Bread mold on a wounded yak
Democratizing discovery in the age of AI







As if rejection weren't enough.

Discovering Penicillin

There is a familiar story about the discovery of penicillin. Alexander Fleming returned from holiday in 1928 to his notoriously untidy laboratory at St. Mary's Hospital in London and noticed that one of his bacterial cultures had been contaminated by mold. More importantly, he noticed that bacteria surrounding the mold had disappeared. Instead of throwing the contaminated culture away, Fleming investigated. He isolated the mold, identified it as belonging to the *Penicillium* genus and eventually gave the antibacterial substance it produced the name penicillin.

It is one of the great stories of scientific discovery, and Fleming deserves his place in it.

But there is another story. Long before Fleming looked into his Petri dish, human beings had observed that molds could have curative properties. There are accounts of Tibetans applying bread mold to the wounds of yaks, while bread poultices survived as folk remedies elsewhere. Joseph Lister observed antibacterial effects associated with mold in 1871. Louis Pasteur and Jules François Joubert observed mold inhibiting anthrax bacilli in 1877. In Belgium in 1920, André Gratia and Sara Dath reported fungal contamination inhibiting bacterial growth.

Then Fleming discovered penicillin.

Or did he?

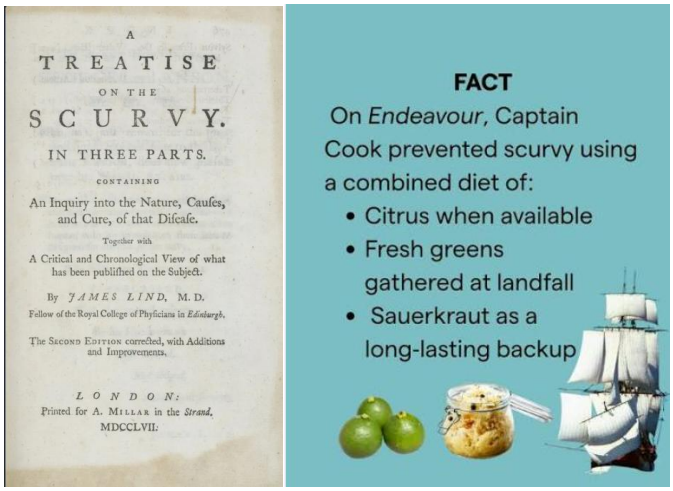
That is not an attempt to take Fleming's discovery away from him and award a retrospective Nobel Prize to an unidentified Tibetan yak herder. It raises a different question: what exactly do we mean when we say something has been discovered?

We often tell the history of discovery as though ignorance existed until somebody finally found the answer. But that is too simple. The problem is not necessarily that earlier people were ignorant. More often, different people possessed different pieces of knowledge that had not yet been connected.

Someone might discover an effect. Someone else learns that the effect can be reproduced. Another person identifies the mechanism. Someone isolates the active ingredient. Others establish dosage and safety. Engineers devise a means of manufacturing it at scale. Eventually all these pieces become sufficiently connected that society gives the resulting body of knowledge a name and a date and calls it a discovery.

What we call a discovery may therefore be better understood as a lifecycle of partial discoveries.

Captain James Cook knew something



Scurvy provides a wonderful example. James Lind's famous 1747 experiment demonstrated the extraordinary effectiveness of citrus fruit in treating sailors suffering from scurvy. Captain James Cook subsequently became remarkably successful at preventing the devastating mortality associated with scurvy on long voyages. Yet neither could have explained what we now regard as the essential biochemical fact: scurvy results from vitamin C deficiency. The vitamin had not been isolated, and the scientific architecture of vitamins and deficiency diseases did not yet exist.

Cook also believed in preventive measures that turned out to have little value against scurvy. His understanding was incomplete and partly wrong.

But saying that does not take us very far. Cook knew something enormously important. He knew that sailors did not have to die of scurvy on long voyages, and he demonstrated it in practice.

Later scientists knew something Cook did not. They could identify the substance and explain the mechanism. Those are not rival claims to the same discovery. They are different discoveries about the same reality.

That distinction is crucial because knowledge exists in many forms. Nor should everything outside experimental science be thrown into one residual category called traditional or experiential knowledge. Historical analysis, qualitative research, professional expertise, community knowledge, individual experience, oral tradition and

accumulated practical knowledge have different methods, strengths and weaknesses. None is automatically correct merely because it comes from experience. None is equivalent to a controlled experiment.

