Sources of Artificial Intelligence

Thomas J. Sargent

March 3, 2023

Abstract

This paper describes artificial intelligence and machine learning and how they were invented.

1 Introduction

This essay is about human and artificial intelligence and machine learning. By *artificial* I mean 'non-human'. Before describing *artificial intelligence* and *machine learning*, I'll define *natural* or *human* intelligence in terms of salient classes of *activities* that a combination of innate and learned skills enable intelligent people to perform: *recognizing patterns* and *making choices*. Other aspects of human intelligence are awareness of time and space, as well as sympathy and empathy with other people. Successive generations of parents pass on to their children tools and perspectives that their parents taught them, along with new ideas that they have learned through during their lives. After describing how Galileo, Darwin, and Kepler combined extraordinary innate talents with their text-book knowledge to create scientific breakthroughs, I'll tell how modern researchers have designed computer programs that recognize patterns and make decisions.[1]

While my principle purpose is to describe the machine-learning "forest", I mention many "trees", i.e., a variety of concepts and technicalities that might be new to a general reader. For readers curious to learn more about a particular "tree", I recommend using a good online search engine or references at the end of this essay.

[1] Hayek (2011, Appendix A) discusses other possible interpretations of *natural* and *artificial*.

2 Human Intelligence

Chapter 13 of *The Blank Slate* by cognitive psychologist Steven Pinker (2003) is titled *Out of Our Depths*. Read it if you are thinking about the purposes of education. Steven Pinker provides advice about what to study in high school and college and why, advice based on his understanding of our cognitive disabilities as human beings. He begins by describing things evolution has hard-wired us to do well. During most of our 100,000 years of human pre-history and history, things that we aren't hard-wired to do well weren't important. But modern life has made some of them important. Pinker identifies four such subjects.

1. *Physics*. Theories and evidence about time, space, mass, motion, energy, heat, and light.

2. *Biology*. Theories and evidence of life, birth, and death.

3. *Statistics*. Methods for describing uncertainties and for recognizing and interpreting relative frequencies.

4. *Economics*. Theories and evidence about work, leisure, families, organizations, production, distribution, markets, prices, and quantities.

Today, private and public decisions rest on understanding these subjects for which our hard-wired "intuitions" often fail us. For working purposes, just define "intuition" as how we think about situations that evolution constructed us to understand quickly. I take "common sense" to be a synonym for intuition, things that we think we understand immediately. Steven Pinker describes how hard-wired prejudices, our theories about these subjects, often lead us astray.

Thus, our common sense doesn't help us understand modern physics. According to Richard Feynman and other distinguished physicists, quantum mechanics make no sense. Neither does the general theory of relativity. Pinker tells how we had evolved to make some statistical calculations that helped us when we were hunters and gatherers. These involved probabilities of events that occurred frequently relative to the incidence of important risks that we have to evaluate today. We are not naturally well-equipped to deal with probabilities of events that occur very infrequently. That has been costly in terms of public policy decisions that involve balancing costs and benefits from accepting low probability risks. Pinker describes how evolution gave our ancestors a set of economic theories about production and exchange that do not equip us to understand the division of labor, distribution, markets, middlemen, intermediaries, stabilizing speculation, and profits. Actually, we naturally *misunderstand* modern economic arrangements, with too often tragic consequences that have occurred during recurrent expropriations and pogroms against middlemen and traders, speculators and liquidity providers, people who were often members of ethnic minorities.

Pinker regards education as a technology for compensating for our innate cognitive limits by taking advantage of our innate abilities to learn. He calls for realigning academic curricula to equip us to make decisions today. This means teaching more biology, statistics, and economics and less of other subjects.

2.1 Innate Cognitive Limits and AI

Can "artificial intelligence" (AI) compensate for our cognitive disabilities? A paradox lurks here because the principal technical tools being used to create artificial intelligence and machine learning come from physics, biology, statistics, and economics, the same areas in which we are innately challenged. Pioneers of machine learning and AI compensated for their natural cognitive deficiencies by thoroughly learning and then imaginatively using the best analytical techniques available to them. In the next section, I'll offer two examples of two examples.

3 Pioneers of Machine learning

3.1 Galileo Galilei

Because he accepted the theory of Nicholaus Copernicus (1473 - 1543) that the earth revolves around the sun, Italian mathematician, scientist, physicist, astronomer Galileo Galilei (1564-1642) was arrested by the Inquisition in 1633.[2] In 1603 Galileo deployed what we know call "machine learning". He (1) used some experiments to generate a data set; (2) searched for patterns; (3) reduced the dimension of his data by fitting a *function*; and (4) generalized by applying that function to ordinates beyond his data set. Galileo's approach offers a beautiful example for what machine learning and artificial intelligence are all about.

I refer to Galileo's "inclined plane" experiments. Galileo wanted to know rates at which balls of different weights fall toward the earth. Perhaps you think: "That's easy, just apply Newton's laws of gravity." Not so fast: Newton was born in 1643 and Galileo died in 1642. The prevailing theory of falling bodies was Aristotle's from 2000 years earlier: heavier bodies fall faster than lighter ones.

[2]Copernicus was also a monetary economist. In 1517 he formulated an early version of Gresham's Law, the notion a commodity money that is overvalued at the mint will drive out a money that is undervalued at the mint.

