The Pretense of Knowledge: On the insidious presumptions of Artificial Intelligence

JORDAN OTT

Web: https://sites.google.com/view/jordan-ott

In 1974 Friedrich Hayek received the Nobel Prize in economics for his theory on how changing prices convey information throughout an economy. In his acceptance speech, *The Pretense of Knowledge*, he argued against a variety of economic approaches. Complex settings for which those in his own field failed to grasp the ineptitude of certain empirical results in relation to the economy as a whole. His timeless declaration proves ever more relevant in modern society, with a special application to the field of Artificial Intelligence. In this article, we examine why it is unlikely for Artificial Intelligence to make real advances while those writing the programs grasp ever tighter for control.

1. INTRODUCTION

The Pretense of Knowledge was a specific application of more general ideas presented previously by Hayek in *The Use of Knowledge in Society* (Hayek 1945). What these declarations share is made explicit by the second chapter of *Law Legislation and Liberty*, Cosmos and Taxis (Hayek 1982). The underlying idea is that of planned or unplanned orders, made or grown systems, top-down control or bottom up emergence, planned chaos or spontaneous order, *τάξης* (taxis) or *κόσμος* (cosmos). This unifying theme has been critically vetted in the domain of social systems and economic markets. However, it has yet to make an impact on the field of Artificial Intelligence.

In this article, we examine the field under the lens of spontaneous orders and complexity. This realization yields that, above all, Artificial Intelligence has engaged in an almost century-long misclassification. What the field purports to solve—intelligence—is in fact impossible given current and historical methods. It is this assumption that we seek to expose in Sections 2 and 3 while providing an alternative direction in Section 4.

2. HISTORY

The history of Artificial Intelligence is rife with presumptions of knowledge and top-down style solutions. An analysis of methodologies employed throughout the field—past and present—will shed light on this problem. The ambition of early pioneers is well captured in the words of Herbert Simon, "Machines will be capable, within 20 years, of doing any work a man can do."

It was 1965 when he made that claim. Shortly thereafter, in 1966, Seymour Papert and others at MIT proposed they would solve vision during that summer (Papert 1966).

These presumptuous claims resulted directly from a pretense of knowledge. The audacious claims of the best and brightest researchers are not the issue. Instead, it is the notion that discretizing high-level, visible, characteristics of complex systems and implementing them via centrally planned rules, heuristics, and cost functions will be synonymous with the system as a whole. This is the τάξης view of intelligence and its artificial creation.

Starting in 1956, Allen Newell and Herbert Simon invented a "Logical Theorist," which leveraged symbolic reasoning to solve geometric proofs (Newell 1956). Simon described it as a "thinking machine." When in reality, it operated on a set of heuristics prescribed by its programmers. The next year produced the "General Problem Solver" (Newell 1959). Peter Norvig thought so highly of this achievement that he said, "there are now in the world machines that think, that learn and create."

Papert's confidence for his Summer Vision Project was so high, undergraduates were assigned to the task. The work conducted that summer, and long after, focused on template schemes and hierarchical models (Papert 1966). This approach translated to humans designing features they believed to be important in recognizing visual objects.

Following this trend, there was a rise in symbolic logic, often termed Classical AI. In programs like CYC (Lenat 1995), engineers explicitly write millions of hand programmed rules in an attempt to codify human knowledge and common sense. The paper boasts, "a person-century of effort has gone into building CYC" and "10^6 commonsense axioms have been handcrafted for and entered into CYC's knowledge base." While humans rely on symbolic logic at an abstract level, this is not synonymous with a system composed of logic statements. The space of information was, and is, too vast to be codified in a set of logic statements. Though CYC is an impressive engineering feat, it offers no clear path to intelligence.

From the geometric puzzle solver, ANALOGY (Evans 1964), to the ELIZA chatbot (Weizenbaum 1966), and SHRDLU language processor (Winograd 1972), the story is the same. Hard-coded rules, heuristics, and symbolic manipulation to achieve the appearance of intelligent behavior. The quarrel is not, as we will see, with heuristics or symbolic logic. These are essential tools in the arsenal of any computer scientist. The trouble derives from the mindset of the implementor.

