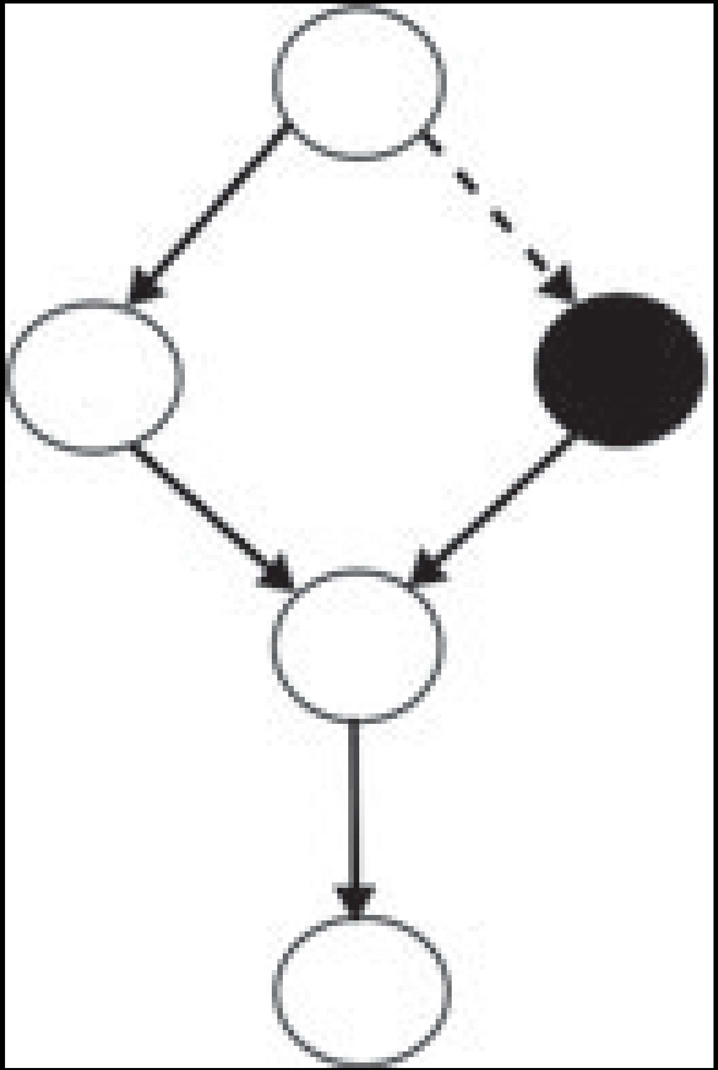


Should Causal AI Rule over Deep Learning?

ALBERTO D. HORNER

CADIS

In which direction will we promote progress?

The general argument

- **'Yes, Causal AI should rule over Deep Learning.'**
  1. There are many reasons just mentioned in this presentation.
  2. Causal Markov Condition can be understood as an advantage.
  3. Cognitive arguments support this thesis.

- This is a general argument that addresses two global approaches to AI.

- Scientific communities need to decide in which direction promote progress, and which methods will the apply.

