Journal of Theoretical and Philosophical Psychology

# Do Minds Have Immune Systems?

Andy Norman, Luke Johnson, and Sander van der Linden

Online First Publication, December 12, 2024. https://dx.doi.org/10.1037/teo0000297

CITATION
Norman, A., Johnson, L., & van der Linden, S. (2024). Do minds have immune systems? *Journal of Theoretical and Philosophical Psychology*. Advance online publication. https://dx.doi.org/10.1037/teo0000297

AMERICAN PSYCHOLOGICAL ASSOCIATION

THEORETICAL & PHILOSOPHICAL PSYCHOLOGY

Journal of Theoretical and Philosophical Psychology

© 2024 The Author(s)
ISSN: 1068-8471

https://doi.org/10.1037/teo0000297

# Do Minds Have Immune Systems?

Andy Norman[1, 2], Luke Johnson[1], and Sander van der Linden[3]

[1] Cognitive Immunology Research Collaborative, Pittsburgh, Pennsylvania, United States

[2] Department of Philosophy, Carnegie Mellon University

[3] Department of Psychology, Cambridge University

Do minds have immune systems? In this article, we remove several obstacles to treating the question in a rigorously scientific way. After giving the hypothesis that minds *do* have such subsystems a name—we call it mental immune systems theory—we show why it merits serious consideration. The issue hinges on our definition of an immune system, so we examine the definition that currently prevails, demonstrate its shortcomings, and offer an alternative that addresses those shortcomings. We then lay out the empirical evidence that minds really do have immune systems in the specified sense. Findings about psychological inoculation, identity-protective cognition, cognitive dissonance, psychological reactance, information diffusion, and cognitive bias all point to the existence of evolved cognitive defenses—informational "immune systems" that function in much the way that bodily immune systems do. Finally, we discuss the prospects of *cognitive immunology*, a research program that (a) posits mental immune systems and (b) proceeds to investigate their functioning.

***Public Significance Statement***

In this article, we show that minds have immune systems of their own: evolved informational defenses that function to ward off disruptive information. The study of these systems—cognitive immunology—promises a deeper understanding of how to cultivate resistance to mis- and disinformation.

*Keywords:* misinformation, susceptibility, inoculation, immune system, cognitive bias

James Cresswell served as action editor.

Andy Norman https://orcid.org/0000-0002-6873-3150

This work was supported by Tyler Stanek, Henry Thorne, and the Bernard and Audre Rapoport Foundation. The authors thank Prosocial World and the Evolutionary Philosophy Circle for hosting the symposium that allowed the authors to road test some of the arguments presented here.

This work is licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International License (CC BY-NC-ND 4.0; https://creativecommons.org/licenses/by-nc-nd/4.0). This license permits copying and redistributing the work in any medium or format for noncommercial use provided the original authors and source are credited and a link to the license is included in attribution. No derivative works are permitted under this license.

Andy Norman played a lead role in conceptualization, investigation, methodology, supervision, writing–original draft, and writing–review and editing. Luke Johnson played a supporting role in conceptualization and writing–review and editing. Sander van der Linden played a lead role in empirical validation and a supporting role in conceptualization and writing–review and editing.

Correspondence concerning this article should be addressed to Andy Norman, Cognitive Immunology Research Collaborative, 7440 Ben Hur Street, Pittsburgh, PA 15208, United States. Email: anorman@cmu.edu

*Do minds have immune systems?* Many find the question puzzling, for talk of mental immune systems remains rare. Among those familiar with the latest research on misinformation susceptibility, though, the question is now taken with real seriousness.[1] Why? Because close investigation reveals something surprising: Our minds acquire resistance to infectious and disruptive information in much the way that animal bodies acquire resistance to infectious and disruptive microbes (Compton et al., 2021; Norman, 2021).[2] The parallels are striking; we think they portend a significant shift in our understanding. For if our minds do in fact have antibody-generating defenses, we need to reexamine some long-standing assumptions—about the best way to protect citizens from manipulative influence campaigns, say, and the best way to educate children.[3]

We have long known that our *bodies* sport sophisticated defenses—evolved subsystems that protect against disease. These systems manufacture biochemical agents—"antibodies"—that work to neutralize infectious pathogens. The network of implicated mechanisms confers a degree of immunity to microbial threats; we refer to them collectively as the "immune system." Immunologists have studied this system for more than a century. Their findings have revolutionized health care (Silverstein, 2009).

Now, we are learning that our *minds* sport comparable subsystems. These subsystems manufacture what might be called "cognitive antibodies" (Norman, 2021; van der Linden, 2024). They do so to neutralize infectious *information*, and often they confer a degree of *mental* immunity (van der Linden, 2022). Nor are the parallels accidental: facing analogous challenges, natural selection seems to have fashioned analogous solutions for both our bodies and our minds (Cacioppo et al., 1999).[4] Moreover, in both cases, you can induce protective resistance via *inoculation*—a process that enhances the subsystem's capacity to recognize and combat parasitic, harmful, and self-replicating agents (Norman, 2021; Traberg et al., 2022).

In this article, we show why the question in our title needs scientific evaluation. We summarize the relevant empirical findings and develop the arguments that knit them together. Several domains of inquiry have unearthed hard evidence of mental immune systems; of these, *psychological inoculation research* is particularly suggestive. This research dovetails with findings about *identity-protective cognition* (IPC), *cognitive dissonance*, *psychological reactance*, *confirmation bias*, *motivated reasoning*, and *belief perseverance*. Meanwhile, *information epidemiology* and *evolutionary biology* provide independent grounds for thinking that informational immune systems are an evolved feature of our minds.

Psychology has a long history of borrowing from other sciences to enhance our understanding of human nature.[5] For example, David Hume's vision for a science of the mind drew concepts and inspiration from Newtonian physics, and Freud modeled psychodynamics on thermodynamics (Hume, 2007; Sabelli et al., 1990). In this case, we show that concepts borrowed from immunology prove remarkably fruitful: Applying them to the mind proves explanatory, unifying, insight-generating, and predictive. The theory that minds do in fact have immune systems, for example, can help explain our astonishing capacity to deny evidence. It helps to explain how belief systems diverge from reality and how worldviews harden into evidence-resistant ideologies. Later, we will also spell out some testable implications of this theory.

We mean to paint a vivid picture of an overlooked natural system. To aid comprehension, we sketch that picture here: We are a communicative species; we form social networks through which information with certain properties propagates. We are individually and collectively prone to misinformation "infection," but thankfully, we are not defenseless. We are equipped with

[1] In 2022, some authors of this article served on a blue-ribbon panel tasked with studying our world's misinformation problem. That panel issued a public declaration highlighting the role cognitive immunology is likely to play in addressing the problem: https://mentalimmunityproject.org/declaration. Over 100 scientists and domain experts have since signed it. The present article provides a more rigorous treatment of this declaration's central contention.

[2] Norman (2021) treated the claim's conceptual foundations; Compton et al. (2021) treated its empirical foundations.

[3] If our minds acquire immunity to misinformation primarily via exposure to it, we need to expose children to a rich mix of reliable and unreliable information—and help them learn how to evaluate and sort it. Classrooms that feed them only vetted, board-certified "content" probably are not doing enough to empower them as critical thinkers. To build misinformation immunity, kids need to play with sketchy ideas and practice the art of testing them.

[4] The general point here is that, when it comes to evolved forms, form follows function.

[5] For a more recent example, see https://www.sciencedirect.com/science/article/abs/pii/S0303264723002459.

evolved cognitive defenses that can and do generate protective antibodies. Sometimes, though, these defenses are more concerned to protect our beliefs than to arrive at the truth. As we will see, signature tendencies of the human mind—confirmation bias and IPC among them—may prove to be artifacts of mental immune function.

These "mental immune systems" can both protect and skew our thinking. They can blind us to evidence. Sometimes, they get confused and attack the very minds they are supposed to defend. Minds, in other words, are subject to *immune deficits, immune disorders,* even *autoimmune disorders* (Norman, 2021). Fortunately, they are also *inoculable*: We can use informational "inoculants" to train them. In this way, we can make them more truth-oriented, more resistant to misinformation, more autonomous, and more resilient (van der Linden, 2023). Assemble the clues, view them through the right lens, and the conclusion practically leaps out: Minds *really do* have immune systems.

We will summarize the relevant empirical evidence presently. The question we are posing, though, pivots on the implicated *philosophical* questions—questions like What is an "immune system," anyway? Who gets to define the phrase? and Should talk of mental immunity be taken literally?

Talk of mental immune systems seems metaphorical. In important respects, it *is* metaphorical.[6] If the concept continues to prove explanatory, though, the phrase will undergo a process of literalization. As we will show, the language of immunology brings important features of human psychology into focus: Train it on puzzling features of our minds, and you see new things. Factor in overdue adjustments to our concept of an immune system, and once-metaphorical claims about cognitive immune systems morph into literal truths.

If we are right, the clues have long been staring us in the face; somehow, though, we failed to connect the dots and give the resulting pattern a proper name. We neglected to make a pivotal natural system a named object of scientific study, and this neglect had consequences—among them an underdeveloped grasp of how best to counter mis- and disinformation.

The emerging science of mental immunity aims to address these omissions. It builds on a substantial body of evidence that minds do in fact have evolved defenses—mechanisms that protect against information disruption (Wardle & Derakhshan, 2017). It deems these defenses worthy of study, gathers the evidence, and finds them to be very like bodily immune systems in the way they function. (Both systems function to limit the harmful effects of disruptive entities—entities that might otherwise inject disorder, turmoil, or dysfunction.) The science opts to refer to these systems as "mental immune systems" or "information immune systems" and proceeds from there.

We call this fledgling science "cognitive immunology." It increasingly looks like a fruitful approach to addressing our world's mis- and disinformation problems (Cognitive Immunology Research Collaborative, 2023). The concepts of *mental immune function* and *mental immune health*, for example, help to explain puzzling variations in misinformation susceptibility. And advances in the field are helping millions develop their resistance to misleading and manipulative kinds of information (Bond, 2022; Roozenbeek et al., 2022; Yee, 2022).

It is important not to conflate *disruptive* information and *mis*information. They are both problematic, but information's tendency to disrupt is one thing; its tendency to mislead is another.[7] A mechanism honed by natural selection to combat the one will, of course, be imperfectly tuned to combat the other. As we will see, the mind's natural tendency to resist disruption may well provide foundational explanations of many of our biases and blind spots. The mind probably evolved to make use of adaptive fictions, and some fictions are hard to dislodge from the mind precisely because they help to stabilize the psyche against the disruptive effects of (some) truths. For example, belief in an afterlife may well protect against the

[6] The idea that minds have immune systems extends the concept of an immune system into a new domain. In this sense, it is metaphorical. Admitting this, though, does not entail that "mental immune system" lacks a referent. For language can be both metaphorical and referential. After all, "the Bible belt" refers to an actual geographical area. Metaphors can direct our attention to real entities.

