Week 5
Understanding under Threat

Quantum Mechanics and Artificial Intelligence

Hans Halvorson

May 2026

Today's plan

Part I: Quantum Mechanics *(lectures 1–2)*

- The framework of QM: states, observables, dynamics
- The measurement problem and the Copenhagen myth
- Interpretations: GRW, Bohm, Everett, ...
- Complementarity: Bohr's response to the crisis
- Bell's theorem and the limits of causal understanding

Part II: Artificial Intelligence *(lecture 3)*

- Models and understanding: what philosophy of science says
- The double meaning of "model"
- Sullivan's argument: link uncertainty vs. implementation opacity
- A deeper challenge: is there an inside?

Group Discussion

From last week, quantum mechanics is our recurring example of a theory where prediction and understanding have come apart.

1. Do you feel that you **understand** quantum mechanics? What would it even mean to understand it?
2. Is "shut up and calculate" a satisfying attitude toward a fundamental theory of nature?

Table of Contents

Two perspectives today

First-order Which interpretation of QM is correct? What is the theory *actually* saying about the world?

Second-order There has been no consensus on the foundations of QM for *100 years*. Should we be concerned? What does this tell us about science as a knowledge-producing enterprise?

Both questions bear directly on scientific understanding — which is why De Regt and Dieks chose quantum non-locality as their motivating example.

The foundational crisis: 1900–1935

- 1900: Planck introduces the quantum of action h
- 1905: Einstein explains the photoelectric effect using photons
- 1913: Bohr's model of the hydrogen atom
- 1925–26: Heisenberg's matrix mechanics; Schrödinger's wave mechanics
- 1927: Born's probability interpretation; Heisenberg's uncertainty principle; Bohr's complementarity
- 1935: The EPR argument — QM must be incomplete

The early pioneers were acutely aware that they did *not* understand what was going on. Bohr thought our ability to understand the physical world was genuinely up for grabs.

Apparent resolution: the Copenhagen school

- By 1927 a working consensus formed around Bohr's school
- Bohr's key insight: the existence of the **quantum of action** calls for a *revision* of what it means to give an objective, unambiguous description of physical phenomena
  - An unambiguous description must specify the experimental arrangement; the conditions of observation become part of the description itself
  - Classical concepts remain indispensable for describing the apparatus, but cannot all be simultaneously applied to the quantum object
- This leads to **complementarity**: a structural feature of unambiguous description, not a limitation on knowledge
- This is a positive philosophical proposal, not a counsel of despair

Revival of the crisis

- 1952: Bohm publishes a working hidden-variable interpretation — shattering the apparent consensus
- 1957: Everett proposes the relative-state (many-worlds) interpretation
- 1964: Bell proves his non-locality theorem
- Today: active foundational debate; the cutting edge of physics proceeds without settling the foundations

The second-order question: is 100 years without consensus acceptable? Does it reveal something deep about QM, or is it a sociological failure of physics?

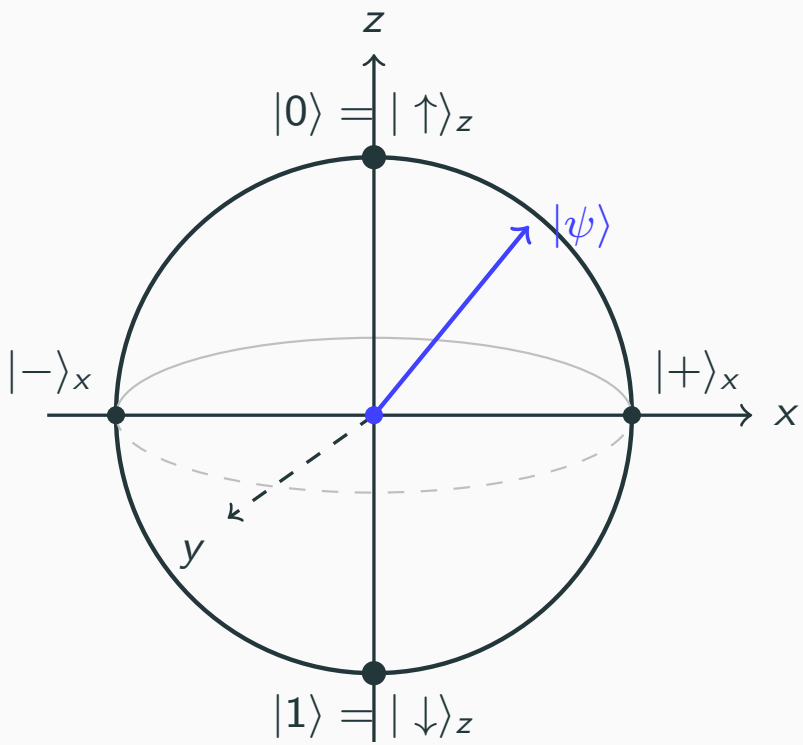
Your reading: Weinberg (2017) — a Nobel laureate expressing exactly this second-order concern from inside physics.