3. PERPETUATED METHODOLOGY

Assumptions made in the past can appear comical in hindsight. Unfortunately, we are in tomorrow's past, and this methodology continues to perpetuate the field. Deep learning, the current method in vogue, is lauded because we no longer assume programmatic declarations of human knowledge. The claim is that because we learn directly from data, we sidestep the aforementioned faux pas. This line of thought is a grave red herring. Deep learning is just as vulnerable as the CYCs or Logical Theorists of the past. Though we no longer rely on rule-based systems, the declarations are still explicit, and the systems centrally planned. This methodological shortcoming is hidden, once again, behind a thick veil—the pretense of knowledge.

The ImageNet challenge contains millions of images and thousands of categories (Russakovsky et al. 2015). Models are trained for weeks and months to achieve optimal error. Autonomous vehicle companies pour millions of dollars into mapping roads and capturing human driver training data. Speech recognition systems rely on manually labeled corpora, with each audible phrase paired with its textual equivalent. Natural language systems consume large chunks of the internet to generate convincing sentences (Radford et al. 2019). These methods share an instilled discretization of intelligent behavior, fragmented from the whole, and explicitly solved in isolation. In terms of system abstractions, Papert's Vision Project, with designed templates and heuristics, is no different from a convolutional network trained on ImageNet. While the Vision Project devised templates, deep networks rely on explicit cost functions. Though the technical details differ, the underlying methodology is the same.

Those that recognize these deficiencies often turn to reinforcement learning as a solution. This methodology offers convincing results in which agents can solve various problems from autonomous vehicles, navigation, and cooperative games. While reinforcement learning is a step in the right direction, its ail-

ments are nonetheless similar (Ott 2019). Artificial agents are constrained to specific tasks, with the sole objective of maximizing rewards over time. These rewards are defined by the environment, or worse by the researcher! If we believe this paradigm is any different from the handcrafted heuristics of the past we are fooling ourselves:

> It seems to me that this failure of the economists to guide policy more successfully is closely connected with their propensity to imitate as closely as possible the procedures of the brilliantly successful physical sciences—an attempt which in our field may lead to outright error. It is an approach which has come to be described as the "scientistic" attitude—an attitude which, as I defined it some thirty years ago, "is decidedly unscientific in the true sense of the word, since it involves a mechanical and uncritical application of habits of thought to fields different from those in which they have been formed" (Hayek 1974).

Much like Hayek's critique of economics, the same is true for Artificial Intelligence. The field relies on methods derived from statistics and numerical optimization, which is decidedly unscientific in its application to intelligence research.

"Scientism" as Hayek calls it, is the desire to abstract a system and precisely quantify aspects of it. Following this approach is understandable from the AI researcher's perspective, given our position in the scientific community as we are surrounded by fields—biology, chemistry, and physics—that make system-level abstractions and give precise predictions about outcomes. Consequently, cost functions are a natural solution, as they provide an exact quantification of the degree to which the system has *learned*. However, through system abstraction and quantification, we are likely to lose critical information so as to be no longer relevant to the original system. This process's technical underpinnings are captured in the discretize and conquer approach, which we detail in the following subsection.

