Week 4
Part C
The AI as Scientist

Hans Halvorson

May 19, 2025

Table of Contents

1. Does ML change the scientific method?

2. Conceptual engineering

Table of Contents

1. Does ML change the scientific method?

2. Conceptual engineering

Foundations of scientific inference

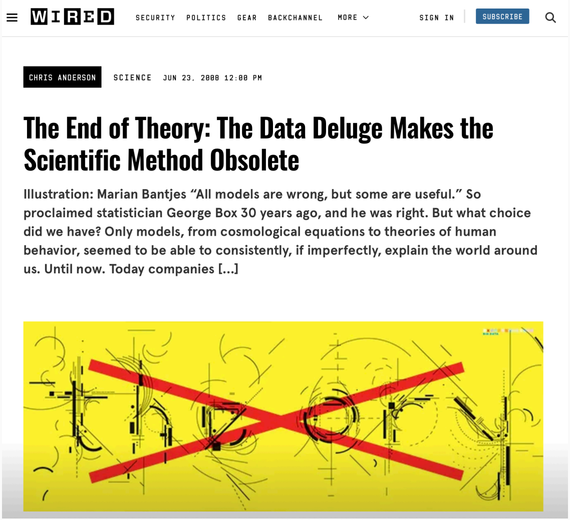
- In the early modern period (17th century and forward), philosophers reflected on the foundations of scientific inference — i.e. how empirical data might justify our belief/acceptance of theoretical hypotheses
- This investigation was especially prominent among English, Scottish, and Irish philosophers — the so-called **British empiricists**
- The earliest model was **straight induction**: a generalization T is abstracted from many repeated instances $E_1, E_2, \ldots$ of a phenomenon.
    - Since the first 1,000 ravens were black, we are justified in believing that the next raven will be black.

Hume's problem of induction

- David Hume (1711–1776) asked: what is the justification for the inductive method? Why believe that this method is rational?
- The link between past instances $E_1, E_2, \ldots$ and future projection T seems to be mediated by an assumption:
  - **Uniformity of Nature:** The future will resemble the past.
- But what is the justification for UN?

- Many philosophers conclude that theorizing requires more creative input from the scientist.
    - Karl Popper: hypothetico-deductive method
- Many philosophers conclude that scientific inference can only proceed against a backdrop of assumptions that are *not* justified by scientific inference.
    - Some (e.g. Bayesians) are hopeful that “inductive learners” would converge in the long run.

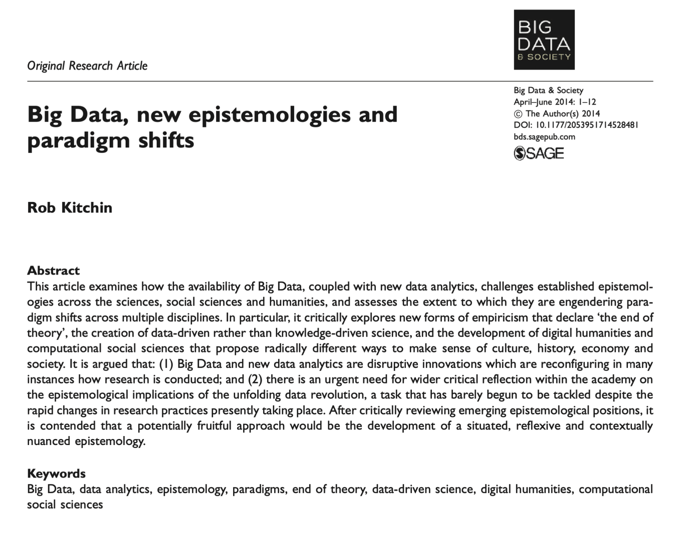
Does Big Data change our understanding of science?



CHRIS ANDERSON SCIENCE JUN 23, 2008 12:00 PM

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

Illustration: Marian Bantjes “All models are wrong, but some are useful.” So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies [...]



