26th International Congress
of History of Science and Technology
www.ichst2021.org

A multiperspective causal analysis of computing in predictive models based on machine learning

Franck Varenne – Associate Professor of philosophy of science
University of Rouen (France) – ERIAC Lab (EA 4705)

Online conference - Managed from Prague - Date : July 29, 2021

Motivations and approach

- Questions about explanation, causality and their mutual relationships in scientific models are notoriously multifarious, complex and subject to many perspectives
- Today, explainability in predictive models used in AI (e.g. in DL) is a related but particular and hot subject matter
- Some interesting questions arise among others:
  - Is it true that predictive models based on machine learning cannot be understood from a causal perspective as it is often said?
  - I.e.: do we have to understand their predictive power only in terms of what contemporary philosophers call “non causal explanations” (Woodward “Scientific Explanation”, *SEP*, 2003)?
  - But, isn’t this claim paradoxical as “non causal explanation” are often said to be formal ones and of a mathematical nature, whereas we precisely often lack a full mathematical comprehension of the predictive power of programs based on machine learning? (black boxes)
- Suggestion: using the Aristotelian 4-causes schema to try to see a bit clearer

Outline

- **I. The 4 causes in digital computing broadly speaking**
- **II. Explanation and causality in short**
- **III. Kinds of learning and predicting from data**
- **IV. “Model-free” predictions, efficient and material causes**
- **V. Predictions with models, final and formal causes**
- **VI. Implicit models in so-called “model-free” prediction**

Conclusions

I. The 4 causes in digital computing broadly speaking

	Definition	Ex 1: A table	Ex 2: A program
Efficient cause	The primary source of change or rest of a thing	Carpenter	Programmer; electric current
Material cause	That out of which a thing is made	wood	Electronic circuits; microprocessor; chip
Final cause	That for the sake of which a thing is done	To eat or to work on it	Resolution of equation; prediction; classification; simulation
Formal cause	The account of what it is to be the thing it is	A kind of furniture	Kind of mathematical resolution; kind of algorithmic method

Sources for definitions and example #1: Aristotle, *Physics*, II, 3 ; Lombrozo & Vasilyeva, “Causal Explanation”, 2017

II. Explanation and causality in short

Explanation: "a justification or a reason for a belief or action" (Miller, 2020)

Cause: "something without which something else would not happen" (Cambridge Dictionary)

Not all explanations are causal…

- e.g. in maths or maths applied to physics →

"A rigid board has a round hole and a square hole. A peg with a square cross-section passes through the square hole, but not the round hole. Why? Putnam suggests that this can be explained by appeal to the ***geometry*** *of the rigid objects (which is* ***not causal****), without appeal to lower-level physical phenomena (which are presumably causal). Is this a case of* ***non-causal explanation****? It depends on whom you ask."*

Lombrozo & Vasilyeva, "Causal Explanation", 2017.

- final cause seems to reverse the time arrow
- formal cause seems to be more like a constitutive relationship than a causal one

... but many explanations tend to be causal in the human psyche (Lombrozo et al., 2017)

2 general kinds of explanation in AI

Source: Mittelstaedt *et al.* *"Explaining explanations in AI",* 2019.

1. For experts (existing xAI)

= approximate scientific modeling of either…

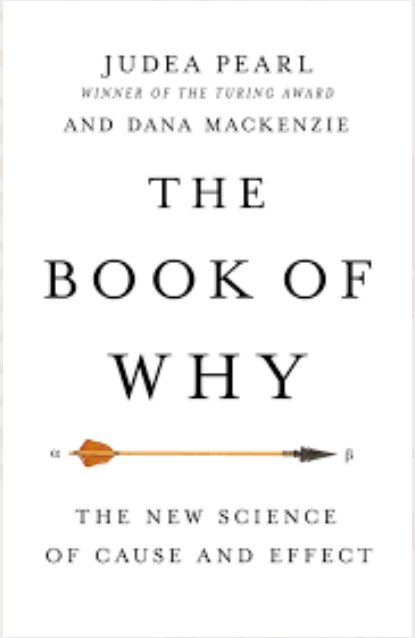
a. the internal functioning of the program: transparency
b. …or the external behavior of the program: post-hoc interpretability