Galileo wanted to check Aristotle's theory empirically. Why not just drop balls of different weights and measure how fast they fall? Galileo couldn't do that because balls of *all* weights fell much faster than the crude clocks then available could accurately measure. Therefore, Galileo constructed a smooth inclined plane and adjusted the angle of incline to slow a falling ball enough so that he could use a crude clock to measure a ball's velocity as it travelled along the plane. For a plane of length ℓ and height h, the ratio $\frac{h}{\ell}$ determines the angle of the plane. Galileo dropped each ball and carefully measured distances d along the plane that the had ball traveled at various times t elapsed after the ball had been dropped. For each ball of a given weight, he made a table with two columns in which he recorded t_i and d_i, $i = 1, \ldots, n$ for his n measurement times. For a given experiment, he then plotted d_i against t_i. He conducted experiments for a variety of balls of different weights with different settings of ℓ and h (i.e., different angles for the inclined plane). He stared at his graphs and noticed a pattern: for *all* of the graphs, d was proportional to t^2 - the distance traveled was proportional to the square of the elapsed time, independently of the weight of the ball. He inferred a formula

$$d = \tilde{g} \left(\frac{h}{\ell} \right) t^2.$$

Remarkably, the weight of the ball does not appear on the right side. So the rates at which balls fall were *independent* of their weight. Thus, by fitting a *function*, Galileo simultaneously accomplished data dimension reduction and generalization. He discovered an empirical regularity that was an essential input into Isaac Newton's thinking.

Galileo's inclined plane experiments have all of the elements of modern machine learning and artificial intelligence. He started without a trusted theory. He conducted experiments and collected tables of numbers, one table for each experiment, indexed by the weight of a ball as well as by the length ℓ and height h of his inclined plane. From his tables, Galileo inferred a quadratic function of elapsed time with a constant of proportionality that depends on the angle of incline but not the weight of the ball.[3,]

I don't know what inspired Galileo to design his experiments, collect those measurements, and reduce the dimensionality of his data by fitting a function. I do know what tools that Galileo actually possessed and other tools that *could* have helped him but that he didn't possess. Galileo knew geometry and algebra. He was thoroughly conversant with Euclid and Archimedes. Without those tools, his intellectual brilliance and his doubts about Aristotle's theory would not have been enough. But Galileo didn't know differential and integral calculus – only decades later would Fermat and Newton and Leibniz invent those tools.

[3]Fast forward to today and watch how scientists use machine learning and AI. You'll see smart people collecting masses of data and fitting functions. For some wonderful examples, please see de Silva et al. (2020) and Brunton and Kutz (2022).

3.2 Charles Darwin

My next story is partly about how economists helped Charles Darwin (1809-1882) construct his theory of "evolution of species by natural selection". The following 1899 statement by Simon N. Patton cited by Hayek (2011, Appendix B) summarizes my message: "... just as Adam Smith was the last of the moralists and the first of the economists, so Darwin was the last of the economists and the first of the biologists." Distinguished game theorists and economists now routinely use evolution as a source of economic and social dynamics. Some of them are inspired by Darwin.It is fascinating that Darwin actually got an essential piece of *his* theory from economists. Thus, Hayek (2011, Appendix A) notes that Darwin's reading Adam Smith in 1838 provided him with key components of his theory of evolution through natural selection. Hayek (2011) documents that theories of cultural evolution were widely accepted by economists and sociologists long before Malthus wrote in 1800.

Darwin used raw empiricism and informal statistical data dimension reduction to construct his theory. He didn't know what a gene was. He didn't know what DNA was. He assembled a huge data set, collected from his having bred pigeons and observed wild animals and plants. From his pigeon data he inferred two of his three fundamental principles.

1. Natural variation.

2. Statistical inheritance of some variations.

As a pigeon breeder, Darwin used these two principles to observe variations, to select desirable traits from them, and then to rely on statistical inheritance to create new varieties of pigeons. Baby pigeons occasionally acquire some characteristics from their parents. "Selection by Charles Darwin" guided his breeding strategy. Darwin sought a source for selection of traits by nature. He tells us that he found that source in a book by Thomas Malthus (2007) entitled *An Essay on the Principle of Population as It Affects the Future Improvement of Society*. Malthus wrote about a competitive struggle for survival that was set off by the propensity of populations of people to reproduce at faster rates than do food sources. This situation created a struggle for existence that aligned surviving population size with available food. This part of Malthus's theory presented Darwin with his missing piece: *natural* selection emerges from a struggle for existence. More babies of every species are born than food sources can feed. The introduction to Darwin (1859) credits Malthus with the third pillar of his theory:

3. Natural selection via a competitive struggle for survival.

Darwin's research strategy stands as an instructive example of reducing a huge data set to extract a low-dimensional model based on three principles that can be applied generally. Data collection, data reduction, and generalization to deduce three principles: what an extraordinary package!