Original Research Article

Big Data & Society
April–June 2014: 1–12
© The Author(s) 2014
DOI: 10.1177/2053951714528481
bds.sagepub.com
SAGE

Big Data, new epistemologies and paradigm shifts

Rob Kitchin

Abstract

This article examines how the availability of Big Data, coupled with new data analytics, challenges established epistemologies across the sciences, social sciences and humanities, and assesses the extent to which they are engendering paradigm shifts across multiple disciplines. In particular, it critically explores new forms of empiricism that declare 'the end of theory', the creation of data-driven rather than knowledge-driven science, and the development of digital humanities and computational social sciences that propose radically different ways to make sense of culture, history, economy and society. It is argued that: (1) Big Data and new data analytics are disruptive innovations which are reconfiguring in many instances how research is conducted; and (2) there is an urgent need for wider critical reflection within the academy on the epistemological implications of the unfolding data revolution, a task that has barely begun to be tackled despite the rapid changes in research practices presently taking place. After critically reviewing emerging epistemological positions, it is contended that a potentially fruitful approach would be the development of a situated, reflexive and contextually nuanced epistemology.

Keywords

Big Data, data analytics, epistemology, paradigms, end of theory, data-driven science, digital humanities, computational social sciences

“Generative modelling offers a way to discern the most credible theory from various explanations for observational data. This is achieved solely through the data, without any predetermined understanding of the potential physical mechanisms operating within the studied system.” (Rodrigues, “Machine learning in physics”, referring to Schawinski, Turp, and Zhang, “Exploring galaxy evolution with generative models” 2018)

A new empiricism?

“Our approach of using generative models like the Fader network to forward model physical processes and test hypotheses in a data-driven way has significant potential in astrophysics and other fields. Its central advantage is its data-driven nature which makes no assumptions on the underlying physics.” (Schawinski et al. 2018)

A new empiricism?

“Deep learning leverages deep neural networks to automatically learn representations from the data.” (Rodrigues, p 5)

Table of Contents

1. Does ML change the scientific method?

2. Conceptual engineering

- “Surprising and creative ideas are the foundation of advances in science” (p 764)
- What kind of thing is a “new idea”?
    - A new conjecture
      Ex. “Perhaps injecting some of this virus will prevent the person from getting a worse case?”
    - An expansion of the conceptual framework

"Ved forskellige lejligheder har jeg forsøgt at vise, at den belæring, som fysikkens nyere udvikling har givet os med hensyn til nødvendigheden af en stadig udvidelse af begrebsrammerne for indordningen af nye erfaringer, fører os til en almindelig erkendelsesteoretisk indstilling, der turde være egnet til at undgå tilsyneladende begrebsvanskeligheder, også på andre af videnskabens områder." (Niels Bohr, Causality and Complementarity 1936)

“. . . at ingen erfaring er definerbar uden en logisk ramme, og at enhver mangel på harmoni synligt i et sådant forhold, kun kan fjernes ved en behørig udvidelse af den begrebslige ramme.” (Niels Bohr, Fysisk videnskab og studiet af religioner 1953)

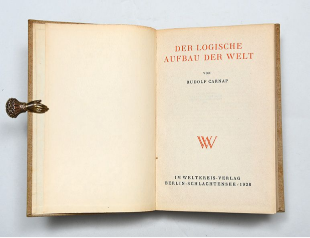
Theoretical concepts

- Modern science is characterized by the introduction of **novel concepts** that enable new levels of understanding to be achieved
- Examples:
    - Genes
    - Forces
    - Electromagnetic fields
    - The quantum of action
    - Spin (quantum two-valuedness)
    - Pauli exclusion principle
    - Lorentz transformation
    - Geometric phases

- “It would be truly exciting to see an AI uncover hidden patterns or irregularities in scientific data previously overlooked by humans, which could lead to new ideas and, ultimately, to new conceptual understanding. As of now, we are not aware of such cases.” (p 765)
- “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 unknown concepts.” (p 766)

Rational novelty?