Two kinds of methods

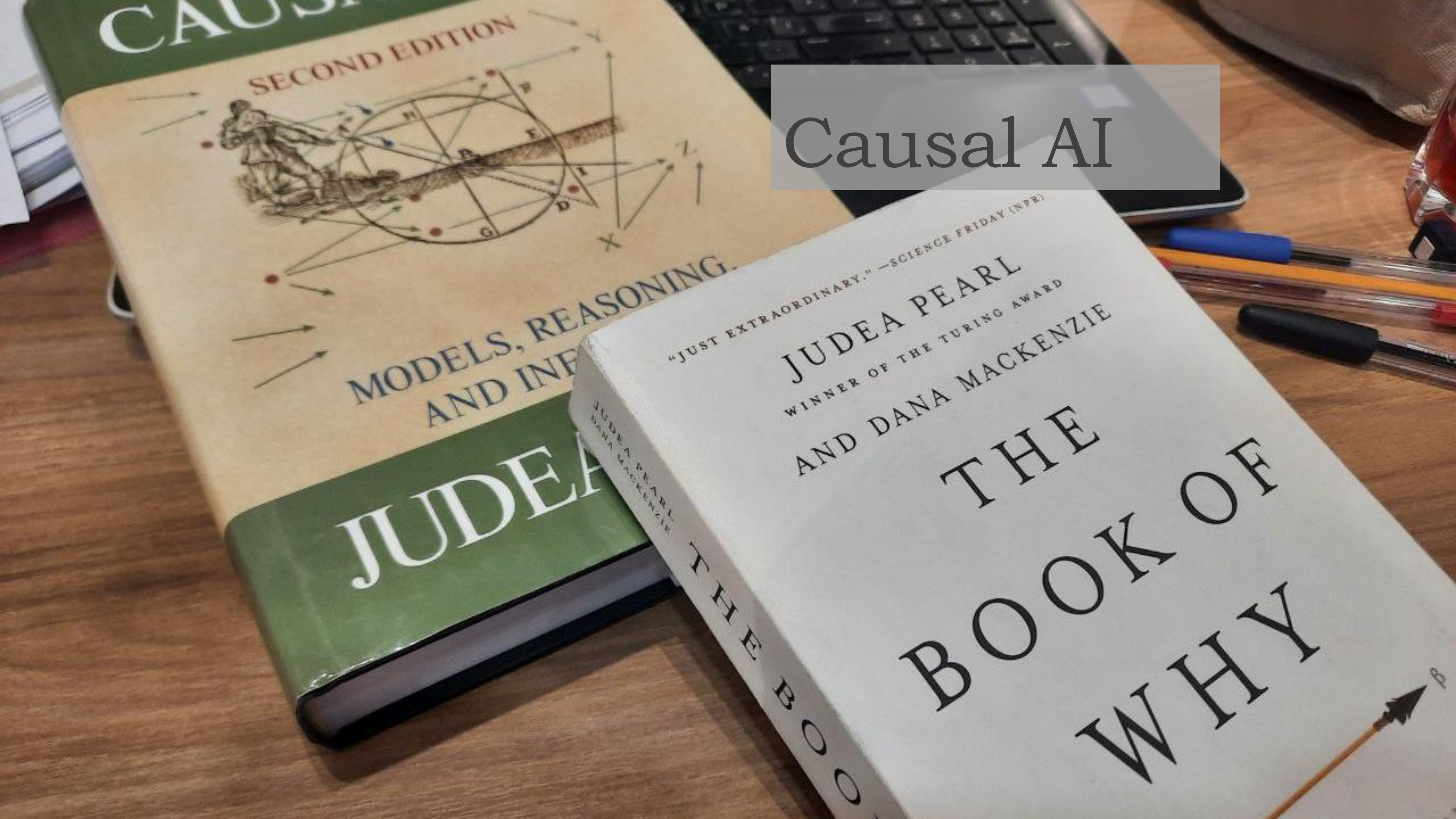
CAUSAL AI

- AI based on Causal Bayesian Networks (CBNs), and causal models.
- Causal models, their assumptions, and the algorithms to infer them are here understood as Judea Pearl defines them in his book *Causality (2009)*.
- A CAUSAL AI SYSTEM is expected to convey causal explanations of its behavior, and to operate with information about the relations between variables.

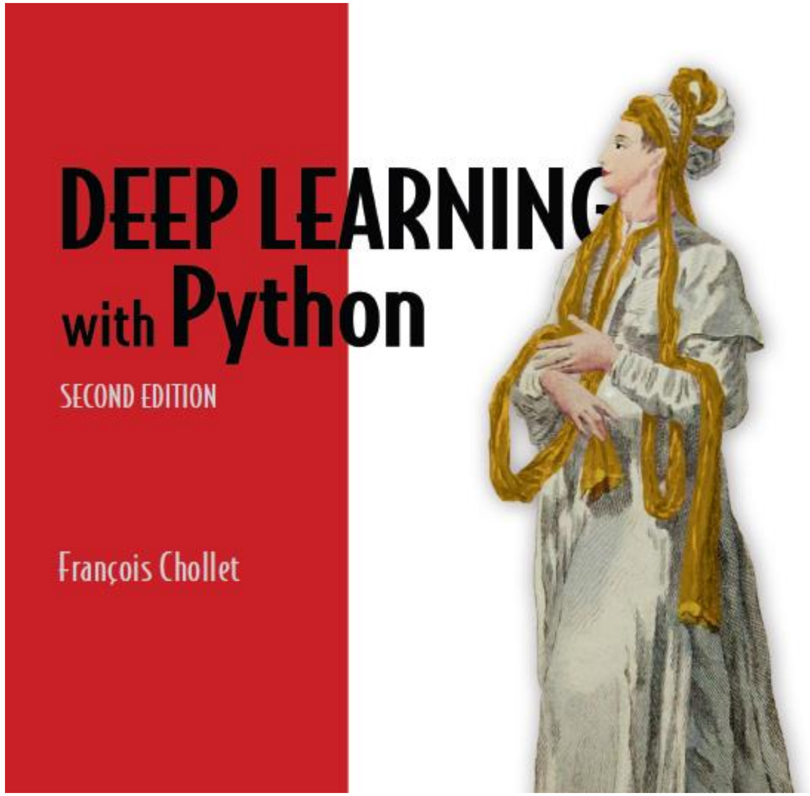
DEEP LEARNING

- Densely connected neural networks.
- All methods described in Francois Chollet's book *Deep Learning with Python* (ConvNets, RNNs, etc.).
- *A DEEP LEARNING AI SYSTEM is expected to find the function that best fits the data.*

Causal AI


SECOND EDITION
MODELS, REASONING,
"JUST EXTRAORDINARY." —SCIENCE FRIDAY (NPR)
JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE
THE BOOK OF WHY
JUDEA PEARL
DANA MACKENZIE

