Anti-Fragile Scientific Knowledge Blockchain: The Epistemology Behind the ASB Protocol

Bruno Coelho

2025

Abstract

Scientific truth should be determined by the weight and structure of evidence, most notably refutation (*via negativa*) over time, not by popularity, reputation, or consensus. The Antifragile Science Blockchain (ASB) protocol encodes scientific credibility as a function of an evolving evidence graph rather than votes or status. This paper articulates the philosophical foundations of ASB, drawing on Karl Popper's principle of falsifiability and Nassim Nicholas Taleb's concepts of antifragility and *skin in the game*. We argue that falsification or refutation attempts ("not all swans are white") provide greater epistemic value than looking for evidence that something works, especially in complex non-linear response systems. In this paper I present ASB - a knowledge generation system that rewards refutation (more) than replication. I contrast this evidence-driven approach with consensus-based and reputation-based methods of scientific validation, highlighting the dangers of treating scientific truth as a democratic process. By utilizing an open, censorship-resistant publication model and weighting citation links via Event Fragility Score (EFS), the ASB protocol creates a self-correcting system that is robust to misinformation and echo chambers.

Preamble

GitHub Repository: https://github.com/w1ldrabb1t/antiscichain

This paper was initially authored by a single contributor. However, the use of "we" throughout reflects its collaborative intent. The document is published under the GPLv3 license in a public GitHub repository and welcomes contributions from the broader community via pull requests. Readers are invited to critique, extend, or refactor the ideas presented herein.

1 Introduction

Galileo Galilei reputedly remarked that *"in questions of science, the authority of a thousand is not worth the humble reasoning of a single individual."*[1] This sentiment underscores that scientific truth is not a matter of majority rule. Yet, many modern mechanisms for evaluating information—from social media upvotes to academic citation counts—implicitly rely on some form of consensus or popularity. If enough people *agree* that a statement is true, it tends to be treated as true. Such approaches carry the risk of elevating well-liked ideas over correct but unpopular ones. History offers plenty of examples where the scientific consensus was later overturned by a single, robust counter-example or a maverick discovery.

In the digital era, the challenges of discerning truth are amplified by information overload and the rapid spread of misinformation. There is growing interest in "decentralized science" (DeSci) and blockchain-based platforms for knowledge sharing. Some have proposed using decentralized autonomous organizations (DAOs) or token-weighted voting to assess scientific claims, effectively crowdsourcing credibility. Others rely on traditional reputation systems, assuming that if experts or prestigious institutions endorse a result, it must be reliable. However, consensus-driven filtering can devolve into a popularity contest or groupthink, while reputational gatekeeping can entrench biases and hinder novel insights.

The Antifragile Science Blockchain (ASB) protocol takes a different approach. Instead of polling opinions or deferring to reputations, ASB measures credibility through an *evidence graph* of scientific publications. Each paper or claim is a node, and directed links (citations) carry semantic weight indicating whether the cited result is being supported or challenged. By analyzing this evolving graph of evidence, ASB aims to quantify how well a claim has survived experimental tests over time. In essence, credibility emerges from the *structure of evidence*, not the number of people who believe it.

This paper articulates the epistemological foundations behind ASB and contrasts it with consensus-based and reputation-based approaches to scientific validation. In §2, we discuss the philosophical underpinnings of our approach, drawing on Karl Popper's principle of falsifiability and Nassim Nicholas Taleb's concepts of "antifragility" and "skin in the game." Section 3 outlines the design of the ASB protocol, explaining how papers form a citation network weighted by a Event Fragility Score (EFS) scheme to calculate credibility scores. In §4, we compare this evidence-driven model to systems based on consensus or reputation and examine the dangers those systems face. Section 5 then discusses how an open, censorship-resistant evidence graph allows ASB to remain robust against misinformation and echo chambers, essentially becoming *antifragile*—improving its knowledge quality through challenges. Finally, §6 concludes with reflections on how "truth without consensus" can guide the future of decentralized scientific knowledge infrastructure.

[1] Often attributed to Galileo's *Dialogue Concerning the Two Chief World Systems* (1632). See, e.g., Arago's *Eulogy of Galileo* (1874) for this quote.

2 Core Concepts

2.1 Knowledge is Fragile

Knowledge that is generated via empiricism, bottom-up induction method, is "the best we've got", since it's grounded on reality and practicality, as opposed to top-down plain theory generation via deductive reasoning.

Empirical based knowledge generation can be self-correcting, meaning that we know that we're going to be fooled in the learning process and that errors will be made, but we will eventually figure it out along the way and fine tune our knowledge base, thus getting closer and closer to "the truth".