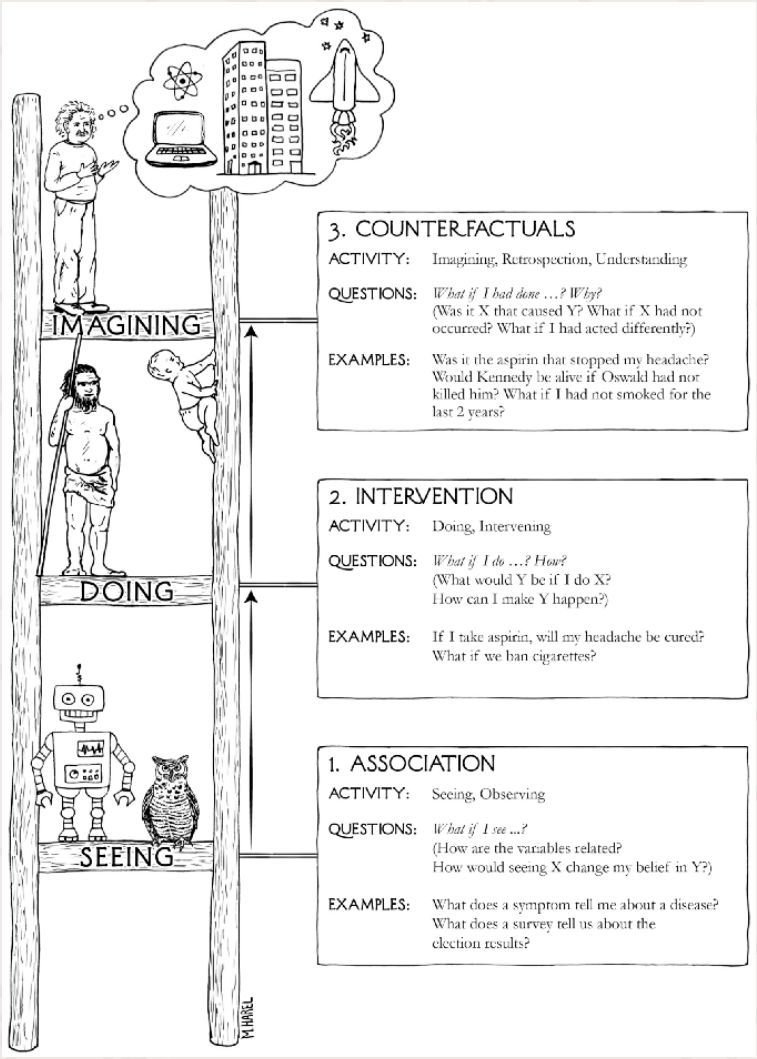
2. For non experts (desired xAI)

Explanations for laymen, users or persons affected by the decisions

a. Contrastive (Why P and not Q?)
b. Selective
c. Socially interactive

The “ladder of causation” according to Judea Pearl (2018)





III. Kinds of learning and predicting from data

- 1. **Classical inferential statistics**: association, regression, multivariate analysis

- 2. **Statistical learning** (to tackle the "curse of dimensionality" arising in parameters estimation for large multivariate problems, Vapnik, 1998): Perceptron (pattern recognition), SVM (Support Vector Machine), Neural Networks, Deep Learning (Vapnik, Hinton, Le Cun, Bengio)

- 3. **Probabilistic bayesian networks**: Prior belief (subjective probability) + New evidence → Revised belief (Pearl)

- 4. **Causal diagrams inference**: subjective causality (Pearl, Schölkopf)

IV. Model-free predictions, efficient and material causes

- According to Pearl, classical statistics (multivariate analysis), Support Vector Machine (SVM) and Neural Networks (whether Deep or not) lead to algorithms that are model-free.

- They only belong to the first rung of his “ladder of causation”: association (“how are the variables related?”)