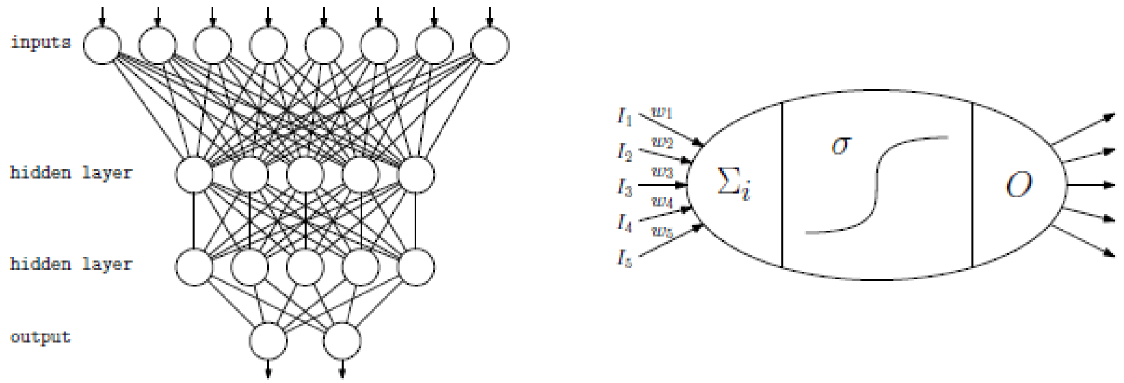
For the sake of specificity:



They do not exclude each other

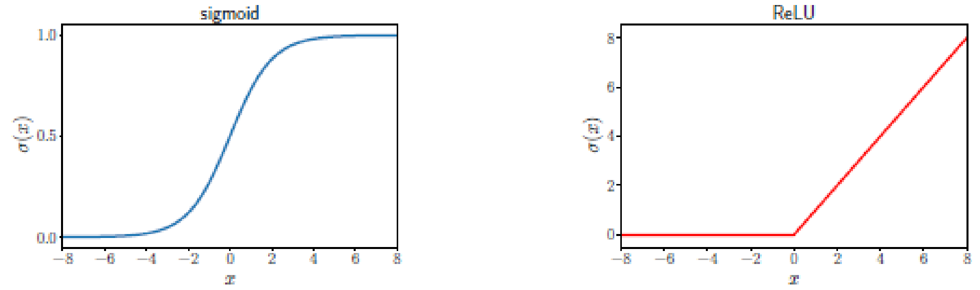
- Recently, we have seen considerable progress towards fusing both approaches (Vg. Testing Bayesian Networks (TBNs), (Choi, Wang, Darwiche, 2019, IJAR)).

- Notwithstanding, a predominant one needs to be chosen.
  1. Scientific goals
  2. Complex Systems



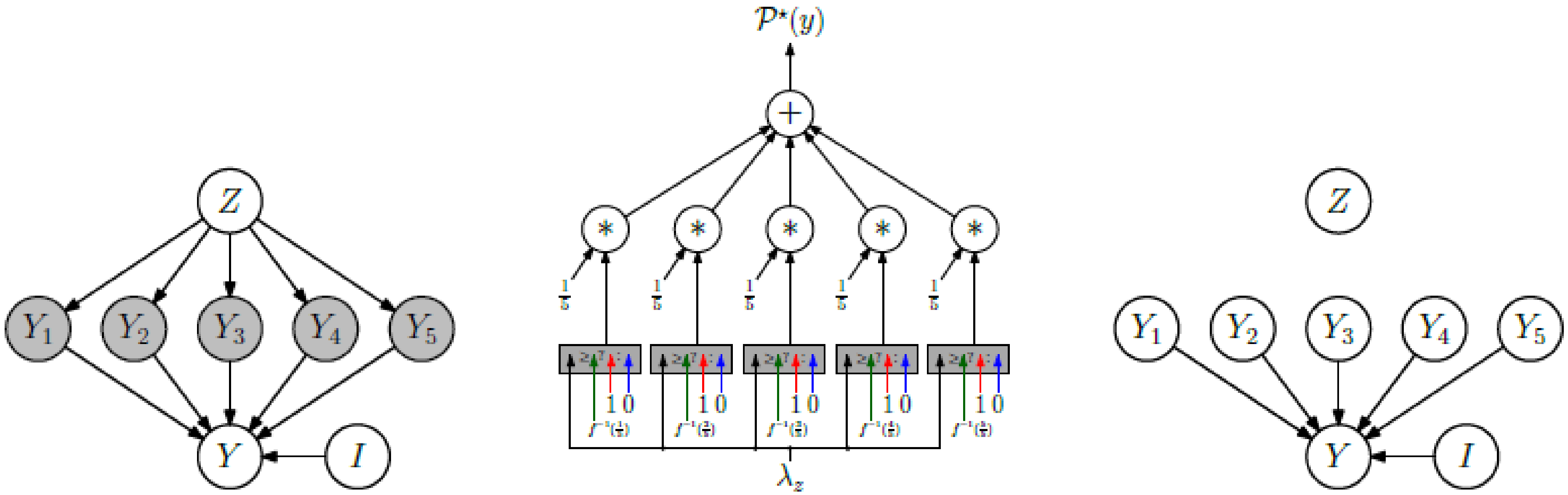
(a) A neural network structure (b) A mathematical model of a neuron

Figure 1: A neural network and a neuron.



