Unpopular Science

DRAFT

David H. Silver

Unpopular Science

Exploring Curious Phenomena

David H. Silver

With illustrations by
Jessica DG Hayes

Publisher/Institution Name
First Edition • 2025
Printed on recycled electrons

Contents

Introduction

This book is not a popular science book. It is not a textbook. It is not an academic book. It is not even a chimera of the above.

It does share some goals with the three: to inspire wonder (like a popular science book), to include some rigour (like textbooks), and to introduce readers to phenomena that might challenge their understanding (as academic works often achieve).

As with some combinations, like a sushi-pizza restaurant, it excels at none. The main exposition isn't long enough for full understanding, the technical part is often too abstract or detailed to follow, and despite claims of rigour, von Neumann's line that *there's no sense in being precise when you don't know what you're talking about* fits too well.

But if my plan works, you'll get the appetite to leave this sushi-pizza diner. Maybe to a textbook because you're intrigued. Maybe to Wikipedia or blogs for more context about this fantastical world. That's the value: developing an appetite to dive deeper.

Few disclaimers.

- The book contains errors ranging from typos to wrong equations. Please report them, and be forgiving of mistakes. While precision is unrealizable, this serves as a more accurate guide to reality than popular science books.
- All chapters can be read independently. The **essence** is accessible to anyone, mostly in the chapter summaries. Some chapters are extremely mathematical and may not appeal to unfamiliar readers. The exponential map and four-dimensional spacetime chapters are the most mathematical.

Most topics have personal stories behind them — I remember how I learned about them. I hope I can infect you with some of that excitement.

Here we go.

DRAFT WARNING

This is a **very early draft**. Parts of it are placeholders. Some claims may be wrong.

Prologue

This book returns to the roots of scientific wonder, combining accessible explanations with rigorous mathematical foundations. Unlike contemporary science communication that oversimplifies or sensationalizes, it highlights the beauty of science as it truly is: both elegant and complex. The focus is understanding, not just exposure.

Too often, modern science communicators rely on a "laugh track" approach — telling readers how they should feel ("This is mind-blowing!") instead of letting wonder arise naturally from the ideas. This cheapens the experience, as though science requires manufactured excitement. Science doesn't need exaggeration; its wonder is self-evident to those who explore it properly.

This book contains 50 stories, each structured to guide readers from the intuitive to the profound. Here's how it will unfold:

Backdrop Each chapter begins with concise background — the people, circumstances, and discoveries behind the phenomenon. These stories ground readers in the scientific journey.

Phenomenon Description The phenomenon is described in straightforward terms, avoiding sensational language for clear, accurate explanations. We make concepts relatable while preserving depth — showing what makes something remarkable rather than declaring it "unbelievable."

Hardcore Analysis For readers ready to dive deeper, the third section provides rigorous academic analysis. Here, the mathematical and technical underpinnings of the phenomenon are laid bare, complete with equations, references, and detailed derivations. This section is unapologetically tough, offering readers the tools to validate the claims, explore further, or simply appreciate the true complexity of the science.