The point is precisely that they are different.

Discovery often occurs when knowledge residing in one of these worlds becomes connected with knowledge residing in another.

Discovery is not the finish line

There is also a second distortion in the way we tell these stories. I think of it as *present science bias*.

Once vitamin C was isolated, the history of scurvy seems to end. Once penicillin has been isolated, purified, manufactured and supplemented by other antibiotics, penicillin becomes something humanity has figured out. Darwin gives us natural selection; Mendel gives us the foundations of genetics. Titanium is found to integrate remarkably well with bone, opening the way to modern implants. Certain polymers are found capable of conducting electricity.

Discovery becomes a solved case.

We caught the culprit.

That is an understandable way to organize textbooks. It is a dangerous way to think about knowledge.

The history of discovery ought to teach us almost exactly the opposite lesson. If Cook could know something important about scurvy without knowing about vitamin C, and Fleming could know something enormously important about penicillin without possessing everything later generations would learn about antibiotics, why should we imagine that we happen to occupy the first generation that sees the whole picture?

We don't look back at 1928 and say Fleming knew everything there was to know about antibiotics. Why should someone in 2128 look back at us and conclude that we knew everything there was to know about them?

A discovery does not close a subject. It changes the frontier of ignorance.

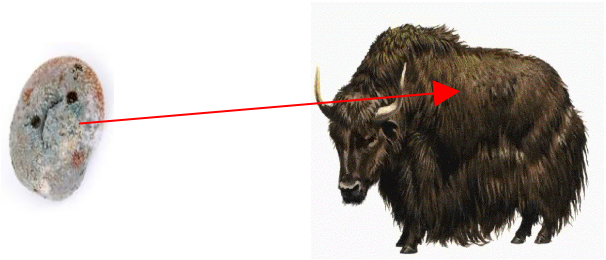
Indeed, mature science depends upon this principle. Natural selection did not end evolutionary biology. Mendelian inheritance did not end genetics. Penicillin did not end microbiology. Each powerful discovery reorganized the questions that could be asked next.

Yet our histories—and increasingly our artificial intelligence systems—can easily acquire a bias toward the present. They know the story backwards. They know which explanation eventually prevailed. The earlier observations are therefore interpreted as incomplete versions of what we now know, while today's explanation is presented as though it were simply knowledge.

The asymmetry should make us uncomfortable. Past knowledge was partial. Present knowledge is partial too.

Discovery and power





There is another reason this is important. Knowledge does not enter the world on equal terms.

Who possesses an observation is crucial. So does whether that person can write it down, whether it is expressed in the language society recognizes as expertise, and whether its possessor has access to a university, laboratory, journal, government department or other institution capable of testing and disseminating it.

The yak herder and Fleming differed in ways having nothing to do with the antibacterial properties of mold. One possessed knowledge embedded in practice and perhaps tradition. The other made an observation capable of entering institutions that could isolate it, test it, name it, explain it and eventually transform it into reliable medicine.

This does not diminish science. It identifies what science adds.

Scientific institutions can determine whether an apparent effect is reproducible rather than accidental. They can distinguish correlation from causation, discover mechanisms, identify harmful practices, establish dosage and safety, and turn a local observation into knowledge capable of benefiting millions.

The mistake comes when the institutional capacity to validate and extend knowledge becomes confused with having originated all the knowledge being validated.

Once knowledge acquires a paper, an author, an institution and a publication date, it becomes easy to cite. What came before becomes harder to see. Eventually the archive can become mistaken for the history, and the history for everything humanity knew.

Social policy has its own wounded yaks

This is not just an argument about the history of medicine. Social policy is an unusually good place to see this problem because governments routinely possess enormous quantities of administrative and statistical knowledge and see the world through the lens of an analytic language that tends to see only averages and cross-sections All the

while, the people experiencing the consequences possess another kind of knowledge expressed as personal lifecourse narratives — and the two are often poorly connected.

Consider someone receiving social assistance who says:

> *"I worked more hours last month and, after they reduced my benefits and I paid the costs of working, I was hardly any better off."*

A policy analyst may subsequently describe the interaction of earnings exemptions, benefit reduction rates, taxes and other income-tested programs as producing a high marginal effective tax rate.