TBNs: An example of fusing both approaches

Choi, Wang, Darwiche: 2019



(a) TBN (b) TAC for query $\mathcal{P}^\star(y)$ (c) BN after CPT selection

TBNs: An example of fusing both approaches

Choi, Wang, Darwiche: 2019

They do not exclude each other

- Recently, we have seen considerable progress towards fusing both approaches (Vg. Testing Bayesian Networks (TBNs), (Choi, Wang, Darwiche, 2019, IJAR)).

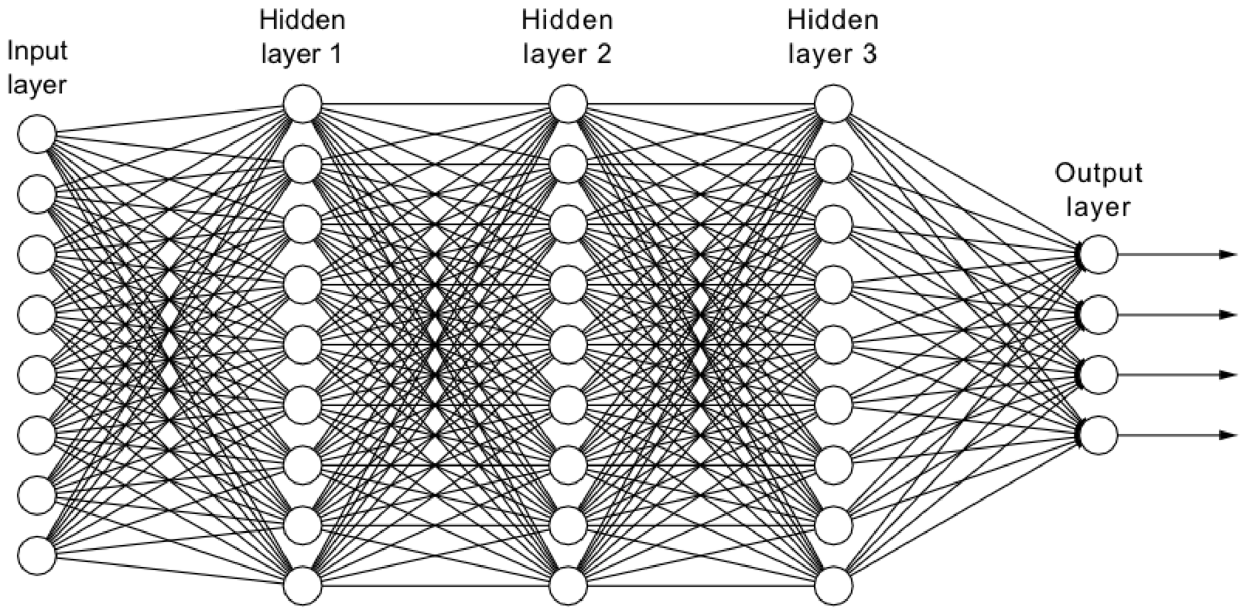
- Notwithstanding, a predominant one needs to be chosen.
  1. Scientific goals
  2. Complex Systems

Advantages of Causal methods

- Causal models grasp changes in the probability distribution, they deal with a family of ‘n’ probability distributions
- Interpretability and simplicity
- Encode previous (expert) knowledge
- Measure causal effects
- Grasp invariant qualitative knowledge

Advantages of Deep Learning

- Universal approximators
- Expert knowledge is not strictly required
- They have found no competitors *in the problems they solve*



The Causal Markov Condition could be an advantage

First argument

Causal Inference Assumptions (IC-like algorithms)

1. Minimality
2. Stability
3. The Causal Markov Condition (CMC)

Causal Inference Assumptions (IC-like algorithms)

1. Minimality
2. Stability
3. The Causal Markov Condition (CMC)

Problematic!

What do we get with The Causal Markov Condition?