3.1 DISCRETIZE AND CONQUER

Neurons in the brain form the cortical substrate from which intelligent behavior is an emergent property. Mathematically, one can regard the brain as a function and intelligence as an output of that function. Such that the manifold of intelligence is described by the function, $M = f(S_t ; \Theta_t)$. Where the manifold, M, is the output from the function f. S_t is all input stimuli to the system (all afferent sensory inputs—touch, vision, sound), and Θ_t is the internal state of the brain at time t (all synaptic weights, voltage differentials of neurons, ion flows, gated channels, protein formations, etc).
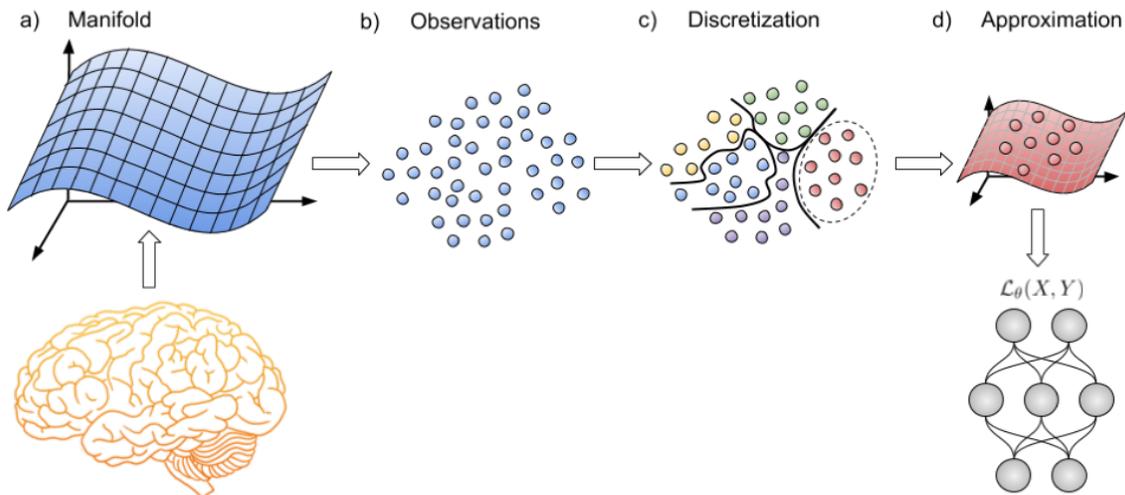
Figure 1: Discretize and conquer a) Interactions within cortical networks produce emergent phenomena—intelligence—described by an unknowable manifold. b) The manifold of intelligence is unknowable, however, we receive observations from it through actions and behavior. c) Arbitrary bounds are placed on the observations discretizing them. d) A cost function is used to describe the observations. And gradient descent approximates the cost function.

Figure 1a conveys this pictorially, with the brain producing some manifold over possible brain states, Θ, and input stimuli, S. In reality this manifold is unknowable from a practical standpoint as well as a computational one. Practically, it is not currently possible to record all biological details—the activity of all neurons, their synaptic weights, electrical and chemical gradients, etc. Computationally, modeling every detail could be done given sufficient computing resources but such intricacy could not run in real time. For all intents and purposes, the manifold is not known.

As a result, we must rely on incomplete observations from the manifold. These observations are high-level attributes or behaviors that are emergent products of the underlying system. Figure 1b depicts this by showing single points that represent observations realized from the full manifold. For example, intelligent systems can perceive through vision, communicate through language, reason through abstractions, and act through planning. These are all visible observations from the manifold. What is not visible is the processes, interactions, and dynamics that produce these high-level attributes. Thus the characteristics we ascribe to intelligent beings are only the byproducts of the system from which intelligence can emerge, they are not indicative or defining features of intelligence but merely the result of it.

With a large collection of observations, it becomes natural to cluster and group them according to kind. This is the discretization stage. For example, one may observe that intelligent agents classify objects, identify their location in space, and label them semantically, all from visual stimuli. Grouping these observations together forms the basis of a vision system. Figure 1c shows boundaries placed discretizing all observations—the green dots may refer to language abilities, red to vision, purple to search, blue to planning, and yellow to speech.

Once the attributes of intelligent agents have been placed in identifiable groups, one seeks to solve each disparate task. This is the conquer stage. A new manifold now describes the red dots clustered together in the discretization stage (Figure 1c). One that a deep learning system attempts to approximate through the use of a cost function and gradient descent. In Figure 1d the vision manifold is approximated by a deep neural network trained with a cost function.

The discretize and conquer (DAC) methodology has been fruitful for narrow, domain-specific, engineered applications—top-down planning, *τάξης*. Much less fruitful is DAC's ability to create general intelligence or any form of emergent phenomena. The DAC methodology shows a complete disregard for the process that generates intelligence—spontaneous order, *κόσμος*. The field of artificial intelligence has focused on discretized subproblems, all the while ignoring that intelligence is the emergent result of a complex system. It is an unplanned order, resulting not from central cost functions or engineered heuristics, but individual agents (neurons) acting on local signals. To only look at visible characteristics of intelligent agents is to ignore the process that makes these attributes possible.

Artificial intelligence is into its eighth decade as a scientific field. Over this period, the field has seen the invention of a great diversity of algorithms, accomplishing tremendous feats in their time. These algorithms have followed a consistent underlying methodology of discretization. The ubiquity of this approach has produced algorithms that are technically diverse but methodologically homogeneous.

