Do Minds Have Immune Systems?

Andy Norman, Barry Mauer, Luke Johnson…? (+??)

Abstract: Do minds have immune systems? In this paper, we remove several obstacles to deciding this question in a rigorously scientific way. First, we show why the scientific community needs to take up the question. Then, we give the hypothesis a name: the *Mental Immune Systems Thesis* (or MIST) is the claim that minds *do in fact* have immune systems. It's tempting to dismiss this claim as "mere metaphor" – and many do – but that stance turns out to be indefensible. It is at best a well-intentioned stopgap: one that postpones a pivotal reckoning. So how to settle the question? Above all, we need clarity about the meaning of "immune system." To that end, we examine candidate definitions, nominate one, and show why it makes sense to embrace that definition. We then consider an evolutionary argument for MIST: mental immune systems, so defined, didn't just evolve, they *had to* evolve – to protect minded creatures from informational threats. We then detail some of MIST's testable implications and summarize the extant empirical evidence. Finally, we discuss the prospects of *cognitive immunology*, a research program that (1) posits mental immune systems and (2) proceeds to examine and explain how they work. MIST, we conclude, is a hypothesis that deserves serious scientific development.

1. A Question That Won't Go Away

For six decades, researchers working in a field known as inoculation theory have amassed evidence that it's possible to "inoculate" minds—that is, prime them in ways that build resistance to problematic or unwanted information.[1] Combine this fact with another – that bodies are inoculable *because they have immune systems* – and the question practically asks itself: Do minds, then, have immune systems of their own?

This question first arose in the 1960s. The scientific community, though, has yet to take it up with real seriousness. We hope to remedy that here.

1

Hundreds of studies now testify to the inoculability of minds.[2] In this literature, many experts see evidence that minds *really do* have immune systems. Understandably, others aren't ready to draw that inference.[3]

Suppose we delve deeper, then, and ask how psychological inoculation works. What mechanisms are involved? Are these mechanisms relevantly similar to those that make up the body's immune system?

When William McGuire discovered "attitudinal inoculation" in 1961, he noticed a curious parallel: attitudinal inoculants stimulate the mind to produce *counterarguments*, in much the way that medical inoculants stimulate the production of antibodies. The similarity here may seem superficial. It isn't. To see this, consider four facts about bodily immunity. One: healthy bodies manufacture the biochemical agents we call *antibodies*. Two: they ship these antibodies to sites of actual or possible infection. Three: these antibodies sometimes bind to and neutralize infectious agents. Four: in this way, they confer a degree of immunity to disease.

Now note the parallels. One: a healthy mind will manufacture the cognitive agents we call *counterarguments*. Two: our minds often ship counterarguments to conscious thought, where they can be paired with (possibly infectious) bids to change one's mind. Three: counterarguments sometimes "bind to" (become mentally associated with) such bids, thereby neutralizing them. Fourth, in this way, they confer a degree of immunity to persuasion.

These are the parallels that inspired McGuire to formulate inoculation theory. Some find them striking and regard them as evidence; others remain skeptical.[4] (By the way, *qualms*, *doubts,* and *objections* seem to function like counterarguments: they counteract or modulate conviction. In the ecology of thought, they play similar roles – even when they function beneath the level of conscious awareness.)

2

3

[4] A working definition of evidence is useful here. Our definition can be summed up as follows: any reliable indicator of the truth of something can be thought of as evidence for that something.

Either way, these facts pose intriguing questions. Are counterarguments, objections, doubts, and qualms cognitive antibodies? If it's their job to neutralize informational threats, can they be responsibly thought of as antibodies of the mind? Love or hate the language of mental immunity, to your taste: if your thinking is fact-based, you'll concede that our minds have subsystems that (sometimes) manufacture the conviction-modulating entities we call doubts and objections.

The implicated neural circuitry, of course, has not yet been identified with great precision. It might even be distributed across the brain. But these aren't reasons not to name the suite of subsystems that performs the function. Do so, of course, and we again confront the question: Does it really make sense to think of this assembly as an immune system? Sadly, the community of scientists has not yet rendered a verdict.

Another body of evidence is relevant here. "Identity-protective cognition" is now a well-documented phenomenon.[5] Turns out we have a distinct tendency to employ conscious and unconscious processes to protect our identities from perceived informational threats. If you challenge an ardent believer with evidence, for example, his or her mind will usually manufacture reasons to discount that evidence. If you question the worth of a person's career, you can expect pushback. (Sinclair Lewis wrote that "It's difficult to get a man to understand something when his salary depends on his not understanding it.") Psychologists call such deformations of thought "motivated reasoning." Some call it "confirmation bias." It's clearly related, also, to the phenomenon we call "rationalizing."