[7] Clarity here can be enhanced by envisioning a two-by-two matrix, with rows labeled "true" and "false" and columns labeled "disruptive" and "non-disruptive." We think that information falls into all four quadrants of this matrix. Similar matrices can be created using other concepts: "maladaptive" and "adaptive," "misleading" versus "non-misleading," "helpful-to-me" versus "helpful-to-us," "short-term-beneficial" versus "long-term-beneficial," and so forth. In the article, we try to use the correct adjective to express our meaning in each context.

anxieties that can attend an honest acceptance of one's mortality.[8]

Adaptive fictions and disruptive truths underscore the fact that information has both causal *and* epistemic properties. A given message's causal properties can pull us in one direction, even as its epistemic properties pull us in another. It follows that information immune systems can be "tuned" to, as it were, different "frequencies"—to "care about" different things. Our evolved cognitive defenses were presumably tuned by natural selection to care, first and foremost, about information's utility, but for many modern purposes, it is important to tune them to care instead (or also) about truth and understanding.[9] Cognitive immunologists will need to bear these distinctions in mind, especially when attempting to apply the science to our world's misinformation problem.

Before proceeding, we need to acknowledge the emotional elephant in the room: Talk of mind infections is deeply unsettling. It is hard to reconcile with comforting beliefs about our cognitive autonomy. But problematic ideas really do spread virally and flourish at human expense (Bak-Coleman et al., 2022; Boudry & Hofhuis, 2018; Kauk et al., 2021). Extremist ideologies really do exploit fears of insignificance to hijack people's minds (Hankir & Hassan, 2019; Kruglanski et al., 2014).[10] Ideologies may lack intention; they may lack the kind of agency we have, but they have tendencies linked to their causal and logical properties. As such, they can *do* things to us, such as trigger, provoke, confuse, incite, and delude (Distin, 2005). Even mainstream ideas can compromise our autonomy and judgment. Another corollary—that we may need to inoculate against certain kinds of mis- and disinformation—strikes many as dangerously Orwellian. It is no surprise, then, that cognitive immunology has critics (Häusler & Baraghith, 2023; Simon & Camargo, 2023).

Evidence of mental immune systems first surfaced more than 6 decades ago in the research of psychologist McGuire (1961). It could be variously interpreted, though, and scholars erred on the side of caution. For half a century, they steered clear of the bold claim that would have given cognitive immunology a focal point. They avoided positing mental immune systems and, consequently, could not ask, *How do they work?* They were not in a position to ask, *Why do these systems so frequently malfunction?* And for the most part, they failed to ask, *How can we make them work better?*[11] As a result, a promising field of inquiry lay mostly fallow.

Such questions can no longer be postponed. Since smartphones became ubiquitous, internet-borne "infodemics" have become a serious threat to public health (Grimes, 2020). Digital influence operations have proliferated, trust in public institutions has plummeted, and mental illness has spiked (Braghieri et al., 2022; Newton & Norris, 2000). Outbreaks of genocidal fervor have spread through social media, and democracies across the globe are struggling to prevent authoritarian uprisings (McKay & Tenove, 2020; Mozur, 2018; Slaughter, 2022). All these problems are rooted in our susceptibility to problematic information, and all are exacerbated by digital connectivity.[12] It is high time we developed a deep understanding of susceptibility and immunity to problematic information.

Two centuries ago, the science of *bodily* immunity began to alter our relationship with troublesome microbes. It gave us vaccines: powerful tools that help us beat back immiserating diseases. Now, the science of *mental* immunity promises to rebalance our relationship with viral misinformation—and prevent some of the suffering that infodemics undoubtedly cause (Rodgers & Massac, 2020). Before we can do that, though, we

[8] While we cannot fully treat the topic in this article, we believe that false beliefs can be short-term adaptive but are almost always maladaptive in the long run. We recognize that false beliefs can serve special interests at the expense of the common good. We suspect, though, that the long-term collective interest is best served by true beliefs.

[9] For any belief we have, it is possible to ask, "Do I believe this because I find the idea useful, or do I believe it because I have reason to think it true?" We think questions like this can train a mind to screen ideas for truth as well as usefulness. We call such questions mental immune boosters and imagine them figuring prominently in future efforts to strengthen mental immune systems.

[10] As we like to put the broader point: Hope and despair can impact mental immune health.

[11] Inoculation theorists, of course, have been asking closely related questions for some time, questions like "Can we prompt minds in ways that make them less susceptible to misinformation?" The strong ontological commitment to mental immune systems, though, can change the character of such inquiries. MIST, we think, adds depth, significance, and perspective to psychological inoculation research.

[12] Heightened digital connectivity makes infodemics more likely, just as the heightened proximity of urban living made epidemics more likely in the 19th century. Digital connectivity reduces "social distance," exacerbating misinformation susceptibility.

need to address a simple, foundational question: *Do minds have immune systems of their own?*

## The Question

Before examining the evidence, it pays to clarify the question. Note that it invites an answer of "yes" or "no." To answer "yes" is to assert that mental immune systems do, in fact, exist. This stance merits a name. We call it mental immune systems theory, or MIST. "MISTics" think that minds really do have information immune systems.

Those drawn to the opposite answer—"no"—deny this. Suspecting that the phrase "mind's immune system" does not refer to an actual *thing*, they argue that mental immunity talk is not literally true. Sometimes, they insist that mental immunity talk is "merely metaphorical." This stance, too, merits a name. We call it the Mental Immunity as Mere Metaphor Allegation, or MIASMA. Defenders of this stance can be called "MIASMists."

As of this writing, the scientific community is divided between MISTics and MIASMists.[13] Presumably, MIASMists constitute the majority. Yet there are reasons for likening today's MIASMists to the 18th-century thinkers who resisted the germ theory of disease. (The original miasmists insisted that disease was caused, not by germs, but by noxious "miasmas"; Kannadan, 2018. They were wrong.) Today's MIASMists—those who reject MIST—also tend to reject what we might call the *germ theory of cognitive contagion* (GTCC): the idea that misinformation and objectively problematic ideas are essentially transmissible mind parasites (Norman & McIntyre, 2023; Williams, 2023).[14]

The GTCC and MIST are aligned: They both employ a largely Darwinian frame to better understand information dynamics. But neither entails the other, and it is not obvious that they stand or fall together. In this article, we focus on evaluating MIST; GTCC deserves separate treatment.[15]

Clearly, 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."

An example serves to illustrate. The question "Does our solar system have nine planets?" is indisputably empirical; it is the kind of question that ought to be settled by facts. But the answer also hinges on our definition of "planet." In 2006, the International Astronomical Union changed its definition of a planet; Pluto was demoted from "planet" to "dwarf planet," and the number of planets in our solar system dropped from nine to eight. With this—a mere definitional change—the correct answer flipped from "yes" to "no." An entirely sensible preliminary answer to the question, "Do minds have immune systems?," then, would be "It depends what you mean by 'immune system.'"

No one should render a verdict on MIST until the meaning of "immune system" has been clarified. In fact, we should all be willing to adjust our answer depending on the definition that happens to be operative. A sensible interlocutor should be able to say things like, "I grant that, given definition X, MIST is merely metaphorical; given definition Y, though, it appears both literal and well-supported by the evidence." A full resolution of the scientific question, then, requires us to ask and answer a prior *philosophical* question, specifically: "How should we define 'immune system'?" Should we define it narrowly, in a way that excludes the possibility of mental immune systems? Or should we embrace a broader definition, one that makes the existence claim true?

The concept of mental immune systems represents a departure from common usage. Consequently, proponents of MIST are obliged to explain and justify that departure. More precisely, they take on a three-part burden. First, they need to explain what they mean by "immune system." Second, they should show why it makes sense to adopt that definition. Last, they should present

[13] A weeklong virtual symposium on the subject made this abundantly evident to all who attended. We thank Prosocial World and the Evolutionary Philosophy Circle for hosting the symposium, which became the basis of a print symposium: https://www.prosocial.world/posts/the-science-of-mental-immunity-has-arrived.

[14] For example, some ideas have properties that incite hatred or contribute to the dehumanization of outgroups. Properties like these are objectively problematic regardless of the idea's truth value. The phrase "objectively problematic" is meant to set this category of ideas apart from those that are merely "subjectively disliked."

[15] The academic literature on memes represents a down payment on this treatment. See Blackmore et al., 2000; Häusler & Baraghith, 2023.

solid empirical evidence that minds *do* have immune systems in the specified sense.

Until MISTics provide such things, it is fair to say that MIASMists have the upper hand. There is more to say, though, on all three counts.

## What Is an "Immune System"?

The British Society for Immunology defines an immune system as a "complex system of structures and processes that … evolved to protect us from disease" (British Society for Immunology, n.d.). This definition has merit. The phrase was coined in or around 1909; ever since, it has been a kind of umbrella concept: a catch-all for biochemical mechanisms that function to protect organisms from a range of threats. Not *all* threats, mind you: Immune systems do not protect against, say, large predators. So, what kinds of threats *do* immune systems handle? The British Society's definition emphasizes "foreign invaders—bacteria, viruses, parasites, and fungi." Other authoritative sources include cancers and invasive objects such as wood splinters (Cleveland Clinic, 2023a).[16]

Cancer cells, of course, are rogue *domestic* agents—microscopic threats of *internal* origin. Splinters and viruses, meanwhile, are tiny but harmful *foreign* entities—dangers of *external* origin. The implication is that immune systems protect against very small *internal* and *internalizable* disruptors.

If you ask an immunologist how immune systems work, they are apt to mention antibodies. This is shorthand. Immune systems dispatch T cells and B cells, macrophages and lymphocytes, cytokines, and chemokines. Why do immunologists place all these things (and dozens more) into the "immune mechanisms" category? Because they function to "protect us from disease." Protective function, then, is the glue that holds the concept together.

Prominent definitions, of course, are not the final word on the subject. For example, Wikipedia seems to imply that, by definition, immune systems employ exclusively *biochemical* means: "[Immune systems] use *molecules* and *cells* to perform their functions."[17] New research, though, shows that organisms also employ *psychological* and *behavioral* mechanisms to defend against microbial threats (Schaller & Park, 2011; Shattuck, 2021).