Like Galileo, Darwin did not start from a blank slate. He was learned not just in biology and geology but also in economics. His deep understanding of existing work in these fields empowered him to step beyond what had been known. He was a Keynesian macroeconomist in the sense that he put no "micro-foundations" under the first two pillars of his theory – variation and inheritance of some of new traits. He was uncertain about how much time would be required for his three pillars actually to produce the paleontological and biological evidence at hand.[4]

3.3 Brahe, Kepler, Newton

While economists can be proud of how Robert Malthus and Adam Smith influenced Charles Darwin, modern economists have actually been net importers of ideas from the natural sciences. Pioneers of 20th century economics set out to remake their subject by abandoning what they disparaged as the "literary" methods that had been used by 18th and early 19th century economists like David Hume, Adam Smith, and Robert Malthus. Instead they embraced *quantitative* methods of Tycho Brahe, Johannes Kepler, and Isaac Newton.

> . . . the decisive break which came in physics in the seventeenth century, specifically in the field of mechanics, was possible only because of previous developments in astronomy. It was backed by several millenia of systematic, scientific, astronomical observation, culminating in an observer of unparalleled calibre, Tycho de Brahe. Nothing of this sort has occurred in economic science. It would have been absurd in physics to expect Kepler and Newton without Tycho, - and there is no reason to hope for an easier development in economics. Von Neumann and Morgenstern (1944, ch. 1)

To find his three laws of planetary motion that lurked within Tycho Brahe's (1546-1601) tables of time-stamped measurements of the positions of the known planets, Johannes Kepler (1571-1630) used a method like Galileo's. Isaac Newton (1642-1727) synthesized, simplified,

[4] Darwin's work was not immediately accepted by leading natural scientists. For example, on the basis of the then prevailing estimates of the age of the earth, Lord Kelvin would soon say that the earth was simply much too young for Darwin's theory to work. Lord Kelvin's doubts impeded diffusion of Darwin's theory for many years.

and generalized Galileo's and Kepler's findings.[5] Founders of modern quantitative economics fashioned their approach after Brahe, Kepler, and Newton.[6] Koopmans (1947) tells how even raw data collection depends on a 'theory".

> When Tycho Brahe and Johannes Kepler engaged in the systematic labor of measuring the positions of the planets, and charting their orbits, they started with conceptions and models of the planetary system which later proved incorrect in some aspects, irrelevant in others. Tycho always, and Kepler initially, believed in uniform circular motion as the natural basic principle underlying the course of celestial bodies. Tycho's main contribution was a systematic accumulation of careful measurements. Kepler's outstanding success was due to a willingness to strike out for new models and hypotheses if such were needed to account for the observations obtained. He was able to find simple empirical "laws" which were in accord with past observations and permitted the prediction of future observations. This achievement was a triumph for the approach in which large scale gathering, sifting, and scrutinizing of facts precedes, or proceeds independently of, the formulation of theories and their testing by further facts.
>
> . . . in due course, the theorist Newton was inspired to formulate the fundamental laws of attraction of matter, which contain the empirical regularities of planetary motion discovered by Kepler as direct and natural consequences. The terms "empirical regularities" and "fundamental laws" are used suggestively to describe the "Kepler stage" and the "Newton stage" of the development of celestial mechanics. It is not easy to specify precisely what is the difference between the two stages. Newton's law of gravitation can also be looked upon as describing an empirical regularity in the behavior of matter. The conviction that this "law" is in some sense more fundamental, and thus constitutes progress over the Kepler stage, is due, I believe, to its being at once more elementary and more general. It is more elementary in that a simple property of mere matter is postulated. As a result, it is more general in that it applies to all matter, whether assembled in planets, comets, sun or stars, or in terrestrial objects - thus explaining a much wider range of phenomena. Koopmans (1947, p. 161)

even for the purpose of systematic and large scale observation of such a many-

[5]Weinberg (2015) offers a spell-binding account of the scientific methods of Kepler and Galileo. Li et al. (2021) use machine learning techniques to extract one of Kepler's laws from Brahe's data.

[6]Newton was gainfully employed as a monetary economist when he Warden and then Master of the Royal Mint from 1697 to 1727. Sargent and Velde (2014) describe Newton's role in controversies about the theory of policy of managing a commodity money. Copernicus was also a monetary economist.

> sided phenomenon, theoretical preconceptions about its nature cannot be dispensed with .... Koopmans (1947, p. 163)

> ... the extraction of more information from the data requires that, in addition to the hypotheses subject to test, certain basic economic hypotheses are formulated as distributional assumptions, which often are not themselves subject to statistical testing from the same data. Of course, the validity of information so obtained is logically conditional upon the validity of the statistically unverifiable aspects of these basic hypotheses. The greater wealth, definiteness, rigor, and relevance to specific questions of such conditional information, as compared with any information extractable without hypotheses of the kind indicated, provides [an argument] against the purely empirical approach. Koopmans (1947, p. 170)

Von Neumann and Morgenstern (1944) urged economists to abandon the imprecise literary methods that Keynes (1936) had used to address widespread macroeconomic and social problems and instead to work on well posed small problems.[7]

> It is necessary to begin with those problems which are described clearly, even if they should not be as important from any other point of view. ... The situation is not different here than in other sciences. There too the most important questions from a practical point of view may have been completely out of reach during long and fruitful periods of their development. This is certainly still the case in economics, where it is of utmost importance to know how to stabilize employment, how to increase the national income, or how to distribute it adequately. Nobody can really answer these questions, and we need not concern ourselves with the pretension that there can be scientific answers at present. ... The great progress in every science came when, in the study of problems which were modest as compared with ultimate aims, methods were developed which could be extended further and further. The free fall [of Galileo] is a very trivial physical phenomenon, but it was the study of this exceedingly simple fact and its comparison with the astronomical material, which brought forth mechanics.[8] Von Neumann and Morgenstern (1944, ch. 1)

[7] For more about Von Neumann and Morgenstern's opinions about Keynes 's approach, see Bhattacharya (2022, ch. 6).