Table of Contents

States as vectors

- The state of a quantum system is a unit vector $|\psi\rangle$ in a complex Hilbert space $\mathcal{H}$
- **Superposition principle**: if $|\psi\rangle$ and $|\phi\rangle$ are states, so is $\alpha|\psi\rangle + \beta|\phi\rangle$ (after normalization)
- For a **qubit** (two-level system): $\mathcal{H} = \mathbb{C}^2$
- Computational basis: $|0\rangle \equiv |\uparrow\rangle, \quad |1\rangle \equiv |\downarrow\rangle$

The Bloch sphere



Every pure state of a qubit is a point on the sphere. Antipodal points are orthogonal.

Observables as operators

- Physical quantities: **Hermitian operators** $A = A^\dagger$
- Possible measurement outcomes: eigenvalues λ_i
- Eigenstates $|a_i\rangle$: states with definite value λ_i

Spin operators (Pauli matrices $\times \hbar/2$):

$$S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

Superpositions and the puzzle

If the system is in $|\psi\rangle = \alpha|a_1\rangle + \beta|a_2\rangle$, observable A has *no definite value* — yet measurement always yields a definite outcome λ_1 or λ_2.

The puzzle: what is the system doing between preparation and measurement? Does the question even make sense?

This tension is the engine of every interpretation of QM.

Uncertainty relations

Non-commuting observables cannot both have definite values:

$$[S_y, S_z] = i\hbar S_x \quad \Longrightarrow \quad \Delta S_y \cdot \Delta S_z \;\geq\; \tfrac{\hbar}{2}|\langle S_x \rangle|$$

Spin-z eigenstate decomposed in spin-y basis:

$$|\uparrow\rangle_z \;=\; \tfrac{1}{\sqrt{2}}|\uparrow\rangle_y + \tfrac{1}{\sqrt{2}}|\downarrow\rangle_y$$

so $\Delta S_y = \hbar/2$ is maximal when S_z is definite.

Conceptually: **Fourier analysis** — narrow in one basis, spread in the conjugate basis.

Composite systems and entanglement

Composite systems: $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, with $x \otimes (y + z) = (x \otimes y) + (x \otimes z)$.

Not all states are product states. The **singlet state**

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}\big(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle\big)$$

cannot be written $|\psi\rangle_A \otimes |\phi\rangle_B$: it is **entangled**.

Discussion (2 min): what features does a system have if it is entangled with another? Can we assign a definite state to particle A alone?

Linear dynamics

Time evolution is governed by the **Schrödinger equation**:

$$i\hbar \frac{d}{dt}|\psi\rangle = H|\psi\rangle$$

The solution is **unitary**: $|\psi(t)\rangle = U(t)|\psi(0)\rangle, \quad U^\dagger U = \mathbf{1}$.

Crucially, U is **linear**:

$$U\big(\alpha|\psi\rangle + \beta|\phi\rangle\big) = \alpha U|\psi\rangle + \beta U|\phi\rangle$$

This linearity is exactly what generates the measurement problem.

