effect sizes were roughly half the size originally reported.

This single result became the most cited real data point in what's now widely called the reproducibility crisis — and it wasn't confined to psychology; similar patterns have since been documented across cancer biology, economics, and other fields.

Why This Happens: A Real, Influential Framework

Physician-researcher John Ioannidis's real 2005 PLOS Medicine paper, "Why Most Published Research Findings Are False," argued the problem is structural, not simply a matter of individual dishonesty. He identified real, specific factors that make a published finding less likely to be genuinely true:

Small Sample Sizes

Smaller studies produce noisier, less reliable real

Small Effect Sizes

A genuinely small real effect is harder to distinguish reliably from statistical noise.

Publication Pressure

Novel, striking, significant results are far more

estimates, more prone to chance findings.

publishable than a careful replication or a null result.

A REAL, STRUCTURAL INCENTIVE PROBLEM

The Open Science Collaboration's own real interpretation pointed toward exactly this cultural pattern: academic incentives reward novelty over reproduction. A researcher who spends real time and resources carefully replicating someone else's already-published finding gains far less career credit than one who publishes something new — even though replication is what actually confirms a finding is real.

The Real, Measured Drop

	ORIGINAL STUDIES (2008)	REAL REPLICATIONS (2015)
Statistically significant result	97%	~36%
Average effect size	Baseline	Roughly half

THE REAL, PRACTICAL LESSON

A single striking published result — especially from a small study in a competitive, "hot" research area — deserves real, genuine caution until independently replicated. Chapter 8 covers real, current reforms (preprints, open data, pre-registration) built directly in response to this exact crisis.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why a drop from 97% to 36% significant results is a more alarming real finding than it might first appear — what does it suggest about how many of the original 2008 findings were likely real effects at all?

View solution

EXERCISE 2

Explain, in your own words, how Ioannidis's real "small sample size" and "publication pressure" factors could combine to produce a specific, concrete false finding — walk through a plausible real scenario.

View solution

EXERCISE 3

A classmate argues the reproducibility crisis proves science itself can't be trusted. Explain, in your own words, why the existence of the 2015 study itself is actually evidence against that conclusion.

View solution

CHAPTER 6 QUICK REFERENCE

- The real 2015 Open Science Collaboration study replicated 100 psychology studies; only ~36% reproduced a significant effect, down from 97% originally, with effect sizes roughly halved
- Ioannidis's real 2005 paper identified small sample sizes, small effect sizes, and publication pressure as structural drivers of unreliable findings
- Academic incentives reward novel findings far more than careful replications
- A single striking result deserves real caution until independently replicated

CHAPTER 7: PSEUDOSCIENCE & THE DEMARCATION PROBLEM IN PRACTICE

Scientific Methodology

Chapter 7 · Pseudoscience & the Demarcation Problem in Practice

Chapter 1 introduced Popper's falsifiability criterion; Chapter 2 introduced the real legal test applied in *Kitzmiller v. Dover*. This chapter applies both tools to real, contested cases — including one genuinely surprising fact: the demarcation problem isn't only a "science vs. obvious pseudoscience" question. It's actively, publicly argued over inside mainstream physics itself.

A Real, Tested Case: Homeopathy

Homeopathy is a useful first case precisely because it is **not** simply unfalsifiable — it makes specific, testable claims (this remedy improves this condition compared with placebo), and those claims have been tested, repeatedly, at real scale.

> **AUSTRALIA'S NHMRC, MARCH 2015**
>
> The National Health and Medical Research Council examined **57 systematic reviews covering 176 individual studies**, across **61 health conditions**, using only controlled trials. Its real, precise conclusion: "*there are no health conditions for which there is reliable evidence that homeopathy is effective.*"

A second, independent demarcation angle comes from homeopathy's own preparation method. Remedies are made by repeated serial dilution — often labelled "X" (1:10 per step) or "C" (1:100 per step). Past roughly **12C** dilution (about 1 part in 10^{24}), Avogadro's number (~6.022×10^{23}) means there is a vanishingly small real probability that even a single molecule of the original substance remains — yet many commercial remedies are diluted to 30C or higher.