[8] "Newton's achievement was based, not only on the regularities observed by Kepler, but also on experiments conducted on the surface of the earth by Galileo." Koopmans (1947, p. 166)

4 Artificial Intelligence

So far we have been discussing human intelligence. Let's now turn to artificial intelligence or machine learning. What is it?

By artificial intelligence I mean computer programs that are designed to do "intelligent" things that people like Galileo, Darwin, and Kepler have done. Much "machine learning" uses data, probability theory, and calculus to infer patterns. Designers of computer chips and algorithms and codes that do machine learning copy Galileo's falling body experiments. Think of a *function* as a collection of "if-then" statements. Think of an "if" part as the abscissa x of a function $y = f(x)$ and think of a "then" part the y ordinate. Using a computer to recognize patterns involves (1) partitioning data into x and y parts, (2) guessing a functional form for f, and then (3) using a statistical method like "least-squares" or "least-lines" to infer f from data on x and y. The discipline called "statistics" provides tools for inferring the function f.

Here's a simple example. Suppose that at a fixed location, each day of the year you record the length of "daytime" from sunrise to sunset. Record the day of the year as an integer running from 1 to 365 on the x axis. Record time from sunrise to sunset on the y axis. Make a table with x and y as the two columns. This table has 365 times 2 equals 730 numbers. Now plot them and stare. Guess that a *function* $y = \cos(\alpha + \beta x)$ can approximate the data. Use calculus to find values $(\hat{\alpha}, \hat{\beta})$ of the two parameters α, β that make the function fit well in the sense that they minimize $\sum_{i=1}^{365}(y_i - \cos(\hat{\alpha} + \hat{\beta} i))^2$. You'll find that this function fits well (though not perfectly). By summarizing the data (also known as performing "data compression" or "data reduction") in this way, you will have "generalized" by discovering a rule of thumb (a function) that you can use to predict lengths of days for days $i > 365$ outside of our sample.

5 Tools for AI

Machine learning and artificial intelligence import ideas from:[9]

1. Physics

2. Biology

3. Statistics

[9] It is not a coincidence that an important inventor of modern computing and AI, John von Neumann, contributed to all four of these fields. See Bhattacharya (2022) for a wonderful account of von Neumann's life and his contributions to all four fields, among others.

4. Economics

Let's take these up one by one.

5.1 Physics

Eighteenth and nineteenth century work by Euler, Lagrange, and Hamilton extended and perfected ways to use calculus to optimize integrals of functions of quantities over time. That put in place building blocks for a twentieth first-century Hamiltonian Monte Carlo simulation technique that powers modern Bayesian estimation and machine learning. Nineteenth century work by Clausius, Maxwell, Boltzmann, and Gibbs created ways to describe thermodynamics statistically. They defined a second law of thermodynamcs in terms of entropy, an expected value of the logarithm of the ratio of one probability distribution to another. One of those probability distributions is a flat uniform distribution that statistically represents complete disorder, the other a distribution that represents "order" in a particular statistical way. Entropy is how many machine learning algorithms measure discrepancies between possible models' probability distributions and an empirical distribution traced out by data. In ways that would contribute further tools for artificial intelligence and machine learning, Paul Samuelson (1947) and his coworkers imported these and other techniques from mathematical physics into economics.

5.2 Mathematical Biology

Biology is about patterns of reproduction and variation of species across time and space. Patterns are detected at "macro" and "micro" levels, depending on the unit of analysis – either an individual person or animal, or smaller units like DNA, RNA, or the molecules composing them. Mathematical theories of biology (e.g., Feldman (2014) and Felsenstein (1989)) are dynamic systems cast as stochastic difference or differential equations. At the micro-level, key ideas involve encoding DNA as a binary string upon which an analyst can perform mathematical operations that represent *mutation* and *sexual reproduction* via cutting and recombining. See Holland (1987).

5.3 Statistics

Modern mathematical statistics uses two distinct meanings of a "probability distribution":[10]

[10] This web site explores these two senses of probability with the assistance of Python code: `https://python.quantecon.org/prob_meaning.html`.

- A *frequentist* interpretation of a probability as a relative frequency to be anticipated after observing a very long sample of independently and identically distributed random variables.

- A *Bayesian* interpretation of a probability as a subjective expression of uncertainty about an unknown "state" or "parameter".

Modern statistics deploys an arsenal of tools for (1) specifying sets of functions that are characterized by vectors of parameters or by hierarchies of vectors of parameters; (2) inferring those parameters from data; (3) characterizing the uncertainty that a reasonable person should ascribe to those inferences; and (4) using probabilistic versions of those fitted functions to generalize by projecting "out of sample". These bread and butter techniques of machine learning in turn rest on differential and integral *calculus*, tools unknown to Galileo.