Spectral decomposition and the Born rule

Every Hermitian operator A has an orthonormal eigenbasis $\{|a_i\rangle\}$ (**spectral theorem**):

$$A = \sum_i \lambda_i |a_i\rangle\langle a_i|$$

Any state expands as $|\psi\rangle = \sum_i c_i |a_i\rangle, \quad c_i = \langle a_i|\psi\rangle$.

Born rule: the probability of outcome λ_i when measuring A on $|\psi\rangle$:

$$\Pr_\psi(A = \lambda_i) = |c_i|^2 = |\langle a_i|\psi\rangle|^2$$

The rule extracts the squared component of $|\psi\rangle$ along each eigenstate. Sanity check: if $|\psi\rangle = |a_k\rangle$ already, then $c_k = 1$ and all others vanish — outcome λ_k with certainty.

Discussion: do we *understand* QM if it is just a recipe for computing these probabilities?

Table of Contents

Definition and derivation

Ideal measurement of A: $|a_i\rangle \otimes |r\rangle \longrightarrow |a_i\rangle \otimes |y_i\rangle$ (each eigenstate paired with a distinct pointer state — unitary).

Now suppose $|\psi\rangle = \sum_i c_i |a_i\rangle$. By **linearity**:

$$\left(\sum_i c_i |a_i\rangle\right) \otimes |r\rangle \longrightarrow \sum_i c_i \left(|a_i\rangle \otimes |y_i\rangle\right)$$

The meter ends in an **entangled superposition** — no definite reading. But we *always* observe a definite outcome.
Contradiction.

The collapse postulate

Standard "solution": $\sum_i c_i|a_i\rangle \longrightarrow |a_k\rangle$ with probability $|c_k|^2$.

Problems:

- Violates unitarity and linearity — the two most fundamental features of the theory
- Creates a two-dynamics theory: Schrödinger evolution *and* collapse
- When does collapse occur? Why does it occur?
- The theory gives no definition of "measurement"

The Copenhagen myth

- Von Neumann (1932) first axiomatized collapse in *Grundlagen der Quantenmechanik*; by the 1950s it was standard textbook orthodoxy
- The label **Copenhagen interpretation** was applied retroactively — Bohr never used it
- In reality there was no single Copenhagen view: Bohr, Heisenberg, Born, Pauli all held different positions
- **HH:** Bohr never claimed the measurement problem is *solved* by collapse — his actual response is philosophically much deeper, and has been systematically misread

Decoherence: not a solution

Decoherence: environmental interaction entangles the system rapidly; the *reduced* state approaches a mixture.

Why this does not solve the measurement problem:

1. The reduced state is not *uniquely* interpretable as a mixture over definite pointer readings
2. The global state (system + environment) is still a superposition — decoherence is a local approximation, not a collapse

Decoherence explains why interference is hard to observe macroscopically. It does not explain why we observe *one* outcome.

Table of Contents

Four strategies

1. **Reject unitary dynamics**: add stochastic collapse *(GRW)*
2. **Add hidden variables**: supplement QM with additional ontology *(Bohm)*
3. **Reject uniqueness of outcomes**: all outcomes occur and the observer splits *(Everett)*
4. **Revise the ontological picture**: resist the demand for a classical reality behind QM *(relational, QBism, pragmatic)*

GRW, Bohm, and their problems

GRW (Ghirardi, Rimini, Weber 1986): add random spontaneous localization events. Free parameters tuned so macroscopic objects localize (no pointer superpositions) while microscopic systems remain unitary. Open: ontology? Relativistic extension?

Bohm (1952): both wavefunction and particle positions are real; the wave guides particles. Particles always have definite positions.

- Guidance equation requires a preferred frame — inconsistent with special relativity
- Violates the spirit of momentum conservation (no particle moves inertially)

Everett / many worlds

Everett (1957): take unitary dynamics completely seriously; no collapse; all branches of the wavefunction are equally real.

We observe one branch because we ourselves are part of the wavefunction — each measurement splits the world.