> **WHY THIS CASE MATTERS FOR DEMARCATION**
>
> Homeopathy isn't disqualified because it can't be tested — it's disqualified because it *was* tested, repeatedly, and failed. When the dilution problem is raised, proponents have proposed new mechanisms in response ("water memory," more recently claims about water's own "coherent

domains") — each one untested and each one arriving only after the previous explanation was challenged. That pattern of continually inventing a new, unfalsified rescue has its own name, covered next.

The Demarcation Problem Inside Mainstream Science: String Theory

A genuinely striking real fact: whether a theory satisfies Popper's own falsifiability criterion is not only debated about pseudoscience — it is a live, unresolved dispute among working physicists, about one of the most prominent research programs in modern theoretical physics.

Lee Smolin, 2006

The Trouble with Physics. Argues string theory's "landscape" of an astronomical number of possible vacuum states (commonly cited around 10^{500}) means it may accommodate almost any observed outcome — casting real doubt on what it actually forbids.

Peter Woit, 2006

Not Even Wrong — the title borrows physicist Wolfgang Pauli's real put-down for an idea "so incomplete that it could not even be used to make predictions to compare with observations." Woit argues string theory has produced no falsifiable predictions at all.

Sabine Hossenfelder

A more recent, vocal critic arguing string theory has never been used to derive the Standard Model's own real physical constants, and that its landscape raises doubt about whether it meaningfully constrains reality at all.

AN HONEST COMPLICATION

This is a real, unresolved dispute, not a settled verdict — plenty of working physicists still consider string theory a serious, legitimate research program, and the debate itself is ongoing. The point for this course isn't who's right; it's that the demarcation question genuinely applies at physics' own frontier, not just to fringe claims.

Ad Hoc Rescue: The Mars Effect

French psychologist and statistician Michel Gauquelin claimed a real statistical correlation between Mars's position at birth and later becoming an elite sports champion — the "Mars effect." Belgium's Comité Para tested the claim in 1967 and initially reported a replication,

though most of the underlying data had still been collected by Gauquelin himself. Subsequent independent replication attempts largely failed.

> **THE REAL RESCUE MANEUVER**
>
> When the effect failed to replicate cleanly, defenders introduced new post hoc statistical constructs — an "eminence test" and an "IMQ bias indicator" — that effectively redefined, after the fact, which athletes counted as sufficiently notable for the effect to apply to. These constructs were tuned to fit the existing data rather than specified in advance, which is exactly what makes them an ad hoc rescue rather than a genuine, independent confirmation.

This is the pattern to watch for generally: a specific, testable claim fails a fair test, and rather than the claim being abandoned or revised in a way that could itself be tested again, a new explanation is invented specifically to explain away that one failure — with no independent evidence of its own.

Why Vague Claims Feel Confirmed Anyway: The Forer Effect

In 1948, psychologist Bertram Forer gave 39 of his students a personality test, telling each they would receive a personalized write-up. Every student actually received the identical, generic description, assembled from a newspaper astrology column. Asked to rate its accuracy, the average score was **4.26 out of 5** — before the trick was revealed. Psychologist Paul Meehl later named this the "Barnum effect" (1956); it's often called the Forer effect today. It doesn't explain the Mars effect's specific failure above, but it explains why astrology's *broader* appeal survives such failures — vague, universal statements feel personally accurate to nearly everyone.

Popper's Own Original Target: Psychoanalysis

Popper's falsifiability criterion has a real, specific origin: in the early 1920s he observed in Alfred Adler's Vienna clinics, and contrasted his growing skepticism of psychoanalysis with his real admiration for Arthur Eddington's 1919 solar-eclipse test of general relativity — a theory that made one specific, risky, falsifiable prediction that could have failed, and didn't (Chapter 1).

> *"As for Freud's epic of the Ego, the Super-ego, and the Id, no substantially stronger claim to scientific status can be made for it than for Homer's collected stories from Olympus. These theories describe some facts, but in the manner of myths... not in a testable form."*
> *— Karl Popper, Conjectures and Refutations*

Popper argued psychoanalysis was **"simply non-testable, irrefutable"** — that there was "no conceivable human behaviour which could contradict" it.

A REAL COMPLICATION, EVEN HERE

Philosopher Adolf Grünbaum later argued Popper's charge went too far — that psychoanalytic claims genuinely *are* testable in principle, but that specific predictions mostly fail when actually tested. That's a different, and in some ways harsher, criticism than pure unfalsifiability — a reminder that even a founding demarcation case isn't unanimously settled among philosophers of science.