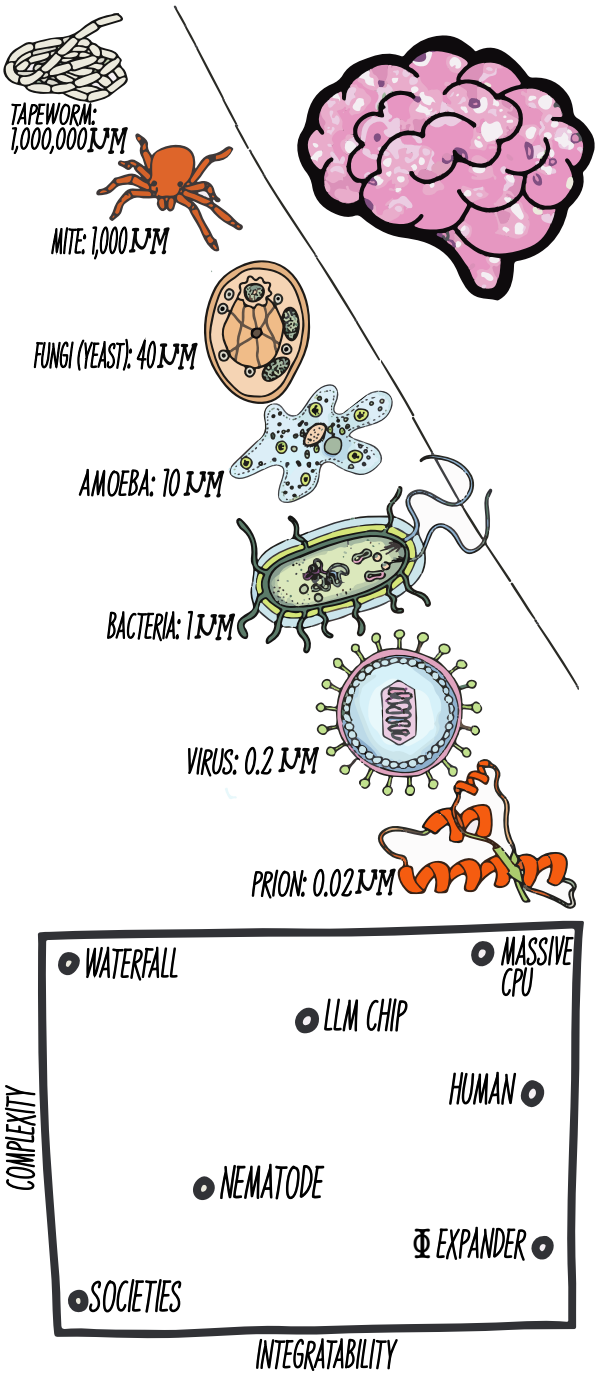
While the Hardcore Analysis is genuinely difficult, it remains essential. Like references in a scientific article, it's not necessary to grasp the main ideas, but it's the foundation on which everything stands. It provides scaffolding, justifies the clarity above it, and reminds us that simplified versions are built on layers of rigor.

The goal of this structure is to respect the reader's intelligence and curiosity. Whether discussing topological insulators, the mechanics of atomic clocks, or the subtleties of time dilation, the chapters will present science as it is: demanding, rewarding, and deeply inspiring.

This book counteracts oversimplified science communication. Science isn't slogans or easy answers — its complexity is a feature to celebrate. Understanding takes effort, but transforms fleeting curiosity into lasting enlightenment.

If you're ready to explore science in its full intellectual glory, I invite you to turn the page.

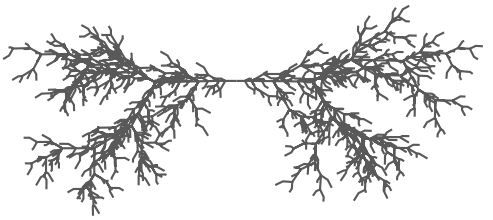
A Freely Willful Ignorance



Top (Scales of Infection and Pathogenic Agents): Infectious agents span eight orders of magnitude — from macroscopic parasites like tapeworms to sub-viral prions. Each has its own transmission pathway, lifecycle, and interaction mode with the host. Top-right: a brain, target of prion neurodegeneration.
Bottom (Post-Hoc Theories of Consciousness): Proposed mechanisms for consciousness map along a complexity–integrability plane: waterfall-like chaos, chip design, nematode circuits, human brains, and societal behavior. But all are descriptive, not explanatory. No theory derives subjective experience from first principles — only correlates it post hoc to system properties and for each example we can find synthetic constructs as a counter-example.

A Freely Willful Ignorance

Milligrams of propofol erase consciousness in seconds. Fatal familial insomnia prevents its cessation for months until death. While we can reliably toggle awareness, no unified mechanism explains why subjectivity vanishes. Consciousness cannot be reduced to neural correlates or fit by classifiers. Any attempt to locate its origin in physical mechanisms presupposes the very phenomenon under study. Free will and physics appear incompatible, but the standoff is asymmetric: agency is the lived fact that makes physics construction possible. Consciousness occupies the apex of a revision hierarchy where, in any conflict with lower-level descriptions, the knower must prevail.