For example, when we react to rotting meat with disgust, we tend not to ingest the meat. This helps to keep certain disease-bearing pathogens outside of our bodies. That, biologists tell us, is ultimately why we find rotting meat disgusting: the tendency to feel revulsion in its presence helped to keep certain pathogens at bay. Presumably, this helped our ancestors survive. The trait is thought to have advantaged them relative to contemporaries who lacked that disgust reflex. (By contrast, vultures have microbiomes that protect them from the microbes in rotting meat; as a result, they did not evolve a disgust reflex like ours.)

For similar reasons, arachnophobia helps to keep spider venom at bay. The mechanisms in question here (disgust and fear of spiders) are as much psychological as biochemical, but they serve a protective function, so biologists include them under the "immune system" umbrella. In fact, biologists now refer to this suite of mechanisms as the "behavioral immune system." The point, of course, is that immunologists have time and again stretched our concept of an immune system to incorporate new insights. The concept has proven elastic. Good thing, too: Its elasticity has helped science progress.

Should we continue to evolve our concept of an immune system? Of course. Should blind deference to a prevailing definition preclude such evolution? Of course not. Dictionary definitions do not decide such questions, for the people who compile them can overlook novel but salutary uses. Nor does de facto usage automatically decide such questions, for prevailing usage can lag behind best linguistic and scientific practices. The real question for working scientists is whether the concept of an immune system, appropriately defined, fits the facts. So defined, does it explain things? Does it help confer understanding?

The concept of an immune system that currently prevails is easily generalized. In fact, it practically begs to be generalized. For there is no good reason to focus exclusively on *biological* threats—not

[16] Wikipedia's cultural influence makes its definition also worth considering in this context: https://en.wikipedia.org/wiki/Immune_system.

[17] This was Wikipedia's definition as of August 15, 2023: https://en.wikipedia.org/wiki/Immune_system. (Note that we are not citing this definition *as* authoritative; the point instead is to provide an example of an influential definition with imperfections.) We acknowledge that Wikipedia's definition was probably not meant to be exclusive of other-than-biochemical means; the point is that, since definitions are commonly taken to be specifications of sufficient *and necessary* conditions, the definition could easily be interpreted as excluding systems that employ, say, behavioral or mental means.

when organisms also ingest infectious and potentially disruptive *information*. Note that our immune systems combat *physical* splinters as well as *biological* microbes: This shows that nature does not care whether an internalized threat is biological; if it impacts fitness, natural selection will start devising protective solutions, even if that means crossing arbitrary definitional boundaries. Put differently, the boundary between biological and informational threats is proving to be every bit as arbitrary as the boundary between biological and physical threats.

So, the question must be asked: Is an evolved capacity to spot and shed problematic information also part of our genetic inheritance? Three elementary considerations make the answer all but inescapable: (a) Minds are evolved structures; (b) to persist, evolved structures must protect themselves from disruptive forces; and (c) information can be a disruptive force.

Nor is there any good reason to focus exclusively on *biochemical* defense mechanisms, not when *psychological* and *behavioral* mechanisms are also known to serve a protective function. Nor should we assume that such mechanisms function only to protect our *bodies*—not if the facts suggest that they also protect our *minds*. As we will show in Exhibits E, F, G, and H, natural selection *has* devised mechanisms that protect our minds from informational threats. We think our concept of an immune system should be stretched to encompass such insights.

The fact that these mechanisms afford imperfect protection, of course, is neither here nor there. Bodily immune systems afford imperfect protection, and that does not prevent us from positing them. Whether the system exists is one question; how well it works to achieve a given outcome (such as filtering out falsehoods) is a separate question. Separating such questions allowed immunology to advance; untangling the corresponding questions should allow cognitive immunology to advance.[18]

Incidentally, the concept of an immune system is now being used to better understand social and political systems. For example, police forces and legal systems function to suppress internal forces—crime, violence, fraud—that tend to unravel societies. Are they evolved features, then, of a society's immune system? Some scholars take the claim quite seriously (Wilson, 2022). They point out that societies that lack such mechanisms are essentially immune-deficient: unable to counteract disruptive internal threats. Such "immune deficient" societies, of course, are less likely to persist. (Not coincidentally, overzealous and unjust policing is also maladaptive—not to mention comparable to autoimmunity.) You need not agree with all of this, of course, to grasp that we should be encouraging enterprising thinkers to explore such connections. We certainly should not reject such work out of blind fealty to inherited language norms.

So, who gets to define "immune system"? We do. The task falls to the community of thinkers concerned to better understand our world. How do we decide? Well, we examine competing definitions and compare what they allow us to do; we examine the evidence and see which definition makes the most sense.

Of course, the authors of this article represent just a tiny sliver of the community that will ultimately decide the question. That is why we put forward the following definition and invite the consideration and commentary of our colleagues:

> An immune system is a suite of traits that functions to protect an evolved entity from internal and internalizable threats, especially parasitic, infectious, and cancerous agents.

Addendum: immune systems typically perform this function by manufacturing and mobilizing "antibodies"—entities that combat and neutralize "in-house" threats.[19]

Note that, unlike some definitions, this one is substrate-neutral. We have generalized it in a way that allows it to be applied across multiple domains. It can be applied to groups as well as individuals (yielding talk of social and cultural immune systems), macroscopic behaviors as well as microscopic biochemistry (justifying talk of behavioral immune systems), and it can be applied to minds as well as bodies (legitimizing talk of mental immune systems).

We think this definition represents an upgrade over traditional, domain-specific definitions. It is

[18] Clearly, the science of immunology had to get past the objection, "If our bodies have immune systems, then why do we get sick?" To do so, it had to separate these questions. We are doing the same here for the corresponding cognitive science.

[19] Here, we opt not to include this addendum within the definition proper. In this way, we make the manufacture of antibodies an inessential quality—a feature of many, but not necessarily all, immune systems. A case can be made though that this provision should be folded into the definition. We postpone this issue for another day.

more disclosive. It highlights important connections. It aids understanding. It is more encompassing. And as we will see, it fosters explanatory coherence.

Importantly, this definition applies across the multiple "levels" of so-called multilevel selection (Wilson & Sober, 1994). Multilevel selection theory holds that the evolutionary algorithm works simultaneously across multiple levels: that it shapes lower level entities like cells and genes while at the same time shaping higher level entities like groups and religions and nations. If this is right, we are likely to find immunity-inducing mechanisms in unexpected places.

Incidentally, our definition can also be applied to "memeplexes"—culturally evolved bundles of information that do clever things to protect themselves from informational disruption. For example, ideologies are prone to turn people into ideologues: That is, they reshape human minds in ways that render them closed to alternatives. They can induce people to rationalize and dismiss evidence (Budiman et al., 2013). Some ideologies have induced their hosts to establish orthodoxies and intimidate heretics. When an ideology does such things, it makes sense to think of it as parasitizing and hijacking minds.

We suspect that ideas can commandeer a mind's immune system and repurpose it: get it to serve, not the individual's interests but the idea's "interests" (Mauer, 2022; Norman, 2021). The rationalizations that protect an ideology can even be thought of as components of the *ideology's* immune system.

Cancerous tumors perform a neat trick: They emit proteins that effectively blind the body's immune system to their presence (Steer et al., 2010). This feature—essentially a cloaking device—allows them to "fly under the radar" and exploit normal, healthy cells. They do this despite lacking intention or any strong kind of agency. We find the same phenomenon in the world of information: self-replicating ideologies blinding the mind's (rational) immune system to better serve their replicative "interests" (Dennett, 2007).

We will turn soon enough to the evidence that minds *do* have immune systems in the sense captured by our definition. First, though, we need to dispel a pervasive misconception about cognitive entities and the kind of evidence needed to establish their existence.

## What Evidence of Abstract Entities Looks Like

Many assume that an entity must take "concrete" form to count as real. How do we know that our bodies have immune systems? Well, their component parts—T cells and B cells, phagocytes and lymphocytes—are tangible hunks of stuff that we can see through microscopes. This gives us very direct evidence that bodily immune systems are real. Before we declare mental immune systems to be real, should we not amass comparably direct evidence of comparably concrete components?

Not necessarily. Think of beliefs and desires, hopes and fears, ideas and theories. None of these entities can be viewed through a microscope. None show up on brain scans. (Correlated brain *activity* shows up on brain scans, but not the entities per se.) Yet still, sober scientists treat these things as real (albeit abstract) entities. You can turn your attention inward and notice that you desire something, but does that give you scientific evidence that desires exist? Scientists tend to discount such "first-person" introspection, often for good reason: Subjective reports of subjective events can be unreliable. Instead, they favor third-person observation. And there is the rub: Third-person observation never turns up direct evidence of cognitive entities. At best, it turns up their behavioral, symbolic, or neural correlates.[20] The entities themselves prove inscrutable to third-person observation.

To get a firm grip on the point, focus on beliefs. Our conviction that they exist is underwritten, in part, by introspection; mostly, though, it rests on indirect third-person evidence. When a person behaves as if they believe X—consistently acting as if X is true, say—we often treat this as sufficient evidence that they *do* believe X. We observe the behavior of concrete things (people) and infer the existence of one or more abstract, behind-the-scenes things (beliefs). There is nothing inherently problematic about such inferences; responsible agents—and responsible scientists—make them all the time.

[20] It is of course possible to present people with falsehoods and ask them how many counterarguments they can generate. The resulting count can be a meaningful measure of mental immune function. Cognitive immunologists are now measuring psychophysiological responses to misinformation. See, for example, Clayton et al., 2023, and Gordon et al., 2019.

Every day, we treat abstract entities as extant. The concept of hunger, for example, has a long track record of helping us understand human and animal behavior. It pays its way. Presumably, it refers to *something* significant—a salient but as-yet-undefined pattern of brain activity that correlates with food-seeking behavior, say—so we *posit* hunger: We admit it into our ontologies and treat it as real. Hunger really happens. It exists.

In this way, we "reify" countless abstract things: norms and rules, rights and values, beliefs and desires, corporations and nations, personalities and character traits. One final example: personality traits do not show up on brain scans. They leave their traces on behavior, though. They leave traces on psychological survey instruments, too. It is useful to think of people as having personality traits—it is explanatory and sometimes predictive—so we infer their existence. We posit them and treat them as real. We are not wrong to do so; without such posits, our understanding of human nature would be greatly diminished.

The point, of course, is that mental immune systems might belong to this rather expansive class of abstract entities—entities we rightly infer from indirect, behavioral evidence. They might be the kind of thing one *posits* rather than *discovers*.