Problems:

- *Probability:* the theory is deterministic — where do the Born-rule probabilities come from?
- *Preferred basis:* what selects the "world" branches?
- No-hidden-variables theorems imply the branches cannot themselves be classical

HH: these approaches are philosophically sophisticated but may relocate rather than dissolve the problem.

Less ontological interpretations

Relational QM (Rovelli 1996): quantum states are always *relative to an observer* — there is no observer-independent quantum state.

The inspiration is special relativity: velocity has no absolute value, only values relative to a reference frame. Rovelli proposes the same for quantum states. When Alice measures a particle and gets a definite outcome, that outcome is real *relative to Alice*. The particle may still be in superposition relative to Bob, who has not interacted with it. Both descriptions are equally valid. The measurement problem is *dissolved*: collapse is real, but only relative to an observer, never absolute.

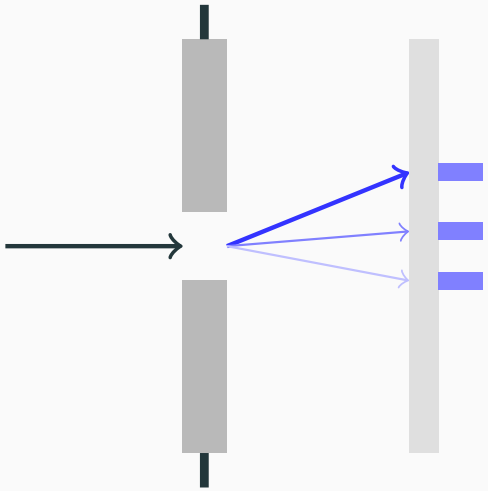
QBism (Fuchs–Schack): quantum states encode an agent's *beliefs* about future experiences — the wavefunction is epistemic, not ontic.

Pragmatic (Healey): QM provides a tool for updating credences, not a description of a quantum reality.

Table of Contents

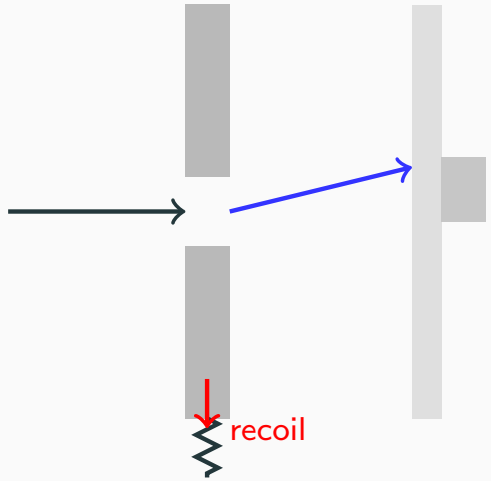
Bohr's single-slit thought experiment

A: bolted screen



position known $\Rightarrow$ spacetime description

B: free screen



recoil *measured* $\Rightarrow$ causal description

Complementarity and visualizability

Two arrangements are **complementary** when they are mutually exclusive and each exhaustive within its domain.

Spacetime description position, trajectory, "which path" — **visualizable** (anskuelig) picture

Causal description momentum, energy, conservation laws — dynamical picture

"The viewpoint of complementarity presents itself as a rational generalization of the very ideal of causality." (Bohr 1948)

This connects directly to De Regt & Dieks: Bohr is *not* saying we cannot understand quantum phenomena. He is proposing a new kind of understanding that does not require a single classical picture.

