



## Searching for the Algorithms Underlying Life

The biological world is computational at its core, argues computer scientist Leslie Valiant. His “ecorithm” approach uses computational concepts to explore fundamental mysteries of evolution and the mind.

*By John Pavlus*



**Video:** Leslie Valiant explains the term “ecorithm”.

To the computer scientist [Leslie Valiant](#), “machine learning” is redundant. In his opinion, a toddler fumbling with a rubber ball and a deep-learning network [classifying cat photos](#) are both learning; calling the latter system a “machine” is a distinction without a difference.

Valiant, a computer scientist at Harvard University, is hardly the only scientist to assume [a fundamental equivalence between the capabilities of brains and computers](#). But he was one of the first to formalize what that relationship might look like in practice: In 1984, his “probably approximately correct” (PAC) model mathematically defined the conditions under which a mechanistic system could be said to “learn” information. Valiant [won the A.M. Turing Award](#) — often called the Nobel Prize of computing — for this contribution, which helped spawn the field of computational learning theory.

Valiant's conceptual leaps didn't stop there. In a 2013 book, also entitled "[Probably Approximately Correct](#)," Valiant generalized his PAC learning framework to encompass biological evolution as well.

He broadened the concept of an algorithm into an "ecorithm," which is a learning algorithm that "runs" on any system capable of interacting with its physical environment. Algorithms apply to [computational systems](#), but ecorithms can apply to [biological organisms](#) or [entire species](#). The concept draws a computational equivalence between the way that individuals learn and the way that [entire ecosystems evolve](#). In both cases, ecorithms describe adaptive behavior in a mechanistic way.

Valiant's self-stated goal is to find "mathematical definitions of learning and evolution which can address all ways in which information can get into systems." If successful, the resulting "theory of everything" — a phrase Valiant himself uses, only half-jokingly — would literally fuse life science and computer science together. Furthermore, our intuitive definitions of "learning" and "intelligence" would expand to include not only non-organisms, but non-individuals as well. The "wisdom of crowds" would no longer be a mere figure of speech.

*Quanta Magazine* spoke with Valiant about his efforts to dissolve the distinctions between biology, computation, evolution and learning. An edited and condensed version of the interview follows.

## **QUANTA MAGAZINE: How did you come up with the idea of "probably approximately correct" learning?**

LESLIE VALIANT: I belonged to the theoretical computer science community, specializing in computational complexity theory, but I was also interested in artificial intelligence. My first question was: Which aspect of artificial intelligence could be made into a quantitative theory? I quickly settled on the idea that it must be learning.

At the time I started working on it [in the 1980s], people were already investigating machine learning, but there was no consensus on what kind of thing "learning" was. In fact, learning was regarded with total suspicion in the theoretical computer science community as something which would never have a chance of being made a science.

On the other hand, learning is a very reproducible phenomenon — like an apple falling to the ground. Every day, children all around the world learn thousands of new words. It's a large-scale phenomenon for which there has to be some quantitative explanation.

So I thought that learning should have some sort of theory. Since statistical inference already existed, my next question was: Why was statistics not enough to explain artificial intelligence? That was the start: Learning must be something statistical, but it's also something computational. I needed some theory which combined both computation and statistics to explain what the phenomenon was.

## **So what is learning? Is it different from computing or calculating?**

It is a kind of calculation, but the goal of learning is to perform well in a world that isn't precisely modeled ahead of time. A learning algorithm takes observations of the world, and given that information, it decides what to do and is evaluated on its decision. A point made in my book is that all the knowledge an individual has must have been acquired either through learning or through the evolutionary process. And if this is so, then individual learning and evolutionary processes should have a unified theory to explain them.

## **And from there, you eventually arrived at the concept of an “ecorithm.” What is an ecorithm, and how is it different from an algorithm?**

An ecorithm is an algorithm, but its performance is evaluated against input it gets from a rather uncontrolled and unpredictable world. And its goal is to perform well in that same complicated world. You think of an algorithm as something running on your computer, but it could just as easily run on a biological organism. But in either case an ecorithm lives in an external world and interacts with that world.

## **So the concept of an ecorithm is meant to dislodge this mistaken intuition many of us have that “machine learning” is fundamentally different from “non-machine learning”?**

Yes, certainly. Scientifically, the point has been made for more than half a century that if our brains run computations, then if we could identify the algorithms producing those computations, we could simulate them on a machine, and “artificial intelligence” and “intelligence” would become the same. But the practical difficulty has been to determine exactly what these computations running on the brain are. Machine learning is proving to be an effective way of bypassing this difficulty.