The fact that we remain unaccustomed to inferring the existence of mental immune systems, of course, is neither here nor there. Epistemically, that should count for nothing. For it could be a mere accident of history: an artifact of our not having named the underlying mechanism until recently. Suppose someone had thought to advance the conjecture that minds *do* have immune systems—decades ago before William McGuire discovered that minds are inoculable. Had history played out in this fashion, his 1961 discovery might be interpreted today as a decisive experiment: one that conclusively demonstrated the existence of mental immune systems. The tendency to draw the inference might have been thoroughly normalized by now. And the existence of mental immune systems would be common knowledge.

But that is counterfactual history. In *our* world, the concept of mental immune systems has had little time to compile an explanatory track record. For we have yet to take the possibility seriously. To change that, we need to entertain the hypothesis that minds have informational defenses. We need to try on the idea and see if it fits. Does it explain things we *can* observe? Do people behave as if their minds have immune systems? What do the evidence and arguments say? Let us find out.

## The Evidence

### Exhibit A: Psychological Inoculation

For 6 decades, researchers working in a field known as inoculation theory have amassed evidence that it is possible to "inoculate" minds—that is, build resistance to false or misleading information via preemptive exposure to weakened or inactivated strains of such information along with instructions on how to neutralize them (McGuire, 1964; van der Linden, 2023). Indeed, psychological inoculants do for the mind what vaccines do for the body: induce them to produce the "cognitive antibodies" we call counterarguments and objections, thereby conferring immunity. Dozens of studies, including systematic reviews and meta-analyses of the empirical literature, now testify to the inoculability of minds in this sense; we offer this body of evidence as Exhibit A (Banas & Rains, 2010; Compton & Pfau, 2005; Lewandowsky & van der Linden, 2021; Lü et al., 2023; Roozenbeek et al., 2022; van der Linden, 2021).

Now combine this fact with another: Our bodies are inoculable *because they have immune systems*. The result? A creditable argument for MIST: Minds are inoculable because they have immune systems too. Put differently, MIST provides a straightforward explanation for the inoculability of human minds. We call this the *argument from inoculability*.

We do not regard this argument as decisive. Not in this simplified form. After all, there might be better explanations of cognitive inoculability. Also, one might question whether language borrowed from immunology is fully warranted in this context.

So let us delve deeper and ask how psychological inoculation works. What mechanisms are involved? Are these mechanisms relevantly similar to those that underlie bodily immunity?

### Exhibit B: Cognitive Antibodies

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 is not. To see this, consider four facts. First, healthy bodies manufacture the biochemical agents we call antibodies. Second, they ship these antibodies to sites of actual or possible infection. Third, these antibodies bind to and neutralize infectious agents. Fourth, in this way, they confer a degree of immunity to disease.

Now note the parallels. One: under certain conditions, a healthy mind will manufacture the cognitive "agents" we experience as counterarguments. Two: it will ship these agents to conscious awareness, where they can be paired with arguments—potentially infectious bids to change one's mind. Three: counterarguments sometimes "bind to" (become mentally associated with) such arguments, thereby neutralizing them. Four: in this way, they confer a degree of immunity to persuasion.

These are the parallels that inspired McGuire to formulate inoculation theory. We find them striking. We submit that they constitute hard evidence of relevant similarity. Mental immune systems work in much the way that bodily immune systems do.

Not ready to draw that inference? Try performing this experiment. Read an obviously wrong headed ideological screed and notice what happens in your mind. You will experience dissonance. Doubts and objections will likely swarm into conscious awareness, effectively neutralizing the persuasive effects of the screed. In the framework of mental immune systems theory, this is the mind's immune system in action. The point, of course, is that you can directly *observe* such activity—via introspection. (First-person evidence must be handled with care, but it is evidence of a kind. It counts.)

Incidentally, this thought experiment is analogous to a famous scientific experiment. In 1882, a Russian zoologist named Elie Metchnikoff injured a starfish larva with a citrus thorn and put it under a microscope. He watched as white blood cells swarmed to the scene of the injury and attacked the invading thorn. He was the first human being to watch antibodies in action. He named them and described their behavior; he called the cells "phagocytes" (Greek for "thing that devours") and called their antigen-devouring behavior "phagocytosis." He put antibodies on the scientific map. For this, he won a Nobel Prize.

Note that *qualms*, *doubts*, *questions*, and *objections* seem to function like counterarguments: They counteract or modulate conviction. In the ecology of thought, they play similar roles. This raises a fascinating question: are counterarguments, qualms, doubts, questions, and objections *cognitive antibodies*? If they function to neutralize unwarranted or maladaptive conviction, are they not essentially *antibodies of the mind?* "Doubts" is a useful shorthand for this broad class of conviction-modulating entities; used this way, it permits a more economical formulation of an idea central to MIST: *Doubts are the antibodies of the mind.*

Like antibodies, doubts can run amok and even attack the minds they are meant to protect. In the MIST framework, this is understood on analogy with autoimmunity, the phenomenon whereby an immune system attacks the body *it* is meant to protect.

Object to the language of mental immunity if you must; fact-based thinkers need to concede this much: our minds have subsystems that manufacture the conviction-modulating entities we call doubts. We call this *the argument from conviction modulation.*

It is not unusual to have qualms about the language of "cognitive antibodies." A well-functioning mental immune system *should* react to such talk; after all, it is a nonstandard way of thinking. It is foreign. It is "other." But just as you can overcome a peanut allergy by gradually exposing your body's immune system to peanuts, you can overcome an allergy to a useful new way of thinking via progressive exposure. Give the concept of cognitive antibodies a chance; in time, you may find it congenial. You may even find what we have found: that it is both scientifically and pedagogically useful.[21]

## Exhibit C: The Coupled Concepts Argument

A third argument enters here. It serves to address qualms about the application of language that appears to "medicalize" thought. The concepts of susceptibility and immunity are used in domains

[21] In our work teaching critical thinking, for example, we urge children to think of doubts as friendly agents that are trying to tell us something. Like the body's antibodies, they are (usually) our friends. (Overweening, exaggerated, and intemperate doubts can be exceptions to this rule.) Learn to respect and listen to judicious doubts, and your immunity to misinformation will grow.

far removed from disease. For example, we can speak of Bob's susceptibility to peer pressure and Claire's immunity to manipulative sales tactics. Susceptibility to impulse buying is a thing; so is immunity to prosecution. The point, of course, is that the concepts of susceptibility and immunity have long since "spilled over"; they have proved indispensable elsewhere and no longer belong to medicine alone.

Note, too, that the concepts of susceptibility and immunity are tightly correlated: By linguistic convention, more of the one means less of the other. "Susceptible" pretty much *means* nonimmune, and "immune" pretty much *means* nonsusceptible. Any claim about something's susceptibility to X, then, can be translated into a claim about that thing's lack of immunity to X. Of course, neither need be an all-or-nothing thing; both come in degrees. You can be *somewhat* susceptible to hypnotic suggestion and *moderately* immune to persuasion. Where susceptibility is partial, we can speak of limited immunity; where immunity is partial, there is some susceptibility.[22]

These are just facts about normal language usage, but in this context, they constitute a kind of evidence. For if it is permissible to speak of misinformation susceptibility (and presumably it is), it is also permissible to speak of misinformation immunity. You cannot allow one usage and object to the other—not without explaining why and how we should decouple the concepts. We call this *the coupled concepts argument*, and it appears to justify mental immunity talk.

Note the larger context here. This tight coupling of susceptibility and immunity has created a conceptual bridge that more and more scholars are willing to cross. That is why we are witnessing more scholarly references to things like "misinformation susceptibility," "misinformation resistance," "mental immunity," "cognitive immunity," "mind inoculants," and "cognitive antibodies" (Gorbis et al., 2019; Maertens et al., 2023; van der Linden, 2022). Meanwhile, investigators are using the concepts of "viral misinformation," "memes," "mind-parasites," and "infodemics" to better understand outbreaks that for all the world look like cognitive contagion (Blackmore, 2001; Blackmore et al., 2000; Boudry & Hofhuis, 2018; Eysenbach, 2020; Larson, 2018; Zarocostas, 2020). Information epidemiology is now a real science, one the World Health Organization is promoting under the label "infodemiology" (Calleja et al., 2021; Gallotti et al., 2020). Simply put, misinformation experts are taking up "medicalizing" language for good, scientific reasons.

These developments raise a serious question: Does language drawn from the study of disease afford systematically better answers to long-standing questions about our minds? (Waldrop, 2023). What if it can deepen our understanding of information dynamics? Can the language of susceptibility and immunity help to explain why some, but not others, fall for conspiracy theories? Could use of these metaphors herald the arrival of a progressive paradigm shift? We need to be giving these possibilities serious consideration, not dismissing them because the language makes us uncomfortable.

## Exhibit D: Infodemics

Information epidemiology furnishes a fourth body of evidence. When scientists study information diffusion, they find that it spreads through populations the way diseases do (Boudry & Hofhuis, 2018; Kauk et al., 2021; Kucharski, 2016). In fact, epidemics and infodemics obey the same mathematical laws (Cinelli et al., 2020; McCormack & Salter, 2010; Ndii et al., 2018). To successfully model the spread of any infectious vector, you have to segment the population into those who *have not* contracted it, those who *have*, and those who have *contracted it but recovered*. Members of the third class are treated differently from members of the second class because they have gained something—a kind of resistance—that makes them relatively immune, hence unlikely to transmit the infectious agent to others. Such resistance *is* immunity.

Simply put, the equations of epidemiology offer eloquent testimony that mental immunity talk refers to something real. Immunity to the solicitations of snake oil salesmen is every bit as real as immunity to chicken pox. Both have a basis in reality.

Of course, mental immunity has a psychological as well as biological basis. We have much to learn about the cognitive and neurological underpinnings of misinformation immunity. But that is precisely the point: We are likely to need the vocabularies

[22] The COVID-19 epidemic underscored that biological inoculation tends to confer limited, rather than absolute, immunity. The same is true of psychological inoculation.

of several disciplines to fully understand such phenomena. It all adds up to an argument *for* positing mental immune systems—and developing the field of cognitive immunology—not an argument *against* doing so.

"Infodemic" is a neologism, but infodemics are nothing new. False and harmful information has spread through human populations for ages. For example, accusations of witchcraft spread like wildfire through early modern Europe, inducing some to persecute, stone, and burn (mostly female) "witches." Detailed historical research shows that this tragic episode is best understood as an infodemic, with witchcraft beliefs playing the part of mind parasite (Hofhuis, 2022). When conditions changed, the infodemic died out. Thankfully, today's Europe has achieved something like herd immunity to witchcraft panics.

