Week 4
Part B
Tools for Understanding

Hans Halvorson

May 19, 2025

Table of Contents

Table of Contents

- De Regt and Dieks: Understanding cannot be reduced to any of these particular accounts of scientific explanation
- HH: De Regt and Dieks do *not* say that understanding requires at least one of these types of explanations. This seems like a defect of permissiveness in their account.

- "Salmon treats causality as a standard for intelligibility" (p 144)
- "Present-day scientific developments cast severe doubt on the alleged privileged status of [Salmon's] model of causal explanation as the way to scientific understanding." (p 145)
- "At the deepest levels of physical reality Salmon's concept of causality is highly problematic." (p 145)
- "Physics is full of examples that show that causal-mechanical explanation is not always the actually preferred manner of achieving understanding." (p 145)

Quantum mechanics

- "Causal connections of this type . . . do not exist according to quantum theory in its standard interpretation." (p 145)
  - Arguments against trajectories
  - Bell non-locality

Lorentz contraction

- "The usual way of making the contractions intelligible is by connecting them deductively to the basic posulates of special relativity (the relativity postulate and the light posulate). . . . Causal reasoning is not involved." (p 146)
- More controversial than De Regt and Dieks make it out to be. See Bell, "How to teach special relativity" or H. Brown, *Physical Relativity*

Deflection of light by matter in GTR

"The just-mentioned example undermines the causal conception of understanding, because no causal chains were identified that are responsible for the deflection of the light." (p 157)

- “But even in pre-twentieth-century physics causal-mechanical explanation was not always the norm.” (p 146)
- “Between 1700 and 1850 action-at-a-distance rather than contact action and causal chains dominated the scientific scene.” (p 146)

- "These facts are sufficient to cast doubt on the core idea that causality has a special status as *the* fundamental, privileged standard of intelligibility." (p 146)
- "It would be erroneous to maintain that visualization is essential for obtaining understanding." (p 156)
- "The various intelligibility standards proposed by philosophers of science (e.g., visualizability, causality, and continuity) find a place in our approach as 'tools' for achieving understanding: they can help to 'see intuitively' the consequences of a scientific theory." (p 157)

Does causality have a privileged role?

- "By the guidance which analysis in terms of cause and effect has offered in many fields of human knowledge, the principle of causality has even come to stand as the ideal for scientific explanation." (Bohr 1948)
- "The viewpoint of complementarity presents itself as a rational generalization of the very ideal of causality." (Bohr 1948)

- "Unification appears to be an effective tool for achieving understanding, but like causality it is one among a variety of tools." (p 149)
- "The various intelligibility standards proposed by philosophers of science (e.g., visualizability, causality, and continuity) find a place in our approach as 'tools' for achieiving understanding." (p 156–157)

Table of Contents

Criterion for Understanding Phenomena

CUP: A phenomenon P can be **understood** if a theory T of P exists that is intelligible.

Criterion for the Intelligibility of Theories

CIT: A scientific theory T is **intelligible** for scientists (in context C) if they can recognize qualitatively characteristic consequences of T without performing exact calculations.

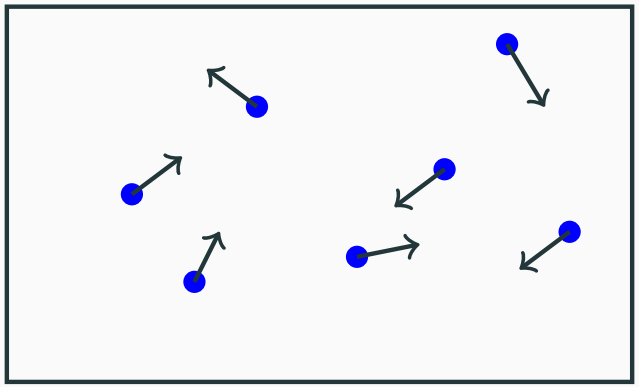
"What one wants in science is the ability to grasp how the predictions are brought about by the theory." (p 151)

Understanding Boyle's Law via the Kinetic Theory i

Illustration of CUP and CIT: How the kinetic theory provides understanding of gas behavior.

- **Qualitative model:** Boltzmann's kinetic theory pictures a gas as a collection of freely moving molecules.
- **Temperature** = average kinetic energy of molecules.
- **Pressure** = cumulative force from molecular collisions with the container walls.

Understanding Boyle's Law via the Kinetic Theory ii



Molecules move randomly and collide with container walls

Understanding Boyle's Law via the Kinetic Theory iii

Key insights (no equations):

- Adding heat $\Rightarrow$ molecules move faster $\Rightarrow$ more forceful, frequent collisions $\Rightarrow$ higher pressure.
- Reducing volume $\Rightarrow$ more collisions per unit area $\Rightarrow$ higher pressure (at constant temperature).

Conclusion: The kinetic theory provides an intelligible, causal-mechanical picture that explains Boyle's law qualitatively — in line with CUP and CIT.

Problems with CIT?