- 20th century philosophers of science: novel scientific concepts must have some rational connection to already-understood concepts
    - Something is needed for these new concepts to be **intelligible**.
    - If I simply introduced a phrase “slithy toves”, you would have no idea what I meant



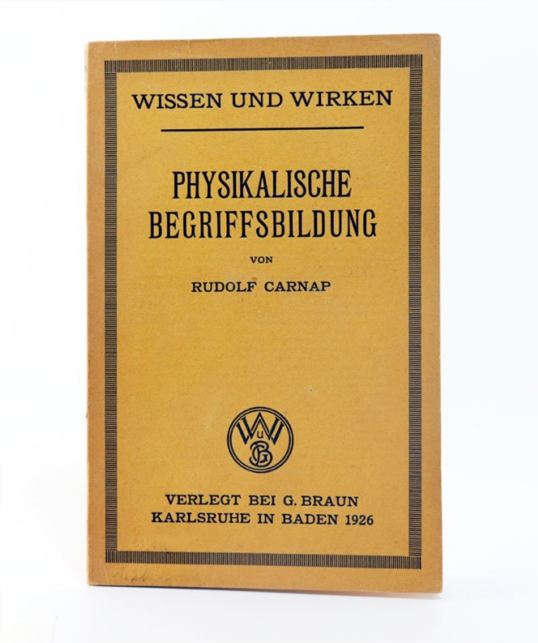
DER LOGISCHE
AUFBAU DER WELT

VON

RUDOLF CARNAP

IM WELTKREIS-VERLAG
BERLIN-SCHLACHTENSEE / 1928

- Rudolf Carnap (1891–1970) originally proposed that the new concepts should all be **logical constructions** from better understood concepts



WISSEN UND WIRKEN

PHYSIKALISCHE BEGRIFFSBILDUNG

VON

RUDOLF CARNAP

VERLEGT BEI G. BRAUN
KARLSRUHE IN BADEN 1926

- Even with a more advanced understanding of “logical construction”, the criterion of logical constructability seems far too strict
- Even according to more advanced accounts of “logical construction”, no genuine novelty arises (and that is the point of the relation being logical)

- A novel concept is similar to a **theoretical posit**, i.e. something whose existence is conjectured to explain the phenomena.
    - Le Verrier's (1846) prediction of the existence of Neptune
    - Gell-Mann's prediction of ...
    - Prediction of the Higgs boson

Can AI Perform Scientific Inference?

Bayesian conditionalizing: Yes, AI systems can update beliefs using Bayes' rule.

- Widely implemented in machine learning and probabilistic modeling.
- But: relies on well-specified prior probabilities and likelihoods.

Inference to the Best Explanation (IBE): Partially

- AI can rank hypotheses based on fit, simplicity, etc.
- But: IBE involves theory choice, explanatory power, and background knowledge — often context-sensitive and informal.
- Current AI lacks deep semantic understanding or explanatory intuition.

Can LLMs Perform Scientific Inference?

LLMs (e.g., GPT-4, Claude): Predict text based on patterns in large datasets.

Bayesian Updating: Not natively

- LLMs are not probabilistic reasoners in the Bayesian sense.
- They can talk about Bayes' rule, but don't maintain internal probabilistic belief states.

Inference to the Best Explanation (IBE): Superficial imitation

- Can generate plausible-sounding explanations.
- Can rank hypotheses based on heuristics (e.g., coherence, simplicity) — if prompted.
- But lacks grounding in actual theory choice, background understanding, or causal modeling.

Toward an Intelligent Scientific Agent

Comparison of AI Systems for Scientific Inference

Capability	LLMs	Probabilistic Models
Natural language fluency	Yes	No
Bayesian updating	No	Yes
Causal explanation	Partial (imitated)	Yes (if built-in)
Theory generation	Partial (pattern-based)	Limited (depends on priors)
Hypothesis revision	No	Yes
Scientific judgment	No	Partial (domain-specific)

Hybrid efforts: Ongoing research aims to combine LLMs with Bayesian, causal, and symbolic reasoning models.

Can AGI Do Science Like a Human?

Current AI systems (e.g., LLMs):

- Simulate scientific discourse and generate plausible hypotheses.
- Summarize theories, analyze data, complete analogies.
- Lack genuine belief representation, explanatory goals, and epistemic norms.

What human-like scientific reasoning requires:

- Formulating and revising hypotheses in light of evidence.
- Understanding causal mechanisms, not just correlations.
- Distinguishing relevance and explanatory depth.
- Participating in social and normative dimensions of inquiry.