Examine history through the lens of information epidemiology, and it becomes clear that outbreaks of problematic information are a perennial source of human misery. Orators inflamed unruly passions in ancient Athens; this led the city into a disastrous war and caused the decline of a magnificent civilization. Hitler used propaganda to delude the German public and murder millions of Jews (Voigtländer & Voth, 2015). The story is as old as time: misinformation (or disinformation) spreads; it inflames partisan passions and stokes division; it fuels pointless conflict and causes needless suffering.

Fortunately, the discipline also teaches that small reductions in aggregate susceptibility can bring a raging infodemic under control. Nudge what epidemiologists call "R-zero"—the basic reproductive rate of an infection—below one, and an epidemic will start to die out (van der Linden, 2023). The same is true for the basic reproductive rate of false narratives that scapegoat witches. Or immigrants. Or Jews.

Surely it behooves us, then, to study misinformation immunity. We need to understand what it is and how it is acquired; why some have it in spades, while others struggle with mental immunity deficits. How do we protect everyone from the worst forms of cognitive contagion? In our digital age, that question needs to be near the top of scientific and policy agendas.

Biological immune activity leaves behind distinct traces. The concept of *inflammation* refers to a particular set of such traces. If mental immune systems exist, they should also leave distinctive traces. So, *does* mental immune activity leave such signs? And if so, might the effects be thought of as a kind of inflammation?

## Exhibit E: IPC

Our minds can perform extraordinary feats of credulity and denial to protect our religious and political identities. Scientists have noticed and named this tendency; they call it identity-protective cognition, or IPC (Kahan, 2017). The concept points to powerful cognitive forces: forces that underlie identity maintenance and, for the most part, operate beneath the level of conscious awareness.

Empirical studies of IPC afford a fifth body of evidence for MIST. You need not pour over the journal articles, though, to appreciate the point. The phenomenon is known to us all: Criticize someone, and usually, they will get defensive. Challenge an ardent believer with evidence, and they will manufacture reasons to discount that evidence. Question the value of a person's career, and you are bound to get pushback.

Clearly, our minds conduct operations that protect our identities, our egos, and our beliefs from perceived informational threats. We grow attached to our beliefs and begin thinking in distorted ways to protect them. Presumably, one or more subsystems carry out such protective—that is to say, defensive—operations.

We submit that IPC constitutes conspicuous evidence that our minds *do in fact have defenses*. It appears that these defenses evolved to protect our *selves:* our bodies, our minds, our social alliances, our self-esteem, the mental models we rely on, and—importantly—our beliefs.

Apparently, they evolved to protect *against* the kinds of information that threaten to disturb these things: identity-threatening information, belief-disrupting information, manipulative signals, information that might fracture the coalitions that matter to us, beliefs that tend to reduce biological fitness, misleading falsehoods, and so forth. (We put falsehood avoidance near the bottom of this list because conspicuous examples of IPC often involve the protection of self-serving misbeliefs.)

Clearly, our minds have informational defenses. Is it really such a stretch to call them mental immune systems?

## Exhibit F: Reactance and Dissonance

Further evidence for MIST can be found in the scholarly literatures on *psychological reactance*

and *cognitive dissonance* (Festinger, 1957; Steindl et al., 2015). For these literatures establish two things: (a) that the human mind has a distinct tendency to react to (and sometimes *over*react to) threatening information and (b) that the protective reaction often takes the form of "dissonant" thoughts and feelings—which is to say, cognitive activity that to some degree counteracts the perceived threat.

We have all experienced psychological reactance. Experts define it as "an unpleasant motivational arousal that emerges when people experience a threat to or loss of their free behaviors" (Steindl et al., 2015). When a teenager overreacts to parenting she feels is controlling, that is reactance; when political discourse triggers an overwrought response from a partisan, that is reactance. The response involves emotions that can be a counterforce to the perceived threat; this makes reactance a fundamentally defensive tendency.

The defensive subsystem responsible for generating psychological reactance deserves a name. "Mental immune system" is as good as any. In fact, it is better than most. For it is no arbitrary linguistic designation; no empty, "black box" explainer. To say that mental immune systems are responsible for reactance is to hazard a substantive and falsifiable conjecture. For if the emotional arousal characteristic of reactance tended to *reinforce* information threats rather than *counteract* them, that would count against MIST. It might even falsify it. MIST links an observable phenomenon (reactance) to a well-developed body of knowledge about evolved defense systems. Moreover, it is easy to see how a psychological immune system with reactant tendencies might evolve, so the posit is genuinely explanatory. It helps explain where reactance comes from. We call this the argument from reactance.

It is worth considering how this argument might land if someone used MIST to predict reactance *before* reactance was empirically observed. Imagine highly intelligent aliens from another planet intercepting early human radio transmissions and deducing that Earth is inhabited by creatures with sophisticated minds. Suppose these aliens are aware that minds are prone to information "infections." Suppose that they also understand evolutionary dynamics. An alien theorist, then, might well conjecture that these Terran minds are probably equipped with evolved informational defenses. Seeking empirical validation, this theorist might predict that, under certain conditions, these strange Terrans would exhibit reactance. Alien experimentalists might then comb through their trove of radio signals and find evidence that, yes, these Terran creatures *are* psychologically reactant. In a scenario like this, the discovery might be viewed as vindicating the theorist's conjecture.

Needless to say, the mere accident of MIST's historically late formulation should not be held against it. The evidential significance of consilient phenomena like reactance should be exactly the same, whether the hypothesis that purports to explain it comes along before or after the fact. The weight given here to a successful *retro*diction should be comparable to the weight given to a successful *pre*diction.

Now consider cognitive dissonance. Cognitive dissonance is the discomfort people feel when they encounter information that contradicts their beliefs, values, or ideas (Cleveland Clinic, 2023b). Often, the discomfort motivates the affected person to reduce the dissonance in one of several ways. For example, cognitive dissonance might motivate someone to deny inconvenient facts or explain away contradictory evidence. On the bright side, it can also motivate a scientist to address a discrepancy between theory and observation or unify two discrepant theories.

There is no doubt that cognitive dissonance is real; the open question is what causes it. The ultimate evolutionary cause was probably identified by Leon Festinger, the American psychologist who named the phenomenon. Festinger understood that inconsistent belief systems can compromise our ability to function well in the world and proposed that we strive for internal consistency so that we can function well (Harmon-Jones & Mills, 2019). On this account, cognitive coherence is to some degree adaptive, so natural selection found a way to promote it.

But this story about cognitive dissonance's *ultimate* cause leaves us with an unanswered question about its *proximal* cause. What neural hardware and/or psychological software generates cognitive dissonance? What on-board equipment is responsible for creating the discomfort needed to motivate real-time coherence-enhancement efforts?

We think cognitive dissonance is an experiential manifestation of mental immune function. MIST is the idea that minds have defenses that monitor for, flag, and attempt to neutralize informational threats. Since information that contradicts on-board beliefs, values, or ideas

does threaten them, *you would expect mental immune systems to flag and attempt to neutralize it.* Embrace MIST, in other words, and you get a straightforward explanation of cognitive dissonance: The latter is just what mental immune function feels like from the inside.

To sum up, psychological reactance and cognitive dissonance are precisely the kind of tendencies you would expect of a creature equipped with an informational immune system. The two phenomena constitute evidence for MIST. In fact, we can think of psychological reactance and cognitive dissonance as two of the traits that make up the mind's immune system. (Recall that we have defined an immune system as a suite of *traits* that protects an evolved entity from "in-house" threats. The point here is that the phenomena and the definition dovetail neatly in the case of the human mind.) Whether these traits are directly adaptive or mere by-products of adaptations need not concern us here; either way, they provide evidentiary support for MIST.

## Exhibit G: Biases

Evidence for MIST can also be found in the extensive literature on cognitive biases. This literature shows that our minds are strangely susceptible to making certain errors. We are prone to confirmation bias, motivated reasoning, belief perseverance, groupthink, base-rate neglect, and dozens of other distortions. One recent count has the total number of documented cognitive biases at 188 (Desjardins, 2021).

Here is the thing to see: The close conceptual coupling of susceptibility and immunity makes every fact about error-susceptibility also a fact about our *lack of immunity* to such errors. This means that a huge literature on cognitive biases can be translated into insights about cognitive immune *vulnerabilities*. Cognitive bias research illuminates gaps in the mind's defenses.

The point should be clarified. Cognitive bias hunters look for mental tendencies that are suboptimal as measured by modern standards of truth-oriented information processing. These tendencies count as biases because they fall short on *epistemic* criteria. They mislead. Even so, they did not prevent our ancestors from becoming ancestors, so they are assumed to have been reasonably adaptive even so. They may be *error-conducive* in important respects, but they had to be sufficiently *survival-conducive*.[23] Otherwise, we would not have inherited those tendencies.

Evolutionary theory predicts that informational defenses should be epistemically suboptimal (Peirce, 1877). Why? Because survival-conduciveness is one thing and truth-conduciveness is another, and the two goals are not always aligned. And this *is* what the empirical evidence shows; our proneness to illusory superiority, for example, is thought to be both deceptive *and* adaptive (DeAngelis, 2003). Nor are adaptive illusions uncommon. (Notice also that survival-conduciveness in *ancestral* environments is not the same as well-being-conduciveness in *today's* environments: This difference underlies the phenomenon known as *evolutionary mismatch*.)

Had it gained acceptance first, MIST could have been used to predict key findings of cognitive bias research. Take confirmation bias. The American Psychological Association defines confirmation bias as the "tendency to gather information in a way that confirms preexisting expectations, typically by emphasizing or pursuing supporting evidence while dismissing or failing to seek contradictory evidence" (American Psychological Association, 2018). Note that this tendency is inherently conservative: It tends to conserve existing beliefs, values, and mental models. Why might this tendency evolve? Well, if your beliefs got you this far, you probably should not abandon them at the first sign of trouble.

An easy way to implement such conservatism in a learning system is this: Use existing beliefs as a reference point to gauge the threat level posed by newly arriving information. The more counterintuitive the information, the more suspicion it merits. For if new information clashes with beliefs that are already on board and those beliefs have a track record of working for you, it pays to assign the new information low initial ("prior") credence.

