Sources of Artificial Intelligence

Thomas J. Sargent

June 25, 2022

Abstract

This essay describes a personal account of how artificial intelligence and machine learning were invented. I name names and ideas.

1 Introduction

This essay is about human and artificial intelligence and learning. I take *artificial* to mean 'non-human'. Before describing *artificial intelligence* and *machine learning*, I'll state my understanding of what *natural* or *human* intelligence is by describing 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 intelligence are awareness of time and space, and also sympathy and empathy with other people. Successive generations of parents pass on to their children tools and perspectives that their parents taught them, as well as new ideas that they have learned. After describing how Galileo and Darwin combined their innate talents with their text-book knowledge to create scientific breakthroughs, I'll tell how modern researchers have designed computer programs that can recognize patterns and make choices.[1]

While I hope that my description of the machine-learning "forest" is clear, I do 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 perplexing "tree", I recommend checking a good online search engine or some of the items in the references at the end of this essay.

2 Human Intelligence

I start with messages from chapter 13 of *The Blank Slate* by cognitive psychologist Steven Pinker. Chapter 13 of Pinker (2003) is titled *Out of Our Depths.* Read it if you are a high school student or college freshman or anyone else who wants to think about the purpose 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 some things evolution has hard-wired us to do well and some other things that we have to learn to do.

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

Things that we aren't hard-wired to do well weren't important during most of our 100,000 years of human history and pre-history. But modern life has elevated the importance of some things we just aren't hard-wired for. Pinker identifies four such subjects.

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

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

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

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

Making wise private and public decisions in modern life requires understanding these four fields. But these are subjects for which our "intuitions" often fail us. For working purposes, just define "intuition" as how we think about situations that evolution constructed us to understand quickly. Maybe "common sense" is a synonym for intuition, things that we think we understand instinctively. Steven Pinker describes how our hard-wired theories in these four fields can lead us astray unless they are improved by education.

Pinker provides fascinating examples from all of these areas. Thus, our common sense and intuition don'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 innately *misunderstand* these things, 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.

These cognitive deficiencies set the stage for Pinker's chapter 13 recommendations for redesigning curricula. Pinker describes education as a technology for compensating for our innate cognitive limits and for exploiting our innate abilities to learn. He calls for significant changes in academic curricula to align them better with what will help us enjoy life and make good decisions today: biology, statistics, and economics. He acknowledges that teaching more of these subjects means teaching less of others.

2.1 AI and our innate cognitive limits

Another lesson that we can learn from chapter 13 is how we want "artificial intelligence" to supplement and surpass our innate natural human intelligence.

A paradox lurks here because the principal technical tools that have been and are being used to create artificial intelligence and machine learning are drawn from physics, biology, statistics, and economics, the same areas in which we are innately limited. Thus, the very fields that we are not naturally equipped to be good at are being used to create artificial intelligence and machine learning. Early pioneers and practitioners 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.

3 Two Pioneers of Machine learning

3.1 Galileo

Because he believed that the earth revolved around the sun, the great early 17th century Italian mathematician, scientist, physicist, astronomer Galileo Galilei (1564-1642) was eventually arrested by the Inquisition. Many years before he was arrested, Galileo did something that I regard as illustrating the essence of the "machine learning" approach. Galileo (1) designed and conducted experiments to collect data; (2) stared at his data and tried to spot patterns; (3) reduced the dimension of the data by fitting a *function*; and (4) interpreted that function as a general law of nature. Galileo's strategy offers a beautiful example, maybe the first, for what machine learning and artificial intelligence are all about.

I am of course writing about Galileo's "inclined plane" experiments and the data processing and data reduction that he performed on his data. Galileo wanted to discover the natural laws that govern the dynamics of falling bodies. Perhaps you are thinking: "That's easy, just apply Newton's laws of gravity." But not so fast: Newton wasn't born yet. The most widely accepted prevailing theory was the one that Aristotle had pronounced 2000 years before: heavier bodies fall faster than lighter ones.

Galileo wanted to study 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 existing clocks could accurately measure. Therefore, Galileo decided to construct smooth inclined planes of different angles, and to adjust the angle so that falling balls slowed down enough so that he could measure their rates of travel along the plane with the clocks he had. For a plane of length ℓ and height h, the ratio $\frac{h}{\ell}$ determines the angle of the plane. Galileo dropped a ball and carefully measured the distance d along the plane that the ball traveled as a function of the time t elapsed after the ball had been dropped. 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 each experiment. 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 then stared at his graphs. He noticed a striking thing: for *all* of the graphs, $d \propto t^2$ - the distance traveled was proportional to the square of the elapsed time, independently of weight of the ball and independently of the angle of the plane. He inferred a formula

$$d = \tilde{g}\frac{h}{\ell}t^2.$$

Notice that, remarkably, the weight of the ball is *not* in the function on the right side. So the rate at which a ball falls is apparently *independent* of its weight. Thus, by fitting a *function* to data from his experiments, Galileo simultaneously accomplished data dimension reduction and generalization. He discovered a law of nature that was an important input into Isaac Newton's thinking 50 years later.