Conclusion: Simulating science $\neq$ doing science. AGI would need more than language fluency — it must integrate causal reasoning, epistemic

- A simpler case: automated theorem proving
    - Even in pure mathematics (i.e. deductive logic), humans rely on intuitions about which proof strategies to pursue
- What about the role of scientists in making value judgments?
    - When should an experiment be re-run?
    - Inductive risk

- After Carnap realized that logical reduction was too strict, he tried various other more liberal relationships (between new concepts and old)
    - Partial reduction
    - Implicit definition
    - Ramsey sentences

- Both camps (realist and antirealist) of late 20th century philosophers of science gave up on the project of understanding conceptual novelty
    - Realists: new concepts are good when they latch onto the joints of reality
    - van Fraassen: the aim is simply to build models that are empirically adequate
    - Kuhn: the concepts of the new paradigm are often incommensurable with those of the old paradigm
- As a result, we have few plausible models of what conceptual innovation/growth might amount to

Conceptual engineering and mathematization

- Conceptual advances in modern physics have often followed the development of new mathematical frameworks
    - Einstein was able to treat gravity as a field by using the resources of Riemannian geometry
    - Heisenberg was (apparently) able to overcome the contradictions of the old quantum theory by employing non-commutative matrix algebra
    - The classification of fundamental particles was enabled by the theory of representations of Lie groups

Extending concepts to new domains

- Einstein: the concept of two events being simultaneous breaks down when velocities are high compared to the speed of light

- With genuinely novel concepts, the framework for inquiry is changed — and it is difficult to imagine how AI could do that
- Can AI only do “normal science”? Or could AI make a revolutionary advance? What would that look like?
    - Compare to description of how AI might try to explain something that is “above our heads”
- Krenn et al. AI is good at “rediscovery tasks”, but not much evidence yet that it can drive discoveries.

References i

- Anderson, Chris (June 2008). **"The End of Theory: The Data Deluge Makes the Scientific Method Obsolete".** In: *Wired Magazine*. URL: https://www.wired.com/2008/06/pb-theory/.
- Barman, Kristian Gonzalez et al. (2023). **"Towards a Benchmark for Scientific Understanding in Humans and Machines".** In: *Minds and Machines*. DOI: 10.1007/s11023-024-09657-1. URL: https://link.springer.com/article/10.1007/s11023-024-09657-1.
- Goldstein, Simon (2024). **"LLMs Can Never Be Ideally Rational".** In: *Philosophy and Technology*. Forthcoming. URL: https://philarchive.org/rec/GOLLCN.

References ii

- Henderson, Leah (2022). **“The Problem of Induction”.** In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2022. URL: https://plato.stanford.edu/archives/win2022/entries/induction-problem/.
- Kitchin, Rob (2014). **“Big Data, New Epistemologies and Paradigm Shifts”.** In: *Big Data & Society* 1.1. DOI: 10.1177/2053951714528481. URL: https://journals.sagepub.com/doi/full/10.1177/2053951714528481.
- Leonelli, Sabina (2014). **“What Difference Does Quantity Make? On the Epistemology of Big Data in Biology”.** In: *Big Data & Society* 1.1. DOI: 10.1177/2053951714534395. URL: https://journals.sagepub.com/doi/10.1177/2053951714534395.

References iii

- Leonelli, Sabina (2020). **“Scientific Research and Big Data”.** In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2020. URL: https://plato.stanford.edu/entries/science-big-data/.
- Rodrigues, Filipe (2023). **“Machine Learning in Physics: A Short Guide”.** In: *arXiv preprint arXiv:2310.10368*. URL: https://arxiv.org/abs/2310.10368.
- Romeijn, Jan-Willem (2022). **Howson on Induction, with Applications to Machine Learning.** Presentation at the Howson Memorial Conference, LSE. URL: https://romeijn.web.rug.nl/presentation/2022_romeijn_-_MachineLearningHowson.pdf.

References iv

Schawinski, Kevin, M. Dennis Turp, and Ce Zhang (2018). **"Exploring Galaxy Evolution with Generative Models".** In: *Astronomy & Astrophysics* 611, A97. DOI: 10.1051/0004-6361/201833800. URL: https://www.aanda.org/articles/aa/full_html/2018/08/aa33800-18/aa33800-18.html.