- This is often called Numerical AI (based on numbers) ≠ Symbolic AI (based on rules)

- To date, there seems to be no full mathematical explanation of the success of Deep learning: hence, I would say, such predictions lack formal cause

- Analogous to Pattern recognition and mimicking (see Generative Adversarial Neural networks)

IV. Model-free predictions, efficient and material causes

	Definition	Ex 1: A table	Ex 2: A mathematical resolution program	Ex 3: A predictive program based on Deep Learning
Efficient cause	The primary source of change or rest of a thing	Carpenter	Programmer ; electric current	**Programmer ; electric current**
Material cause	That out of which a thing is made	wood	Electronic circuits; microprocessor; chip	**Electronic circuits; microprocessor; chip**
Final cause	That for the sake of which a thing is done	To eat or to work on it	Resolution of equation	**Prediction; mimicking a model (as reference)**
Formal cause	The account of what it is to be the thing it is	A kind of furniture	Kind of mathematical resolution; kind of algorithmic method	**None; no mathematical explanation; lack of mathematical model**

IV. Model-free predictions, efficient and material causes

	Definition	Ex 1: A table	Ex 2: A mathematical resolution program	Ex 3: A predictive program based on Deep Learning
Efficient cause	The primary source of change or rest of a thing	Carpenter	Programmer ; electric current	Programmer ; electric current
Material cause	That out of which a thing is made	wood	Electronic circuits; microprocessor; chip	Electronic circuits; microprocessor; chip
External configuration (Aristotle: *DA*, II, 1; *PA*, I, 1)	**Material form, shape, spatial structure, temporal structure ("sensible form": Thomas Aquinas)**	**The sensible organization of the top and legs of the table**	**The temporal and spatial structures of its output values**	**The sensible and measurable reproduced or recognized pattern**
Final cause	That for the sake of which a thing is done	To eat or to work on it	Resolution of equation	**Prediction; mimicking a model (as reference)**
Formal cause	The account of what it is to be the thing it is	A kind of furniture	Kind of mathematical resolution; kind of algorithmic method	**None; no mathematical explanation; lack of mathematical model**

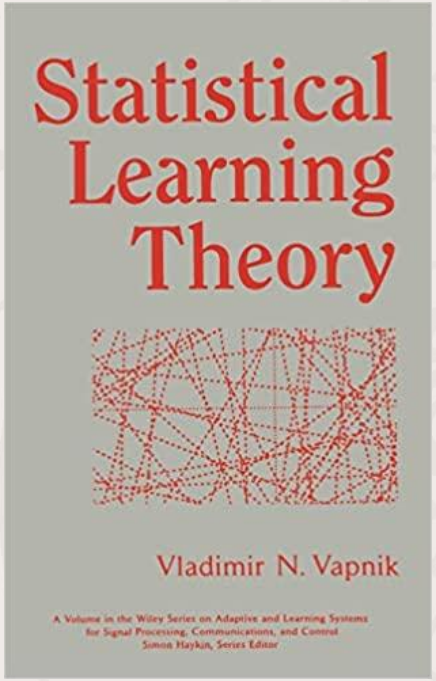
IV. Model-free predictions, efficient and material causes

	Definition	Ex 1: A table	Ex 2: A mathematical resolution program	Ex 3: A predictive program based on Deep Learning	Ex 3: A predictive program based on Deep Learning	Ex 3: A predictive program based on Deep Learning
Efficient cause	The primary source of change or rest of a thing	Carpenter	Programmer ; electric current	**Efficient cause of the program as predictor**	Efficient cause of the output as pattern	Programmer ; electric current; values and data used for training ; observed pattern
Material cause	That out of which a thing is made	wood	Electronic circuits; microprocessor; chip	**Material cause of the program as predictor**	Material cause of the output as pattern	Electronic circuits; microprocessor; chip
External configuration (Aristotle: *DA*, II, 1; *PA*, I, 1)	**Material form, shape, spatial structure, temporal structure ("sensible form": Thomas Aquinas)**	**The sensible organization of its top and legs**	**The temporal and spatial structures of its output values**	**The temporal and spatial structures of its output values**	**The temporal and spatial structures of its output values**	**The sensible and measurable reproduced or recognized pattern**
Final cause	That for the sake of which a thing is done	To eat or to work on it	Resolution of equation	**Prediction; mimicking a model (as reference)**	**Prediction; mimicking a model (as reference)**	**Prediction; mimicking a model (as reference)**
Formal cause	The account of what it is to be the thing it is	A kind of furniture	Kind of mathematical resolution; kind of algorithmic method	**None; no mathematical explanation; lack of mathematical model**	**None; no mathematical explanation; lack of mathematical model**	**None; no mathematical explanation; lack of mathematical model**