5.4 Economics

Economics is about how groups of people choose can allocate scarce resources. Modern economic theory is multi-person decision theory within coherent environments. The abstract artificial people inside a coherent economic model are "rational" in the sense that all of them solve well posed constrained optimization problems that include a common correct understanding of the environment.[11] The two leading classes of such multi-person decision theories are[12]

- Game theory

- General equilibrium theory

Components of these theories include

- Preferences

- Constraints

- Uncertainties

- Accounting systems that track stocks of assets and flows of payments

[11]When economists speak of "rational expectations" they are referring to an assumed "common correct understanding of an environment". The phrase "rational expectations" modifies "model", not "people".

[12]See Kreps (1997) for an account of common features of these two classes of models, as well as interesting opinions about possible new directions that have accurately foretold some subsequent applications of AI to economics.

- Decentralizations and parallel computations
- Ledgers that describe histories of networks of exchanges of goods and services
- Prices
- Competition

In such models, each agent's decision rules appear within constraint sets are parts of other agents' choice problems. The solution of each agent's constrained optimization problem produces a personal *value* that contains useful information for allocating resources. There is decentralized decision making and widespread "parallel processing". Interconnected constraints arise via a model's "equilibrium conditions". A social arrangement called an equilibrium serves to reconcile diverse selfish decisions with each other and with limits on physical resources.

Within both of the dominant frameworks – game theory and general equilibrium theory – precise notions of an *equilibrium* prevail. Defining an equilibrium is one thing. Computing one is another. Prominent economic theorists for years have wrestled with curses of dimensionality as they sought fail-safe methods to *compute* a competitive equilibrium allocation and price system. Landmark contributions to that enterprise were Arrow and Hurwicz (1958), Arrow et al. (1959), Arrow (1971), and Nikaidô and Uzawa (1960) as well as Scarf (1967), Scarf et al. (2008). Some of these authors developed algorithms that deploy accounting schemes that keep track of individual and social *values* and *gaps* between quantities of goods and activities that people want and quantities that the social arrangement provides.

Such work eventually discovered an intimate connections between *computing* an equilibrium and *convergence to* an equilibrium by a collection of boundedly rational agents. Bray and Kreps (1987) and Marcet and Sargent (1989) distinquished between "learning *within* an equilibrium" and "learning *about* an equilibrium". Marcet and Sargent (1989) and Sargent (1993) studied convergence to a rational expectations equilibrium by using the mathematics of stochastic approximation (e.g., see Gladyshev (1965)). Robbins and Monro (1951) introduced stochastic approximation as a recursive way of solving the following problem. Let $M(x)$ denote the expected value at level x of the response to a certain experiment. Assume that $M(x)$ is monotone but unknown to an experimenter. The experimenter wants to find an x_o that solves $M(x_o) = \alpha$. Robbins and Monro describe a recursive for generating a random sequence $\{x_n\}$ that converges to x_o in probability.[13]

[13] Hotelling (1941) and Friedman and Savage (1947) had proposed a statistical sampling method to accomplish closely related purposes associated with the design of experiments. such work ultimately led to today's "Bayesian optimization" machine learning techniques. For example, see Snoek et al. (2012).

Related work by Shubik (2004) and Bak et al. (1999) formulated *games* that can be used to think about equilibrating processes that are facilitated by *price setters.* (Inside a general equilibrium model, there are no price-*setters, only price*-takers.) Game theorist Shubik's work exploited his expertise about another subject at the frontier between general equilibrium theory and game theory with important lessons for machine learning and AI:

- Monetary theory

In the spirit of Shubik (2004), a way to think about monetary theory is to view one of its tasks as being to provide s a theory about how an equilibrium price vector could be set by agents who actually live inside a general equilibrium model. The classic general equilibrium model of Arrow and Debreu tells properties of an equilibrium price vector, but is silent about *who* sets that price vector and how: a *deus ex machina* outside the model mysteriously announces a price vector that simultaneously clears all markets.[14]An equilibrium price vector assures that every agent's budget constraint is satisfied. In a general equilibrium model, trade is multilateral and budget constraints for all agents are reconciled in a single centralized exchange with a comprehensive set of accounts. The model is *static* in the sense that all exchanges occur once-and-for-all in a single place.[15] Monetary theory is instead about a decentralized system populated by people who meet only occasionally in a sequence of bilateral meetings and who exchange goods and services by using "media of exchange". Media of exchange can be durable metals (gold or silver), tokens (pennies or paper "dollars" or "pounds"), circulating evidences of indebtedness, or entries in a ledger of a bank or clearing house or central bank. Ostroy and Starr (1974), Kiyotaki and Wright (1989), Ostroy and Starr (1990), and most recently Townsend (2020) describe work in this tradition that leads directly to theories about crypto currencies.

A few more words about how studying games has contributed to machine learning. Economists constructed algorithms to compute an equilibrium of a game. Key tools that underlie these calculations include backward induction (dynamic programming) and extensive form tree search. Because the number of possible states to be investigated grows exponentially, reducing the number of situations to be investigated is essential to making headway. Here the minimax algorithm and the alpha-beta pruning tree search algorithm are mainstays. See Knuth and Moore (1975) and `https://www.youtube.com/watch?v=STjW3eHOCik` for descriptions of alpha-beta tree search and watch for the appearance of an accounting system

[14]Keisler (1992, 1995, 1996) used ideas from statistical mechanics to construct processes that converge to a competitive equilibrium price vector. His approach is connected in interesting ways to the applications of stochastic approximation mentioned earlier.