Some immunologically informed cognitive scientists suspect that these phenomena can be modeled as aspects of mental immune function. Mightn't knee-jerk dismissals of identity-threatening information be a kind of auto-immune over-reaction? The jury remains out, but the hypothesis deserves its day in court. Its case deserves to be heard, without prejudice. Meanwhile, the fact remains: our minds conduct operations that "protect" our identities, our egos, and our beliefs from perceived informational threats.

5

Again, these operations are carried out by barely understood subsystems that, for the most part, still lack names. So why not name them? Why not model their behavior? At least cognitive immunologists (those that take the immune metaphor seriously) are hazarding a testable conjecture.

Will analogies with the body's immune system help us understand the workings of the mind's informational "defenses"? Here's one reason to think they might: Bodily immune systems evolved to protect us from a host of dangers, among them infectious microbes, manipulative pathogens, and the rogue cells we call cancers. (Yes, bodily immune systems take on enemies both foreign and domestic.) The human *mind* evolved in environments full of analogous hazards: infectious information, manipulative messaging, and the (domestic and often subversive) states we call mistaken beliefs. The evolutionary challenges appear structurally similar, so the mechanisms that natural selection crafted to solve them may also prove structurally similar.

Meanwhile, proliferating scholarly references to "mental immunity," "psychological inoculation," "misinformation inoculation," and the like ensure that the question won't go away. As a matter of fact, our minds are equipped with defenses that evolved to handle a range of informational threats. *The question that remains is this: Are these defenses relevantly similar to those our bodies evolved to handle biological threats?*

Bold scientific conjecture and diligent empirical testing are surely the best way to find out.

2. MISTics and MIASMists

Our collective answer to the question "Do minds have immune systems?" should be evidence-based. The facts should decide. But this too is true: the correct answer hinges on the meaning of words. Define "immune system" one way, and the facts might dictate a verdict of "no"; define it another, and they might require us to answer "yes." A responsible resolution of the question, in other words, must examine the definitions on offer. This entails philosophical idea-testing.

Consider an analogous case: the question "Does our solar system have nine planets?" is indisputably empirical, but the answer hinges on our definition of "planet." In 2006, the International Astronomical Union changed its definition of a planet and the correct answer flipped from "yes" to "no." An entirely sensible preliminary answer to *"Do minds have immune systems?,"* then, would be: *"It depends what you mean by 'immune system'!"*

We've talked with dozens of scholars who claim, with great confidence, that the concept of mental immunity is "mere metaphor." This is not a stance one should take before one has examined the evidence. Nor is it a stance one should take before the phrase "immune system" has been defined. For there are in fact scientifically respectable definitions that require us to answer, "Yes, minds do indeed have immune systems in that sense." We'll show that presently.

To gain clarity about such matters, it helps to examine the question. Cearly, it's a yes/no question. It calls for a verdict. But a verdict on what? Well, a verdict on a hypothesis. Let's give that hypothesis a name. The Mental Immune Systems Thesis, or MIST, is the idea that *Our minds have immune systems of their own, much as our bodies do.* Return an answer of "yes" and you commit to defending MIST.

To answer "no" is to *deny* that minds have immune systems. This denial carries discursive burdens as well. Those who take this stance typically defend it by arguing that *mental immunity-talk is merely metaphorical*. This view also deserves a name. We call it the Mental Immunity As Mere Metaphor Allegation, or MIASMA. As of this writing, the scientific community is divided between MISTics and MIASMists.

3. How to Defend MIASMA Theory

Defenders of the "mere metaphor" consensus have their reasons. It *is* possible to take a metaphor too far. (This is true of all metaphors, even ones we rely on every day.) If mental immune systems don't really exist, references to them are bound to mislead. They might produce faux understanding. They could fuel the scientific equivalent of a unicorn hunt. They could generate bad predictions. Analogous features might dazzle us and cause us to overlook important differences. The concept of mental immunity could be

weaponized. Or the concept of mental immune systems might crowd out a more promising model.