V. Predictions with models, final and formal causes

	Definition	Ex 1: A table	Ex 2: A mathematical resolution program	Ex 3: A predictive program based on Deep Learning	Ex. 4: A predictive program based on a causal model (with explicit agents)
Efficient cause	The primary source of change or rest of a thing	Carpenter	Programmer ; electric current	Programmer ; electric current	Programmer ; electric current; representations of elements and activities (mechanisms)
Material cause	That out of which a thing is made	wood	Electronic circuits; microprocessor; chip	Electronic circuits; microprocessor; chip	Electronic circuits; microprocessor; chip
Final cause	That for the sake of which a thing is done	To eat or to work on it	Resolution of equation	**Prediction**	**Approximately realistic description and explanation of the functioning of the target system**
Formal cause	The account of what it is to be the thing it is	A kind of furniture	Kind of mathematical resolution; kind of algorithmic method	**None; no mathematical explanation; lack of mathematical model**	**Sets of rule; sometimes uncompressible in their interactions (not formally mathematizable, hence necessitating computers)**

VI. The implicit models in "model-free" predictions (1): statistical learning (Vapnik)



- Whereas classical statistical inference is based on the "classical law of large number" **assuming** the existence of **independent and identically distributed** (i.i.d.) Lebesgue integrable random variables *("the frequency of any event converges to the probability of this event with an increasing number of observations"* (Vapnik, 1998, p. 9)…

- …"the theory of induction is based on the **uniform** law of large numbers" (ibid., p. 13): *"for a given **set** of events the sequence of probabilities of events with the smallest frequency converges to the smallest possible values for this **set**"* (ibid.).

VI. The implicit models in "model-free" predictions (2): deep learning (Le Cun)

1. Le Cun's assumption: material reality is hierarchically organized in robust levels that can be perceived and interpreted as levels of abstraction (Simon)

2. Existing living sensory systems are organized in layers so as to (="final" cause) match these levels (through a causal (efficient) evolution-selection process)

3. These living sensory systems reproduces (mimics) the objectively layered structure of reality

4. In its turn, by mimicking these living sensory nervous systems, deep-learning finds this constraint again, hence mimics the true layered-structure of reality

Some first conclusions…

- **1. There is no predictive programs nor algorithms deprived of any ontological commitment regarding the constitution or structuration of reality**

- **2. The current general theory of statistical induction (Vapnik, Le Cun, etc.) surely, can be said to be model-free (Pearl) viewed from the directly causalist standpoint but cannot be said to assume no model of structuration for reality**

- **3. Its minimal ontological commitment relies on the assumption of the applicability of the "uniform law of large numbers", which is not deductively proven but only relies on a fruitful bet**

- **4. Hence, current predictive models do not model reality, but they model the structuration of its signals. It is more the molding of patterns or shapes (where the fifth Aristotelian "cause" plays a central role: "external configuration") that a modeling of phenomena or events**

- **5. Making the distinction between causes in Machine Learning is more a question of perspective as Aristotle already noted when he noticed that the distinction between matter and forms is also to a certain extent a matter of standpoint**

Thank you for your attention!

Selected references:

Aristotle: *Physics, On the Soul, Parts of Animals.*

Denis, C. Varenne, F., « Interprétabilité et explicabilité pour l'apprentissage machine : entre modèles descriptifs, modèles prédictifs et modèles causaux. Une nécessaire clarification épistémologique », *Actes Conférence Nationale en Intelligence Artificielle (CNIA),* PFIA 2019 (https://www.irit.fr/pfia2019/wp-content/uploads/2019/07/actes_CNIA_PFIA2019.pdf)

Le Cun, Y., *Quand la machine apprend,* Paris, Odile Jacob, 2019.

Lombrozo, T., Vasilyeva, N., "Causal Explanation", *in* M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 415–432), Oxford University Press, 2017.

Miller, T., "Contrastive Explanation: A Structural-Model Approach", arXiv preprint arXiv:1811.03163, 2019. (Under review)

Mittelstadt, B., Russell, C., Wachter, S., "Explaining Explanations in AI", FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency, January 2019 Pages 279–288https://doi.org/10.1145/3287560.3287574

Pearl, J., Mackenzie, D., *The Book of Why*, Basic Books, 2018

Vapnik, V.N., *Statistical Learning Theory*, Wiley, 1998.

Woodward, J., "Scientific Explanation", *Stanford Encyclopedia of Philosophy*, 2003