[15]By indexing the commodities traded by calendar time and contingencies, respectively, the general equilibrium model makes dynamics and risk special cases of that static model.

and of a competitive process involving "survival of the fittest". A related line of research studied whether a collection of players who naively optimize against histograms of their opponents past actions converges to a Nash equilibrium. For examples, see Monderer and Shapley (1996), Hofbauer and Sandholm (2002), Foster and Young (1998), and Fudenberg et al. (1998). When convergence prevails, such "fictitious play" algorithms provide a way to compute an equilibrium (see Lambert Iii et al. (2005)).

5.5 John Holland's Model of a Mind

Now back to how economics and other fields have influenced AI. Computer scientist John Holland[16] combined ideas from all four of our technical fields to construct computer models of decision makers living in environments in which they must "learn by doing" in the sense of Arrow (1971). Holland (1987, 1992) described his approach. Marimon et al. (1990) applied it to multi-person economic environments of types that Kiyotaki and Wright (1989) had used to analyze classic questions in monetary economics. An essential component of Holland's machinery is a global search algorithm that he called a "genetic algorithm". It searches "rugged landscapes" by representing arguments of functions as strings of binary numbers that can be randomly matched into pairs, then cut and recombined. This is Holland's mechanical way of representing "sexual reproduction". Holland's "genetic algorithm" comprised part of what he called a "classifier" system. Holland's classifier system consists of (1) a sequence of if-then statements, some of which compete with each other for the right to decide on-line (i.e., in real time); (2) a way to encode if-then statements as binary strings to be subjected to random mutation, cutting, and recombining; (3) an accounting system that assigns rewards and costs to individual if-then statements; (4) procedures for destroying and creating new if-then statements that include random mutations and sexual reproduction based on cutting and recombining strings; and (5) a sequence of competitive auctions that promotes survival of fit decision rules. All of this goes on inside the mind of a single decision maker. Systems of Holland classifiers learned how to be patient in dynamic settings, a subtle outcome that illustrates Ramon Marimon's dictum that "patience requires experience". Holland classifiers succeeded in computing a "stable" Nash equilibrium for a dynamic economic model that the model's authors had not anticipated although they could verify the "guess" that the Holland classifier system eventually handed to them. See Marimon et al. (1990).

[16]Please see https://en.wikipedia.org/wiki/John_Henry_Holland, https://www.nytimes.com/2015/08/20/science/john-henry-holland-computerized-evolution-dies-at-86.html.

5.6 AI today

In a remarkable achievement, DeepMind's computer program called AlphaGo succeeded in learning how to play Go well enough to defeat champion human players. See Wang et al. (2016). AlphaGo's algorithm reminds me of how to cook well – add a touch of this to a handful of that, taste, and add something else .... Among the ingredients combined to cook up AlphaGo were ideas gathered from dynamic programming; Thompson (1933) sampling; and stochastic approximation (Robbins and Monro (1951) and Wolfowitz (1952)); alpha-beta tree search (Knuth and Moore (1975)); Q-learning (sWatkins and Dayan (1992)); and Monte Carlo tree search (Browne et al. (2012)). A rule of thumb tunes a parameter that balances "exploration" against "exploitation".[17]

Other recent advances in machine learning also import heavily from economics and statistics. Thus, computational optimal transport (e.g., Peyré et al. (2019)) uses a linear program of Dantzig, Kantorovich, and Koopmans to measure discrepancies between a theoretical probability and an empirical measure. It then uses that measure to guide computationally efficient ways to match data to a theory. Economist Harold Hotelling (1930) used Riemannian geometry to represent parameterized families of statistical models. That idea inaugurated computational information geometry, an approach systematized by Amari (2016).

6 Imitation and Innovation

Galileo, Kepler, Newton, and Darwin discovered new laws of nature by somehow combining mastery of findings and methods of their predecessors with unprecedented flashes of insight. Respect for precedents, and their ability to venture beyond them, characterized their work. Subsequent scientists used the same general approach. A source of other examples is supplied by the sequence of discoveries by Franklin, Davy, Faraday, Maxwell, and Michaelson and Morley about the science of electricity and magnetism. Their work set the stage for Einstein's theory of special relativity. Each of those scientists began, not from a *Blank Slate* (not coincidentally the title of Pinker (2003)), but equipped with deep understandings and respect for the ideas of their predecessors. All saw something that their predecessors hadn't. Somehow they figured out how to deploy improved ways of observing or reasoning. Thus, by unleashing mathematics that Faraday did not know, Maxwell organized a breathtaking unification, generalization, and reduction of the laws governing electro-magnetic dynamics into twelve equations that Heaviside would soon reduce to four equations. That set the stage

[17] Also see March (1991) and Fudenberg and Kreps (1993, 1995).

for Einstein's special theory of relativity.[18]

Seemingly unrelated, purely theoretical work in mathematics preceded and then coincided with those discoveries about electro-magnetism. Descartes (1596 –1650) invented a coordinate system that enabled him to transform geometry problems into algebra problems and to write down *functions*. Fifty years later Newton and Leibniz (1646 – 1716) used Cartesian coordinates to invent differential and integral calculus. In the first half of the nineteenth century, Gauss (1777 – 1855) and his student Riemann (1826 – 1866) refined geometries for curved spaces and parallel lines that meet. Ricci (1853 – 1925) added a sharp notion of curvature.