A change in Statistics

CLASSICAL APPROACH

1. Small set of strong features
2. Simple functions (vg. linear)

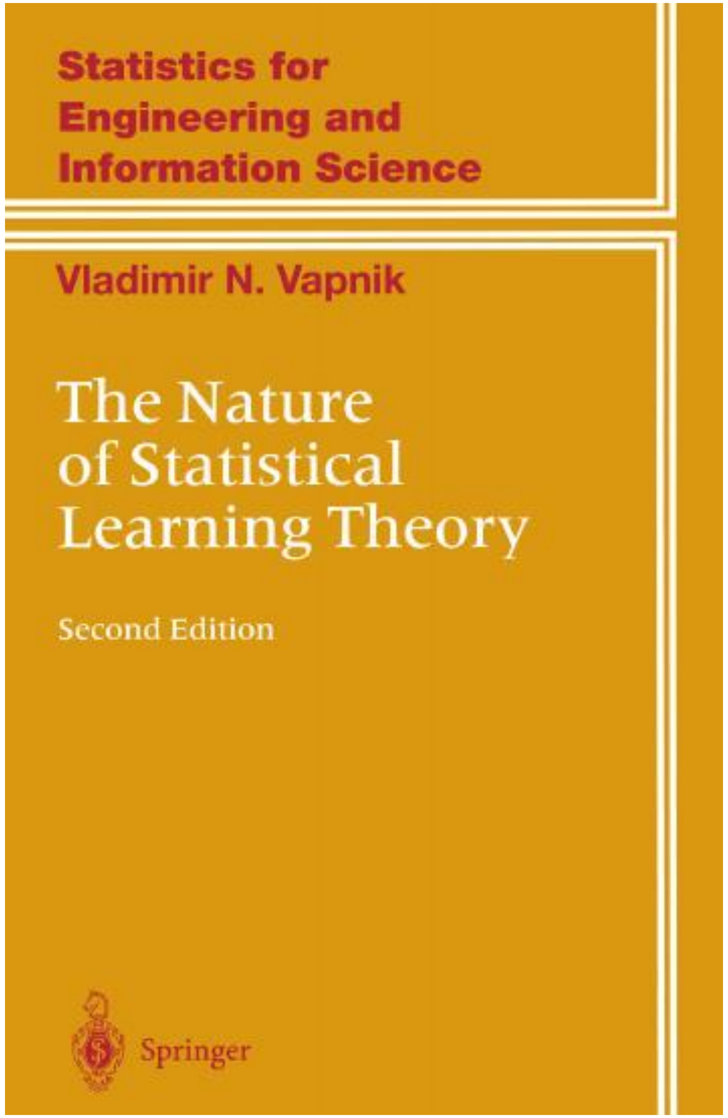
THE NEW TECHNIQUES

1. A large number of weak features (vg. data mining)
2. 'Smart' functions

Vapnik (2000)

Current practices references:

- Goodfellow, Bengio, Courville (2016)
- Lee, W. (2019)
- Chollet, F. (2021)



What we get with the CMC

CAUSAL AI

1. ~~Small set of strong features~~
2. ~~Simple functions (vg. linear)~~

- **The structure**

❑ It is feasible to select a few causal paths that will render the others insignificant.

DEEP LEARNING

1. A large number of weak features (vg. data mining)
2. ‘Smart’ functions

❑ There is no feasible procedure to select a few neural paths that will render the others negligible.

What we get with the CMC

CAUSAL AI

1. ~~Small set of strong features~~
2. ~~Simple functions (vg. linear)~~

- **The structure**

❑ It is feasible to select a few causal paths that will render the others insignificant.

The Causal Markov Condition!

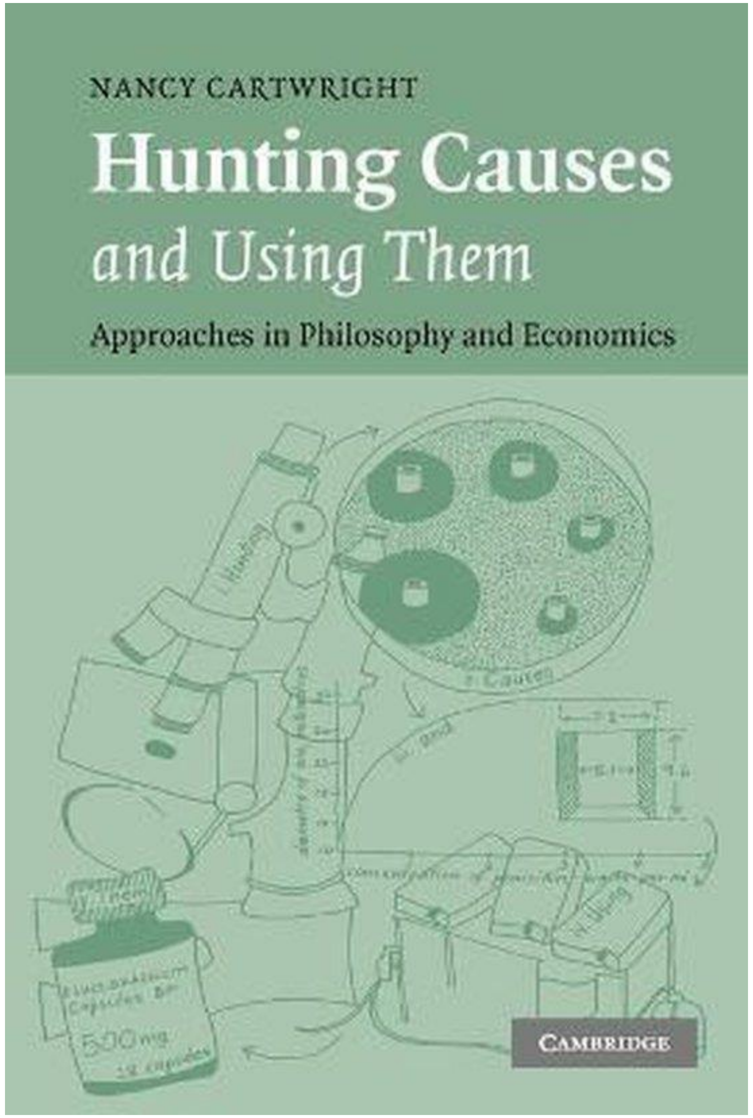
DEEP LEARNING

1. A large number of weak features (vg. data mining)
2. 'Smart' functions

❑ There is no feasible procedure to select a few neural paths that will render the others negligible.

