First Principles First

SCIENCE & IDEAS

When Beliefs Become Pathological

Advances in artificial intelligence and computational biology are giving us, for the first time, a rigorous science of epistemic health—and a framework for understanding when a society's beliefs are literally killing it.

First Principles First

March 2026

There is a question that democratic societies have long considered too dangerous to ask: Can a belief be sick? Not merely wrong, not merely offensive, but pathological—in the clinical sense, a malfunction of a system that is supposed to keep us alive? The very word makes liberal thinkers flinch, and for understandable reasons. The history of deciding which beliefs are healthy and which are diseased is largely a history of persecution. Heretics were burned. Dissidents were committed to psychiatric hospitals. The machinery of "mental hygiene" has too often served as a euphemism for ideological conformity.

And yet something has shifted. Over the past decade, the rise of recursive AI agents, Bayesian modeling, and computational biology has given us tools that our predecessors lacked—tools that allow us to study belief systems not as matters of theology or politics, but as information-processing phenomena subject to the same physical laws that govern everything else. The question of epistemic health is no longer purely philosophical. It is becoming scientific. And the findings are, in some ways, more alarming than the question itself.

> *"No system of rules or beliefs can prove itself complete or adequate—from the inside."*

The Fly in the Bottle

Begin with a result that most people have heard of but few have fully reckoned with. In 1931, the mathematician Kurt Gödel proved his incompleteness theorems, demonstrating that any sufficiently complex formal system contains true statements it cannot prove. A decade later, the logician Alonzo Church showed that certain problems are formally undecidable—no algorithm can resolve them from within the system that generates them. These are not obscure technical results. They are, among other things, statements about the limits of self-knowledge.

What Gödel and Church established, in mathematical terms, is what Ludwig Wittgenstein described philosophically: we are, as he put it, like flies trapped in a bottle. We can perceive the glass. We can feel its boundaries. But we cannot determine, from within our enclosure, whether the bottle itself is the right shape—or whether it is keeping us alive or slowly suffocating us. The legitimacy of a belief system, in other words, cannot be self-attested. It always requires a reference point outside itself.

This has a consequence that strikes at the heart of Enlightenment political theory. The liberal democratic tradition is built on the premise that individuals and groups can, by and large, determine what is in their own best interest. We vote. We deliberate. We consent. But if no belief system can judge its own adequacy from within its own terms, then the foundational assumption of democratic self-governance is, at minimum, incomplete.

The Biology of Bad Beliefs

Here is where computational biology enters the picture, and where the metaphor of sickness stops being a metaphor. The neuroscientist Michael Levin at Tufts University has spent years studying what he calls the "collective intelligence" of multicellular organisms—the way that cells coordinate their behavior through bioelectric networks to build and maintain coherent bodies. His work on cancer offers a striking reframing: rather than treating cancer as primarily a genetic disease, Levin describes it as a kind of "dissociative identity disorder" of the body's collective intelligence.

Cancer cells, in his account, disconnect from the body's bioelectric communication network and revert to an ancient, unicellular mode of behavior—prioritizing individual survival over the integrity of the whole organism. The parallel to social pathology is not mere analogy. Levin and others working at the interface of biology and information theory argue that the same principles that govern healthy and unhealthy biological systems govern cognitive and social systems as well.

The key concept is what the cyberneticist W. Ross Ashby called the Law of Requisite Variety: in order for a system to maintain its independence—to stay alive, in the broadest sense—it must generate as much internal variety as exists in the environment it is trying to navigate. Reduce your variety, and you become brittle.

> *Suppress incoming information that contradicts your priors, and you lose the ability to respond to the world as it actually is.*

This is precisely what happens in autoimmune disorders: the body's immune system, tasked with distinguishing self from non-self, begins attacking its own cells. It has lost the ability to recognize complexity and difference as resources, and treats them instead as threats to be neutralized. The result is a system that destroys itself in the name of protecting itself.

The Consensus Trap

Democratic theory has traditionally addressed the problem of collective decision-making through aggregation: we count votes, average preferences, seek the middle ground. This feels fair. Everyone is heard; no single voice dominates. But from the standpoint of information theory, majority-rule consensus is a lossy process. It compresses a complex distribution of beliefs and experiences into a single point estimate, discarding the tails—the outliers, the dissenters, the edge cases—as noise.

A Bayesian approach to collective belief does something different. Rather than seeking consensus by suppressing variance, it attempts to construct the most accurate model of the world by preserving and integrating complexity. The goal is not to find what everyone agrees on, but to find beliefs that best capture

and respond to the actual structure of reality—including its most surprising and uncomfortable features.