4. A *COMPLEX* SOLUTION

Intelligence—much like economies—is a result of the interactions within large complex systems. Often we cannot measure the system in a broad sense. The definitions for intelligence are inadequate, and the metrics used to assess it are even worse. As a consequence, the field results to the DAC approach. Abstractions come in the form of direct cost functions that our models optimize. Intelligent behaviors are enumerated, and corresponding cost functions are designed for each one. Vision, speech, audition, navigation, and planning all follow this paradigm. The discretization approach yields models with the appearance of intelligence but without understanding. We must consider what is fundamentally different about the economy of neural circuits in which intelligence can emerge.

Much like the macro level, shortcomings are evident in the micro-level as well. Neuroscience generates enormous amounts of detailed observational data. Where regions are discretized and studied in isolation. Unfortunately, the whole cannot be understood by observing the individual. This principle is true of the economy, of ant colonies, and as well as of brains. We will not be able to understand intelligence by observing single actors. Neurons are individual agents in a local-decentralized system. They compete for resources with their neighbors while cooperating in order to achieve beneficial results for the whole. This concept is perfectly summarized in the words of Friedrich Engels, "For what each individual wills is obstructed by everyone else, and what emerges is something that no one willed." Engels said this in reference to an economy, however, the application to neuroscience and the emergence of intelligence are equally satisfying.

Just as markets coordinate large groups of actions without any one individual being in control, neuronal-markets do the same. No one neuron, or group of neurons, is in charge of all the others.

In his closing remarks, Hayek states:

... he cannot acquire the full knowledge which would make mastery of the events possible. He will, therefore, have to use what knowledge he can achieve, not to shape the results as the craftsman shapes his handiwork, but rather to cultivate a growth by providing the appropriate environment, in the manner in which the gardener does this for his plants. (Hayek 1974).

What Hayek's words translate to is a change in how we think about building intelligent systems. In no way can we achieve the human brain's complex behavior by top-down control—whether from hard-coded heuristics of the past or explicit cost functions of the present. Instead, we must cultivate an environment comprised of decentralized local actors (Ott 2020) where each actor pursues his own self-interest—governed by local rules. It is the incentives provided to each individual, which give rise to appropriate dynamics in which intelligent behavior can emerge.

At the current moment in history, it is clear we are capable of designing solutions to particular problems—vision, speech, audition. Given sufficient training data, a deep network or statistical model performs

remarkably well. What remains unclear is the path toward human-level intelligence, that offers a breadth of diversity across all aspects of life. As intelligence is the result of a complex system, it is unlikely for the field to make real advancements while those writing the programs grasp ever tighter for control.

REFERENCES

Evans, T. G. 1964. A heuristic program to solve geometric-analogy problems. In: *Proceedings of the April 21-23, 1964, spring joint computer conference* (pp. 327-338).

Hayek, F. A. 1945. The use of knowledge in society. *The American economic review, 35*(4): 519-530.

_____. 1974. The pretence of knowledge. https://www.nobelprize.org/prizes/economic-sciences/1974/hayek/lecture/

_____. 1982. Cosmos and Taxis. In: *Law Legislation and Liberty*, Vol. 1. London: Routledge.

Lenat, D. B. 1995. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM, 38*(11): 33-38.

Newell, A., and Shaw, J. C. 1959. A variety of intelligent learning in a general problem solver. *RAND Report P-1742*, July 6.

Newell, A., and Simon, H. 1956. The logic theory machine—A complex information processing system. *IRE Transactions on information theory*, 2(3):61-79.

Ott, J. 2019. Questions to Guide the Future of Artificial Intelligence Research. *arXiv preprint arXiv:1912.10305*.

_____. 2020. Giving Up Control: Neurons as Reinforcement Learning Agents. *arXiv preprint arXiv:2003.11642*.

Papert, S. A. 1966. The summer vision project. https://dspace.mit.edu/handle/1721.1/6125

Radford, A., et al. 2019. Language models are unsupervised multitask learners. *OpenAI Blog,* 1(8): 9.

Russakovsky, O., et al. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211-252.

Weizenbaum, J. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.

Winograd, T. 1972. SHRDLU: A System for Dialog. https://hci.stanford.edu/~winograd/shrdlu/