Putting the Cases Side by Side

CASE	MAKES TESTABLE CLAIMS?	TESTED?	RESULT
Homeopathy	Yes	Yes, extensively (NHMRC 2015)	No reliable evidence found, across 61 conditions
String theory	Disputed	Not yet, per critics	Unresolved — a live dispute inside physics itself
Mars effect	Yes	Yes (Comité Para, 1967 onward)	Failed replication; rescued with post hoc statistical constructs
Psychoanalysis (Popper's charge)	Popper: no	Grünbaum: yes, and it mostly fails	Genuinely disputed even among philosophers of science

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why the homeopathy case is a stronger illustration of the demarcation problem than a case that's simply unfalsifiable from the start — what does the NHMRC review actually demonstrate that a pure falsifiability critique alone couldn't?

View solution

EXERCISE 2

Explain, in your own words, why the real string theory dispute shows the demarcation problem isn't confined to fringe claims — and what would have to happen for the dispute to actually be resolved one way or the other.

View solution

EXERCISE 3

Explain, in your own words, precisely what made the Mars effect's "eminence test" and "IMQ bias indicator" an ad hoc rescue rather than a legitimate methodological refinement — what would have made it legitimate instead?

View solution

CHAPTER 7 QUICK REFERENCE

- Homeopathy: NHMRC's 2015 review of 57 systematic reviews/176 studies/61 conditions found no reliable evidence — a testable claim that was tested and failed, not an unfalsifiable one
- String theory's falsifiability is a real, unresolved dispute inside mainstream physics itself (Smolin, Woit, Hossenfelder)
- The Mars effect failed independent replication and was rescued with new, post hoc statistical constructs — the ad hoc rescue pattern
- The Forer effect (1948) explains why vague personality claims feel personally accurate to nearly everyone
- Popper's own original demarcation target was psychoanalysis — though even that verdict is disputed by philosopher Adolf Grünbaum

CHAPTER 8: PEER REVIEW, PREPRINTS & THE OPEN SCIENCE MOVEMENT

Scientific Methodology

Chapter 8 · Peer Review, Preprints & the Open Science Movement

The sibling course Research Methodology already covers how traditional peer review works, how to assess source credibility, and real cases of a broken peer-review pipeline. This chapter picks up a different thread: real, current reforms — preprints and open science — built directly in response to problems like the reproducibility crisis from Chapter 6.

Preprints: Publishing Before Formal Review

A preprint is a completed research paper posted publicly **before** undergoing formal peer review — trading the months-long review timeline for immediate, open visibility.

> **ARXIV: THE REAL ORIGINAL**
>
> Physicist **Paul Ginsparg** launched arXiv at Los Alamos National Laboratory on **August 14, 1991**, as a simple email-based server for high-energy physics theory papers — expected to handle roughly 100 submissions a year, it received about 400 in its first six months alone. Ginsparg moved it to Cornell University in 2001, where it's hosted today. It has since expanded well beyond physics into math, computer science, and more. By mid-2026 arXiv had passed **3 million hosted articles**, with monthly submissions repeatedly setting new records — over 30,000 papers in a single month by 2026, up from roughly 120,000/year in 2017.

bioRxiv (2013) extended the same model to biology; **medRxiv** (2019) extended it to medicine — with one real, deliberate difference: medRxiv applies stricter screening and more prominent disclaimers than bioRxiv. The reason is direct real-world stakes — an unvalidated biology finding rarely changes anyone's behavior overnight, but an unvalidated clinical claim can influence real treatment decisions and public policy the moment it reaches the news. The case below shows exactly that risk playing out.

A Real Case: The Santa Clara Antibody Study

In April 2020, a Stanford-affiliated team — including Eran Bendavid, Jay Bhattacharya, and John Ioannidis (whose 2005 paper Chapter 6 already covered) — posted a preprint to medRxiv testing 3,330 Santa Clara County residents for COVID-19 antibodies.

THE REAL CLAIM

50 of 3,330 tests (1.5%) came back positive. The preprint's headline conclusion: the true infection rate was **50–85 times higher** than confirmed case counts — implying COVID-19 was far more widespread, and far less deadly, than official numbers suggested. Released amid an active lockdown-policy debate, it drew immediate, massive public and media attention.