Bohr's deeper point

- The quantum of action requires a revision of what counts as an **objective description**:
  - Objectivity does not disappear — but it now requires specifying the experimental conditions as part of any unambiguous statement about a quantum phenomenon
  - We cannot separate "how Nature is" from "the experimental arrangement in which it is observed"
- This is not about abandoning objectivity but about *reconceiving* it for a domain where the interaction between object and instrument cannot be neglected or controlled
- The misreading of Bohr as "shut up and calculate" has distorted the foundations debate for 70 years
- **HH:** Bohr is working in a tradition running from Kant through Høffding: the conditions of unambiguous description are themselves a philosophical achievement

Table of Contents

EPR, von Neumann, and Kochen-Specker

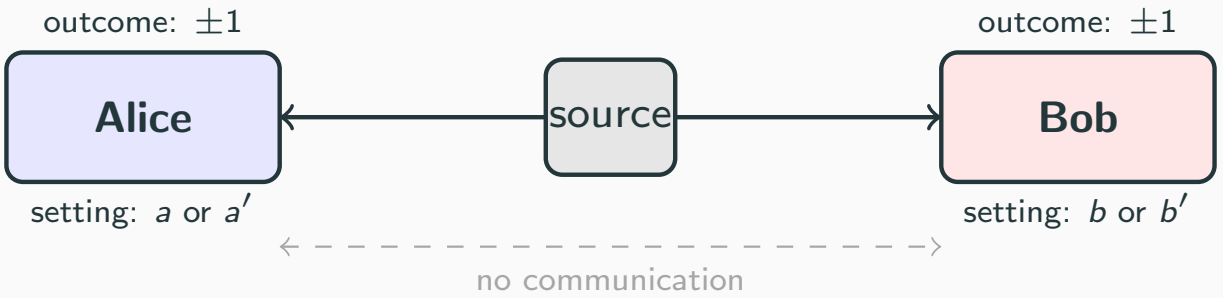
EPR (1935): QM is incomplete — quantum correlations must be explained by predetermined hidden variables λ.

Von Neumann (1932): no hidden-variable theory can reproduce QM given additivity $\mathbb{E}_\lambda(A+B) = \mathbb{E}_\lambda(A) + \mathbb{E}_\lambda(B)$ for non-commuting A, B. Bell (1966): this assumption is physically unmotivated.

Kochen–Specker (1967): no assignment $v(A) \in \mathrm{spec}(A)$ to all observables can satisfy functional consistency $v(f(A)) = f(v(A))$ and the sum rule for commuting pairs. **Value definiteness** is impossible.

Kochen–Specker is more compelling than von Neumann: the assumptions are operationally motivated.

The CHSH scenario



The entangled pair is measured far apart. Each party independently chooses a setting and records a ± 1 outcome.

Local hidden variables and the CHSH inequality

Suppose outcomes are determined by $\lambda \sim \rho(\lambda)$, with **reality** ($A, B$ definite) and **locality** (Alice independent of Bob's setting). Define the **correlation** between outcomes:

$$E(a,b) = \int A(a,\lambda)\, B(b,\lambda)\, \rho(\lambda)\, d\lambda$$

i.e. the expectation value of the product of Alice's and Bob's outcomes (± 1 each) for settings a and b. Perfect correlation: $E = +1$; perfect anti-correlation: $E = -1$; no correlation: $E = 0$.

Under local hidden variables:

$$\left| E(a,b) - E(a,b') + E(a',b) + E(a',b') \right| \;\leq\; 2$$

Proof sketch: for each λ, the combination $A(a)[B(b) - B(b')] + A(a')[B(b) + B(b')]$ takes values in $\{-2, 0, +2\}$.

Quantum violation

For $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$: $\quad E(a,b) = -\cos(\theta_a - \theta_b)$

Choose $a = 0°$, $a' = 90°$, $b = 45°$, $b' = 135°$:

$$E(a,b) = -\cos(45°) = -\frac{1}{\sqrt{2}}$$

$$E(a,b') = -\cos(135°) = +\frac{1}{\sqrt{2}}$$

$$E(a',b) = -\cos(45°) = -\frac{1}{\sqrt{2}}$$

$$E(a',b') = -\cos(45°) = -\frac{1}{\sqrt{2}}$$

$$S = -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} = -\frac{4}{\sqrt{2}} = -2\sqrt{2}$$

$$\Rightarrow \quad \boxed{|S| = 2\sqrt{2} \approx 2.83 > 2}$$

The Jarrett analysis and what Bell means

Jarrett (1984): local hidden variables = two conditions: **PI (parameter independence)** Alice's outcome does not depend on Bob's *setting*

OI (outcome independence) Alice's outcome does not depend on Bob's *outcome*

Bell violation $\Leftrightarrow$ at least one of PI or OI fails.

