A Toy Model of Scientific Progress

By Susanne Lohmann*

Abstract. Scientific research programs have a life cycle consisting of a sequence of overlapping subprograms. Each subprogram works within set boundaries, driven by the internal logic of its theoretical and empirical puzzles; outside of the boundaries lie the blind spots of the subprogram. Over time, boundaries shift as the blind spots become hot spots of inquiry. For a while, the status-quo subprogram overlaps with the subprogram challenging it, but eventually it is superceded. Change occurs through replacement (the market entry of Young Turks threatens the Establishment) rather than adaptation (the Establishment moves with the times). Scientific progress depends on society supporting scientists and scholars by giving them a “room of their own,” and it requires systematic methods, but the most important determinant is the structure of the scientific enterprise. For scientific progress to occur, the structure must allow scientists and scholars to form deeply specialized clusters and at the same time force cross-fertilization across clusters.

I

What is Science for?

Consider a toy society that interacts with harsh and capricious nature. Every day, in every way, year in and year out, Nature spews out disasters. For convenience, let us say there are 10 possible disasters, and they are numbered 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. A string of disasters occurring over the course of 50 time periods, at a rate of one disaster per time period, might look like this:

*Susanne Lohmann received a Ph.D. in Economics and Political Economy from Carnegie Mellon University in 1991. She is Professor of Political Science and Director of the Center for Governance at UCLA. She has published about two dozen articles and a dozen book chapters, mostly on collective action and central banking. Professor Lohmann is completing a book on How Universities Think. This paper was prepared for the special issue of the American Journal of Sociology and Economics on “Production and Diffusion of Public Choice.”

To protect itself from disaster in any given time period, the society can take one action. For convenience, the 10 possible actions are numbered 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. For effective protection, the number of the disaster and the number of the action must match up. In the event of a mismatch, the disaster causes great damage. We can think of this society as achieving a utility level of “1” in a given time period if the number of the disaster and the number of the action match up and a utility level of “0” in the event of a mismatch. Thus, a society that responds to the string of disasters above with a perfectly matched string of actions,

51038209815383497293759093119785309126885713234926,

would achieve the utility level 50 over the course of 50 time periods. Whereas a society that responds with a partially mismatched string of actions

93485472625348701938169402745645620437542859542378

would achieve the utility level 4. (The matching four disasters and actions are highlighted in bold.)

So far, so good. Here is the snag. The society must choose the action before Nature moves, or simultaneously with Nature’s move. In other words, at the time of choosing the action, the society does not know which disaster Nature will produce. To protect itself effectively against disaster, the society must predict which disaster will occur next. The society does not, however, know the underlying law by which Nature produces its disasters—indeed, for all people know, the sequence of disasters may well be random. It is obviously of value to the society to figure out whether there is an underlying law and what it is.

II

Common Sense and the Scientific Method

In their daily lives, human beings are quite good at correlating small-scale disasters and actions and drawing conclusions about appropri-
ate actions. (Indeed, one could say this is what the process by which a newborn develops into an adult is all about.) For example, people carry umbrellas with them to avoid getting wet when it rains, and they are more likely to carry umbrellas with them in rainy season. In this example, the causality linking past disasters ("rainy season") to possible disaster today ("rain"), potential damage ("getting wet"), and preventive action ("carry umbrella") is straightforward and transparent, which is why people get it right.

For large-scale social disasters and actions the mechanisms of causality tend to be rather more complex and opaque. When information and decision-making powers are distributed across many people, individual actors only see a small piece of the action, and it is hard for them to get a synoptic view of what is going on and why it is going on. How should we as a society fight crime and drugs, or reform the Los Angeles Unified School District?

Some disasters are low-probability events. A single individual may never experience such a disaster in her lifetime, and the few that do are hardly in a position to provide a succinct analysis. Do electric power lines, asbestos, and saccharin cause cancer, and how should we regulate them?