THE REAL STATISTICAL CRITICISM

Statistician Andrew Gelman and others pointed out a critical flaw: the antibody test's own reported specificity had a confidence interval of roughly **98.1%–100%**. At a raw positive rate of only 1.5%, that matters enormously — if true specificity were, say, 98.5% (well inside the stated interval), essentially **all 50 positive results could be false positives**, statistically indistinguishable from zero real infections detected. Gelman called the errors "the kind of screw-ups that happen if you want to leap out with an exciting finding and you don't look too carefully at what you might have done wrong." A separate, real documented concern: two authors had published a Wall Street Journal op-ed against stay-at-home orders a month before running the study.

The study was **not retracted**. It went through substantial revision and, after formal peer review, was published in 2021 in the *International Journal of Epidemiology* — with the estimate revised down to a seroprevalence of 2.8%, translating to roughly a 54-fold undercount, closer to (but still below) the original claim. This is a genuinely nuanced real outcome: peer review functioned as intended, just applied to a claim that had already reached wide public attention before that scrutiny happened — exactly the structural risk a preprint carries.

The Other Side: When Speed Genuinely Helped

In January 2020, early COVID-19 epidemiological preprints posted via medRxiv provided some of the first real R0 (reproduction number) estimates for the Wuhan outbreak — roughly in the 2–3 range — while very little else was known about the virus. Those early estimates were broadly consistent with later, more rigorous peer-reviewed work, and gave public health bodies real, usable numbers during the most information-starved phase of the pandemic. This is the structural benefit preprints exist to deliver — directly contrasted against Santa Clara's own structural cost of that same speed.

Registered Reports & the Center for Open Science

The **Center for Open Science (COS)** was founded in January 2013 by psychologist **Brian Nosek** and then-PhD student Jeffrey Spies, with a mission to "increase the openness, integrity, and reproducibility of scientific research" — a real, direct response to problems like Chapter 6's own reproducibility crisis.

COS's **Registered Reports** format changes the order of peer review itself: a journal reviews and provisionally accepts a study's introduction and methodology *before* data is collected or results are known. Publication no longer depends on getting a "positive," statistically significant result — which removes the structural incentive to p-hack a null finding into something publishable, or file-drawer it away entirely (Chapter 6).

A REAL, PRECISE MEASURED EFFECT (SCHEEL, SCHIJEN & LAKENS, 2021)

Comparing every published Registered Report in psychology at the time (71 studies) against a random sample of standard psychology papers (152 studies): **96%** of results in the standard literature supported the authors' original hypothesis, versus only **44%** under the Registered Reports model — a **52-percentage-point drop**.

WHY THIS NUMBER MATTERS

A 96% "success rate" in ordinary science publishing is itself a red flag, echoing Chapter 1's Ioannidis material — real hypotheses shouldn't be confirmed nearly universally. Once publication was decoupled from getting a positive result, that rate collapsed to something much closer to what honest hypothesis testing should actually produce. This is a real, measured demonstration that a structural fix can work, not just a plausible-sounding proposal.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, why the Santa Clara study's real outcome (revised and eventually published, not retracted) is a more nuanced lesson than "the preprint was simply wrong" — what does the 2021 revised figure actually tell you about the original claim?

View solution

EXERCISE 2

Explain, in your own words, the real structural tradeoff preprints represent, using the Santa Clara case and the January 2020 R0-estimate case as your two contrasting examples.

View solution

EXERCISE 3

Explain, in your own words, precisely why decoupling publication from a "positive" result is what caused the real drop from 96% to 44% in the Scheel et al. study — what does the 96% figure suggest was happening under the traditional model?