Consciousness & Anesthesia ◦ General Anesthetic Mystery ◦ Multiple Molecular Mechanisms ◦ Fatal Familial Insomnia ◦ Free Will vs Physics ◦ First-Person Experience ◦ Libet Readiness Potential ◦ Revision Cost Hierarchy ◦ Cogito Ergo Sum ◦ Direct Self-Knowledge ◦ Hard Problem

الاعتقاد ليس هو المعنى المقصود ، بل المعنى المتصور في النفس

"Belief is not the intended meaning, but the meaning conceived in the soul."

— Maimonides, circa 1191 CE

"Reason will prevail."

— The Gang, 2005

A Freely Willful Ignorance

Ether-era anesthesia began in 1846 with Morton's public demonstration in Boston; within months, ether and chloroform spread worldwide. By the early 20th century, Meyer and Overton independently observed a correlation: anesthetic potency scaled with lipid solubility across diverse compounds. This supported the idea that consciousness could be turned off by a nonspecific action on neuronal membranes. Yet the correlation cracked under scrutiny: highly lipophilic yet inert molecules failed to anesthetize, while effective agents deviated from the predicted potency.

Mid-to-late 20th century work shifted toward specific molecular targets. Volatile agents were shown to prolong inhibitory currents at $GABA_A$ receptors, while nitrous oxide and ketamine disrupted glutamatergic signaling via NMDA antagonism. Parallel findings implicated two-pore K^+ (K2P) channels and hyperpolarization-activated cyclic nucleotide–gated (HCN1) currents in setting neuronal excitability under anesthetics. Still, no single pathway unified the class.

In prion disease, a different historical thread exposed the opposite failure mode. In 1986, Prusiner proposed prions — proteinaceous infectious particles — as agents of neurodegeneration. A rare PRNP mutation producing fatal familial insomnia (FFI) was later traced to selective thalamic degeneration, abolishing sleep despite otherwise preserved wakeful function. The San Giovanni pedigree in Italy provided the defining clinical arc: onset with fragmented sleep, inexorable insomnia, autonomic failure, cognitive collapse, and death within months. Where anesthesia induced obliviousness, FFI prevented it.

General anesthesia abolishes subjectivity itself. Other drugs alter perception, mood, or pain. Anesthetics suspend the condition for all perception and mood. A standard intravenous dose of propofol — two milligrams per kilogram — eliminates awareness in less than a minute. The transition is sharp. One moment the subject tracks voices and surroundings; the next moment there is no report, no continuity of thought, and no subsequent memory. The effect is reliable, reversible, and indispensable to surgical practice. Yet it remains unexplained.

Different drugs converge on this endpoint through divergent and sometimes contradictory mechanisms. Propofol potentiates γ-aminobutyric acid type A ($GABA_A$) receptors, amplifying inhibitory currents and reducing excitability across the cortex. Isoflurane, sevoflurane, and other volatile anesthetics bind to potassium and sodium channels, producing generalized dampening of neuronal firing. Nitrous oxide and xenon inhibit N-methyl-D-aspartate (NMDA) receptors, reducing excitatory drive. Ketamine blocks NMDA receptors yet increases cortical activity globally, producing electroencephalographic patterns closer to wakefulness than sleep while still abolishing awareness. Distinct molecular actions — some silencing neurons, some exciting them — terminate consciousness with similar reliability.

The search for an underlying model for general anesthesia once looked promising. At the turn of the twentieth century, Hans Meyer and Charles Ernest Overton noted a correlation: anesthetic potency scales with lipid solubility. The Meyer–Overton rule suggested that anesthetics dissolved into neuronal membranes, altering their physical properties. For

decades this correlation dominated, reinforced by its simplicity. Yet the correlation broke. Non-immobilizers — molecules with high lipid solubility — fail to anesthetize. Others, poorly soluble in fat (such as etomidate), work effectively. The membrane theory could not account for exceptions.

The focus moved to receptors. Different anesthetic classes bind to distinct proteins: $GABA_A$, NMDA, and two-pore domain potassium channels among prime candidates. Yet receptor theories also encounter anomalies. No single target is necessary. Mice engineered with $GABA_A$ subunits resistant to volatile anesthetics still lose consciousness when exposed. No single target is sufficient: receptor agonists or antagonists with precise effects on candidate pathways often fail to produce general anesthesia. What remains is a map of partial correlates, not a law specifying why awareness vanishes.