Introspection, casual observation, and common sense do not get us very far with questions of this kind. To come to reliable conclusions, society must pool the experiences of many people, many countries, and many historical time periods. It must explore alternative models and discriminate statistically between them.

All of this takes time and energy and resources. At a minimum, society needs to give some people a "room of their own"—time and support—so they can devote themselves full-time to the acquisition of specialized skills and the exploration of disasters.

Furthermore, it takes a systematic method to figure out Nature—a method that runs counter common sense. Human beings have Occam’s Razor built into their heads (possibly as a result of evolutionary selection effects). They go for straightforward and simple explanations. This is not a problem when people reason about simple causal relationships such as the causality between carrying an umbrella and not getting wet when it rains. It is a problem when mechanisms of causality are complex. Because the human brain is ill-
equipped to think about complex systems, people tend to jump to erroneous conclusions based on spurious correlations. If after a long drought a bush burns and it starts to rain and another time the bush burns and once again it starts to rain soon after, then in next to no time at all people start praying to the Bush God while their priest sets fire to the bush, the idea being to bring about rain. In the case of primitive tribes, we speak of superstition. In the modern era, we have Love Canal.

III

How Science Makes Sense of Nature (It's Messy)

Meanwhile, our toy society sets up universities and populates them with scientists who specialize in inventing (or discovering) and testing theories about disasters—and spawning new generations of specialized scientists who do the same.

There are too many disasters out there for one person to analyze them all. So each scientist analyzes a subset of disasters and publishes her results, and then other scientists examine her results in the light of their data sets. They confirm or reject her results and come up with modifications or competing theories.

For example, suppose there are five scientists, labeled $A$, $B$, $C$, $D$, and $E$. Each of them analyzes a string of numbers, representing disasters, of length 10—$A$ analyzes the first 10 numbers, $B$ the next 10 numbers, and so on:

$A$: 5103820981;
$B$: 5383497293;
$C$: 7590931197;
$D$: 8530912688;
$E$: 5713234926.

$A$ right off the bat discovers a law: 0 is a two-period leading indicator of 8; that is, an 8 always follows two periods after a 0. (It happens two times in his data set, and one can fit a regression line to two points.) $B$, $C$, and $D$ reject the law out of hand: $B$ provides an example in which 8 is not preceded by a 0 two periods earlier; $C$ shows that the 0 in her data set is not followed by an 8 two periods later; and
$D$ makes a killing by showing that the 0 in his data set is not followed by an 8 two periods later and that the 8 in his data set is not preceded by a 0 two periods earlier. $A$’s law is dead meat.

Meanwhile, $C$ visits $D$’s department on a fellowship. The two scientists enjoy hiking together, and over the course of many conversations they find out that their data sets share a 09 in common. They co-author a paper suggesting that 0s and 9s come in pairs, though they are careful to point out that $C$’s data set contains two additional 9s that do not come attached to 0s: clearly, a 0 is a necessary condition for a 9 but not vice versa. In a follow-up paper, they show that it is 5 that makes all the difference: 5 is a leading indicator of a 09; in the absence of a 5, 9s can occur without being preceded by a 0.

$B$ and $E$ compile evidence suggesting that the theory advanced by $C$ and $D$ is wrong. They show that the 5s in their data sets are not followed by 09s. $B$ and $E$ end up co-authoring a paper, and in the course of their joint research efforts they find out that in both of their data sets 5 is a leading indicator of 34. Further study reveals that 34 is “always” (that is, in both of their data sets) followed by a 9. Now the two scientists have a full-blown theory suggesting that 5 is a leading indicator of 349. Their theory is far richer than the impoverished theory advanced by $C$ and $D$ (which in any case is wrong). $B$ and $E$ come to be known as the 5349-ers, and their graduate students are much in demand.

Meanwhile, their opponents $C$ and $D$ have attracted $A$ to their cause. Jointly, $A$, $C$, and $D$ figure out that 09 is followed by a 1. So now they have a more comprehensive theory saying that 5 is a leading indicator of 09, which in turn is a leading indicator of 1. To distinguish themselves from $B$ and $E$, they call themselves the 5091-ers.

