

13

Epilogue

Unknown outcomes often follow patterns found in past observations. But when do they not? As powerful as statistical patterns are, they are not without limitations. Every discipline built on the empirical law also experiences its failure.

In fact, Halley's contemporaries already bore witness. Seeking to increase revenue still despite the sale of life annuities, King William III desired to tax his citizens in proportion to their wealth. An income tax appeared too controversial and unpopular with his constituents so that the king's advisors had to come up with something else. In 1696, the king introduced a property tax based on the number of windows in a house. It stands to reason that the wealth of a family correlated strongly with the number of windows in their home. So, the window tax looked quite reasonable from a statistical perspective.

Although successful on the whole and adopted by many other countries, the window tax had a peculiar side effect. People adjusted. Increasingly, houses would have bricked-up window spaces. In Edinburgh an entire row of houses featured no bedroom windows at all. The correlation between the number of windows and wealth thus deteriorated.

The problem with the window tax foretold a robust limitation of prediction. Datasets display a static snapshot of a population. Predictions on the basis of data are accurate only under an unspoken stability assumption. Future observations must follow the same data-generating process. It's the "more of the same" principle that we call generalization in supervised learning.

However, predictions often motivate consequential actions in the real world that change the populations we observe. Chemist and technology critic Ursula Franklin summarizes the problem aptly in her 1989 book called *The Real World of Technology*:

[T]echnologies are developed and used within a particular social, economic, and political context. They arise out of a social structure, they are grafted on to it, and they may reinforce it or destroy it, often in ways that are neither foreseen nor foreseeable.¹

Franklin continues:

[C]ontext is not a passive medium but a dynamic counterpart. The responses of people, individually, and collectively, and the responses of nature are often underrated in the formulation of plans and predictions.

Franklin understood that predictions are not made in a vacuum. They are agents of change through the actions they prompt. Decisions are always part of an evolving environment. It's this dynamic environment that determines the merit of a decision.

Predictions can fail catastrophically when the underlying population is subject to unmodeled changes. Even benign changes to a population, sometimes called distribution shift, can sharply degrade the utility of statistical models. Numerous results in machine learning are testament to the fragility of even the best performing models under changing environments.

Other disciplines have run into the same problem. In his influential critique from 1976, economist Robert Lucas argued that patterns found in historical macroeconomic data are an inadequate basis of policy making, since any policy would inevitably perturb those statistical patterns. Subsequently, economists sought to ground macroeconomics in the microeconomic principles of utility theory and rational behavior of the individual, an intellectual program known as microfoundations dominant to this day. The hope was that microfoundations would furnish a more reliable basis of economic policy making.

It is tempting to see dynamic modeling as a possible remedy to the problem Lucas describes. However, Lucas critique *was* about dynamic models. Macroeconomists at the time were well aware of dynamic programming and optimal control. A survey of control-theoretic tools in macroeconomics from 1976 starts with the lines:

In the past decade, a number of engineers and economists have asked the question: "If modern control theory can improve the guidance of airplanes and spacecraft, can it also help in the control of inflation and unemployment?"²

If anything, the 60s and 70s had been the heyday of dynamic modeling. Entire disciplines, such as *system dynamics*, attempted to create dynamic models of complex social systems, such as, corporations, cities, and even the western industrial world. Proponents of system dynamics used simulations of these models to motivate consequential policy propositions. Reflecting on these times, economist Schumacher wrote in 1973:

There have never been so many futurologists, planners, forecasters, and model-builders as there are today, and the most intriguing product of technological progress, the computer, seems to

offer untold new possibilities. [...] Are not such machines just what we have been waiting for?³

It was not the lack of dynamic models that Lucas criticized, it was the fact that policy may invalidate the empirical basis of the model. Lucas' critique puts pressure on how we come to *know* a model. Taking action can invalidate not just a particular model but also disrupt the social and empirical facts from which we derived the model.

If economics reckoned with this problem decades ago, it's worth taking a look at how the field has developed since. Oversimplifying greatly, the ambitious macroeconomic theorizing of the 20th century gave way to a greater focus on microeconomics and empirical work. Field experiments and causal inference, in particular, are now at the forefront of economic research.

Fundamental limitations of dynamic models not only surfaced in economics, they were also called out by control theorists themselves. In a widely heralded plenary lecture at the 1989 IEEE Conference on Decision and Control, Gunter Stein argued against "the increasing worship of abstract mathematical results in control at the expense of more specific examinations of their practical, physical consequences." Stein warned that mathematical and algorithmic formulations often elided fundamental physical limitations and trade-offs that could lead to catastrophic consequences.

Unstable systems illustrate this point. A stable system has the property that no matter how you disturb the system, it will always come back to rest. If you heat water on the stove, it will always eventually return to room temperature. An unstable system on the other hand can evolve away from a natural equilibrium exponentially quickly, like a contagious pathogen. From a computational perspective, however, there is no more difficulty in mathematically solving a sequential decision making problem with unstable dynamics than in solving one with stable dynamics. We can write down and solve decision making problems in both cases, and they appear to be of equal computational difficulty. But in reality, unstable systems are dangerous in a way that stable systems are not. Small errors get rapidly amplified, possibly resulting in catastrophe. Likely the most famous such catastrophe is the Chernobyl disaster, which Stein described as the failure to "respect the unstable" inherent in the reactor design.

As the artificial intelligence and machine learning communities increasingly embrace dynamic modeling, they will inevitably relearn these cautionary lessons of days past.

Beyond pattern classification?

Part of the recent enthusiasm for causality and reinforcement learning stems from the hope that these formalisms might address some of the inherent issues with the static pattern classification paradigm. Indeed, they might. But neither causality nor reinforcement learning are a panacea. Without hard earned substantive domain knowledge to guide modeling and mathematical assumptions, there is little that sets these formalisms apart from pattern classification. The reliance on subject matter knowledge stands in contrast with the nature of recent advances in machine learning that largely did without—and that was the point.

Looking ahead, the space of machine learning beyond pattern classification is full of uncharted territory. In fact, even the basic premise that there is such a space is not entirely settled.

Some argue that as a practical matter machine learning will proceed in its current form. Those who think so would see progress coming from faster hardware, larger datasets, better benchmarks, and increasingly clever ways of reducing new problems to pattern classification. This position isn't unreasonable in light of historical or recent developments. Pattern classification has reemerged several times over the past 70 years, and each time it has shown increasingly impressive capabilities.

We can try to imagine what replaces pattern recognition when it falls out of favor. And perhaps we can find some inspiration by returning one last time to Edmund Halley. Halley is more well-known for astronomy than for his life table. Much of astronomy before the 17th century was more similar to pattern recognition than fundamental physics. Halley himself had used curve-fitting methods to predict the paths of comets, but found notable errors in his predictions for the comet Kirch. He discussed his calculations with Isaac Newton, who solved the problem by establishing a fundamental description of the laws of gravity and motion. Halley, so excited by these results, paid to publish Newton's magnum opus *Philosophiæ Naturalis Principia Mathematica*.

Even if it may not be physics once again or on its own, similarly disruptive conceptual departures from pattern recognition may be viable and necessary for machine learning to become a safe and reliable technology in our lives.

We hope that our story about machine learning was helpful to those who aspire to write its next chapters.

Bibliography

¹ Ursula Franklin. *The Real World of Technology*. House of Anansi, 1999.

² David Kendrick. Applications of control theory to macroeconomics. In *Annals of Economic and Social Measurement*, volume 5, pages 171–190. NBER, 1976.

³ Ernst Friedrich Schumacher. *Small Is Beautiful: A Study of Economics as if People Mattered*. Random House, 2011.