Network hypotheses move up a level. Thalamic "switch-off" models propose that sensory relay and intralaminar nuclei disengage cortical broadcasting. Alternatives hold that long-range cortico-cortical integration degrades: effective connectivity fragments, ignition-like reverberation collapses, and fronto-parietal synchrony decouples. Empirically, anesthetic depth tracks changes in spectral power, complexity, and coherence. But counterexamples persist. Ketamine increases cortical activity and high-frequency power yet abolishes consciousness. Dexmedetomidine reduces thalamic throughput yet permits vivid dreams.

The opposite extreme clarifies the boundary. Fatal familial insomnia, a prion disease, destroys neurons in the thalamus, especially in the anteroventral and mediodorsal nuclei. These nuclei regulate sleep architecture. As they degenerate, the subject loses the ability to enter non-rapid eye movement sleep. Ordinary fatigue accumulates, but sleep never arrives. Patients remain in escalating wakefulness until death, often within a year of symptom onset. Other neurodegenerative diseases erase memory, language, or motor control. This one erases the possibility of unconsciousness. Consciousness persists compulsively until the body collapses under uninterrupted wakefulness.

Anesthesia and prion disease bracket the same mystery. Milligrams of a synthetic molecule suspend awareness entirely. Widespread neuronal loss fails to interrupt it. Consciousness is too easy to subtract and, simultaneously, impossible to eliminate. This indicates that manipulations reach only the conditions under which consciousness manifests. They do not specify what consciousness is. Practitioners can toggle the switch without knowing what is being switched.

Measuring consciousness remains harder than turning it off. Clinical scales rely on responsiveness; neurophysiology adds proxies: cross-regional EEG (Electroencephalography) coherence, perturbational complexity from TMS-evoked (transcranial magnetic stimulation) responses, and theoretical constructs like Integrated Information Theory's Φ. Each stumbles. Some unresponsive patients process speech. High Φ can be assigned to systems with no plausible subjectivity. EEG signatures of wakefulness can appear under amnestic sedation. Competing theories — Global Workspace, Integrated Information, Recurrent Processing — disagree on what makes a state conscious, and experiments often adjudicate proxies rather than experience itself.

The working picture is that multiple molecular routes converge on a few network-level

motifs — reduced ignition, impaired integration, altered thalamocortical gating — sufficient to block access to a reportable workspace. That picture explains much of practice and little of essence.

The gap between control and understanding demands a different frame entirely. Consciousness is singular. Treating it as a parameter vector to be fit by a support-vector machine or a deep network condescends to the phenomenon. A classifier extracts invariants and separates classes. Consciousness is first-personal presence and deliberative control. No change of basis, no margin optimization, no loss function turns one into the other. The distinction is categorical.

Any research program that seeks to locate the origin of consciousness in physical mechanisms presupposes the very phenomenon it attempts to explain. The attempt uses consciousness to investigate consciousness. You deploy attention, select among hypotheses, compare results, and conclude. Each of those acts exercises the thing under study. The circularity constitutes an epistemological wall. Every explanation collapses back into the standpoint it aimed to eliminate.

Free will and physics appear incompatible. If physics is a complete description — deterministic or stochastic, local or quantum, simulated or fundamental — then every decision reduces to a trajectory in state space. Free will becomes an illusion, a narrative that complex systems tell themselves about their own deterministic unfolding. But if free will exists, then physics is either incomplete or itself simulated. The standoff seems symmetric: pick your side.

The symmetry is false. Free will is the lived fact. Physics is the constructed model. If physics denies free will, physics has misclassified its own status. Constructing, testing, and revising physical theories requires a subject that directs thought, selects among candidate explanations, and exercises judgment. To declare that subject an illusion saws off the branch on which the declaration sits. Illusions presuppose a subject that misperceives. If the subject is deleted, the word ”illusion” loses reference. The sentence ”free will is an illusion” requires a subject that can contrast seeming with being. That requirement reinstates free will.

Superdeterminism attempts to dissolve the conflict by denying the independence of measurement choices. In this view, the experimenter’s decision to measure spin-up versus spin-down correlates with the particle’s prior state through a common past. Bell’s theorem assumes measurement settings can be freely chosen. Superdeterminism rejects this assumption by claiming that every choice traces back to initial conditions that also determined the particle’s properties. The loophole preserves determinism at the cost of denying that experimenters select their measurements. Yet the superdeterminist must still choose which papers to write, which theories to propose, which objections to raise. Experiencing the act of advocating superdeterminism, exercises the agency that superdeterminism denies.