- **Orthodox (collapse):** OI fails — collapse makes Alice's probability depend on Bob's *outcome*
- **Bohm:** PI fails — the quantum potential makes Alice's probability depend on Bob's *setting*
- **Everett:** unclear — no unique outcomes, so the framework is hard to apply. (Everettians typically *claim* the theory is local — but making this precise is notoriously difficult.)
- **Retrocausal models:** PI fails — future settings influence

Taking stock — and a question

- QM achieves extraordinary predictive precision: the anomalous magnetic moment of the electron agrees with experiment to 12 significant figures
- Yet the theory is arguably *inconsistent as stated* (the measurement problem), and after 100 years there is still no consensus on how to resolve it
- Each interpretation pays a serious price: non-locality, many worlds, anti-realism, new fundamental dynamics

Some physicists say: "QM works, so who cares if we don't understand it?" If we accept this, we have weakened our ground against:

> *"AlphaFold predicts protein structures better than any human theory — so who cares if we don't understand it?"*

HH: The goal of science is not mere prediction but understanding.

Table of Contents

The New York Times

Graduates Boo Commencement Speech About A.I.

Humanities students made their displeasure known at the University of Central Florida.

Listen · 5:54 min

Share full article

Students at the University of Central Florida booed a commencement speaker after she said that "artificial intelligence is the next industrial revolution." University of Central Florida

AI and the battle for knowledge

The AI tools entering universities are not neutral: they serve corporate interests, not the growth of knowledge.

Digital feudalism: the tools of intellectual work come under the control of a few corporations, creating hard-to-exit dependencies and extracting value from those who depend on them.

There is a genuine conflict of interest:

- **Universities** exist to produce and transmit knowledge; they require critical independence
- **Tech corporations** exist to generate returns on capital; they benefit from opacity and dependency

The students who booed were not merely being reactionary. They were intuiting something real.

Group Discussion

1. Has AI helped you **understand** something you did not understand before? Or has it only helped you **produce** things more efficiently?
2. Do you see a genuine prospect of AI helping science achieve deeper understanding — or are we being sold a lie?

Two perspectives on AI in science

Theoretical Science has always used tools to extend understanding. But AI tools seem borderline *intelligent* themselves. Does that change what the task of science *is*?

Practical Science produces two kinds of good: **knowledge** about the world, and **power** over the world. AI threatens to tip the balance away from knowledge.

What kind of science do we want to do?