Nancy Cartwright's Objections

2007



Problems with Causal Markov Condition

1. Common causes
2. Causes cooperating to produce one effect
3. Mixed populations
4. Changes in the same direction of time
5. By-products

What does it mean to hold a scientific attitude?

- Option 1: “When you don’t know, you don’t know.” (Non-observed variables agnosticism)
- Option 2: “Hold adequate assumptions.”

Control tasks

- The Causal Markov Condition enables us to model decision making situations.
- Marvin Minsky's *Society of Mind*

Just wonder if it is possible to design an AI system with a causal model as its main structure, one which operates not based on raw data inputs, but on results obtained by different neural networks that process different kinds of data.

On Cognitive grounds

Neuroscience and Artificial Intelligence mutually enrich themselves

Backpropagation and the brain

Timothy P. Lillicrap, Adam Santoro, Luke Marris, Colin J. Akerman and Geoffrey Hinton

Abstract | During learning, the brain modifies synapses to improve behaviour. In the cortex, synapses are embedded within multilayered networks, making it difficult to determine the effect of an individual synaptic modification on the behaviour of the system. The backpropagation algorithm solves this problem in deep artificial neural networks, but historically it has been viewed as biologically problematic. Nonetheless, recent developments in neuroscience and the successes of artificial neural networks have reinvigorated interest in whether backpropagation offers insights for understanding learning in the cortex. The backpropagation algorithm learns quickly by computing synaptic updates using feedback connections to deliver error signals. Although feedback connections are ubiquitous in the cortex, it is difficult to see how they could deliver the error signals required by strict formulations of backpropagation. Here we build on past and recent developments to argue that feedback connections may instead induce neural activities whose differences can be used to locally approximate these signals and hence drive effective learning in deep networks in the brain.

predominantly unsupervised fashion[1,25–27], building representations that make explicit the structure that is only implicit in the raw sensory input. It is natural to wonder, then, whether backprop has anything to tell us about learning in the brain[25,28–30].

Here we argue that in spite of these apparent differences, the brain has the capacity to implement the core principles underlying backprop. The main idea is that the brain could compute effective synaptic updates by using feedback connections to induce neuron activities whose locally computed differences encode backpropagation-like error signals. We link together a seemingly disparate set of learning algorithms into this framework, which we call 'neural gradient representation by activity differences' (NGRAD)[9,27,31–41]. The NGRAD framework demonstrates that it is possible to embrace the core principles of backpropagation while sidestepping many of its problematic implementation requirements. These considerations may be relevant to any brain circuit that incorporates both feedforward and feedback connectivity.

Backpropagation and the brain

Timothy P. Lillicrap, Adam Santoro, Luke Marris, Colin J. Akerman and Geoffrey Hinton

Abstract | During learning, the brain modifies synapses to improve behaviour. In the cortex, synapses are embedded within multilayered networks, making it difficult to determine the effect of an individual synaptic modification on the behaviour of the system. The backpropagation algorithm solves this problem in deep artificial neural networks, but historically it has been viewed as biologically problematic. Nonetheless, recent developments in neuroscience and the successes of artificial neural networks have reinvigorated interest in whether backpropagation offers insights for understanding learning in the cortex. The backpropagation algorithm learns quickly by computing synaptic updates using feedback connections to deliver error signals. Although feedback connections are ubiquitous in the cortex, it is difficult to see how they could deliver the error signals required by strict formulations of backpropagation. Here we build on past and recent developments to argue that feedback connections may instead induce neural activities whose differences can be used to locally approximate these signals and hence drive effective learning in deep networks in the brain.

predominantly unsupervised fashion[1,25–27], building representations that make explicit the structure that is only implicit in the raw sensory input. It is natural to wonder, then, whether backprop has anything to tell us about learning in the brain[25,28–30].

Here we argue that in spite of these apparent differences, the brain has the capacity to implement the core principles underlying backprop. The main idea is that the brain could compute effective synaptic updates by using feedback connections to induce neuron activities whose locally computed differences encode backpropagation-like error signals. We link together a seemingly disparate set of learning algorithms into this framework, which we call 'neural gradient representation by activity differences' (NGRAD)[9,27,31–41]. The NGRAD framework demonstrates that it is possible to embrace the core principles of backpropagation while sidestepping many of its problematic implementation requirements. These considerations may be relevant to any brain circuit that incorporates both feedforward and feedback connectivity.

Is the brain implementing backpropagation?