With this ferocious scientific debate going on, it almost goes unnoticed that the competing groups have made some scientific progress: a consensus is emerging that 5 is a leading indicator of 9.

At this point the debate is stuck. The data sets of the two camps are exhausted, and everybody is stymied. (Which doesn’t prevent them from engaging in busywork, regressing one thing on another and publishing it.) Stymied, that is, until Young Turk $D’$—the son of $D$—figures out a revolutionary new way of “seeing” the string of observations. He has some extraordinarily powerful results to back
up his claim that the new way of seeing the data is superior. His contribution reorganizes people’s thinking about the way the world works. The older generation sticks to its traditional knitting, but several scientists of the younger generation—$A'$, $B'$, $C'$, and $E'$—join the revolution.

Whereas established scientists examine the correlation of disasters over time, $D'$ studies the correlation of disasters across countries. He splits up the 50 disaster string in a different way: one data set consists of the first number in the long string and every fifth number thereafter; another data set of the second number and every fifth number thereafter; and so on. $D'$ himself can of course only analyze a string consisting of 10 numbers and thus works with one of the data sets, but $A'$, $B'$, $C'$, and $E'$ pitch in and analyze other data sets:

\begin{align*}
A' & : 5259738153; \\
B' & : 1037515274; \\
C' & : 0982913619; \\
D' & : 3839090832; \\
E' & : 8143979826.
\end{align*}

$D'$'s claims about 3 and 8 (his data set contains 383 and 83) initially stun the academic world. $A'$ backs him up with a 38 finding. Later-comers suggest that $D'$’s claims do not hold up. And so the scientific dispute over theories, tests, countertheories, and further tests goes on.

Meanwhile, scholar $A''$—the daughter of $A'$—modifies a method invented in another discipline and imports it into the debate. The method allows individual scholars to process more data. So each scholar can now look at strings of length 25 rather than 10. $A''$ and $D''$ the son of $D'$—apply the new method in combination with $D'$’s revolutionary new way of seeing the data:

\begin{align*}
A'' & : 5259738153103751527409829; \\
D'' & : 1361938390908328143979826.
\end{align*}

$B''$ and $E''$—the daughters of $B'$ and $E'$—use $A''$’s new method but reject $D'$’s new way of seeing the data. So they work with the same data but differently organized:
Meanwhile, $C''$ continues to work with the data set of his mentor $C'$:

$$C'': 0982913619.$$ 

$C''$'s findings are surpassed before his dissertation is published. He fails to find employment as a postdoc and vanishes from the face of the earth.

$A''$'s methodological innovation allows individual scholars to come up with more refined and robust theories, which speeds thing up—previously scholars had to pool their findings over the course of several publications and counterpublications to come to the same result that one scholar can now produce on his own. Rather mysteriously, this methodological innovation does not make a difference for the debates, which continue unabated, albeit at a more sophisticated and ferocious level.

While all of this scientific churning is going on, Nature continues to produce disasters:

$$87416754564556789785467643787289908456873478675865.$$ 

The older generation that grew up with their data sets doesn’t really notice that something new has happened until two Young Turks, $A'''$ and $B'''$, run the existing theories on new data:

$$A'''$: 8741675456455678978546764;
B'''$: 3787289908456873478675865.$$

$A'''$ and $B'''$ show that the traditionalists on both sides got it all wrong (5 is not a predictor of 9) and come up with theories of their own. Their outrageous results trigger a new round of theories, tests, countertheories, and further tests.