View solution

CHAPTER 8 QUICK REFERENCE

- arXiv (1991, Paul Ginsparg) pioneered preprints; bioRxiv (2013) and medRxiv (2019) extended the model, with medRxiv screening more heavily given real clinical stakes
- The April 2020 Santa Clara antibody preprint claimed a 50-85x infection undercount; critics flagged the test's own specificity confidence interval as capable of explaining nearly all "positive" results as false positives
- The study wasn't retracted — it was revised and published in 2021 at a lower, still-elevated estimate, showing peer review working after the fact rather than before public exposure
- Early January 2020 R0 preprints show the real structural benefit preprints are designed to deliver: timely, usable numbers when nothing else exists yet
- The Center for Open Science's Registered Reports (reviewed before results exist) measurably cut psychology's own suspiciously high 96% "positive result" rate down to 44%

CHAPTER 9: WHEN SCIENCE CORRECTS ITSELF: RETRACTION & SELF-CORRECTION

Scientific Methodology

Chapter 9 · When Science Corrects Itself: Retraction & Self-Correction

The sibling course Research Methodology used the Wakefield MMR fraud as its own retraction case — a real 12-year gap between publication (1998) and retraction (2010). This chapter uses a genuinely different, far faster case to show self-correction actually working close to the speed it's supposed to.

A Real, Fast Self-Correction: The Surgisphere Scandal

In May 2020, two major papers — one in *The Lancet* (22 May, on hydroxychloroquine and COVID-19 mortality), one in the *New England Journal of Medicine* (1 May, on cardiovascular drugs and COVID-19) — drew on a purported multinational hospital registry from a small analytics company, **Surgisphere**.

> **THE CLAIMED DATASET**
>
> The Lancet paper claimed real data from **96,032 patients across 671 hospitals on six continents** — an extraordinary registry no one in the research community had previously known to exist.

Independent researchers quickly flagged real implausibilities — patient and hospital counts, and demographic/dosing breakdowns, that didn't square with what was independently known about hospital records in the claimed regions. Surgisphere refused to grant independent auditors access to the underlying patient-level data or identify the participating hospitals. On **28 May 2020**, over 180 researchers, clinicians, and statisticians signed a real open letter to the Lancet's editor stating the numbers "seem unlikely" and demanding an independent audit.

CASE	PUBLISHED	RETRACTED	REAL GAP
Wakefield MMR (Research Methodology)	1998	2010	~12 years
Surgisphere — Lancet	22 May 2020	4 June 2020	~13 days
Surgisphere — NEJM	1 May 2020	4 June 2020	~34 days

WHY THE CONTRAST MATTERS

Both cases involved a real, serious flaw eventually caught and formally retracted — but the real gap between them is enormous. Wakefield's fraud survived over a decade of real, ongoing public and medical consequences before retraction; Surgisphere collapsed within weeks once independent scientists actually scrutinized the data. Self-correction isn't guaranteed to be fast, but this case shows it genuinely can be, especially when a claim reaches enough scrutiny quickly.

The Formal Escalation: Correction, Expression of Concern, Retraction

The **Committee on Publication Ethics (COPE)**, founded in 1997 and followed by most major journals and publishers today, defines a real, three-tier framework for how a journal responds to a flawed paper:

Correction

An honest error that doesn't undermine the paper's core results or conclusions — a fix, not a retraction.

Expression of Concern

Real, unresolved doubt about a paper's reliability, often issued mid-investigation, without yet formally invalidating it.

Retraction

Reserved for errors or misconduct serious enough to genuinely invalidate the paper's own results and conclusions.

The Surgisphere case is a clean, real, worked example of this exact escalation. NEJM issued an Expression of Concern on **2 June 2020**; The Lancet followed with its own on **3 June**. Both papers were formally retracted the very next day, **4 June 2020** — a real two-step escalation, from suspicion to formal invalidation, completed in under 48 hours once independent verification of the underlying data proved impossible.

Retraction Watch: Tracking Self-Correction Itself

Retraction Watch was founded in August 2010 by science journalists Ivan Oransky and Adam Marcus, specifically to track and report on scientific retractions. Its structured **Retraction Watch Database** (launched 2015) now tracks over 30,000 retractions and counting — a real, running public record of exactly the kind of self-correction this chapter is about.

Is Science Getting Worse? A Real, More Nuanced Trend

A REAL, MEASURED INCREASE

The retraction rate for European biomedical-science papers roughly **quadrupled between 2000 and 2021**, with research misconduct — data or image manipulation, authorship fraud — accounting for roughly two-thirds of retractions and a growing share of the total.