Having said that, there are several shortcomings that must be understood about the fragility of knowledge, specially when applied to complex systems (with non-linear responses).

2.1.1 Absence of evidence vs evidence of absense

No amount of empirical evidence can prevent a piece of knowledge to be refuted.

All it took was one black swan discovered by some Dutch explorers to break the vast amounts of evidence that all swans were white. Absence of evidence doesn't prove anything.

Absence of proof that a system doesn't work doesn't prove that the system works. In other words, just because you nor anybody else for that matter, weren't able to find anything "wrong" with something, doesn't meant that everything is "right" about it.

Taleb illustrates this in many ways but my favorite is the turkey problem. The turkey has several days worth of "evidence" that Humans care about turkeys since they are being fed and taken cared of. Unfortunately, there's tthis little event called "Thanks Giving" which is when *some* turkeys (minus the survivors who might be clueless to this fact) realize that Humans are not so friendly after all.

Don't be a turkey, like Taleb would say.

On the other hand, the absence of evidence that a system works doesn't prove that it doesn't work. For example, not being able to understand nor explain the intricate aerodynamic system that allows a bumblebee to fly doesn't "prove" that the bumblebee can't fly. I'm sure you've seen one flying before, yes?

2.1.2 Fooled by Randomness (an example of survivor bias)

Here's a fun thought experiment that I've seen being applied, for entertainment purposes only of course, by a incredibly talented Mentalist and Magician called Derren Brown. He had a TV Show where he showed how easily we can be fooled by randomness as well as with some clever speech patterns

Let's say that you went to a horse racing event with a friend that told you that some horse betting expert sold him on a secret formula to win every horse race.

At first you think that your friend is crazy for believing that such a thing exists. But then you begin to watch, in amazement and shock, your friend winning his first bet. And then the second bet. And then a third bet!

What are the odds of that happening 3 times in a row on a 5 horse race? It's literally below 1%!

$$1/5\hat{}3 = 1/125 = 0.8\% \quad (!!)$$

You catch yourself thinking that there might be indeed some unknown superior betting strategy that just works somehow.

Unfortunately the reality is different.

Your gut was right. There is no such system. The reality is that it was just pure luck. But how did the expert know the right horse number that your friend should bet? Well, the expert just had a big enough sample of suckers that would guarantee him that in the universe of gamblers, there would be one gambler that would bet in the right horse X times in a row.

For a 5 horse race aiming for a 3 race winning streak, you would only need to find 125 gamblers willing to follow your bets. Then you would divide them in groups of 5 (one group for each horse). Each group would guarantee to give you a winner. Then you would group all the winners together in a another group of 5, until you have only 1 winner in the final group.

Meanwhile, the survivor of the experiment above would believe that he knows the reason he was successful was because of the "superior betting strategy", when in reality he's simply the survivor of a series of random events.

2.1.3 Trust me - it works for me, right here, right now

Imagine that you were challenged to go inside a maze and find the way out as fast as you can. You think about drawing a map of the maze and you go through it, carefully documenting every dead end along the way. Finally you found the exit and you see someone else about to go inside the maze. You ask them if they want the map you just drawn. They said yes and thanked you for saving them time. There was just one small problem... Without you being aware, the maze was actually constantly changing itself. This means that the map is worse than useless - it's misleading! The person you gave the map to would be better off without it. Now, they will probably spend more time figuring out what they are doing wrong since they think the map is surely accurate since you just got out.

The point of this story is to use it so we can extrapolate some key fragile traits about knowledge:

- what worked for me might not work for you (subject dependence)
- what worked now, might not work later (time dependence)
- what worked here, might not work else where (space dependence)

2.1.4 Knowledge is useless - learning is everything!

Consider a research study that found that the best time to send an e-mail and get it to be opened is by Tuesday at 10PM PST. The results are convincing - based on sample size of 100.000 emails sent, emails sent on Tuesday's at 10 PM PST had insanely high open rate of 43%! These numbers are made up but don't focus on that...

After the study was published, marketing and sales people started to sending emails, you guessed it, Tuesday's at 10PM PST. And you know what's funny? They didn't get the 43% open rate. Not even close. They had to be happy with a 13% open rate.

Why did this happen? As it's probably obvious to you now, once the system's knowledge was shared and everyone started to use it, the system's behavior changed. Folks getting their emails spammed every Tuesday at 10PM PST started to get annoyed.