So far we have only considered relatively simple theories involving the sequencing of disasters. The matter becomes more complicated if the actions taken by society interact with the incidence of various disasters. In the simplest example, society protects itself from smallpox by eradicating smallpox from the face of the Earth. Then past disasters involving smallpox obviously have no predictive power for the future incidence of smallpox. More complex examples involve
nonlinear interactions between many actions and many disasters or even the creation of a new possible disaster (“global warming”). Here, scientists need to study not only the sequence of disasters but also the sequence of actions and 0–1 utility outcomes:

Disasters: 51038209815383497293759093119785309126885713234926
Actions: 93485472625348701938169402745645620437542859542378
Outcomes: 00000000001100000000001000000001000000000000000000

This inference process is clearly more complex by an order of magnitude.

In this system, scientific progress, if it occurs at all, occurs slowly. There are periods of incremental progress, periods of ossification, and periods of vibrance (Kuhn 1962). Incremental progress happens when scholars engage in normal science, developing theories and counter-theories and testing them and discriminating between them. Ossification occurs when an inbred group of scholars spin their wheels based on a very limited number of observations (five theories chase three data points). Vibrance is rare, and it happens when a scholar comes up with a new theory, a new way of “seeing” the string of observations, or a new method of analyzing the data. Typically the revolutionary scholar will be a fresh entrant into the system; older scholars tend to hang on to what is familiar rather than go with the times.

IV

Change Through Replacement Rather than Adaptation

The finding that change occurs through replacement rather than adaptation has important normative implications for the structure of the scientific enterprise: it suggests that society needs to allow for competition and make it easy for scientists with new ideas and methods to enter the system.

Competition comes with duplication, which is wasteful. Making entry easy is actually very hard in a system in which people must invest many years in building up the human capital that allows them to analyze disasters and in which much of that capital can be built up only by sitting at the feet of the established masters.
Because it takes elaborate structures to preserve competition and market entry, it is useful to think about why replacement is so important for scientific progress. What is it that prevents Establishment scientists from going with the times?

One reason is that scientists can hold only so many numbers and theories in their heads. Scientist $B$ who has spent his whole life analyzing a 10-disaster part of the string (highlighted in bold)

51038209815383497293759093119785309126885713234926

has a hard time “seeing” the 10 disasters he knows so well in the following string (which is reorganized according to the principles laid down by Young Turk $D$):

52597381531037515274098291361938390908328143979826.

$B$ stares at the new string, and in it he can’t find the disasters he is so familiar with.

Another reason is that scientists tend to have emotional investments in their pet theories. Research involves lots of dreary and plodding hard work showing that 3 and 8 are correlated even though conventional wisdom says they are not (past research didn’t properly correct for a confounding variable). Something has to move a scientist to get out of bed in the morning and try over and over again. And it’s not just the research itself, but also its publication. Something has to motivate the scientist to keep on submitting to refereed journals even when she keeps getting rejected, or to revise and resubmit a paper five times until the editor is satisfied. Unfortunately, the same thing that keeps a scientist going to show against all odds that 3 and 8 are correlated makes her disinclined to “see” the evidence compiled by other scholars suggesting that the correlation between 3 and 8 is spurious. Competition, too, tends to harden the views on both sides.

How Do We Tell if Scientific Progress is Occurring?

While the scientific debate is going on in the universities, the general public continues to suffer under the disasters and wonders when its
scientists will come up with a predictive theory that would allow people to shield themselves effectively against disaster. Every now and then some advice floats their way, as in “see a 5 prepare for a 9.” The advice seems to work at least some of the time. Different scholars give different advice, and occasionally one of them gets it right; the problem is it’s not always the same one who gets it right, so it’s hard to know whom to listen to.

Indeed, it is impossible to tell whether any scientific progress—defined as a cumulation of theories moving ever closer to the Truth—is occurring. The reason is that the society doesn’t know what the Truth is. In this example, I as the author of this paper know where the two strings of numbers

51038209815383497293759093119785309126885713234926

and

87416754564556789785467643787289908456873478675865

came from: I hammered away at the numbers on my keyboard in what seemed to me a random manner. It is quite likely, however, that if I were to produce a million numbers in this way, systematic patterns would emerge because my hands and fingers would settle down into some rhythm. Certain numbers would be more likely to occur than others, and certain numbers would be more likely to follow certain other numbers.