Einstein brought together these two independently motivated and seemingly "unrelated" lines of work, the first about physical phenomena, the second about purely abstract mathematics. To extend special relatively to accelerating observers, Einstein learned how to use Riemannian geometry and Ricci curvature to construct a coherent general theory of relativity.[19]

Scientific advances illustrate an interaction between "imitation" and "innovation" featured in modern theories of economic growth (for example, see Benhabib et al. (2014) and Benhabib et al. (2020)). For pioneers in electro-magnetism and relativity, the "imitation" phase was their copying techniques of their predecessors and teachers; the "innovation" phase was stepping beyond what they had learned because they understood more than their teachers.

7 Concluding Remarks

My survey of applications from physics, biology, statistics, and economics illustrates how the subjects in which Pinker (2003) tells us we are all innately cognitively challenged are being deployed to create artificial intelligence and machine learning. This is one good reason to study these subjects. Another reason is their intrinsic beauty, an attribute of good science emphasized by Weinberg (2015, ch. 11):

> The theory of Copernicus provides a classic example of how a theory can be selected on aesthetic criteria, with no experimental evidence that favors it.

Weinberg's book has much to say about the hazards of overfitting and the virtues of tightly parameterized, theoretically interpretable structural models. These topics are central to understanding and implementing machining learning and AI.

[18]Einstein kept a photo of Maxwell on his office wall.

[19]See Farmelo (2019, ch. 3) for an account of these events.

References

Amari, Shun-ichi. 2016. *Information geometry and its applications.* Springer.

Arrow, Kenneth J and Leonid Hurwicz. 1958. On the stability of the competitive equilibrium, I. *Econometrica* 522–552.

Arrow, Kenneth J, Henry D Block, and Leonid Hurwicz. 1959. On the stability of the competitive equilibrium, II. *Econometrica* 82–109.

Arrow, Kenneth Joseph. 1971. The economic implications of learning by doing. In *Readings in the Theory of Growth*, 131–149. Springer.

Bak, Per, Simon F Nørrelykke, and Martin Shubik. 1999. Dynamics of money. *Physical Review E* 60 (3):2528.

Benhabib, Jess, Jesse Perla, and Christopher Tonetti. 2014. Catch-up and fall-back through innovation and imitation. *Journal of Economic Growth* 19 (1):1–35.

Benhabib, Jess, Éric Brunet, and Mildred Hager. 2020. Innovation and imitation. *arXiv preprint arXiv:2006.06315* .

Bhattacharya, Ananyo. 2022. *The Man from the Future: The Visionary Life of John von Neumann.* New York: W. W. Norton & Company.

Bray, Margaret and David M Kreps. 1987. Rational learning and rational expectations. In *Arrow and the ascent of modern economic theory*, 597–625. Springer.

Browne, Cameron B, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. 2012. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games* 4 (1):1–43.

Brunton, Steven L and J Nathan Kutz. 2022. *Data-driven science and engineering: Machine learning, dynamical systems, and control.* Cambridge University Press.

Darwin, Charles. 1859. *The origin of species by means of natural selection.* Pub One Info.

Farmelo, Graham. 2019. *The Universe Speaks in Numbers: How Modern Maths Reveals Nature's Deepest Secrets.* Faber & Faber.

Feldman, Marcus W. 2014. Mathematical evolutionary theory. In *Mathematical Evolutionary Theory.* Princeton University Press.

Felsenstein, Joe. 1989. Mathematics vs. Evolution: Mathematical Evolutionary Theory. Marcus W. Feldman. Ed. Princeton University Press, Princeton, NJ, 1989. x, 341 pp. 60;paper, 19.95. *Science* 246 (4932):941–942.

Foster, Dean P and H Peyton Young. 1998. On the nonconvergence of fictitious play in coordination games. *Games and Economic Behavior* 25 (1):79–96.

Friedman, Milton and L.J. Savage. 1947. Planning Experiments Seeking Maxima. In *Selected Techniques of Statistical Analysis for Scientific and Industrial Research, and Production and Management Engineering*, edited by Millard W. Hartay Churchill Eisenhart and W. Allen Wallis, 363–372. New York and London: McGraw-Hill.

Fudenberg, Drew and David M Kreps. 1993. Learning mixed equilibria. *Games and economic behavior* 5 (3):320–367.

———. 1995. Learning in extensive-form games I. Self-confirming equilibria. *Games and Economic Behavior* 8 (1):20–55.

Fudenberg, Drew, Fudenberg Drew, David K Levine, and David K Levine. 1998. *The theory of learning in games*, vol. 2. MIT press.

Gladyshev, EG. 1965. On stochastic approximation. *Theory of Probability & Its Applications* 10 (2):275–278.

Hayek, Friedrich August. 2011. The fatal conceit. In *The Fatal Conceit*. University of Chicago Press.

Hofbauer, Josef and William H Sandholm. 2002. On the global convergence of stochastic fictitious play. *Econometrica* 70 (6):2265–2294.