Some of the biggest challenges that remain for machines are those computations which concern behaviors that we acquired through evolution, or that we learned as small children crawling around on the ground touching and sensing our environment. In these ways we have acquired knowledge that isn’t written down anywhere. For example, if I squeeze a paper cup full of hot coffee, we know what will happen, but that information is very hard to find on the Internet. If it were available that way, then we could have a machine learn this information more easily.

## **Can systems whose behavior we already understand well enough to simulate with algorithms — like solar systems or crystals — be said to “learn” too?**

I wouldn’t regard those systems as learning. I think there needs to be some kind of minimal computational activity by the learner, and if any learning takes place, it must make the system more effective. Until a decade or two ago, when machine learning began to be something that computers could do impressively, there was no evidence of learning taking place in the universe other than in biological systems.

## **How can a theory of learning be applied to a phenomenon like biological evolution?**

Biology is based on protein expression networks, and as evolution proceeds these networks become modified. The PAC learning model imposes some logical limitations on what could be happening to those networks to cause these modifications when they undergo Darwinian evolution. If we gather more observations from biology and analyze them within this PAC-style learning framework, we should be able to figure out how and why biological evolution succeeds, and this would make our understanding of evolution more concrete and predictive.

## **How far have we come?**

“We will understand the intelligence we put into machines in the same way we understand the physics of explosives.”

We haven't solved every problem we face regarding biological behavior because we have yet to identify the actual, specific ecorithms used in biology to produce these phenomena. So I think this framework sets up the right questions, but we just don't know the right answers. I think these answers are reachable through collaboration between biologists and computer scientists. We know what we're looking for. We are looking for a learning algorithm obeying Darwinian constraints that biology can and does support. It would explain what's happened on this planet in the amount of time that has been available for evolution to occur.

**Imagine that the specific ecorithms encoding biological evolution and learning are discovered tomorrow. Now that we have this precise knowledge, what are we able to do or understand that we couldn't before?**

Well, we would understand where we came from. But the other extrapolation is in bringing more of psychology into the realm of the computationally understandable. So understanding more about human nature would be another result if this program could be carried through successfully.

**Do you mean that computers would be able to reliably predict what people will do?**

That's a very extreme scenario. What data would I need about you to predict exactly what you will be doing in one hour? From the physical sciences we know that people are made of atoms, and we know a lot about the properties of atoms, and in some theoretical sense we can predict what sets of atoms can do. But this viewpoint hasn't gone very far in explaining human behavior, because human behavior is just an extremely complicated manifestation of too many atoms. What I'm saying is that if one has a more high-level computational explanation of how the brain works, then one would get closer to this goal of having an explanation of human behavior that matches our mechanistic understanding of other physical systems. The behavior of atoms is too far removed from human behavior, but if we understood the learning algorithms used in the brain, then this would provide mechanistic concepts much closer to human behavior. And the explanations they would give as to why you do what you do would become much more plausible and predictive.

**What if the ecorithms governing evolution and learning are unlearnable?**

It's a logical possibility, but I don't think it's likely at all. I think it's going to be something pretty tangible and reasonably easy to understand. We can ask the same question about fundamental unsolved problems in mathematics. Do you believe that these problems have solutions that people can understand, or do you believe that they're beyond human comprehension? In this area I'm very confident — otherwise I wouldn't be pursuing this. I believe that the algorithms nature uses are tangible and understandable, and won't require intuitions that we're incapable of having.

**Many prominent scientists are voicing concerns about the potential emergence of artificial "superintelligences" that can outpace our ability to control them. If your theory of ecorithms is correct, and intelligence does emerge out of the interaction between a learning algorithm and its environment, does that mean that we ought to be just as vigilant about the environments where we deploy AI systems as we are about the programming of the systems themselves?**

If you design an intelligent system that learns from its environment, then who knows — in some environments the system may manifest behavior that you really couldn't foresee at all, and [this](#)

[behavior may be deleterious](#). So you have a point. But in general I'm not so worried about all this talk about the superintelligences somehow bringing about the end of human history. I regard intelligence as made up of tangible, mechanical and ultimately understandable processes. We will understand the intelligence we put into machines in the same way we understand the physics of explosives — that is, well enough to be able to render their behavior predictable enough that in general they don't cause unintended damage. I'm not so concerned that artificial intelligence is different in kind from other existing powerful technologies. It has a scientific basis like the others.

*This article was reprinted on [Wired.com](#).*