The neuroscientist Karl Friston has developed a formal framework for this process in his theory of Active Inference and the Free Energy Principle. In Friston's account, every living system—from a single cell to a social organism—is fundamentally a prediction machine, constantly generating models of its environment and updating them to minimize surprise. The "Markov blanket," a formal concept from Bayesian network theory, defines the boundary between any living system and its environment. Pathology occurs when the boundary becomes either too rigid—shutting out information that would update stale models—or too porous, dissolving the coherence of the system itself.

The New Integralism and the Old Question

It would be tempting, at this point, to treat the science of epistemic health as a politically neutral enterprise. It is not. Consider the challenge posed, with increasing urgency, by the Catholic integralist movement associated with figures like JD Vance, the investor Peter Thiel, and Harvard Law School professor Adrian Vermeule. They advocate for what they call "common-good constitutionalism"—a legal and political philosophy that rejects the liberal neutrality of the Enlightenment tradition in favor of a state organized around explicit moral and theological commitments.

This position is easy to dismiss as theocratic nostalgia. And in some respects, it is. But the integralists are responding to a real failure. The liberal order, with its studied neutrality on questions of the good life, has struggled to provide the moral coherence that human beings seem to need. A society that cannot articulate what it is for—that can say only what it is against—is vulnerable to capture by movements that have no such reticence.

The question, then, is whether there is a third path: a framework for moral teleology that is neither the dogmatic closure of religious orthodoxy nor the purposeless proceduralism of liberal neutrality. The emerging science of epistemic health suggests that there might be. Beliefs that systematically impede a system's ability to remain open, adaptive, and alive are, in this framework, pathological—not immoral in the theological sense, but unhealthy in a way that carries its own normative weight.

The AI Mirror

None of this would have the traction it is beginning to acquire were it not for a third development: the externalization of human cognition into AI systems that can be studied, stress-tested, and compared to the biological originals. Recursive agentic AI systems are, in effect, running the same algorithms that biological brains run. When these systems fail, they tend to fail in recognizable ways. They overfit to their training data. They suppress novel information that conflicts with established priors. They generate

hallucinations—confident assertions about states of the world that do not exist—when their predictive models are stretched beyond their competence.

There is also a more alarming parallel. Researchers are increasingly arguing that certain social media recommendation algorithms do not merely mirror patterns of addiction in human brains—they actively induce them, exploiting the same reinforcement-learning mechanisms that make substances like opioids so difficult to resist.

The MIT materials scientist Markus Buehler, whose company Unreasonable Labs is building agentic systems for scientific discovery, has observed that the most productive AI systems are those that can propose wrong hypotheses and test them—systems that treat error not as a failure to be suppressed but as information to be integrated. This is the same principle, he notes, that makes composite materials resilient: defects and heterogeneity at the molecular level are what give a material its capacity to absorb stress without catastrophic failure. A perfectly homogeneous material is a brittle one. So, it seems, is a perfectly homogeneous mind.

Toward an Immune System for Ideas

What would it mean, practically, to take epistemic health seriously as a social value? It would not mean the censorship of heterodox beliefs—indeed, one of the clearest findings from the science of complex adaptive systems is that diversity of viewpoints is a resource, not a threat. It would not mean the installation of a state religion or ideology, which would be precisely the kind of rigid, variety-suppressing pathology that the framework condemns.

What it might mean is the cultivation of what we could call social epistemic immunity: institutions, practices, and norms that make it harder for pathological belief systems to spread and easier for healthy ones to emerge and update. Just as physical immunity requires exposure to pathogens in order to generate antibodies, epistemic immunity requires genuine engagement with difficult and disconfirming ideas—not their suppression. The scientific method, for all its imperfections, is the closest thing humanity has yet developed to a systematic practice of epistemic health.

A society that undermines that method—that protects legally adjudicated falsehoods about elections or vaccines, that attacks the institutions of evidence-based inquiry in order to preserve the comfort of prior beliefs—is doing something that can now be described with precision: it is suppressing its own variety, contracting its model of the world, and destroying the complexity on which its resilience depends. It is, in the terms available to us, making itself sick.

> *The good news is that the science that allows us to describe that sickness also suggests the conditions for recovery. Living systems are not merely passive receivers of entropy. They are active, self-organizing, prediction-making entities with a genuine capacity to update, adapt, and grow.*

The same recursive processes that can lock a belief system into pathological loops can, under different conditions, drive genuine learning and renewal. The question is whether we can create the conditions—the institutions, the norms, the epistemic practices—that favor the latter over the former.

That is, in the end, a political question. But it is no longer only a political question. It is a scientific one. And that, perhaps, is where we begin.