Galileo's inclined plane experiments have all of the elements of modern machine learning and artificial intelligence. He started out not knowing how the world works and not having a good theory. What he did was entirely *atheortical*. So he conducted a set of 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 of many numbers he deduced (i.e., "fit") a function that turns out to be determined by only one new number, the "parameter" $\tilde{g}$.[2,]

I don't fully understand what inspired Galileo to design his experiments, collect those measurements, and reduce the dimensionality of his measurements by fitting a function. I do know the tools that Galileo possessed and the tools that *could* have helped him but that he didn't possess. In particular, he didn't know differential and integral calculus – only decades later would those tools be invented by Fermat and Newton and Leibniz. But Galileo did know geometry and algebra very well. 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. [3]

3.2 Charles Darwin

This next story is partly how an economic theory helped Charles Darwin (1809-1882) complete 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."

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. What he did "know" was

[2] 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] To find his three laws of planetary motion that were buried in Tycho Brahe's (1546-1601) tables of time-stamped measurements of the positions of the known planets, Johannes Kepler (1571-1630) used a method similar to Galileo's. Li et al. (2021) follow in Kepler's footsteps by using machine learning techniques to extract one of Kepler's laws from Brahe's data. See Weinberg (2015) for spell-binding accounts of the scientific methods of Kepler and Galileo.

a huge data set, collected from his breeding pigeons and observing animals and plants in nature. From his pigeon data alone he could deduce two of his three fundamental principles.

1. Natural variation

2. Statistical inheritance of new variations.

As a pigeon breeder, Darwin used these two principles to select desirable traits and then rely on statistical inheritance to produce new varieties of pigeons. Baby pigeons acquire some characteristics from their parents. "Selection by Charles Darwin," not *natural selection* guided his breeding or pigeons. For a long time Darwin lacked a source for selection in nature. Then he read 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 struggle that was set off by the propensity of people to reproduce at faster rates than food sources. This situation created a struggle for existence that aligned surviving populations with available food. This part of Malthus's argument presented Darwin with the missing piece: *natural* selection that emerges from a struggle for existence. More babies were born than food sources could feed. The introduction to Darwin (1859) credited Malthus with the third pillar of his theory.

3. Selection via Competition – the struggle for existence.

Some distinguished game theorists and economists now routinely use evolution as a source of economic and social dynamics. Maybe some of them think that they got this idea from Darwin. But Darwin actually got an essential piece of his theory from economists. Thus, Hayek (2011, Appendix A) notes that Darwin's study of Adam Smith in 1838 provided him with a key component of his theory of biological evolution natural selection. Hayek (2011) also documents that theories of cultural evolution were widely accepted by economists and sociologists long before 1800.

Darwin's research strategy stands as a wonderful 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 via three principles, and then generalization: what an extraordinary package!

Like Galileo, Darwin did not start from a blank slate. He was well read 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 and understood. He was a "macro" person in the sense that he had no "micro-foundations" for the first two pillars of his theory – variation and inheritance of some of new traits. He was vague about how much time would be required for his three pillars to produce the paleontological and biological evidence at hand.[4]

[4]Darwin's work was not immediately accepted by leading 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.

4 Artificial intelligence

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

By artificial intelligence I mean computer programs that are designed to perform some "intelligent" things that some humans do. Much "machine learning" deploys calculus and statistics to infer patterns. Designers of the computer chips and programs that do machine learning and AI copy Galileo's inclined plane experiments for measuring speeds of falling bodies. Thus, think of a *function* as a collection of "if-then" statements. Think of an "if" part being the abscissa x in 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 statistics to infer f from data on x and y. The discipline called "statistics" provides tools for inferring or "fitting" the function f.

Here's a simple example. Suppose that at a given location on the earth, 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* $y = \cos(\alpha + \beta x)$ might summarize the data well. 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, we have "generalized" by discovering a rule of thumb (a function) that we 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 their essential methods from the following fields:[5]

1. Physics
2. Biology
3. Statistics
4. Economics

Let's take these up one by one.

[5] Thus, it was not a coincidence that an important inventor of modern computing and AI, John von Neumann, studied and substantially contributed to all four of these fields. See Bhattacharya (2022) for a wonderful account of von Neumann's work and life.