Holland, John. 1992. Holland. genetic algorithms. *Scientific american* 267 (1):44–50.

Holland, John H. 1987. Genetic algorithms and classifier systems: foundations and future directions. Tech. rep., Univ. of Michigan, Ann Arbor, MI (United States).

Hotelling, Harold. 1930. Spaces of Statistical Parameters. *Bulletin of American Mathematical Society* 36 (3):191.

———. 1941. The Experimental Determination of the Maximum of a Function. *Annals of Mathematical Statistics* 12:20–45.

Keisler, H Jerome. 1992. A law of large numbers for fast price adjustment. *Transactions of the American Mathematical Society* 332 (1):1–51.

———. 1995. Approximate tatonnement processes. *Economic Theory* 5:127–173.

———. 1996. Getting to a competitive equilibrium. *Econometrica: Journal of the Econometric Society* 29–49.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money.* London: Macmillan.

Kiyotaki, Nobuhiro and Randall Wright. 1989. On money as a medium of exchange. *Journal of political Economy* 97 (4):927–954.

Knuth, Donald E and Ronald W Moore. 1975. An analysis of alpha-beta pruning. *Artificial intelligence* 6 (4):293–326.

Koopmans, Tjalling C. 1947. Measurement without theory. *The Review of Economics and Statistics* 29 (3):161–172.

Kreps, David M. 1997. Economics: the current position. *Daedalus* 126 (1):59–85.

Lambert Iii, Theodore J, Marina A Epelman, and Robert L Smith. 2005. A fictitious play approach to large-scale optimization. *Operations Research* 53 (3):477–489.

Li, Zelong, Jianchao Ji, and Yongfeng Zhang. 2021. From Kepler to Newton: Explainable AI for Science Discovery. *arXiv preprint arXiv:2111.12210* .

Malthus, Thomas Robert. 2007. *An essay on the principle of population, as it affects the future imporvement of society, with remarks on the speculations of Mr. Godwin, M. Condorcet, and other writers.* The Lawbook Exchange, Ltd.

Marcet, Albert and Thomas J Sargent. 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. *Journal of Economic theory* 48 (2):337–368.

March, James G. 1991. Exploration and Exploitation in Organizational Learning. *Organization Science* 2 (1):71–87.

Marimon, Ramon, Ellen McGrattan, and Thomas J Sargent. 1990. Money as a medium of exchange in an economy with artificially intelligent agents. *Journal of Economic dynamics and control* 14 (2):329–373.

Monderer, Dov and Lloyd S Shapley. 1996. Fictitious play property for games with identical interests. *Journal of economic theory* 68 (1):258–265.

Nikaidô, Hukukane and Hirofumi Uzawa. 1960. Stability and non-negativity in a Walrasian Tatonnement process. *International Economic Review* 1 (1):50–59.

Ostroy, Joseph M and Ross M Starr. 1974. Money and the Decentralization of Exchange. *Econometrica: Journal of the Econometric Society* 1093–1113.

———. 1990. The transactions role of money. *Handbook of monetary economics* 1:3–62.

Peyré, Gabriel, Marco Cuturi, et al. 2019. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning* 11 (5-6):355–607.

Pinker, Steven. 2003. *The blank slate: The modern denial of human nature.* Penguin UK.

Robbins, Herbert and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* 400–407.

Samuelson, Paul A. 1947. *Foundations of Economic Analysis.* Cambridge, Massachusetts: Harvard University Press.

Sargent, Thomas J. 1993. *Bounded rationality in macroeconomics: The Arne Ryde memorial lectures.* Oxford and New York: Oxford University Press.

Sargent, Thomas J and François R Velde. 2014. *The big problem of small change.* Princeton University Press.

Scarf, Herbert E. 1967. On the computation of equilibrium prices. Cowles Foundation Discussion Paper No. 232, Yale University.

Scarf, Herbert E, John B Shoven, et al. 2008. Applied general equilibrium analysis. *Cambridge Books* .

Shubik, Martin. 2004. *The theory of money and financial institutions*, vol. 1. Mit Press.

de Silva, Brian M, David M Higdon, Steven L Brunton, and J Nathan Kutz. 2020. Discovery of physics from data: universal laws and discrepancies. *Frontiers in artificial intelligence* 3:25.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams. 2012. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25.

Thompson, William R. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25 (3-4):285–294.

Townsend, Robert M. 2020. *Distributed Ledgers: Design and Regulation of Financial Infrastructure and Payment Systems*. MIT Press.

Von Neumann, John and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton, New Jersey: Princeton university press.

Wang, Fei-Yue, Jun Jason Zhang, Xinhu Zheng, Xiao Wang, Yong Yuan, Xiaoxiao Dai, Jie Zhang, and Liuqing Yang. 2016. Where does AlphaGo go: From church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica* 3 (2):113–120.

Watkins, Christopher JCH and Peter Dayan. 1992. Q-learning. *Machine learning* 8 (3):279–292.

Weinberg, Steven. 2015. *To explain the world: The discovery of modern science*. Penguin UK.

Wolfowitz, Jacob. 1952. On the stochastic approximation method of Robbins and Monro. *The Annals of Mathematical Statistics* 457–461.