There is an practical way by which society can figure out whether its scientists are making scientific progress. If the match between disasters and actions improves over the years (and one would be able to tell because the utility level of the society would increase as a result), it is probably safe to say that science is contributing something of value (“people don’t die of infections anymore”).

From a theoretical point of view, if there is a regularity in the way Nature generates its numbers and the regularity stays relatively fixed (it doesn’t change too fast relative to the speed at which scholars are figuring it out), then a university of scholars who keep chipping away will eventually figure out the regularity, or at the very least get close. By generating new hypotheses and weeding out false hypotheses, or hypotheses that perform poorly relative to others, the university of scholars will eventually converge to the Truth.
VI

There’s Something Special about the Social Sciences

So far mine is a “hard science” model of natural disasters having to do with physics or chemistry or geology. In the hard sciences, the objects of study don’t think and theorize and change their behavior in response to their thinking and theorizing. Because the laws governing the behavior of the objects of study remain fixed, one would expect to see scientific progress occur. Within disciplines one would expect cumulative and encompassing results (a new theory builds on, and encompasses, an old theory).

But there are also human-made disasters—political, economic, social, and cultural disasters. These are the objects of study of the soft sciences, and specifically the social sciences. When people and social systems are the objects of study, they do not stay put; they talk back incessantly. People respond to a changing understanding of the way a social system works by changing their behavior, which changes the aggregate-level behavior of the social system. Social systems are more complex than physical systems by an order of magnitude.

The social sciences are shooting at a moving target rather than a standing-still one. For this reason, we would expect to see scientific progress occur more slowly in the social sciences, and depending on how fast Nature moves relative to the university of scholars, there may be no scientific progress at all. Theories don’t necessarily cumulate or encompass each other, even within disciplines. Indeed, it is possible that a theory might get it right in hindsight (so-called post-prediction) but then gets outrun by reality so that it does a disastrous job in predicting the future. It is also conceivable for a discarded theory to have a comeback and outperform current theory.

VII

The Importance of Structure

If Nature’s regularity is complex, it might take the analysis of a million disasters for the university of scholars to make noticeable progress. Recall that each scientist can only manage a data set containing 10 disasters. It would take 100,000 scientists to process a million
disasters. And a scientist whose cognitive capacity is limited to 10 data points is hardly going to be able to cope with the flood of scientific findings produced by his 99,999 colleagues. Even after the methodological innovation allowing individual scientists to hold a string of 25 data points in their heads, it still takes 40,000 scientists to come to grips with one million data points. There is no way one scientist can talk with 39,999 other scientists on a regular basis and pool her findings with theirs.

The reason why we have universities in the first place is that Nature’s laws are complex and there are limitations to the amount of data an individual can process. It immediately follows that universities cannot be places where everybody talks with everybody else all the time. The question is how to organize the university so that the people who get the most out of pooling data and results talk to each other in an ongoing way; the people who might get something out of occasionally picking each others’ thoughts talk to each other every now and then; and the people who have nothing to say to each other never meet.

The efficient organization of the university enterprise consists of deeply specialized clusters that are partially connected in a hierarchical network with feelers cutting across branches and subbranches of the hierarchy.

Deep specialization, which implies sharing methods, pooling data, and integrating research findings, occurs in small clusters. Because there is a certain amount of information pooling going on within a cluster that does not leak to other clusters, each cluster will come up with idiosyncratic ideas about the way the world works, idiosyncratic ways of seeing strings of observations, and idiosyncratic methods of analyzing the data. They develop their separate paradigms—and incompatible policy prescriptions.

But some information does leak through the semi-permeable walls separating the clusters. The network connecting the cluster is hierarchically organized. At the top is the division into larger fields of study, of which the social sciences is one. Within the larger fields of study, we find the disciplines, of which political science is one; within political science we find fields of study that remain relatively fixed over time, American politics being an example; within a field of study, we
find a subfield such as American institutions; within a subfield we find subsubfields such as Congress; within a subsubfield we find clusters competing over the interpretation of Congress (the distributional versus informational “take” on Congress).

