
What it means for a machine to learn or think like a person?

Abhishek Tiwari 

Citation: A. *Tiwari*, "What it means for a machine to learn or think like a person?", Abhishek Tiwari, 2017. [doi:10.59350/h096j-qce52](https://doi.org/10.59350/h096j-qce52)

Published on: June 18, 2017

In a recent article, Lake et al.¹ examine what it means for a machine to learn or think like a person. They argue that contemporary AI techniques are not biologically plausible hence not scalable to the extent that will enable a machine to learn or think like a person. For instance, most neural networks use some form of gradient-based (e.g., backpropagation) or Hebbian learning. Trying to build human-like intelligence from scratch using backpropagation, deep Q-learning or any stochastic gradient descent weight update rule is unfeasible regardless of how much training data are available. When comparing the human learning and the current best algorithms in AI i.e. Neural Networks – human learn from fewer data and generalize in richer and more flexible ways. Zero-shot and one-shot learnings are a hallmark of human learning and thinking. On the other end, Neural Networks are notoriously data and compute hungry.

Core cognitive ingredients

Human-like intelligence requires perception that builds upon and integrates with intuitive physics, intuitive psychology, and compositionality glue together by causality. Perception without these key ingredients, and absent the causal glue that binds them together, can lead to revealing errors as illustrated in the following image. In the following example, image captions are generated by a deep neural network which detected key objects in a scene correctly but failed to understand the physical forces at work, the mental states of the people, or the causal relationships between the objects – in other words, it does not build the right causal model of the data.



Figure 1: How a machine perceives scenes without intuitive physics, intuitive psychology, compositionality, and causality? Image credits: Gabriel Villena Fern´andez (left), TVBS Taiwan / Agence France-Presse (middle) and AP Photo / Dave Martin (right).

Lake et al. propose a set of core cognitive ingredients for building more human-like learning and thinking machines. They argue that these machines should,

¹[Building Machines That Learn and Think Like People](#)

- develop early developmental cognitive capabilities such as understanding of intuitive theories of physics and psychology
- build causal models of the world that support explanation and understanding and not just recognizing patterns
- harness compositionality and learning-to-learn to rapidly acquire and generalize knowledge to new tasks and situations

In addition, authors also cover prospects for integrating deep learning with these core cognitive ingredients. Some of the integration are already happening. For instance, recent work fusing neural networks with lower-level building blocks from classic psychology and computer science (attention, working memory, stacks, queues) etc.

Intuitive physics and psychology

Infants have a rich knowledge of intuitive physics and this is acquired without any formal understanding of physics or mathematics behind these concepts. For instance, in their first-year infants have already developed different expectations for rigid bodies, soft bodies, and liquids. It may be difficult to integrate object and physics-based primitives into deep neural networks (Deep Q-learning), but the payoff in terms of learning speed and performance could be great for many tasks. Similarly, in terms of intuitive psychology, an infants' ability to distinguish animate agents from inanimate objects is remarkable. There are several ways that intuitive psychology could be incorporated into contemporary deep learning systems.

Causality

Authors suggest that ability to explain observed data through the construction of causal models of the world makes human learning and thinking so unique. These casual models are used to explain intuitive physics and psychology. In this perspective, the purpose of learning is to extend and enrich these casual models. Incorporating causality may greatly improve these deep learning models, but again most of these efforts are in early stage and do not reflect the true casual model building process. For instance, the attentional window is only a crude approximation to the true causal process of drawing with a pen.

Compositionality and learning-to-learn

People's ability to compose concepts as well objects to create infinite representations using primitive building blocks is a key element in the process to acquire and generalize knowledge on new tasks and situations. To achieve human-like intelligence machines need a similar type of compositionality.

The human brain can make inferences that go far beyond the data and strong prior knowledge. People acquire this type of prior knowledge through a process called *learning-to-learn*. Learning-to-learn approaches enable learning a new task (or a new concept) through previous or parallel learning of other related tasks (or other related concepts) in an accelerated manner. In machine learning, transfer learning and representation learning are closest concepts which can we relate to the human learning-to-learn experience.

Generally, compositionality and learning-to-learn apply together, but there are forms of compositionality that rely less on previous learning, such as the bottom-up parts-based.

To capture more human-like learning-to-learn process in deep networks and other machine learning approaches, we need to adopt the more compositional and causal forms of representations. Current approaches of transfer learning and representational learning have not yet led to systems that learn new tasks as rapidly and flexibly as humans do. For instance, Deep Q-learning system developed for playing Atari games have had some impressive successes in transfer learning, but it still has not come close to learning to play new games as quickly when compared to a human being².

Cognitive and neural inspiration

Finally, authors suggest that theories of intelligence should start with neural networks. The human brain seems to follow both model-based and model-free reinforcement learning process. The role of intrinsic motivation and drive in human learning and behavior is considered important. Deep reinforcement learning, which until now was very much focused on external rewards and punishment, is starting to address intrinsically motivated learning³⁴ by reinterpreted rewards according to the internal value of the agent. Unfortunately, at this stage neuroscience does not offer any definitive evidence for or against the backpropagation.

²[Actor-Mimic: Deep Multitask and Transfer Reinforcement Learning](#)

³[Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation](#)

⁴[Variational information maximisation for intrinsically motivated reinforcement learning](#)