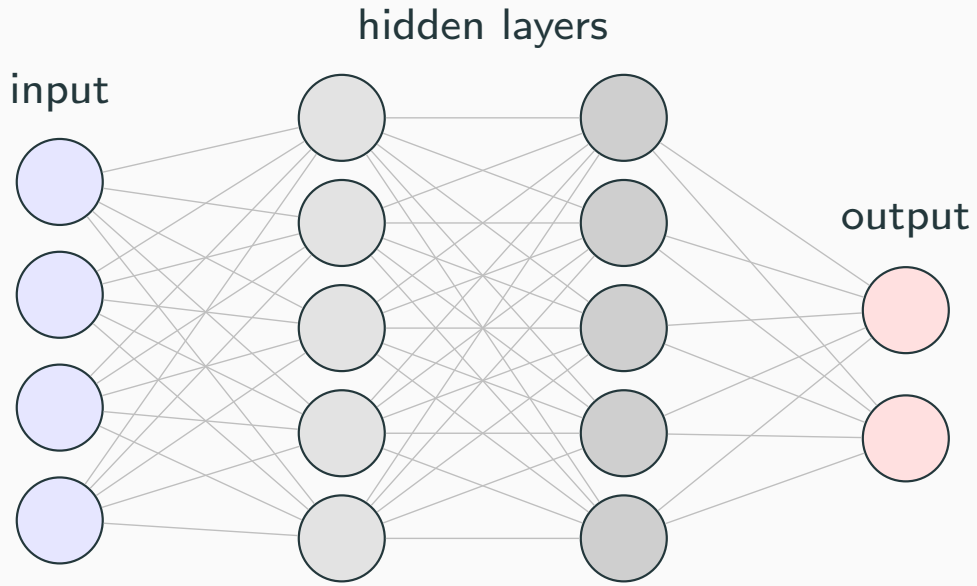
A fact about AI: opacity

Deep Neural Networks (DNNs) — including the transformer models behind AlphaFold and large language models — share one crucial feature:

Even the experts who build them do not know how they work

- There is a fundamental **mismatch** between human reasoning and the path a DNN takes from inputs to outputs
- A DNN learns by adjusting billions of numerical weights through an optimization process; the resulting weights are not human-readable
- Joel Dudley (co-creator of the Deep Patient model): "We can build these models, but we don't know how they work"

What a DNN looks like inside



each edge carries a weight $w_{ij} \in \mathbb{R}$, learned by gradient descent — not designed to represent any physical quantity

The "knowledge" lives in the weights. These are the output of gradient descent over training data — not designed to represent any physical mechanism or causal structure.

Table of Contents

Models and understanding

Last week we discussed DN, causal-nomological, and unificationist accounts of explanation. Despite their differences, they share one assumption:

A genuine explanation says something about what is going on in the world — it does not merely predict.

Contemporary philosophy of science stresses that explanations are typically underwritten by **models**: representations with interpretable content.

Such a model is not a black box for computing predictions. It has an *inside* that scientists can inhabit, reason about, and transfer to new situations.

Two examples

Kinetic theory gas = molecules in random motion; explains temperature, pressure, and effusion rates by appealing to real physical processes.

Schwarzschild metric explains why light bends *twice* the Newtonian prediction, because spatial curvature is a genuine feature of the geometry near a mass.

Pointer to the literature: van Fraassen, *Scientific Representation: Paradoxes of Perspective* (Oxford, 2008) — the central philosophical study of how models represent their target systems.

Table of Contents

Sullivan's question

> *"Are scientists trading understanding for some other epistemic or pragmatic good when they choose an opaque and complex machine learning model?"*
>
> (Sullivan 2022, p. 110)

Discuss: do you think scientists are making such a trade? Is that acceptable?

The double meaning of "model"

One terminological trap: the word **model** does double duty.

Sense 1 — scientific model

Interpretable content: states a mechanism; has an inside that scientists can inhabit; satisfies CIT.

Examples: kinetic theory, Schwarzschild, BCS

Sense 2 — ML model

A parameterized function *fitted to data by optimization*: weights set by gradient descent, not designed to represent anything.

Examples: AlphaFold, GPT, Deep Patient

Why the ambiguity matters

The word "model" in ML descends from statistics, where it *did* once have Sense 1 content (a Gaussian model, a linear model makes readable claims about a data-generating process). Neural networks inherited the term while quietly abandoning the interpretability.

Sullivan's article uses "model" in both senses without clearly distinguishing them. She asks whether a **Sense 2** model can provide the understanding we normally expect from a **Sense 1** model.

This is the right question — but answering it requires acknowledging that there is a genuine gap to be bridged, not just a terminological overlap.

Sullivan's main claim

Sullivan's answer to her own question: it is *not* the complexity or opacity of a DNN that limits the understanding it can provide.

The real culprit is **link uncertainty**:

> *"It is not implementation black-boxing that gets in the way of understanding, it is link uncertainty."*

(Sullivan 2022, p. 116)

- Black-boxing implementation details does *not* undermine understanding in general — this is commonplace in science
- What blocks understanding is the absence of scientific evidence linking the model's outputs to real causal structure in the world

Discuss: do you think Sullivan is right about this? Can you give examples of acceptable and unacceptable black-boxing from physics?

Black boxes

A **black box**: a system whose inputs and outputs are known but whose internal workings are not accessible.