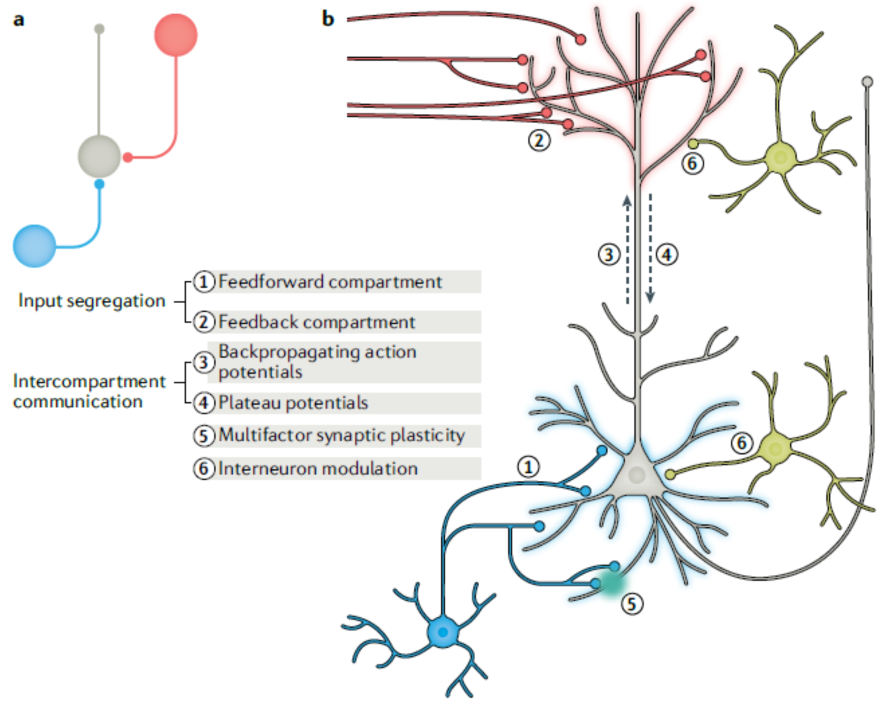
THE NGRAD HYPOTHESIS

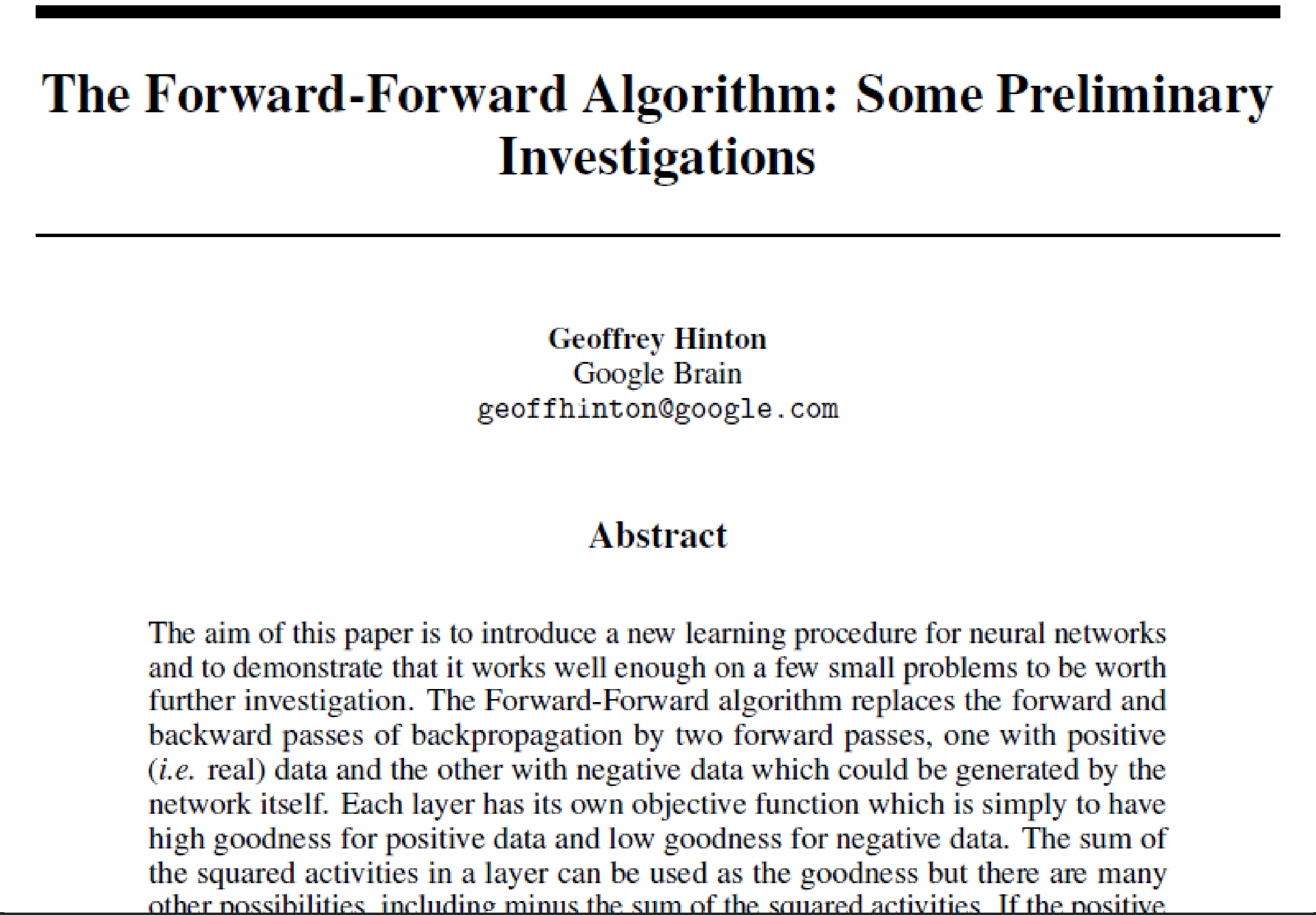
Hinton et al. (2020)

The NGRAD Hypothesis Backprop-like algorithms

It remains physiologically implausible

- Do we compute the derivatives and perform a backwardpass without altering the state of our neurons?