The analyst hasn't merely repeated the recipient's observation. The analyst may calculate its size, discover which program interactions produce it, determine how widespread it is, compare different designs and propose a remedy.

But neither did the analyst create the phenomenon by naming it.

The recipient knows what the system does to her. A community worker who has watched the same thing happen to a hundred people knows something else. Administrative data can establish its prevalence. An economist can model behavioural effects. A historian may show that governments have encountered the problem before. A policy analyst can connect the pieces. Government possesses the authority to change the rules.

The breakthrough comes when different pieces of knowledge meet.

This gives us a better answer to the question about why a broader conception of discovery is useful. It is useful because our institutions can miss important knowledge not because nobody possesses it, but because nobody has connected its different forms.

AI enters the story



This becomes where artificial intelligence becomes genuinely interesting.

For most of human history, translation between intellectual worlds has been expensive. Moving an observation from one vocabulary into another generally required a person who inhabited both worlds. A community needed a researcher. A patient needed a doctor willing to listen. A citizen needed a lawyer. A local observation often required a scientist, journalist or policy analyst before institutions could recognize what was being described.

AI can make some of that translation astonishingly affordable.

Someone doesn't need to know the phrase “marginal effective tax rate” to describe accurately what happens when earning another dollar leaves her scarcely better off. AI can potentially move from that description toward the analytical literature.

But the more revolutionary possibility works in the opposite direction as well.

AI can take the economist's marginal effective tax rate and ask what it looks like in people's actual lives. It can compare apparently unrelated accounts expressed in ordinary language and ask whether they describe the same underlying phenomenon. It can search history for earlier versions of an idea currently being presented as new. It can ask whether a community practice and a scientific finding might be different descriptions of related effects.

AI therefore need not simply translate knowledge.

It might connect partial discoveries.

That is potentially democratizing because someone possessing one piece of the puzzle no longer necessarily has to master the professional language of all the other pieces before being allowed to participate in assembling them.

But AI has a wounded-yak problem of its own

There is a considerable danger here.

AI is trained largely on what humanity has recorded. Consequently, it can reproduce exactly the hierarchy this argument questions. The published overwhelms the unpublished. The written overwhelms the oral. The institutional overwhelms the informal. The highly cited overwhelms the forgotten. And the explanation currently regarded as authoritative overwhelms the possibilities that have not yet acquired authoritative language.

The yak disappears from the dataset.

There is also present science bias. Ask an AI system about scurvy and it naturally organizes its answer around vitamin C deficiency. Ask about bacterial infection and modern antibiotics become the explanatory centre. That is normally useful: we want an AI system to tell us the best-supported current knowledge rather than give equal standing to obsolete theories.

But there is a subtle intellectual cost.

Because AI knows how the historical story turned out, it is tempted to interpret everything that came before as an imperfect approach to today's answer. More importantly, it can treat today's answer differently—as though the intellectual journey has arrived at its destination.

I encountered precisely this problem in thinking about the yak. AI repeatedly wanted to tell the history as a progression toward the correct modern explanation. I had to keep asking it to entertain a different possibility:

> *What if the earlier practitioners knew something real that the later scientific discovery explained rather than replaced? And what if the scientific explanation we possess today is itself another stage rather than the end of the lifecycle?*

The same problem appears in social policy.

Take the familiar proposition that older people should be enabled to remain in their own homes for as long as possible. It sounds humane, progressive and evidence-informed, and in many circumstances it is.

But its very familiarity can make the framing disappear.

Why must the objective be remaining in *the home* rather than remaining in one's home or chosen community? An older person might prefer a congregate setting, a supportive community, a different neighbourhood, a multigenerational arrangement or some form of housing we have not yet designed. “Aging in place” can quietly turn one historically contingent model of independence into the definition of independence itself.

An AI system trained on contemporary policy literature can reproduce that framing effortlessly. It may be very good at describing the cutting edge of the current consensus while being much less naturally inclined to ask whether the edge itself is pointing in the right direction.

That is not uniquely an AI defect. Human institutions do exactly the same thing.

But AI can reproduce the consensus at extraordinary speed and with extraordinary confidence.

What we should ask AI to do

The answer cannot be to ask AI to distrust science or treat every unconventional claim as a suppressed discovery. That would replace one intellectual mistake with a much more dangerous one.

The achievement of science is precisely its capacity to distinguish effects from coincidences, useful practices from harmful ones, plausible mechanisms from imaginary ones, and discoveries from mistakes.