Neuroscience experiments probe the timing of conscious will. Benjamin Libet (1985) measured electrical readiness potentials (RP) beginning 550 milliseconds before subjects reported awareness of their intention to move. The brain initiates action before conscious decision registers. Subsequent experiments refined this: Schurger (2012) showed that RP

reflects general motor preparation rather than specific decision; Fried (2015) recorded individual neurons firing up to 1.5 seconds before reported awareness. Brain activity predicts choice before the subject knows what they will choose.

These findings constrain but do not eliminate agency. The readiness potential precedes awareness of specific intention, not the capacity for veto. Libet himself noted that subjects retain "free won't" — the ability to cancel incipient actions after becoming aware of them. More fundamentally, experimental paradigms that measure spontaneous movements capture only a subset of willing. Deliberative decisions — weighing options, comparing outcomes, selecting among complex alternatives — unfold over seconds to hours, not milliseconds. The neuroscience of snap judgments does not generalize to the neuroscience of reflection. Lastly, but most importantly, the timing of awarness of agency isn't necessarily the same as the timing of the will itself.

All of this however, is moot, as we don't need the experiment to tell us that free will is incompatible with physics - not merely with the current standard model, but with any logically consistent model of the universe (i.e. causal models).

Consciousness is the exercise of will on one's own stream of thought. Hold, release, redirect, compare, adopt, reject. Deliberate selection among candidate continuations. The stream is the ordered sequence of contents available for such selection. The subject is the locus at which selection is enacted.

We define commitment as the act of believing in a proposition. In order to be able to rank which commitments prevail when there is a conflict between them, we will do so by analyzing the revision cost of different commitments. We define a revision cost as the loss of the apparatus of knowing incurred by abandoning a commitment.

Let's rank several commitments by revision cost.

I know the sky is blue. If tomorrow I learn it is an optical illusion — scattering, refraction, atmospheric tricks — fine. Mildly interesting. Nothing essential breaks.

I know there is gravity. If someone pulls the plug and reveals the simulation, forces redraw, mass no longer bends spacetime — I am stunned for days. But then I rebuild the catalog of causes and move on. The capacity to model persists.

I know there are atoms and electrons. If the simulation ends and someone unplugs me from "The Matrix", mind blown. Weeks to recover. But recovery is possible. I can still compare, infer, and correct.

I know $2 + 3 = 5$. If someone demonstrates that arithmetic itself is wrong — that I had a cognitive shortcut, and really $2 + 3 = 11$ — the machinery of thought disassembles. Counting, comparison, consistency all rest on that foundation. Without it, reasoning collapses.

I know I have free will. I know I exist as the thing that directs its own thoughts. If this turns out to be false — then there is no "I" left to register the failure. Incompatible with the standpoint from which acceptability is judged.

The highest commitment dominates. Every statement, inference, or model presupposes a

subject that can assert, doubt, compare, and revise. That presupposition is the content of the highest tier. Lower tiers describe states of affairs in the world. The highest tier secures the existence of the knower to whom the world appears. In any conflict, the knower wins. Without the knower, conflict is unintelligible.

Write the revision cost as $C(\cdot)$. Then:

$$C(\text{appearances}) \ll C(\text{physics}) \ll C(\text{mathematics}) \ll C(\text{agency}).$$

The last inequality is decisive. If agency conflicts with physics, agency prevails. Agency is the condition for there being importance at all.

Neural correlates, receptor binding, thalamic gating, and network fragmentation describe *when* consciousness appears or vanishes and *how* physiology couples to report. That scope is exact and valuable. *What it is to be* the subject for whom appearance and vanishing matter lies elsewhere. "When does awareness switch off?" asks about timing and mechanism. "What is it to direct one's own thought?" asks about the standpoint that makes timing intelligible. Neuroscience answers the first. Philosophy addresses the second. Conflating them produces the reduction error: mistaking access conditions for the subject to whom access matters.

Research that maps brain states to behavioral outputs achieves correlation. Intervention studies that disrupt nodes and track changes achieve mechanism. Both are genuine progress. Constitution — the precondition without which correlation and mechanism cannot be stated — remains distinct. Consciousness sits at the constitutional level. Adding more parameters or finer imaging cannot bridge the gap. The gap is categorical.