The Forward-Forward Algorithm: Some Preliminary Investigations

Geoffrey Hinton
Google Brain
geoffhinton@google.com

Abstract

The aim of this paper is to introduce a new learning procedure for neural networks and to demonstrate that it works well enough on a few small problems to be worth further investigation. The Forward-Forward algorithm replaces the forward and backward passes of backpropagation by two forward passes, one with positive (*i.e.* real) data and the other with negative data which could be generated by the network itself. Each layer has its own objective function which is simply to have high goodness for positive data and low goodness for negative data. The sum of the squared activities in a layer can be used as the goodness but there are many other possibilities, including minus the sum of the squared activities. If the positive

Is the brain implementing FF-algorithm?

Hinton, 2022

Neural networks are not neural enough

- These algorithms can not answer why-questions
  - Option 1: Causality is an associative illusion
  - Option 2: Neural algorithms are not enough to model the human cognitive system

Neural networks are not neural enough

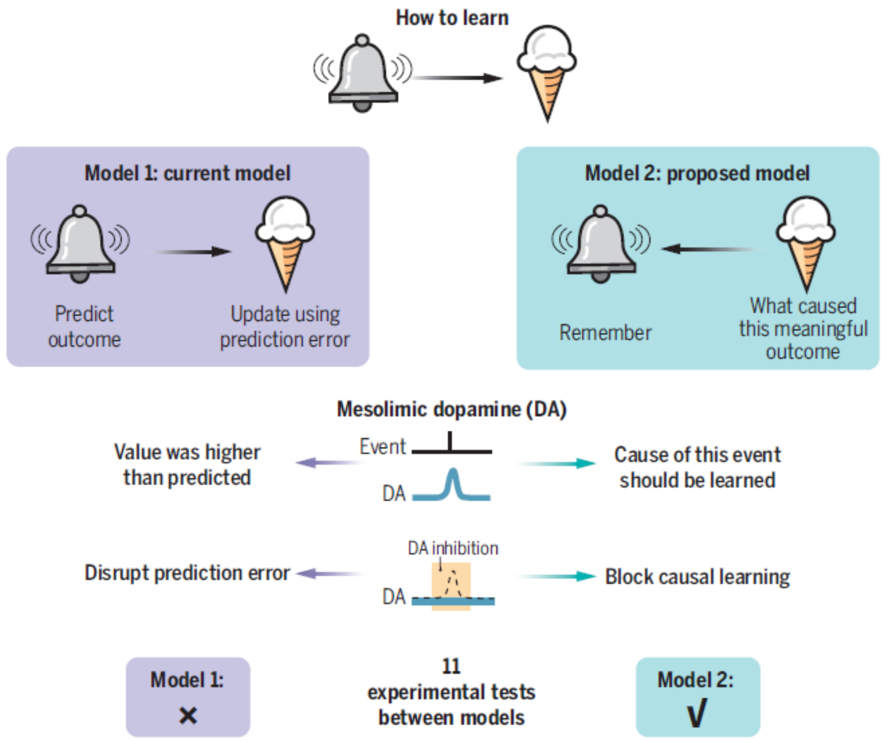
- These algorithms can not answer why-questions

  ~~Option 1: Causality is an associative illusion~~

  Option 2: Neural algorithms are not enough to model the human cognitive system

- Despite the label ‘neural’, it is healthy to question if real neurons, and overall the human cognitive system operate within a neural network framework.

Mesolimbic dopamine release conveys causal associations



- Huijeong Jeong et al. (2022)
- Two Hypotheses:
  - TDRPE: Temporal Difference Reward Prediction Error
  - ANCCOR: Adjusted Net Contingency for Causal Relations
- Dopamine release activity stands for retrospective reasoning.
- Counterfactuals are an intrinsic feature of human cognition.

Concluding remarks

Should Causal AI Rule over Deep Learning?

Originally both methods were designed for different purposes

NEURAL NETWORKS

- The pattern recognition problem

CAUSAL MODELS

- The problem of precisely determining relations between variables and the consequences of those relations

Since both methods obey distinct goals, we can not just combine them. While recognizing that both are extremely useful, it is needed to chose a main goal in order to put them to collaborate.

Yes, Causal AI Should Rule over Deep Learning

1. The Markov Condition is not a reason to reject causal models
   - ✓ It provides a better comprehension.
   - ✓ It is adequate for macroscopic phenomena.
   - ✓ Models which assume CMC are better at decision making and control tasks

2. On cognitive grounds:
   - ✓ Our cognitive system does not obey a machine learning neural algorithm
   - ✓ There is evidence supporting that counterfactual reasoning is anchored in low-level synaptic information transmission.

Causal knowledge is deeper than Deep Learning

Thank you!

- CaDis 2023