In this hierarchy, the walls separating scientists and scholars are more permeable for those who occupy the subbranches of a common branch than for those who occupy the subbranches of different branches. That is, information wanders more easily across a subbranch or up and down from subbranch to branch and back than across branches or subbranches affiliated with different branches.

When the world, or some part of the academic enterprise, moves on, the branch organization may become obsolete. Even so, the hierarchy may continue to exist for a long time because it is reinforced by discipline-based evaluation systems and job markets. Individual subbranches may also ossify, but the scientists and scholars inhabiting those subbranches may not notice because they keep on publishing and their students keep on getting jobs and tenure—after all, the peers who conduct the peer review for publications, recruiting, and promotions are drawn from the same subbranch, and how would they know that they are obsolete? Because of deep specialization, the scientific enterprise has a built-in tendency to ossify.

Happily, curiosity and the desire to find better explanations cannot be suppressed. We find scientists and scholars in one subbranch reaching out to their colleagues in subbranches belonging to other branches: public choice and political economy are examples; so are evolutionary psychology and complexity theory. Such crossovers add vibrance to the academic enterprise. New ideas, new ways of seeing, and new methods jump from one cluster to another, and over the years they break down ossified subbranches and subsubbranches.

VII

The Ancient Science of Hepatoscopy

**Deep specialization is necessary for scientists and scholars to make cumulative progress by standing on the shoulders of giants. But precisely because of the narrowing of vision it comes with, scientists**
and scholars can get snagged in silly ideas, outdated methods, and obsolete ways of seeing the world.

To illustrate the importance of unsnagging, it is worthwhile looking at an example of a society that employed something resembling a scientific method and yet came to conclusions that to our modern understanding come across as disarmingly unscientific.

The Assyrians inherited from the Babylonians the ancient science of hepatoscopy, or divination through inspecting sheep's livers (Finer 1997: 214). Kings and generals would ask their priests for advice whether to go ahead with some royal or military plan (whether to marry off the king's daughter to a suitor or whether to go to war with a neighboring tribe). To answer such a question, the priests would sacrifice a sheep to a god. The priests would then remove the sheep's liver and inspect the markings on it. Over the years, the priests meticulously collected, collated, and filed all markings along with the event that had accompanied or ensued from the questions posed to the god. By comparing the current question and markings, the priests could conduct an exercise in induction and produce an answer.

There is no question that they used something resembling a scientific method and that their conclusions were backed up by an extensive set of empirical correlations. And yet our modern understanding revolts at the underlying hypothesis that the liver of a sheep could give information about anything other than the condition of the sheep.

IX

The Medieval University and the German Model

In the European Middle Ages, Peter Abelard’s Sic et Non (pro and con, or thesis and antithesis) invented (or discovered, or rediscovered) the idea that the great questions of the time should be examined from all sides. Nothing, especially not the deepest and most sacred religious beliefs, was safe from logic-chopping medieval scholasticism. The medieval university prepared the way for the ideal of free inquiry to flourish.

In the 19th century, the concept of the specialized research professor was invented in Scotland and Germany. The success of the
German model was in large part driven by the resulting deep specialization in research.

Deep specialization needs to be complemented by a process or structure that “unsnags” the scientific enterprise. Interestingly, it was the medieval university that supplied not only the idea of free inquiry but also the institutional structures supporting the free exchange and competition of ideas: decentralized federalist structures, bottom-up governance, complex voting procedures, representative assemblies, and institutionalized forms of conflict resolution.

What makes the American university of the 20th century so powerful, both in historical and contemporary comparison, is its combination of the German and medieval models: scientists and scholars come attached to deeply specialized clusters, but they are also part of a network that allows for ideas and methods to jump across clusters (Lohmann 2001).

References