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 set in place essesntial building blocks for a twentieth first-century Hamiltonian Monte Carlo simulation technique that is now powering sophisticated Bayesian estimation and machine learning techniques. Nineteenth century work by Clausius, Boltzmann, and Gibbs created concepts for describing thermodynamics statistically. They defined a second law of thermodynamcs in terms of entropy, an expected value of a likelihood ratio, i.e., the ratio of one probability distribution to another. One of those probability distributions was a flat uniform distribution that statistically represented complete disorder, the other a distribution that represented "order" in a precise statistical way. In the late twentieth century and early twenty first, entropy provided a way for many machine learning algorithms to measure discrepancies between a fitted model's probability distribution and an empirical distribution traced out by data. In ways that would pave the way to producing 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 can be 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)) formalize these things by constructing dynamic systems in the form of 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 of *mutation* and *sexual reproduction* via cutting and recombining. For example, see Holland (1987).

5.3 Statistics

Modern mathematical statistics deploys two possible meanings of "probability":[6]

- A *frequentist* interpretation that a probability is a relative frequency to be anticipated after observing a very long sample of independently and identically distributed random variables.
- A *Bayesian* interpretation that a probability is a subjective expression of uncertainty about an unknown hidden "state" or "parameter".

Modern statistics deploys an arsenal of tools for (1) specifying sets of functions that are characterized by vectors of parameters, or sometimes by a hierarchy of vectors of parameters; (2) inferring

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

or "estimating" those parameters from observations; (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, as we have noted earlier.

5.4 Economics

Economics is about how collections of individuals choose to utilize and to allocate sets of 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 constrained optimization problems that take into the account their common correct understandings of their environment.[7] Two leading classes of such multi-person decision theories are[8]

- Game theory
- General equilibrium theory

Components and forces in these theories include

- Constraints
- Uncertainties
- Decentralizations and parallel optimizations
- Ledgers for exchange networks
- Prices
- Competition

In these models, one agent's decision rules form part of the constraint set of other agents' choice problems. Such constraints arise via a model's "equilibrium conditions". A solution of an agent's constrained optimization problem produces a personal *value* that contains useful information for allocating resources.

These economic models describe "parallel processing" and decentralized decision making. An arrangement called an equilibrium serves to reconcile diverse selfish decisions with each other and with limits on physical resources. Within both of these dominant frameworks, precise notions of

[7]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".

[8]See Kreps (1997) for an account of common features and shortcomings of these two classes of models, as well as some thoughtful opinions and conjectures about new directions that seem to me to have forecasted subsequent incursions of AI into economics.

an *equilibrium* prevail. Defining an equilibrium is one thing. Computing one is another. Thus, 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). These algorithms 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.

Work on computing an equilibrium eventually discovered an intimate connection 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) present important distinctions between "learning *within* an equilibrium" and "learning *about* an equilibrium". Marcet and Sargent (1989) and Sargent et al. (1993) study convergence to a rational expectations equilibrium by using the mathematics of stochastic approximation (e.g., see Gladyshev (1965)). So far as I know, initial work on stochastic approximation began with the quest of Hotelling (1941) and Friedman and Savage (1947) to construct a statistical sampling method to find the maximum of an unknown function that could nevertheless be accurately evaluated at a given abscissa.[9]

Related work by Shubik (2004) and Bak et al. (1999) formulated *games* that could be used to think about equilibriating processes facilitated by the presence of *price setters*. (Inside a general equilibrium model, there are only price-*takers*, no price-*setters*.) Shubik's work exploited his expertise about a topic with important lessons for machine learning and AI, a topic that lives in the interstices between the general equilibrium theory and game theory:

- Monetary theory

In the spirit of Shubik (2004), a good way to think about monetary theory is to notice that its aim is to provide a theory about how an equilibrium price vector could be set by the 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. Instead, *deus ex machina* outside the model mysteriously announces a price vector that simultaneously clears all markets. An equilibrium price vector assures that every agent's budget constraint is satisfied. In a general equilibrium model, trade is multilateral and budget constraints are reconciled in a single centralized account. Monetary theory is instead about a decentralized system populated by people who meet only occasionally in a sequence of bilateral meetings and 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), Ostroy and Starr (1990), and most recently Townsend (2020) describe work in this tradition that leads directly to a theory for crypto currencies.

[9]The work of Hotelling and Friedman and Savage ultimately led to the "Bayesian optimization" machine learning technique. For example, see Snoek et al. (2012).