The last of these possibilities merits special attention. For about a century, efforts to understand how the mind winnows information have centered on the concept of *critical thinking*, a deliberate activity meant to catch errors in reasoning.[6] Significant questions have been raised about this approach (after all, a good bit of information-vetting is unconscious),[7] but in the end, the critical thinking paradigm could prove to be the best path forward. And there are other possibilities. The mind's information-winnowing might be more comparable to *filtration*. (The Scottish philosopher David Hume likened idea-winnowing to "*sifting,*" a process that employs a sieve. We're not aware of ongoing efforts to develop a predictive filtration model, but surely, it is among the possibilities worth considering.)

Our best bet might be Daniel Kahnemann's *dual process model*, whereby slow, effortful, "system 2" thinking occasionally modifies the deliverances of fast, intuitive, "system 1" heuristics. (The former is thought to function as a kind of appeals court: one that can review and in some cases overturn the deliverances of the latter.) Meanwhile, Hugo Mercier and Dan Sperber have argued that the concept of *epistemic vigilance* belongs at the center of such efforts, implying that an alert, watchful concern lies at the heart of the mind's discriminatory powers.[8] Any of these models might prove superior to the mental immunity model.

4. An Emerging Paradigm?

So who's right, the MISTics or the MIASMists? A casual examination suggests that MIASMists have the upper hand. After all, MIST does stretch the traditional concept of an immune system. This stretching results in an extraordinary claim, and extraordinary claims require extraordinary evidence.[9] Moreover, it's hard to imagine finding the kind of direct evidence one would need to validate such a claim. If mental immune systems existed,

---

[6]

[7] Norman

[8]

[9] Quote Carl Sagan.

wouldn't their component parts show up on (say) brain-scans or microscope slides? Given that they haven't, how can one responsibly claim that such systems exist?

Reasoning like this makes it easier to go along with the prevailing MIASMist consensus: MIST pushes the analogy too far.[10] Mental immunity-talk has its uses, but shouldn't be taken literally.[11] Or too seriously.[12] The language is suggestive, but not denotative. "Mind's immune system" (the thinking goes) doesn't refer to an actual, you know, *thing*.

For reasons like these, scholarly references to "psychological immune systems" remained rare until about 1998. Then, scholars began using it to explain the contortions our minds undergo to protect our egos and sense of self-worth.[13] Since roughly 2000, though, domain experts have been using "psychological immune system" in a broader way: to understand how our brains spot and neutralize, not just identity-threatening information, but problematic information of all kinds.[14] How do we spot everyday falsehoods, confusions, and misconceptions? How do we ward off manipulative messaging and shed problematic beliefs? Why are we so vulnerable to conspiracy theories, pseudoscience, and fake news?

Hundreds of researchers now study how the mind grapples with such hazards. And some see patterns suggestive of an immune system at work.[15] Extremism, conspiracy thinking, delusion, identity-protective cognition: each of these may be rooted in mental immune dysfunction.[16] Confirmation bias and motivated reasoning are probably a consequence of natural selection favoring a certain conservatism when it comes to working mental models. ("My beliefs got me this far; does it really make sense to discard them at the first sign of trouble?") An easy way to implement this conservatism in an evolving cognitive system is this: use existing beliefs as a reference-point to gauge the threat posed by newly

---

10

11

12

13

14 For example, Ariely.

15

16 https://onlinelibrary.wiley.com/doi/10.1111/j.1740-1461.2007.00097.x

arriving information. That's why surprising information is generally met with skepticism, even when it's correct. (As a result, belief-systems can become self-reinforcing, and diverge from reality.) Confirmation bias and motivated reasoning, in other words, may be direct consequences of default mental immune function.

Two authors of this paper have argued that ideological fixation involves cognitive immune *hijacking:* a bad idea can quite literally hotwire the mind's immune system – and get it to treat potentially destabilizing evidence as a "foreign" threat.[17] In this way, an idea can protect *itself* against replacement. The biological parasites that pull such stunts are legion; mightn't the same be true of *information* parasites?

It's a deeply unsettling thought, and many dislike it intensely. After all, it threatens our sense of cognitive autonomy. But if our world *is* populated by information parasites, and minds *do in fact* have immune systems, these would be important things to know. We'd need to study these entities, the same way we study other natural phenomena. We'd need to examine them, plumb their inner workings, and model how they work. And we'd need to test those models for empirical adequacy.

Especially in a world riven by mis and disinformation. Indeed, understanding the mind's defenses could afford critical protections. We can't mitigate vulnerabilities we don't understand – not reliably, at any rate. Meanwhile, the truth could set us free.