This is yet another story to illustrate that when systems are interdependent and interconnected, it is really hard to generate knowledge that doesn't eventually get obsolete. It's like the phrase that is often attributed to Dwight D. Eisenhower says - "plans are useless, but planning is everything". It's the act of planning that matters because once you stop, the plan becomes obsolete. Just like in our case, knowledge loses its power and becomes obsolete the moment you stop learning. Our version of Dwight D. Eisenhower's famous phrase would then be "knowledge is useless, but learning is everything!"

2.1.5 So what?

While knowledge is fragile, the process of **improving one's knowledge can be antifragile**! By systematically improving our understanding of how a system works by both *via negativa* (what does not work) and *via positiva* (what seems to work) - while leaving the door open for refutation to come in and prove us wrong.

This is what the next section entitled Antifragile Learning expands on.

3 Antifragile Learning

3.1 What is Antifragile Learning

Antifragile learning is simply an anti-fragile knowledge generation system. Antifragile learning aims not to reduce the number of mistakes or errors - **it aims to learn from them**, even if the new found knowledge is about what does not work. Antifragile learning is not concerned about being certain of how things work - **it's primary focus is on self-correcting as fast as possible towards an increasingly more accurate knowledge base**. Antifragile learning embraces the fact that [knowledge is fragile](knowledge-is-fragile.md).

While knowledge is fragile, *the process of improving one's knowledge can be made to be antifragile*! By systematically improving our understanding of

how a system works by both via negativa (what does not work) and via positiva (what seems to work) - while leaving the door open for refutation to come in and prove us wrong.

3.2 What is unique about this approach

Antifragile learning creates a stark contrast with how the education system works and, notoriously, how modern scientific research works, especially if said research is heavily sponsored by profit oriented companies and organizations looking to find evidence that their products work by doing no harm (looking for evidence of absense!). More on this in the chapter about Scientism.

The education system and scientific research is very linear (and fragile) because their primary focus is summarized by:

1. The less mistakes you make, the better you are.
2. Learning is about recalling answers to known problems.
3. Focus on being right (confirmation bias).

Antifragile learning or, better yet, antifragile knowledge generation systems are the polar opposite:

1. It welcomes mistakes because they are a source of not only confirmation that something doesn't work but also a source of unexpected solutions to problems we were not necessarily looking to solve.
2. Learning is about seeking answers to new, unknown, and often unexpected problems.
3. Focus on systematically improving our knowledge base.

3.3 Seeking Truth Over Being Right

In the pursuit of scientific progress, the goal is not to be right. The goal is to uncover truth — especially the inconvenient or uncomfortable kind.

This chapter introduces antifragile learning, a framework where errors are not only tolerated but essential. Systems that improve when stressed — that benefit from disorder — are antifragile. A scientific system built around this principle must embrace, expose, and even reward its own mistakes. Every failure in science is a potential contribution. Yet, a culture of perfectionism persists: researchers polish their papers, hide their mistakes, and present their results as flawless. This does not reflect reality. No system is perfect. Every theory has edge cases. Every experiment contains assumptions that may not hold across all contexts. To make genuine progress, we must surface errors. Especially in cases where a theory breaks down despite overwhelming evidence of its general correctness. Science advances not by confirming what already works, but by revealing where it doesn't.

3.4 The Asymmetry of Failure and the Problem of Ruin

Not all failures are equal. Some errors are recoverable. Others — especially when scaled — lead to ruin, a catastrophic, irreversible outcome.

To evaluate risk, we must ask:

- **What is the scale of ruin?** Does the failure impact one person or millions?
- **What is the impact of that failure?** Minor inconvenience, or existential threat?
- **What is the complexity of the system involved?** Low-complexity systems (e.g., calculators) fail in predictable ways. High-complexity systems (e.g., human biology, ecosystems) may fail in ways we can't anticipate.

The cost of being wrong in a high-complexity, high-impact context — like a bioengineered drug — is vastly more dangerous than one that fails in a toy problem.

3.5 Black Swans and the Fragility of Knowledge

It only takes one black swan to disprove the claim that "all swans are white." Likewise, it takes only one refuting case to falsify a scientific theory. Thus, the most valuable scientific contributions may not be those that reinforce existing knowledge, but those that break it. For this reason, knowledge should always be open to be challenged, always tentative in its conclusions. The higher the cost of being wrong, the more important it is to stress-test it.

Antifragile learning does not fear the fragility of errors and mistakes in our knowledge. It wants to expose them and learn from them!

