

Probably Approximately Correct. By Leslie Valiant, *Basic Books, New York, 2013, 208 pages, \$26.99.*

In 2010, Leslie Valiant won the Turing Award, the “Nobel prize of computer science.” The citation mentioned a number of contributions, the best known of which is the theory of probably approximately correct (PAC) learning, which Valiant first proposed in the early 1980s. In his new book *Probably Approximately Correct*, Valiant discusses the theory of PAC learning and its applications to artificial intelligence; he also introduces a more specialized concept—“evolvable learning”—which, he claims, characterizes Darwinian evolution.

PAC learning is an elegant framework that characterizes the objective and the computational difficulty of carrying out certain kinds of learning. Imagine a person or a computer program that is trying to determine the characteristics of mammals by viewing a sample of individual animals, each correctly labeled either “a mammal” or “not a mammal.” Suppose, as is likely, that the sample does not include any platypuses or echidnas. The learner might then come up with a rule like “hair, warm-blooded, suckles their young, live-bearing.” Although not quite correct, because platypuses and echidnas are mammals that lay eggs, this rule is *approximately* correct, because it works correctly for every animal except platypuses and echidnas.

Given a large enough random sample, can we develop a learning algorithm that always generates an approximately correct answer? No, because we could be very unlucky in the constitution of our sample. For instance, there is a small but non-zero chance that in a random sample of animals, all the mammals will be platypuses and echidnas. If so, the learner might reasonably posit the additional condition that mammals are egg-laying; that rule would not be approximately correct. So what we can hope for is an algorithm that is *probably* approximately correct—an algorithm that, given a random sample, will probably come up with a rule that is approximately correct.

One particularly elegant feature of PAC learning is that, because the same distribution is used in the “probably” and the “approximately,” the algorithm does not have to make any prior assumptions about the distribution used to select the sample. Consider, for example, an anomalous distribution over mammals that chooses platypuses and echidnas 99.9% of the time. If samples are being generated according to that distribution, the learner is likely to come up with a rule that requires mammals to be egg-laying. But now that rule is approximately correct, relative to this anomalous distribution.

Another attractive feature of the definition, from the standpoint of mathematical analysis, is that it provides a number of numerical parameters to play with. There is $1 - \delta$, the accuracy of the rule; $1 - \epsilon$, the probability of attaining an approximately correct rule; n , the number of samples; s , the size of an individual instance; and t , the running time. *PAC-learnability* is a property of a collection \mathcal{C} of categories in some space of instances \mathcal{X} . The collection \mathcal{C} of all categories definable by a simple conjunction of unary properties like “warm-blooded and hairy and young-suckling” is *PAC-learnable* if, given large enough values of n and t , there is an algorithm that will probably find an approximately correct definition of each category in \mathcal{C} for any distribution over \mathcal{X} and for arbitrarily small δ and ϵ . Finally, following standard practice in complexity theory, Valiant characterizes a problem as “tractable” if it is solvable by an algorithm whose running time is bounded by a polynomial function of the parameters. Thus, the collection \mathcal{C} is *efficiently PAC-learnable* if the running time t and the number of samples n are bounded by a polynomial in $1/\epsilon$, $1/\delta$, and s . Valiant and other researchers have proved a rather

small number of learning problems to be efficiently PAC-learnable.



Despite the title, the focus of Valiant’s book is not PAC learning, but rather a question motivated by Darwinian evolution: Is it plausible that life could have evolved to its current state within the time that the earth has existed, if the mechanism for evolution is Darwinian? Valiant approaches the analysis of this problem by identifying Darwinian evolution with a narrow class of algorithms called “statistical query” algorithms. Algorithms in this class can use only statistical information about the sample as a whole; they are not allowed to use information about individual instances. A class of learning problems is considered evolvable if it satisfies the conditions of PAC learning when the class of algorithms under consideration is limited to statistical query algorithms. The theory of evolvable learning is much more recent and less developed than the theory of PAC learning, but some theoretical results have been obtained.

Valiant applies this theory to the question of the plausibility of Darwinism as follows. The terrestrial environment confronts each species with the challenge of maximizing its “performance” (essentially its fitness). The adaptation of a species through evolution is conceptualized as the execution of a learning algorithm (Valiant calls this an “ecorithm”) to solve that problem. The terrestrial environment, indeed, has a collection of problems that it could set to the species, and the ecorithm must be such that the species will successfully use it to solve any problem in the collection. Valiant is not very explicit about the exact relation of the complexity classes to the theories of evolution. Presumably, if the class of problems set by the environment is found to be evolvable, then Darwinism is plausible; if the collection is PAC-learnable but not evolvable, then some other evolutionary mechanism, such as Lamarckianism, should be sought; if it is not even PAC-learnable, then we must fall back on intelligent design, or on the anthropic principle. This kind of analysis presumably could not distinguish an evolutionary theory that requires 40 million years from one that requires 4 billion years, but it might distinguish these theories from those requiring $10^{20,000}$ years.

I am not at all an expert, but it does not seem to me that Valiant makes a very cogent case that this is a useful abstract model for this question. To what extent is it reasonable to view adaptation as the execution of an algorithm, and specifically a statistical query algorithm? In particular, to what extent is this a reasonable view of the early stages in the emergence of life, the formation of self-replicating biological chemicals? To what extent is it reasonable to view the environment as potentially posing any of the problems in a collection, and to require that the ecorithm be able to solve all of them? How is one to find a characterization of the class of problems that the environment could potentially present? Moreover, the devil is very much in the details in this kind of analysis: Small changes to the formulation of a problem can make a large difference in its learnability. Unless the abstraction reflects reality with high fidelity, an analysis based on that abstraction may well be useless.



The book also includes an extended discussion of machine learning and artificial intelligence, although the view of machine learning is disappointingly narrow. PAC learning applies in a natural way to only a rather restricted, though very important, class of learning problems: learning categories from labeled data (“supervised” learning), in cases in which there is an exact definition (or at least arbitrarily good definitions). If PAC learning is to be viewed as

a universal framework for learning, other forms of learning must then either be shoe-horned into this form, declared unimportant, or ignored. Valiant does some of each. Most strikingly, he specifically denies that unsupervised learning (learning from unlabeled examples, as in grouping examples into clusters) is a useful concept in studying natural learning, though he concedes that it may have some value in machine learning. In one case an errant form of learning sneaks in behind Valiant’s back, when he remarks in passing, in a discussion of learning from positive examples, “We may have figured out that each animal belongs to one species.” On the whole, one suspects that the way we figured this out was by learning it, although the problem of learning propositions like “each animal belongs to one species” does not fit well into the framework of supervised learning of categories.*

PAC learning is only one of many general frameworks that have been proposed for learning, and in many ways, it is one of the most limited. Other frameworks include Bayesian learning, minimum description length learning, the Vapnik–Chervonenkis theory of learning, and classical frequentist statistics. V–C theory is mentioned in an endnote, the others not at all.

PAC learning is an elegant and useful approach to a specific class of learning problems. I am doubtful that an approach of this kind is useful in characterizing evolution; it is certainly not a universal framework applicable to all forms of either natural or machine learning.

*An observer in a state of nature does not directly perceive species; he perceives animals, their features, and their behaviors. Species are a higher-order construct. Learning the proposition “each animal belongs to one species” involves determining that the class of animals is partitioned into categories, that features and behavior of two animals in the same category are generally much more similar than features and behavior of two animals in different categories, and that animals breed only within the same category.