Why is it that our minds automatically assign a level of credence to newly arriving information? Could this be the mind's defenses doing what they are designed to do: gauging threat levels and tagging newly arriving information as more or less likely to disrupt? And why do we tend to attach less credence to information that is strange, surprising, and counterintuitive? Well, that information represents the biggest threat, and the low credence reduces its power to disrupt. Confirmation

[23] "Sufficiently survival conducive" here means something like "not too much of a drag on biological fitness."

bias and Bayesian priors, then, may both be explicable as artifacts of mental immune function.

Here is the point: *Confirmation bias is exactly the kind of tendency you would expect a creature equipped with a psychological immune system to exhibit*. After all, it protects a mind's core beliefs and values from potentially disruptive information. Epistemically speaking, confirmation bias is a bug, but from a self-preservation standpoint, it is a feature.

Notice that a cognitive system prone to confirmation bias is also a system prone to ideological fixation. For unreliable information can infect a mind and skew the way it evaluates subsequent information. Problematic ideas can hijack the mind's immune system—and get it to treat later-arriving evidence as a "foreign" threat (Mauer, 2022). In this way, an idea can protect *itself* against replacement. Many biological parasites pull such stunts; why should information parasites not do the same?

And confirmation bias is just one of 188. We also need to explain why we are prone to motivated reasoning, belief perseverance, myside bias, egocentric bias, status quo bias, and the sunk cost fallacy. Could these too be artifacts of mental immune function (defined, again, as the things minds do to protect themselves from perceived informational threats)? Arguably, yes: creatures equipped with informational defenses *should* be prone to just such errors.[24]

Cognitive psychologists are already looking for a way to bring unity and explanatory coherence to our understanding of human biases (Marshall et al., 2013; Oeberst & Imhoff, 2023). Why is there such a bewildering profusion of them? Can we group them and discern higher level patterns? Can we achieve some kind of explanatory unification here? Speculative? Yes, but the conjecture merits development.

In the meantime, we can mine cognitive bias research for insights into how mental immune systems work. And why they are prone to fail. We think such research will shed light on the human mind's signature vulnerabilities, among them those that peddlers of disinformation exploit so easily in our highly networked world.[25]

## Exhibit H: Evolution

We have learned in recent decades that natural selection is a special case of a more general evolutionary algorithm (Dawkins, 2006; Dennett, 1995). Whenever entities replicate and environments differentially favor variants with certain characteristics, evolution will happen.[26] The entities in question do not have to be biological. They can be informational. Take the entities we call ideas. Given the existence of systems capable of replicating them (language-networked brains, say) and environments that favor some over others, ideas *will* evolve.[27]

The same goes for other culturally transmissible entities. Words, sayings, maxims, songs, stories: all of them are cultural replicators. Theorists call this class of replicators "memes." The concept of memes faced early backlash, but there is no longer any doubt that memes exist.[28] Nor is it possible to deny that they replicate and evolve (Boudry & Hofhuis, 2018).

And here we face a hard truth: Blind memetic evolution does not just happen; it shapes our minds. (In saying this, we do not deny that minds also shape memes in ways that are anything but blind; the causal arrow between minds and ideas goes both ways.) Internalize this insight, and you will be brought up against an inescapable fact: Just as our bodies need defenses to protect them from internal disruptors, our minds need defenses to protect *them* from internal disruptors. Given the existence of parasitic informational replicators, our minds *had* to evolve immune systems. The same logic applies in both domains. At some level, MIST must be true.

The axioms of evolutionary thinking, of course, are massively explanatory. And it seems MIST is among their corollaries.

[24] Myside bias, for example, is plausibly a mental immune adaptation "designed" by natural selection to protect standing in a coalition. For an insightful treatment of myside bias, see Stanovich (2021).

[25] A. Norman's "broken fulcrum model" provides a more fine-grained account of how thinking becomes ideological (Norman, 2021), Chapter 9.

[26] Given large enough populations and sufficient time. The mathematics of large populations ensures this.

[27] Some information environments favor (and thereby select for) diverting content. Hence clickbait. Others select for information that is ideology-reinforcing (Fox News) or thinking that is well-argued (this journal, e.g.).

[28] No one doubts that internet memes exist. These alone prove that the class of memes is not empty. We might quibble over the boundaries of this class—asserting or denying, say, that cultural norms are self-replicating entities that deserve to be called "memes"—but the existence claim is settled.

## Exhibit I: Susceptibility Variance

People vary enormously in their susceptibility to misinformation. Some are taken in by false narratives; others shed them easily. Some are drawn to conspiracy theories; others see right through them. The very propaganda that addles some is laughed off by others. But how best to explain such variation?

Note two things about this question. First, our understanding of misinformation susceptibility is quite limited.[29] For why else would the credulity of, say, QAnon believers astonish us so? How is it even *possible* that flat Earth theory and national socialism win adherents these days? The fact that we find such infodemics baffling is a sign that there is much mystery yet to dispel: Our model of the human mind needs refinement. Second, notice that this question is of critical importance for a digitally connected world; for in such a world, the susceptibility of some (to hateful bigotry, say) can metastasize into a danger for all. In such a world, mental immunity cultivation becomes essential.

Many attempts to account for variance in misinformation susceptibility employ the idea that some possess critical thinking skills, while others lack them. This story has some merit, and it enjoys wide acceptance within academe, but it is increasingly clear that, as an explanation, it does not suffice.[30] Critical thinking skills can be used to rationalize beliefs and preferences, thereby *worsening* one's misinformation susceptibility. Also, thinking can become *hyper*critical: you can learn the basics of critical thinking and have it morph, over time, into indiscriminate suspicion of, say, "the mainstream media" or "Big Pharma."

More critical is not always better; we can be too critical for our own good. Put differently, "critical" thinking can detract from mental immune health (Schwitzgebel & Ellis, 2017). Moreover, the concept of critical thinking does little to shed light on the subconscious processes that go into capable information assessment. Much information filtration is subconscious, so references to a conscious, deliberate evaluation process (however salutary) cannot be the whole story.

The concept of epistemic vigilance has the same defects (Sperber et al., 2010). Epistemic vigilance is generally a good thing, and it can reduce misinformation susceptibility (Roozenbeek et al., 2022). But hypervigilance is also a thing, and it too can distort thinking. Also, vigilance denotes a conscious, deliberate process, something that requires effort and focus. The problem is that misinformation susceptibility also depends on *subconscious* processes: The unconscious processes that flag certain information as worth a closer look, for example. Epistemic vigilance, like critical thinking, cannot be the whole story.

Nor can references to innate heuristics and cognitive biases explain the variance. Why? Because species-universal traits cannot account for intraspecies variation. Not by themselves. Nor can references to innate traits account for variation over the course of a lifetime. Clever people can fall down conspiracy rabbit holes. Open-minded people can become more ideological over time. Why? Built-in cognitive biases may play a role, but *variation* must be traceable to environmental variables not (just) biological constants. Something else is going on.

According to MIST, variation in misinformation susceptibility is largely a product of *mental immune function*—the things minds *do*, at both conscious and subconscious levels, to vet new information and update old beliefs. A mind's defenses can perform these tasks well, thereby keeping the mind relatively misinformation-free, or they can perform them poorly, allowing misbeliefs to form.[31] On this account, understanding the mind's immune system is the key to mitigating our misinformation problem—just as understanding the body's immune system was key to mitigating our disease problem.

We do not claim that this constitutes a full explanation. Many details need to be filled in. The concept of mental immune function, though, has the right shape. As our understanding of healthy

[29] Limited, but cognitive immunologists *have* developed measures of misinformation susceptibility. See Maertens et al., 2023.

[30] On some concepts of critical thinking, the imparted skills can be employed to rationalize favored conclusions rather than reason one's way to truth. Also, the concept of critical thinking sheds little light on the subconscious processes that go into information assessment. Third, the concept of critical thinking is in some ways a black box: a placeholder for a better way to think about thinking. We think the language of mental immunity represents a nice down payment on this better way. The arguments here are spelled out more fully in (Norman, 2021) Chapters 1 and 3.

[31] "Well" and "poorly" can be used to mean "fitness-enhancing" and "fitness-reducing." In this context, though, our usage incorporates a distinct value judgment: We think that mental immune activity should screen out even adaptive falsehoods. We see our value judgment here as comparable to the axiom at the heart of the medical sciences, which unapologetically favors good health over ill health. By analogy, applied cognitive immunology might choose to valorize truth-conduciveness. See Footnotes 7 and 8.

mental immune function grows, we will begin to name and catalogue the different kinds of mental immune dysfunction. Over time, this should lead to a deep understanding of misinformation susceptibility: What it is, what causes it, and what works to mitigate it.

## Conclusion: Cognitive Immunology and Its Prospects

Our reluctance to posit mental immune systems has long inhibited the science of mental immunity. Cognitive immunology attempts to throw off these shackles. It defines "immune system" in a suitably encompassing way and embraces a straightforward consequence of that definition: that minds have immune systems of their own. We need not allow vague metaphysical qualms to hamstring the science; instead, we can posit mental defenses and explore that posit's explanatory potential.

The discipline of cognitive immunology will draw from several more established fields. The empirical foundation was laid by inoculation theorists, but in the future, cognitive immunologists will draw also from information science. It will draw from philosophy (particularly epistemology), anthropology, and immunology. It will leverage evolutionary thinking and the principles of information epidemiology.

The language of immunology opens many doors to deeper understanding. Consider the questions it allows us to pose: What does healthy mental immune function look like? What environmental conditions disrupt such functioning? What habits, ideas, and attitudes qualify as mental immune disruptors? What are the various species of mental immune disorder? Are there acquired mental immune deficiencies? What about autoimmune disorders of the mind? Are doubts and questions cognitive antibodies? Can learning how to wield such antibodies make a mind more flexible, more open, and more resilient? Can exposure to the Socratic method reduce susceptibility? What environmental conditions, habits, ideas, and attitudes boost mental immune performance? What works to inoculate minds? What would a mind vaccine look like? And what ideas, if any, should we "vaccinate" against? Each of these questions promises to deepen our understanding of the mind.

We think cognitive immunology has a bright future. Imagine our understanding of the mind's immune system expanding until it rivals our understanding of the body's immune system. Imagine how much better our treatments for misinformation susceptibility could become. (Think of such treatments as taking the form of next-level critical thinking instruction for the willing, not forced inoculation of the unwilling.) Imagine how much rarer outbreaks of mass irrationality could become. What if we could reduce toxic polarization by 35%? Or make everyone 15% less susceptible to ideological fixation? What if we could make angry, hateful delusions uncommon? Imagine taming the worst infodemics the way we tamed the worst epidemics: by patiently building herd immunity to the nastiest infectious agents.