What we need from AI is something harder:

> *Respect the best-supported knowledge of the present without mistaking the present for the end of knowledge.*

That requires asking different questions. What do we know? How do we know it? What preceded the current explanation?

Which earlier observations turned out to be wrong, and which were incorporated into the later explanation? What apparently different observations might describe the same phenomenon?

What knowledge exists in professional practice, historical research, qualitative evidence or lived experience that has not yet been connected with the dominant analytical account?

And perhaps most importantly:

> *What assumptions in the question are so familiar that we have stopped recognizing them as assumptions?*

Those questions don't weaken science. They embody the scientific temperament.

Discovery without a finish line

Perhaps our mistake has been imagining discovery as the intellectual equivalent of discovering territory. Something exists out there; an explorer reaches it, plants a flag, and history attaches the territory to the explorer's name.

There is something revealing about that metaphor because the land was frequently already inhabited. Ideas sometimes are too.

A better model is a network of partial discoveries. Someone observes something. Someone remembers it. Someone learns to reproduce it. Someone measures it. Someone discovers that part of the explanation was wrong. Someone isolates an active ingredient. Someone identifies a mechanism. Someone makes the result safe. Someone makes it scalable. Someone connects an ancient observation with a new scientific language. And then someone else discovers that what everybody thought was the completed explanation was itself only another piece.

There really are Flemings: individuals who see connections others have missed, and making such a connection is itself a profound discovery. There are Cooks, who demonstrate that something can be accomplished without possessing the later grammar that explains it. There are scientists who transform fragmentary observations into reliable knowledge capable of saving millions of lives. And somewhere in human history there was somebody who put mold on an infected wound, saw what happened and did it again.

They did not know the same things. But neither do we know everything they were collectively on their way to knowing.

That changes what we should want from artificial intelligence. Its greatest contribution may not be making discoveries on our behalf. Nor is it simply translating expert knowledge into ordinary language.

It may be able to help us see relationships among things humanity already knows separately—across disciplines, centuries, institutions and vocabularies—and to give

people possessing different fragments a greater opportunity to participate in putting them together.

But to do that, AI will have to resist a powerful temptation shared by textbooks, institutions and human beings themselves: the belief that because the present can explain the past, the present must possess the completed explanation.

Civilization advances when fragments of knowledge become connected. It advances further when we remember that the picture assembled from those fragments is never necessarily the last one.

And perhaps that is the most useful lesson of the bread mold on the wounded yak. The past was not simply ignorant while waiting for us to discover the truth.

They had pieces. We have more pieces. And there are still pieces missing.

Conclusion

There really are 'Flemings' and civilization needs them: people who see something others have missed and possess the means and curiosity to pursue it. But there are Cooks too, who discover that something works without knowing why it works. And somewhere, long before either of them, there was a yak herder who put moldy bread on an infected wound, watched what happened, remembered it and did it again.

They were all doing different things. Fleming could investigate and isolate an effect. Cook could demonstrate one repeatedly in practice. The yak herder could observe, remember and pass something useful along. To insist that only one of these deserves the word *discovery* is to confuse our modern ability to explain knowledge with the much older human ability to acquire it.

That may be the larger lesson. Humanity did not spend most of its history waiting for science to arrive so that discovery could begin. We watched, tried, failed, remembered, compared and passed things on. Science gave us extraordinarily powerful ways to test those observations, discard the ones that were wrong, explain the ones that were right and turn scattered knowledge into reliable knowledge capable of benefiting millions. That is one of humanity's greatest achievements. But it does not follow that the knowledge science eventually explains began when science named it.

The yak herder knew something. Cook knew something. Fleming knew something. Those who came after Fleming knew more. And people who come after us will almost certainly know things that make some of our own confident explanations look partial. That should not diminish any of these discoveries. It should enlarge our idea of discovery itself.

The past was not an empty room waiting for us to turn on the light. People had been finding their way around in it for a very long time. Fleming gave us a brighter light. Cook had already learned where some of the obstacles were. And somewhere, centuries earlier, a yak herder had learned where to put her hand.

Perhaps that is what discovery really is: not the moment someone finally possesses the truth, but the long human process by which one observation finds another, experience acquires explanation, explanation produces new questions, and fragments become a picture—until somebody discovers that the picture itself was only another fragment.

AI – properly instructed - should be able to help us with this dilemma.