MIST puts other important questions on the front burner. How does immersion in digital environments affect mental immune *function*?[18] Can mental immune systems, like bodily immune systems, become *compromised*? What are the various species of mental immune *disorder*? Is cognitive *autoimmunity* a thing? Can a mind's defenses become *hyperactive*? Can the concept of mental immune *system failure* help explain wholesale and delusional departures from reality?

MIST also poses practical questions. What does mental immune *health* look like? Can we cultivate it? How? Can we build our immunity to extremism, fake news, and

---

[17]

[18] Deep immersion in, say, social media can cause depression, anxiety, and despair, three disorders that themselves

propaganda? How about cult belief and spurious forms of climate denial? Is it really a good idea to give kids 24/7 access to an information environment largely populated by "viral" content? What would it take to develop "herd immunity" to the worst strains of viral nonsense?

In this scenario, an entire discipline – an immunology of the mind – awaits development. (Actually, three disciplines wait in the wings: cognitive immunology, the science of mental immunity; cognitive parasitology, the science of infectious information; and information epidemiology, the science of information spreading.) A potentially significant paradigm shift would be gathering, as we screw up the courage to confront some uncomfortable truths.

The germ theory of disease confronted us with uncomfortable truths, but we rose to the challenge, invented immunology, and re-negotiated our relationship with infectious microbes. We fought back against smallpox and polio – and won. Cognitive parasitology is perhaps even more unsettling. If we can summon the courage, we could fundamentally re-negotiate our relationship with infectious information. We could fight back against climate denial, conspiracy thinking, and propaganda. We could neutralize the digital "influence operations" that currently threaten democracies around the globe. We might even find ways to mitigate confirmation bias and toxic polarization.

MIASMist indecision only postpones a much-needed reckoning. It's time we took a serious look at a long-neglected, and potentially pivotal, ontological question.

5. On What It Means to Claim that Something Is Real

A rigorous treatment of the question must also examine a common misconception. Many assume that something must take "concrete" form to count as real. The components of bodily immune systems – T-cells and B-cells, phagocytes and lymphocytes – are concrete hunks of stuff, and we can see them through microscopes. This gives us very direct evidence that bodily immune systems are real. If mental immune systems were real, wouldn't microscopes or brain scans reveal comparably concrete components?

Not necessarily. Think of beliefs, desires, ideas, and scientific theories. None of these show up on brain-scans, but we consistently treat them as real. Setting aside first-person introspection, we never observe such entities directly. Suitably scientific, third-person observation turns up only the behavioral, symbolic, or neural correlates of such things. The entities themselves are inscrutable to third-person observation.

Take beliefs, for example. Our conviction that they exist is partially introspective, but otherwise rooted in indirect third-person evidence. When a person behaves *as if* they believe that X, we often treat this as sufficient evidence that they *do* believe that X. We observe the behavior of concrete things and infer the existence of one or more "abstract," behind-the-scenes things. There is nothing inherently problematic about such inferences; responsible scientists make them all the time.

Every day, we treat abstract entities as extant. Again, take beliefs. The concept of belief has a long track record of helping us understand human behavior. It pays its way. It probably wouldn't if it didn't refer to *something* significant – a salient but as-yet-undefined pattern of brain activity, say – so we *posit* the entities the term refers to and admit beliefs into our ontologies.

In this way, we "reify" countless abstract things: numbers and norms, rules and conventions, rights and values, personalities and character traits. Personality traits, for example, leave their traces on behavior. They also leave traces on psychological survey instruments. We infer their existence, treat them as real, and are not wrong to do so.

What if mental immune systems belong to a similarly abstract class of entities? Admittedly, the concept has had little time to compile an explanatory track record. But what if we gave it a chance? After all, each of us is susceptible to various kinds of misinformation, and tools exist for discerning a person's position on the spectrum between wildly susceptible and fully immune.[19] If people behave *as if* their minds have immune systems, couldn't that constitute evidence that such systems exist?

6. Linguistic Evolution: A Potted History of a Key Concept

19

The phrase "immune system" was initially defined by its association with *antibodies*, chemical structures thought to bond with pathogenic *antigens*, thereby neutralizing them.[20] Empirical study soon revealed a more complicated picture. Immunologists identified an *innate* immune system, which is found in nearly all forms of life, and an *adaptive* immune system, which is found only in jawed vertebrates. Omnivores have immune mechanisms that herbivores lack. We now know that immune response can be mediated by phagocytes and lymphocytes, T-cells and B-cells, enzymes and antimicrobial peptides. The elements of bodily immune systems now include chemokines, interferons, interleukins, lymphokines, and tumor necrosis factors. These elements form a network that can function in astonishingly adaptive ways.