Of course, we must take care not to abuse our understanding of the mind's immune system. The findings of cognitive immunology should be used to enhance, never diminish, cognitive autonomy. We must use cognitive immunology to *free* minds, not manipulate them.

Twentieth century biologists named the body's immune system and went on to develop a stunningly beneficial discipline. Immunology has made our lives immeasurably better. It has saved hundreds of millions—probably billions—of lives and prevented untold suffering. It falls to us, in the 21st century, to do the same with the mind's immune system.

We conclude with a table describing a set of experiments. Some could yield a decisive demonstration of MIST. Others could deepen our understanding of mental immune systems or extend the theory's explanatory and predictive reach. We invite colleagues—theorists and experimentalists alike—to help us plumb the mysteries of the mind's immune system (Table 1).

If the mind *did* have an immune system, what empirical indicators would we expect to find? We propose a program of research that combines psychological/behavioral, physiological, neurological, and epidemiological indicators that could jointly evidence the presence of a cognitive immune system. For example, research is already starting to show that processes such as psychological inoculation and reactance are associated with distinct physiological signatures (e.g., Clayton et al., 2023). Though it is unlikely that cognitive immunology is associated with a single biochemical marker or neurological substrate given that "many areas of higher cognition are likely involved in assessing the truth value of linguistic propositions" (Harris et al., 2008, p. 1), there is already exciting work on the neural

correlates of counterarguing (Weber et al., 2015) and belief resistance in the face of counterevidence (e.g., Kaplan et al., 2016) where changes in key regions of interest are predictive of responses to future campaign messages (Weber et al., 2015). Jointly, such a research program could provide evidence that mental immune activity has distinct physiological manifestations and neurological signatures. This table presents some ideas for future experimental work.

**Table 1**
*Experimental Tests of Mental Immune Systems Theory*

| Experiment | Goal | Method |
|---|---|---|
| | Psychological/behavioral indicators | |
| Experimental study of psychological inoculation and other cognitive immunology-informed training | Determine (long-term) effects of cognitive immunology training on misinformation identification and resistance. | Randomized experiments assessing behavioral evidence of mental "antibodies" in the inoculation group (relative to control) such as increased counterarguing, better discernment of misinformation strategies, stronger memory of inoculation content over time, and higher motivation to protect oneself from manipulation. |
| Experimental manipulation of cognitive immune system functions. | Determine the psychological and behavioral manifestations of cognitive immune response to informational threats (e.g., people clinging on to false beliefs despite clear counterevidence). Strong reactance might be an indicator of "cognitive inflammation." | Causal experiments and psychological scales that can evidence and measure increased psychological reactance, cognitive dissonance, identity-protective cognition, proneness to self-deception, and other cognitive biases in response to exposing people with strong group identities to belief-challenging information. |
| | Physiological indicators | |
| Physiological correlates of cognitive inoculation and immunity. | Measure physiological manifestations of heightened cognitive immune activity. | Measure heart rate deceleration (focused attention), corrugator muscle activation (emotional valence), eye-tracking and dwell time, galvanic skin response, and so forth during exposure to psychological inoculation material and/or challenging information. |
| Physiological correlates of cognitive immune function. | To closely examine physiological response data for patterns related to the (hypothesized) state of cognitive immune function (e.g., reactance might elicit observable negative emotional reactions). | Assess heart rate variability, cortisol levels, skin conductance, and so forth to see if physiological data corresponds to behavioral indicators of cognitive immune function in the treatment (exposure) versus control (no exposure to threatening material) group. |
| | Neurological indicators | |
| Neuroimaging studies of cognitive immunity to misinformation. | To identify neural correlates and predictors of cognitive immunity in the brain using observational and experimental fMRI/EEG studies. | Differential neural activity in regions of interest such as the hippocampus (memory), the ACC, and the medial prefrontal areas (cognitive control and conflict monitoring). Brain-as-predictor studies indexing fMRI changes to predict future immunity to misinformation. |
| Neural correlates of cognitive immune function. | Explore neural activity associated with the (hypothesized) state of cognitive immune response. | Conduct fMRI or EEG studies where participants are confronted with identity-reassuring versus identity-threatening facts using predefined regions of interest. |
| | Modeling and simulation | |
| Modeling the evolution of mental immune systems. | To show that mental immune systems can evolve under certain conditions. | Model the conditions under which representative agents evolve cognitive "defenses" against disruptive information. |
| Modeling infodemics: How misinformation spreads through communities of agents with cognitive defenses? | To determine whether the spread of information through networks contains evidence that human minds have the capacity to develop resistance to misinformation | Model communities with and without cognitive defenses and compare how information spreads through them. Which model better explains the data gathered by information epidemiologists? |

*Note.* fMRI = functional magnetic resonance imaging; EEG = electroencephalogram; ACC = anterior cingulate cortex.

## References


American Psychological Association. (2018). *Confirmation bias*. Retrieved November 16, 2023, from https://dictionary.APA.org/confirmation-bias&lang

Bak-Coleman, J. B., Kennedy, I., Wack, M., Beers, A., Schafer, J. S., Spiro, E. S., Starbird, K., & West, J. D. (2022). Combining interventions to reduce the spread of viral misinformation. *Nature Human Behaviour*, *6*(10), 1372–1380. https://doi.org/10.1038/s41562-022-01388-6

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281–311. https://doi.org/10.1080/03637751003758193

Blackmore, S. (2001). The meme's eye view. In R. Aunger (Ed.), *Darwinizing culture: The status of memetics as a science* (pp. 24–42). Oxford University Press.

Blackmore, S., Dugatkin, L. A., Boyd, R., Richerson, P. J., & Plotkin, H. (2000). The power of memes. *Scientific American*, *283*(4), 64–73. https://www.jstor.org/stable/26058899

Bond, S. (2022). False information is everywhere. "Prebunking" tries to head it off early. *NPR*. https://www.npr.org/2022/10/28/1132021770/false-information-is-everywhere-pre-bunking-tries-to-head-it-off-early?utm_campaign=storyshare&utm_source=twitter.com&utm_medium=social

Boudry, M., & Hofhuis, S. (2018). Parasites of the mind. Why cultural theorists need the meme's eye view. *Cognitive Systems Research*, *52*, 155–167. https://doi.org/10.1016/j.cogsys.2018.06.010

Braghieri, L., Ro'ee, L., & Makarin, A. (2022). Social media and mental health. *The American Economic Review*, *112*(11), 3660–3693. https://doi.org/10.1257/aer.20211218

British Society for Immunology. (n.d.). *What is immunology?* Retrieved November 15, 2023, from https://www.immunology.org/public-information/what-immunology

Budiman, A., Roan, A., & Callan, V. J. (2013). Rationalizing ideologies, social identities and corruption among civil servants in Indonesia during the Suharto Era. *Journal of Business Ethics*, *116*(1), 139–149. https://doi.org/10.1007/s10551-012-1451-y

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*(5), 839–855. https://doi.org/10.1037/0022-3514.76.5.839

Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracín, D., Altieri, E., Anoko, J. N., Arcos, R., Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S., Brooks, I., Bucci, L. M., Burzo, S., Czerniak, C., De Domenico, M., Dunn, A. G., … Purnat, T. D. (2021).A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. *JMIR Infodemiology*, *1*(1), Article e30979. https://doi.org/10.2196/30979

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, *10*(1), Article 16598. https://doi.org/10.1038/s41598-020-73510-5

Clayton, R. B., Compton, J., Reynolds-Tylus, T., Neumann, D., & Park, J. (2023). Revisiting the effects of an inoculation treatment on psychological reactance: A conceptual replication and extension with self-report and psychophysiological measures. *Human Communication Research*, *49*(1), 104–111. https://doi.org/10.1093/hcr/hqac026

Cleveland Clinic. (2023a). *Immune system*. Retrieved November 15, 2023, from https://my.clevelandclinic.org/health/articles/21196-immune-system

Cleveland Clinic. (2023b). *What is cognitive dissonance? Distress and discomfort can occur when your behaviors or actions conflict with your values*. Retrieved November 16, 2023, from https://health.clevelandclinic.org/cognitive-dissonance/

Cognitive Immunology Research Collaborative. (2023). *Declaration of CIRCE's blue ribbon panel on mental immunity*. https://drive.google.com/file/d/1S7pTB7THyi6s9eIbx4Tjm85ckC-nnn7i/view

Compton, J., & Pfau, M. (2005). Inoculation theory of resistance to influence at maturity: Recent progress in theory development and application and suggestions for future research. *Annals of the International Communication Association*, *29*(1), 97–146. https://doi.org/10.1080/23808985.2005.11679045

Compton, J., van Der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass*, *15*(6), Article e12602. https://doi.org/10.1111/spc3.12602

Dawkins, R. (2006). *The selfish gene*. Oxford University Press.

DeAngelis, T. (2003). *Why we overestimate our competence* [Dataset]. PsycEXTRA Dataset. https://doi.org/10.1037/e300052003-031

Dennett, D. (1995). *Darwin's dangerous idea*. Simon and Schuster.

Dennett, D. (2007). *Breaking the spell: Religion as a natural phenomenon*. Penguin Books.

Desjardins, J. (2021). *Every single cognitive bias in one infographic*. Visual Capitalist. https://www.visualcapitalist.com/every-single-cognitive-bias/

Distin, K. (2005). *The selfish meme: A critical reassessment*. Cambridge University Press.

Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. *Journal of Medical Internet Research*, *22*(6), Article e21820. https://doi.org/10.2196/21820

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. https://doi.org/10.1515/9781503620766

Gallotti, R., Valle, F., Castaldo, N., Sacco, P., & De Domenico, M. (2020). Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nature Human Behaviour*, *4*(12), 1285–1293. https://doi.org/10.1038/s41562-020-00994-6

Gorbis, M., Joseff, K., Frauenfelder, M., & Pescovitz, D. (2019). *Building a healthy cognitive immune system: Defending democracy in the disinformation age*. Institute for the Future. Retrieved November 16, 2023, from https://www.iftf.org/projects/building-a-healthy-cognitive-immune-system/

Gordon, A., Quadflieg, S., Brooks, J. C. W., Ecker, U. K. H., & Lewandowsky, S. (2019). Keeping track of 'alternative facts': The neural correlates of processing misinformation corrections. *NeuroImage*, *193*, 46–56. https://doi.org/10.1016/j.neuroimage.2019.03.014

Grimes, D. R. (2020). Health disinformation & social media: The crucial role of information hygiene in mitigating conspiracy theory and infodemics. *EMBO Reports*, *21*(11), Article e51819. https://doi.org/10.15252/embr.202051819

Hankir, A. Z., & Hassan, S. (2019). Psychological determinants and social influences of violent extremism. In H. Moffic, J. Peteet, A. Hankir, & R. Awaad (Eds.), *Islamophobia and psychiatry* (pp. 391–406). Springer. https://doi.org/10.1007/978-3-030-00512-2_32

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (pp. 3–24). American Psychological Association. https://doi.org/10.1037/0000135-001

Harris, S., Sheth, S. A., & Cohen, M. S. (2008). Functional neuroimaging of belief, disbelief, and uncertainty. *Annals of Neurology*, *63*(2), 141–147. https://doi.org/10.1002/ana.21301

Häusler, L., & Baraghith, K. (2023). Pandemic and infodemic: The spread of misinformation about COVID-19 from a cultural evolutionary perspective. *Biology & Philosophy*, *38*(5), Article 42. https://doi.org/10.1007/s10539-023-09928-8

Hofhuis, S. T. H. (2022). *Qualitative darwinism: An evolutionary history of witch-hunting* [Unpublished doctoral dissertation]. Utrecht University Repository.