WHY THIS ISN'T SIMPLY "SCIENCE IS GETTING WORSE"

A rising retraction rate could mean more misconduct is genuinely happening — or that more of it is being caught. Research-integrity specialists attribute the real rise to **both**: paper mills and image manipulation have genuinely grown at scale, but so has detection — tools like **PubPeer** (a post-publication scrutiny site launched in 2012), statistical forensics, and plagiarism/image-duplication software have made misconduct far harder to hide than it was decades ago. Most experts believe improved detection is doing a large share of the real work behind the rising number — which means, echoing this chapter's own theme, a rising retraction count can itself be a sign self-correction is working better, not proof science is failing.

Hands-On Exercises

EXERCISE 1

Explain, in your own words, what specifically allowed the Surgisphere case to be caught and retracted so much faster than Wakefield's — what real factor made the difference between roughly two weeks and roughly twelve years?

View solution

EXERCISE 2

Explain, in your own words, why the Surgisphere case moved through an Expression of Concern before a full retraction rather than going straight to retraction — what real purpose does that intermediate step serve?

View solution

EXERCISE 3

Explain, in your own words, why a rising retraction rate is genuinely ambiguous evidence about whether science is "getting worse" — what two real, competing explanations does it require you to weigh?

View solution

CHAPTER 9 QUICK REFERENCE

- The Surgisphere/hydroxychloroquine scandal (May-June 2020) saw two major papers retracted within ~13-34 days of publication, versus Wakefield's ~12-year gap
- COPE's three-tier framework — correction, expression of concern, retraction — reflects increasing severity; Surgisphere moved from expression of concern to full retraction in under 48 hours
- Retraction Watch (founded 2010, Oransky & Marcus) tracks over 30,000 real retractions in its public database
- Biomedical retraction rates roughly quadrupled 2000-2021 — driven by both genuinely more misconduct and genuinely better detection tools (e.g., PubPeer, 2012)
- A rising retraction rate is ambiguous evidence: it can reflect a worsening problem or a self-correction system working better than before

CHAPTER 10: CAPSTONE — DESIGNING AND CRITIQUING A REAL EXPERIMENT

Scientific Methodology

Chapter 10 · Capstone: Designing and Critiquing a Real Experiment

This capstone works in two directions. First, it designs a genuinely falsifiable experiment from scratch, applying every tool from Chapters 1-9 in the order a real researcher would actually need them. Second, it turns that same toolkit backward — onto the real Santa Clara antibody study from Chapter 8 — to ask a concrete question: which of these nine chapters' tools, applied *before* publication, would have caught the flaw that only became visible afterward?

Part 1: Designing a Falsifiable Experiment From Scratch

Working hypothesis: *does studying with background instrumental music change next-day recall test scores compared with studying in silence?*

STEP 1 — CH.1: STATE IT FALSIFIABLY

The hypothesis is framed in a form that could genuinely fail: "instrumental background music produces no significant difference in next-day recall scores compared with silence." A real result — in either direction — can disconfirm it, satisfying Popper's own falsifiability criterion.

STEP 2 — CH.2: DISTINGUISH HYPOTHESIS FROM THEORY

This single study tests one specific hypothesis. It is not, by itself, a theory — it would need to be one small, repeated piece of evidence feeding into a broader theory (e.g. cognitive load theory) only after independent replication, never treated as confirming that broader theory on its own.

STEP 3 — CH.3: DEFINE THE REAL VARIABLES

Independent variable: music vs. silence during study. **Dependent variable:** next-day recall test score. **Controlled:** identical study material, identical study duration, identical

test. **Control group:** the silence condition. **Randomization:** participants randomly assigned to condition, not self-selected.

STEP 4 — CH.4: BLIND WHAT CAN BE BLINDED

Participants can't be blinded to their own condition — they know whether they're hearing music. An honest limitation, not a flaw to hide. But the researcher scoring each recall test *can* be blinded to which condition each participant was in, preventing the same kind of expectation-driven scoring bias Chapter 4's placebo material covered.

STEP 5 — CH.6: DESIGN FOR REPRODUCIBILITY BEFORE RUNNING IT

Sample size is set in advance via a real power calculation, not by "running until significant" — the exact flexible-analysis pattern behind the Open Science Collaboration's own 36% replication rate. The full analysis plan is fixed before data collection begins.