Anesthesia deletes awareness in seconds. Fatal familial insomnia prevents its deletion for months. Both manipulate conditions. Neither touches essence. We can flip the switch without knowing what is being switched. The circuit diagram remains incomplete because the subject that the diagram aims to explain is also the subject examining the diagram. That reflexive loop cannot be broken by adding variables. It can only be acknowledged. Consciousness occupies the apex of the revision hierarchy, where it cannot be dissolved into the mechanisms it observes. Attempting dissolution commits category error at the highest cost.

Agency as Axiomatic Ground

Epistemic Formalism

Let S be a knowing subject, $\mathcal{P}$ the set of propositions, and $K_S \subseteq \mathcal{P}$ the commitment set of propositions S holds true. For $p, q \in K_S$, write $p \vdash q$ if q logically follows from p.

Define revision cost:

$$C(p) := |\{q \in K_S \mid p \vdash q\}|$$

This induces partial order $(K_S, \preceq)$ where $p \preceq q \iff C(p) \leq C(q)$.

Hierarchy with Revision Costs

- p_1: "Sky is blue"
  If false: Mildly interesting. Nothing breaks.
- p_2: "Not in The Matrix"
  If false: Stunned. Rebuild ontology. Days to recover.
- p_3: "Gravity exists"
  If false: Physics rebuilds. Weeks to recover.
- p_4: "$2 + 3 = 5$"
  If false: Arithmetic collapses. Reasoning disassembles.
- p_5: "$P \vee \neg P$" (excluded middle)
  If false: Logic fails. Cannot reason about contradictions.
- A: "I direct my thought"
  If false: No subject remains to register the failure.

Strictly: $C(p_1) \ll C(p_2) \ll C(p_3) \ll C(p_4) \ll C(p_5) \ll C(A)$.

Agency as Maximal Element

Agency (A): capacity to perform operations on K_S (selecting, comparing, affirming, rejecting propositions). This is control over thought, not physical action.

To revise K_S by removing A requires performing an operation on K_S, which presupposes A. Thus revision of A is self-undermining:

$$A \vdash p \quad \forall p \in K_S \quad \Rightarrow \quad C(A) = |K_S|$$

Agency is the maximal element in $(K_S, \preceq)$.

Resolution: Physics vs. Agency

Let P_{phy} = "Universe fully determined by physical law." If $P_{\text{phy}} \vdash \neg A$, we face contradiction: $\{P_{\text{phy}}, A, P_{\text{phy}} \to \neg A\}$.

Comparing revision costs:

- $C(P_{\text{phy}})$: finite. Subject persists, rebuilds model.
- $C(A)$: total. No subject remains.

Since $C(P_{\text{phy}}) < C(A)$, agency dominates. Physical models invalidating agency are epistemologically incoherent.

Category Error Formalized

Let $\mathcal{C}_{\text{phys}}$ be category of physical processes with morphisms as causal relations. Let $\mathcal{C}_{\text{agent}}$ be category of first-person agency with morphisms as deliberative acts.

No structure-preserving functor $F : \mathcal{C}_{\text{phys}} \to \mathcal{C}_{\text{agent}}$ exists satisfying:

$$F(\text{correlation}) = \text{constitution}$$

Classifiers extract invariants; subjects exert control. These are distinct logical types. Classification $\ncong$ first-personal presence.

Philosophical Grounding

Descartes' Cogito ergo sum (1641): The act of doubting presupposes the existence of a doubter. Even radical skepticism cannot eliminate the thinking subject. This establishes the subject as the foundation of knowledge, not a conclusion derived from it.

Kant's Transcendental Apperception (1781): The unity of consciousness is not empirically observed but is the logical precondition for any structured experience. The "I think" must accompany all representations. Without a unified subject, no comparison, judgment, or synthesis of data is possible.

Modern Parallel: These establish that agency is axiomatic, not a theorem derived from lower-level descriptions. The subject is the condition for there being theorems, descriptions, and derivations at all.

References:

Descartes, R. (1641). Meditations on First Philosophy.
Kant, I. (1781/1787). Critique of Pure Reason.