Hume, D. (2007). *An enquiry concerning human understanding and other writings. Cambridge texts in the history of philosophy* (S. Buckle, Ed.). Cambridge University Press.

Kahan, D. M. (2017). *Misconceptions, misinformation, and the logic of identity-protective cognition*. Cultural Cognition Project Working Paper Series No. 164, Yale Law School. https://doi.org/10.2139/ssrn.2973067

Kannadan, A. (2018). History of the miasma theory of disease. *ESSAI*, *16*(1), Article 18. https://www.semanticscholar.org/paper/History-of-the-Miasma-Theory-of-Disease-Kannadan/a067a8d17f504883c6378e411f476d1a7b29f101

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports*, *6*(1), Article 39589. https://doi.org/10.1038/srep39589

Kauk, J., Kreysa, H., & Schweinberger, S. R. (2021). Understanding and countering the spread of conspiracy theories in social networks: Evidence from epidemiological models of Twitter data. *PLOS ONE*, *16*(8), Article e0256179. https://doi.org/10.1371/journal.pone.0256179

Kruglanski, A., Gelfand, M., Belanger, J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, *35*(Suppl. 1), 69–93. https://doi.org/10.1111/pops.12163

Kucharski, A. (2016). Post-truth: Study epidemiology of fake news. *Nature*, *540*, Article 525. https://doi.org/10.1038/540525a

Larson, H. J. (2018). The biggest pandemic risk? Viral misinformation. *Nature*, *562*, Article 309. https://doi.org/10.1038/d41586-018-07034-4

Lewandowsky, S., & van der Linden, S. (2021). Countering Misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology*, *32*(2), 348–384. https://doi.org/10.1080/10463283.2021.1876983

Lü, C., Hu, B., Li, Q., Bi, C., & Ju, X. D. (2023). psychological inoculation for credibility assessment, sharing intention, and discernment of misinformation: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *25*, Article e49255. https://doi.org/10.2196/49255

Maertens, R., Götz, F. M., Golino, H. F., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., He, J., & van der Linden, S. (2023). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods*, *56*, 1863–1899. https://doi.org/10.3758/s13428-023-02124-2

Marshall, J. A., Trimmer, P. C., Houston, A. I., & McNamara, J. M. (2013). On evolutionary explanations of cognitive biases. *Trends in Ecology & Evolution*, *28*(8), 469–473. https://doi.org/10.1016/j.tree.2013.05.013

Mauer, B. (2022). Bad ideas recruit the mind's immune system to protect themselves. *This View of Life Magazine*. Prosocial World. Retrieved November 15, 2023, from https://www.prosocial.world/posts/bad-ideas-recruit-the-minds-immune-system-to-protect-themselves

McCormack, R., & Salter, W. J. (2010). An application of epidemiological modeling to information diffusion. In S. K. Chai, J. J. Salerno, & P. L. Mabry (Eds.), *Lecture notes in computer science* (pp. 382–389). Springer. https://doi.org/10.1007/978-3-642-12079-4_48

McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry*, *24*(2), 184–197. https://doi.org/10.2307/2786067

McGuire, W. J. (1964). Some contemporary approaches. *Advances in Experimental Social Psychology*, *1*, 191–229. https://doi.org/10.1016/S0065-2601(08)60052-0

McKay, S., & Tenove, C. (2020). Disinformation as a threat to deliberative democracy. *Political Research Quarterly*, *74*(3), 703–717. https://doi.org/10.1177/1065912920938143

Mozur, P. (2018). A genocide incited on Facebook, with posts from Myanmar's military. *The New York Times*. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

Ndii, M., Carnia, E., & Supriatna, A. (2018). Mathematical models for the spread of rumors: A review. *Issues and trends in interdisciplinary behavior and social science* (pp. 65–73). CRC Press.

Newton, K., & Norris, P. (2000). Confidence in public institutions: Faith, culture, or performance? In S. J. Pharr & R. D. Putnam (Eds.), *Disaffected democracies: What's troubling the trilateral countries?* (pp. 52–73). Princeton University Press. https://doi.org/10.1515/9780691186849-007

Norman, A. (2021). *Mental immunity: Infectious ideas, mind-parasites, and the search for better way to think*. HarperCollins. https://andynorman.org/mental-immunity

Norman, A., & McIntyre, L. (2023). Are the critics of cognitive immunology in denial? *Psychology Today*. Retrieved November 15, 2023, from https://www.psychologytoday.com/us/blog/mental-immunity/202306/are-the-critics-of-cognitive-immunology-in-denial

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*, *18*(6), 1464–1487. https://doi.org/10.1177/17456916221148147

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, *12*(1), 1–15. https://philarchive.org/rec/PEITFO

Rodgers, K., & Massac, N. (2020). Misinformation: A threat to the public's health and the public health system. *Journal of Public Health Management and Practice*, *26*(3), 294–296. https://doi.org/10.1097/PHH.0000000000001163

Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, *8*(34), Article eabo6254. https://doi.org/10.1126/sciadv.abo6254

Sabelli, H. C., Carlson-Sabelli, L., & Javaid, J. I. (1990). The thermodynamics of bipolarity: A bifurcation model of bipolar illness and bipolar character and its psychotherapeutic applications. *Psychiatry*, *53*(4), 346–368. https://doi.org/10.1080/00332747.1990.11024519

Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, *20*(2), 99–103. https://doi.org/10.1177/0963721411402596

Schwitzgebel, E., & Ellis, J. (2017). Rationalization in moral and philosophical thought. In J.-F. Bonnefon & B. Trémolière (Eds.), *Moral inferences* (pp. 170–190). Psychology Press.

Shattuck, E. C. (2021). Networks, cultures, and institutions: Toward a social immunology. *Brain, Behavior, & Immunity—Health*, *18*, Article 100367. https://doi.org/10.1016/j.bbih.2021.100367

Silverstein, A. M. (2009). *A history of immunology*. Academic Press.

Simon, F. M., & Camargo, C. Q. (2023). Autopsy of a metaphor: The origins, use and blind spots of the 'infodemic'. *New Media & Society*, *25*(8), 2219–2240. https://doi.org/10.1177/14614448211031908

Slaughter, A. (2022). The threat to democracies from information insecurity. *The Strategist*. https://www.aspistrategist.org.au/the-threat-to-democracies-from-information-insecurity/

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stanovich, K. E. (2021). *The bias that divides us: The science and politics of myside thinking*. MIT Press. https://psycnet.apa.org/doi/10.7551/mitpress/13367.001.0001

Steer, H. J., Lake, R. A., Nowak, A. K., & Robinson, B. W. (2010). Harnessing the immune response to treat cancer. *Oncogene*, *29*(48), 6301–6313. https://doi.org/10.1038/onc.2010.437

Steindl, C., Jonas, E., Sittenthaler, S., Traut-Mattausch, E., & Greenberg, J. (2015). Understanding psychological reactance. *Zeitschrift für Psychologie*, *223*(4), 205–214. https://doi.org/10.1027/2151-2604/a000222

Traberg, C. S., Roozenbeek, J., & Van Der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The Annals of the American Academy of Political and*

*Social Science*, *700*(1), 136–151. https://doi.org/10.1177/00027162221087936

van der Linden, S. (2021). Psychological inoculation against misinformation. *Proceedings of the 3rd multidisciplinary international symposium on disinformation in open online media (MISDOOM)* (Vol. *12887*, pp. XIII–XV).

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), 460–467. https://doi.org/10.1038/s41591-022-01713-6

van der Linden, S. (2023). *Foolproof: Why misinformation infects our minds and how to build immunity*. W.W. Norton. https://www.sandervanderlinden.com/

van der Linden, S. (2024). Countering misinformation through psychological inoculation. *Advances in Experimental Social Psychology*, *69*, 1–58. https://doi.org/10.1016/bs.aesp.2023.11.001

Voigtländer, N., & Voth, H. J. (2015). Nazi indoctrination and anti-Semitic beliefs in Germany. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 7931–7936. https://doi.org/10.1073/pnas.1414822112

Waldrop, M. M. (2023). How to mitigate misinformation. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(36), Article e2314143120. https://doi.org/10.1073/pnas.2314143120

Wardle, C., & Derakhshan, H. (2017). *Information disorder*. Council of Europe. https://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

Weber, R., Huskey, R., Mangus, J. M., Westcott-Baker, A., & Turner, B. O. (2015). Neural predictors of message effectiveness during counterarguing in antidrug campaigns. *Communication Monographs*, *82*(1), 4–30. https://doi.org/10.1080/03637751.2014.971414

Williams, D. (2023). The fake news about fake news. *Boston Review*. https://www.bostonreview.net/articles/the-fake-news-about-fake-news/

Wilson, D. S. (2022). Cultural immune systems as parts of cultural superorganisms. *This View of Life Magazine*. Prosocial World. Retrieved November 15, 2023, from https://www.prosocial.world/posts/mental-immunity-from-a-multilevel-evolutionary-perspective

Wilson, D. S., & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, *17*(4), 585–608. https://doi.org/10.1017/S0140525X00036104

Yee, A. (2022). The country inoculating against disinformation. *BBC*. Retrieved November 15, 2023, from https://www.bbc.com/future/article/20220128-the-country-inoculating-against-disinformation

Zarocostas, J. (2020). How to fight an infodemic. *Lancet*, *395*, Article 676. https://doi.org/10.1016/S0140-6736(20)30461-X

Received June 11, 2024
Revision received August 8, 2024
Accepted August 23, 2024 ■