STEP 6 — CH.7: PRE-SPECIFY WHAT WOULD COUNT AS FAILURE

What result would count against the hypothesis is written down in advance — and if the result comes back null or unexpected, no new, untested sub-group or post hoc statistical construct will be invented to rescue it, the exact ad hoc pattern that discredited the real Mars effect.

STEP 7 — CH.8: SUBMIT AS A REGISTERED REPORT

The introduction and methodology are submitted for real peer review *before* data collection — the Center for Open Science's own Registered Reports format, which measurably cut psychology's suspicious 96% "positive result" rate down to 44% by removing the incentive to chase a significant finding.

STEP 8 — CH.9: COMMIT TO PUBLISHING REGARDLESS OF OUTCOME

Because the study is a Registered Report, publication doesn't depend on the result — closing off the file-drawer problem before it can happen. In advance, the plan also names what would trigger a real correction (a fixable error), an expression of concern (a serious,

unresolved doubt), or a full retraction (a genuinely invalidated result) if a flaw were discovered after the fact.

Part 2: Critiquing a Real Experiment After the Fact

Chapter 8 covered the real Santa Clara antibody study (April 2020) in detail — a preprint claiming COVID-19 infections were undercounted 50-85x, criticized for not properly accounting for its antibody test's own real specificity uncertainty, and eventually revised and published in 2021 at a lower estimate. Applying this capstone's own nine-chapter toolkit backward asks: how much of that criticism could have been caught *before* the preprint ever reached the public?

TOOL	APPLIED TO SANTA CLARA IN ADVANCE
Ch.1 Falsifiability	The core claim was genuinely falsifiable — this was never the problem
Ch.3 Variable design	The test's own real false-positive rate wasn't fully incorporated into the reported confidence — a design-stage gap, not a data-collection error
Ch.6 Pre-registered analysis	The gap this chapter exists to close: no pre-registered plan specifying in advance how test-specificity uncertainty would be handled
Ch.7 No ad hoc rescue needed	To the study's credit, no post hoc rescue was invented once flaws were raised — the estimate was genuinely revised, not defended by new untested constructs
Ch.8 Registered Report	Not used — a regular preprint reached massive public and policy attention before independent statisticians could review the specificity math
Ch.9 Self-correction	Not retracted — revised and formally published in 2021 at a lower, still-elevated estimate, closer to a large-scale correction than a full retraction

THE REAL, CONCRETE FINDING

Had the Santa Clara design gone through a Registered Report process (Ch.8) — requiring the antibody test's own real specificity uncertainty to be pre-specified and reviewed in the analysis plan (Ch.6) before data collection — the exact criticism that eventually forced a public revision could plausibly have been raised and resolved at the design stage, before the preprint ever reached policy-relevant public attention. The tools this course covers aren't only useful in hindsight; several of them exist specifically to catch a problem like this one *before* it happens.

Chapter Attribution

CAPSTONE STEP	CHAPTER
Falsifiable hypothesis	Chapter 1
Hypothesis vs. theory	Chapter 2
Variables, control group, randomization	Chapter 3
Blinding	Chapter 4
Reflecting on paradigm resistance to a null result	Chapter 5
Pre-specified sample size and analysis plan	Chapter 6
Pre-specified falsification criteria, no ad hoc rescue	Chapter 7
Registered Report submission	Chapter 8
Publish-regardless-of-outcome, correction/EoC/retraction plan	Chapter 9

WHAT THIS COURSE DOESN'T COVER

This course covers the philosophy and mechanics of scientific methodology itself — demarcation, experimental design, paradigm change, reproducibility, and self-correction. It doesn't cover statistical analysis techniques in depth (see this Subject's own sibling course, Research Methodology, for source evaluation and study-design basics), formal logic and argument evaluation (see Critical Thinking), or the mathematics of probability and inference (see the site's own Maths for Programmers Subject). This course is the philosophy-and-practice layer underneath both sibling Study Methodologies courses, not a replacement for either.

COURSE COMPLETE — STUDY METHODOLOGIES SUBJECT

- **Research Methodology** (10 chapters) — sourcing, credibility, and study design fundamentals
- **Critical Thinking** (10 chapters) — everyday reasoning, fallacies, and cognitive biases
- **Scientific Methodology** (10 chapters) — demarcation, experimental design, paradigm shifts, reproducibility, and self-correction

- The Study Methodologies Subject is now complete: 30 chapters across three courses