What unites all these elements? What justifies our classifying them as elements of one and the same thing? Clearly, it's their *functional role:* each of them plays a part in the body's ceaseless effort to ward off disease. (Ceaseless, of course, until it ceases.) Notice that the elements of a body's immune system needn't be co-located. They are distributed throughout the body, interspersed into dozens of bodily tissues. Yet still they constitute an entity worth naming. (Note the consequence: a thing needn't be a contiguous chunk of matter to be a thing.)[21]

Immune systems are by no means unusual in this regard. Science countenances thousands of things that aren't contiguous chunks of matter. Epidemiologists know that pandemics are real. Psychologists speak of beliefs and desires as real, despite not knowing how they are coded in the brain. Economists treat demand as real (after all, it has causal powers), and political scientists treat nations as real. Anthropologists accept norms and practices as real, and physicists treat force fields as real. The idea that only material things are real is at best a simplistic conceit: it's time we retired it.

For decades, our concept of an immune system has been evolving *away from* its initial association with a particular physical mechanism (antigen-neutralizing antibodies)

[20] Paul Ehrlich's work on humoral immunology and Elie Metchnikoff's work on cellular immunology arguably anchored our initial concept of an immune system.

[21] Immune systems are by no means unusual in this regard. Science countenances thousands of things that aren't contiguous chunks of matter: pandemics, beliefs, economies, nations, currencies, norms, practices, force fields….

and *toward* a broader, more functional definition. "Immune system" now means something like "a complex network of processes that functions to defend an organism against the threats posed by a broad class of infectious agents."

7. On Who Gets to Define "Immune System"

We fix the meaning of expressions in different ways. Sometimes, we take prevailing usage at face value and simply use the expression "as one does." That is, we take normal usage as normative. On this approach, "Minds have immune systems" only becomes true when enough people treat it *as* correct usage. Until then, we must bow to prevailing norms and treat such talk as false—or as "merely metaphorical." Is this the best way to fix the meaning of "immune system"? Probably not. Slavish obedience to *de facto* usage norms is not a recipe for scientific progress.

Another option: we can take the expression and simply stipulate that, for our purposes, it's going to mean X. (For example, one might say: "For argument's sake, let's define an 'immune system' as *anything that fights infections*.") We can do this, then expect subsequent usage to conform to the stipulated definition. This can be a useful way to reduce ambiguity and help structure a conversation. And generally speaking, each of us is entitled to propose novel definitions. None of us, though, is authorized to dictate for all what "immune system" shall mean. (The italicized definition above, of course, is vulnerable to the objection that it is too broad; after all, it would classify a flu vaccine as an immune system – a clear error.)

A third option is to find an authoritative source – the Oxford English Dictionary, say – look up how it defines the expression, and treat *that* as the final word. The trouble with this is that lexicographers – the people who study words and compile dictionaries – may not be aware that the expression is being used in novel ways that advance the cause of science. Dictionaries are never the final word on what an expression *should* mean.

Imagine examining all the empirical evidence, looking up Wikipedia's definition of "immune system," and allowing *that* to decide the question. (As I write this, Wikipedia defines an immune system as "a network of biological processes that protects an organism

from diseases." Note that this definition rules out the possibility of minds having immune systems, and does so in advance: "Minds aren't organisms, so they *can't* have immune systems. QED.") Would this be sound scientific practice? Of course not. For it might be time to revise Wikipedia's definition. An alternative conception might have superior disclosive power. Science itself – the project of understanding our world – might need Wikipedia to upgrade its working definition.[22]

8. An Evolutionary Argument for MIST

[Mental immune systems, it seems, had to evolve. After all, minds evolved, and evolved entities must solve a complex problem: how do we protect ourselves from disruptive forces? If infectious agents attempt to parasitize us, how will we fight them off?]

9. MIST's Testable Implications

10. The Evidence to Date

11. Cognitive Immunology and its Prospects

[22] This is in fact the current situation vis-a-vis the concept of cognitive immunology. Wikipedia refuses to allow an article devoted to it on the grounds that champions of the science have a "conflict of interest." Ironically, Wikipedia's page-vetting process has itself been likened to an immune system: an evolved complex of processes meant to protect the integrity of a body of knowledge. increasingly defined by Wikipedia's authority as an arbiter.