I offer a few more words about how studying games has contributed to machine learning. For decades, applied economists have constructed algorithms to compute an equilibrium of a game. Key tools that underlie these calculations include backward induction (dynamic programming) and tree search. Because the dimension of possible states to be investigated grows exponentially, reducing the number of situations to be investigated is essential to making headway on approximating equilibria. 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=STjW3eH0Cik for descriptions of alpha-beta tree search and watch for an accounting system and a "survival of the fittest" idea. 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), Fudenberg et al. (1998). When convergence prevails, such "fictitous play" algorithms provide a way to compute an equilibrium (see Lambert Iii et al. (2005)).

5.5 John Holland's circa 1985 Vision for AI

The renowned computer scientist John Holland[10] was a pioneer. He combined ideas from all of the technical fields we have mentioned to construct computer models of decision makers living in environments in which they have no choice but to "learn by doing" in the sense of Arrow (1971). See Holland (1987, 1992) for descriptions of Holland's approach and Marimon et al. (1990) for an application to a multi-person economic environment. A substantial piece of Holland's approach was a global search algorithm that he called a "genetic algorithm". It searched "rugged landscapes" by representing arguments of functions by strings that could be randomly matched into pairs of strings, cut, and recombined. This was Holland's mechanical way of representing "sexual reproduction". Such a "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 must 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 that can 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 includes random mutations and sexual reproduction based on DNA cutting and recombining; and (5) a competitive struggle that promotes survival of fit decision rules. Systems of Holland classifiers have been shown to learn how to be patient in dynamic settings, a subtle outcome summarized by Ramon Marimon's phrase "patience requires experience" in a world of Holland's artificially intelligent agents. Holland classifiers succeeded in computing a "stable" Nash equilibrium for a dynamic economic model that the model's authors had not recognized *a priori*,, although they could verify the "guess" that the Holland classifier handed to them. See Marimon et al. (1990).

[10] 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 mastering the game of Go so well that defeated champion human players of Go. See Wang et al. (2016). The approach that the creators of AlphaGo deployed reminds me of cooking delicious food – 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 sampling (see Thompson (1933)) and stochastic approximation (see Hotelling (1941) and Friedman and Savage (1947)); alpha-beta tree search (see Knuth and Moore (1975)); Q-learning (see Watkins and Dayan (1992)); and Monte Carlo tree search (see Browne et al. (2012)). A rule of thumb choice for tuning a parameter that trades "exploration" off against "exploitation" is important (as it also is in March (1991),Fudenberg and Kreps (1993) and Fudenberg and Kreps (1995).)

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 construct a computationally efficient way to match data to a theory. An economist 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 Sources of Creativity: Imitation and Innovation

I described how Galileo 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 precedent, and their ability to venture beyond, characterized the work of both geniuses. Many subsequent geniuses used the same general approach. A source of other examples is electricity and magnetism and the sequence of discoveries by Franklin, Davy, Faraday, Maxwell, Michaelson and Morley, and then Einstein. Each of them began not from a *Blank Slate* (not coincidentally the title of Pinker (2003)) but from their deep understanding and respect for their predecessors. Each saw something that their predecessors hadn't, often because they deployed 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. Those four equations set the stage for Einstein's special theory of relativity.[11]

Seemingly unrelated, purely theoretical work in mathematics preceded and then coincided with those discoveries about electro-magnetism. Descartes invented a coordinate system that enabled him to convert geometry into algebra and to write down *functions*. Fifty years later Newton and Leibniz used Cartesian coordinates to invent differential and integral calculus. In the first half of

[11]A photo of Maxwell hung on Einstein's office wall.

the nineteenth century, Gauss and his student Riemann refined geometries for curved spaces and parallel lines that meet. Ricci added a sharp notion of curvature.

Einstein brought together these two independently motivated and seemingly "unrelated" lines of work, the first about practical physical phenomena, the second about some purely abstract mathematics. Struggling to extend his special theory of relatively, Einstein learned how to use Riemannian geometry and Ricci curvature in order to construct a coherent general theory of relativity.[12]

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

7 Concluding Remark

My survey of ideas from physics, biology, statistics, and economics confirms my claim that the subjects in which Pinker (2003) tells us we are all innately cognitively challenged are the ones that are being deployed to create artificial intelligence and machine learning. This is just one more reason to study these subjects in school and after school too. I think that their intrinsic beauty is another.

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* .

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

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* 12:2045.

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

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.

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

Sargent, Thomas J et al. 1993. Bounded rationality in macroeconomics: The Arne Ryde memorial lectures. *OUP Catalogue* .

Scarf, Herbert E. 1967. On the computation of equilibrium prices .

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.

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.