Implementation black-boxing is ubiquitous in science — and usually *unproblematic*:

- A climate model's factorial subroutine may be iterative or recursive — it makes no difference to the climatology
- Schelling's model gives the same result on a computer, a real board, or a Go board

A familiar case: many physicists treat QM itself as a **quantum statistical algorithm** — preparations go in, measurement statistics come out, do not ask what is happening inside. Part I has been arguing this attitude is unsatisfactory.

Black boxes in machine learning

DNNs are black boxes at a *deeper* level: even the designers cannot identify which features the model has latched onto.

The opacity goes beyond implementation details:

- Weights are set by optimization, not by design
- The modeller cannot predict which data features will be most salient
- Indirect probing (e.g. saliency maps) gives only approximate, high-level information

Sullivan's claim: even this deeper opacity does not block understanding. The real obstacle is something else.

The Schelling analogy

Sullivan's central example: Schelling's **checkerboard model** of residential segregation.

- Simple simulation: two types of agent, each moves if $> 70\%$ of neighbours differ; result: a segregated board
- Provides *how-possibly* understanding when the semantic mapping (coin = agent, move = choice) is read off
- Becomes *how-actually* understanding only when empirical evidence connects the model to real populations

Sullivan's claim: the same structure applies to DNNs. Implementation opacity is not the obstacle — link uncertainty is.

HH: But notice the key difference. With Schelling, the semantic mapping was *built in by the modeller from the start*. With Deep Patient, there is no such mapping. The very question of what the nodes and weights *refer to* is unanswered — and it is not obvious it is even askable.

Sullivan's optimism and a challenge

> *"Once the link uncertainty is resolved, the deep patient model is able to explain and enable understanding of disease development."*
>
> (Sullivan 2022, p. 125)

HH: How would the link uncertainty ever be resolved in this case?

- With Schelling: we look for evidence that individual preferences actually drive segregation — this is a clearly statable empirical question
- With Deep Patient: what physical objects or processes would the learned weights be *linked to*?
- The weights emerged from optimization over medical records. There is no design that maps them onto causal factors in disease. The link question cannot be stated in the same form.

Table of Contents

The internal/external distinction

When a physicist derives a result, her brain is — at the physical level — just as opaque as a DNN.

If you traced the neural mechanisms, you would not *see* the reasoning. To understand why she gets the right answers, you have to get inside the **structure of the reasoning itself**.

Leibniz's mill (*Monadologie*, §17): even if the brain were enlarged to the size of a mill, and you could walk around inspecting all its machinery, you would never encounter *thought* anywhere in the mechanism.

Sullivan's blind spot?

Sullivan asks: how opaque is the implementation?

But there is a prior question:

Is there an internal perspective to be had at all?

- With the physicist: yes — her reasoning has a logical structure you can inhabit
- With a DNN: *open question*

Van Fraassen: a model must be a "way the world could be"

Van Fraassen is the arch empiricist: he does not ask that our theories be *true*, only *empirically adequate*.

Yet even van Fraassen would not be satisfied with Deep Patient as a final scientific result.

His **constructive empiricism** is a *semantic* empiricism, not an eliminativist one. A scientific theory presents models — structures interpretable as *ways the world might be*. The scientist may be agnostic about truth, but she must be able to say: this is a *possible description of reality*.

A DNN does not present a way the world could be. It presents a function from inputs to outputs. The question of truth or falsity about disease mechanisms does not even arise.

For reflection

1. Should scientists rest content with having something as predictive as Deep Patient? Or does science have a further goal that remains unfulfilled?
2. Does it matter whether we are talking about *applied* science (predicting disease, protein structure) vs. *fundamental* science (understanding the laws of nature)?
3. We saw in Part I that QM proceeds without consensus on what it means. Does this weaken or strengthen the case for demanding understanding from AI?
4. Intimation: the problem of understanding another human being — and why that seems different from understanding a DNN — points toward questions that philosophy has been wrestling with for centuries.

Selected References i