- CIT is ambiguous. Imagine a mechanical arm that pulls slips of paper out of a barrel, and always pulls out correct predictions. Would we be satisfied knowing how the mechanical arm operates?
- It seems that we would want to know how the mechanical arm manages to get the predictions right. We would want an explanation not of how it generates predictions, but of why it generates the right predictions

Open questions

- Does the account of De Regt and Dieks have any normative content? Or does it just point out the fairly obvious fact that scientific sub-communities have ideals for understanding?
- They point out that the Copenhagen-Göttingen physicists found matrix mechanics to be intelligible, while most other physicists disagreed. (p 141)

Table of Contents

- So far we have been talking about the abstract theory of scientific understanding — i.e. what, in theory, is "understanding", and the claim (of de Regt and Dieks) that understanding is a primary goal of science
- We now turn to talking about (a) the role of **tools** in understanding, and (b) with specific reference to machine learning and AI

Tools for increasing scientific understanding

- Writing!
- Diagrams
- Illustrations
- Scientific instruments
  - Microscope
  - Telescope
  - Thermometer
  - Barometer
- Physical models
- Simulations

source: Hauchs physiske cabinet

Why should (future) scientists think about AI?

- Corporations such as OpenAI, Microsoft, and Google claim that AI will lead to new knowledge — and not simply by eliminating tedious and repetitive tasks
- Will AI take your jobs?

Table of Contents

Michie's classification of machine learning

Weak ML: improved prediction quality with larger amounts of training data (algorithm is treated as a black box)

Strong ML: provides a symbolic representation of its hypothesis

Ultrastrong ML: the algorithm teaches the human operator such that the human performance is improved compared with the human learning from the data alone

When we talk about ML or AI helping science, what exactly is it that we are talking about?

Neural network A computer system modelled on the human brain and nervous system
`https://www.ibm.com/topics/neural-networks`

Deep learning "Deep learning is a subset of machine learning that uses multi-layered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain." (From the IBM website)

Dimension 1: Computational microscope

AI might be able to "see" things that are invisible to the "naked eye"

- "AI can act as an instrument revealing properties of a physical system that are otherwise difficult or even impossible to probe" (p 761)
- "[AI] can provide information not (yet) attainable through experimental means" (p 763)
- "...computational microscopes enable the investigation of objects or processes that cannot be visualized or probed in any other way, for example, biological, chemical or physical processes that happen at length and time scales not accessible in experiments." (p 764)

- What kinds of things can be seen, and how is AI supposed to do this?
  - "... the new computer-generated data"
  - "... new ways to analyze these systems without the need to perform full computations"
  - "... without the need for simulating the entire system"

"Not only is training a neural network much faster and computationally less expensive than running a hydrodynamical simulation, it also does not rely on strong assumptions about the underlying physics, or suffer from limitations arising from coarse resolution." (Schawinski et al. 2018, p 3)

Improvements of computational microscopes?

1. Increasingly complex systems: size, timescale, number of interactions
2. Advances in data representation

Dimension 2: Resource of inspiration

- Throughout human history, people have tried to find a recipe for inspiration
- Much scientific innovation seems to happen by luck, or even as if by magic
- In many cases, macro-level circumstances produced ideal conditions for innovation
    - E.g. Vienna circa 1900
- Social privilege and freedom from other worries
    - E.g. Niels Bohr's childhood (see Favrhold, *Filosoffen Niels Bohr*)

1. Identifying surprises in data
   "Data anomalies can manifest themselves in a more involved combination of variables, which might be very difficult for humans to grasp." (p 764)
   autonomous anomaly detection

2. Identifying surprises in the scientific literature
   "Researchers have to specialize in narrow subdisciplines, which makes finding new interdisciplinary ideas difficult." (p 765)
   a. Unsupervised word embedding of a large corpus of scientific papers
   b. Semantic networks built on large bodies of scientific literature

3. Surprising concepts by inspecting models
   "... rationalizing what AI algorithms have learned in order to solve a specific problem"
   "... understand the model's internal worldview"

"The concepts rediscovered in all of those works were not new and, thus, the most important challenge for the future is to learn how to extract previously unkown concepts." (p 766)

4. New concepts from interpretable solutions

5. Probing the behavior of artificial agents

- "...an artificial muse, expanding the scope of human imagination and creativity"
- Can an AI inspire in a more directed way than, say, a walk in the forest?

Dimension 3: AI as an agent of understanding

- "Algorithms that can autonomously acquire new scientific understanding, and ultimately explain these insights to humans." (p 766)
- An extreme scenario is that AIs become our peers or even our superiors as scientists.
- Granted, such a possibility is in the realm of speculation. However, what could this possibility actually look like?
- It seems that the best model we have at present is the teacher-student (or parent-child) relationship
- "Both require that the machine gets new insights and teaches them to the human." (p 766)

Additional resources

- N. Bohr. "On the notions of causality and complementarity" *Dialectica*
- J. Faye. "Niels Bohr's experimentalist approach to understanding quantum mechanics"