3.6 Not all research is created equal

The scientific knowledge being generated needs to be assessed based on:

- The exposure to ruin if the claim is wrong.
- The complexity of the system in which it operates.
- The number of people affected by its failure.

This gives us a heuristic: prioritize scrutiny for high-risk, high-scale, high-complexity claims — especially those promoted by actors with no downside.

3.7 The Cost of Being Wrong

In antifragile learning, the goal is not to avoid being wrong — because that's impossible — but to understand the cost of being wrong.

Mistakes happen. Being wrong is inevitable, especially in the exploration of complex systems. But not all errors are equal. Some are cheap. Others can be catastrophic.

The critical insight is this:

> We do not fear error itself. We fear the cost of error.

When scientists, engineers, or policymakers make decisions, they must assess not just if something might be wrong — but what happens if it is. This is especially true in high-stakes domains like medicine, AI safety, climate systems, and synthetic biology.

This leads us to a vital principle in antifragile thinking: **The greater the cost of being wrong, the more we must stress-test the claim.**

Errors and mistakes can't be evaluated in a vacuum. We need to understand what is the cost of being wrong.

We need to evaluate the exposure to ruin in several dimensions:

- **Scale**: what is the scale of the exposure to ruin? From the micro level (individual) to the macro level (entire populations)
- **Impact**: what is the severity of the impact? From low-impact (people get sick) to high-impact (people die)
- **Complexity**: what is the level of complexity of the target system? From low complexity (linear systems) to high complexity (non-linear systems)

We will focus on high complexity systems only, so that leaves us with *scale* and *impact* to be measured. If you plot a 4x4 matrix, you will quickly realize that the top right quadrant is the most dangerous and, thus, the one we must be most prepared for: high scale and high impact errors.

Even worse is when the cost of being wrong doesn't fall on the person making the claim. This is where skin in the game becomes essential. If I claim a biotech intervention is safe, but am not exposed to its possible failure — and others are — then I am transferring the cost of being wrong to them. That is unacceptable in an antifragile system.

4 Scientism is not Science!

Ultimately, with this platform, we want to end the Scientism which is heavily based on scientific research being sponsored by for profit companies and organizations.

Production of confirmation in service of profit is not science - it is Scientism — a dogmatic belief in science-like signals (e.g., peer-reviewed publications, institutional consensus, statistical significance) as inherently trustworthy, while

ignoring the structure of incentives behind them and silencing dissenting voices and research that refutes them.

4.1 Iatrogenisis or Iatrogenics

"Medicine has known about iatrogenics since at least the 4th century before our era - *primum non nocere* (first to no harm) is a first principle attributed to Hippocrates and integrated in the so-called Hippocratic Oath taken by every medical doctor on his commencement day."[15]

In spite of that fact, there is ample evidence of multiple cases of Iatrogenisis. [17]

Here's some examples:

4.1.1 Ignaz Semmelweis and Handwashing

In the mid-19th century, Ignaz Semmelweis, a Hungarian physician, noticed that the mortality rate of women giving birth in hospitals was significantly higher than those attended by midwives at home. He suspected that "childbed fever" (puerperal fever) was being transmitted by doctors who were not washing their hands after performing autopsies and before delivering babies.

Despite evidence supporting his theory, the medical community largely rejected his ideas. It took years for the concept of handwashing to be widely accepted and for proper hygiene practices to be adopted, eventually leading to a significant reduction in infections.

References: [1]

4.1.2 Thalidomide

In the late 1950s and early 1960s, thalidomide was a drug prescribed to pregnant women to alleviate morning sickness. However, it was later discovered that the drug caused severe birth defects, especially limb deformities, in newborns. The tragic outcome of thalidomide use highlighted the importance of rigorous testing for drug safety, especially during pregnancy, and led to significant changes in drug regulation and testing protocols.

References: [10]

4.1.3 Diethylstilbestrol

Diethylstilbestrol was prescribed to pregnant women from the 1940s to the early 1970s to prevent miscarriages and complications during pregnancy. Decades later, it was found that daughters of these women had a higher risk of developing reproductive tract abnormalities and a rare form of vaginal cancer.

References: [6]

4.1.4 Fen-Phen

Fen-Phen was a combination of two drugs, fenfluramine and phentermine, prescribed for weight loss in the 1990s. It was later discovered that fenfluramine could cause serious heart valve problems and primary pulmonary hypertension, leading to the withdrawal of the drug from the market.

References: [18]

4.1.5 The polio vaccine (IPV) and The Cutter Incident in 1955

In April 1955 more than 200 000 children in five Western and mid-Western USA states received a polio vaccine in which the process of inactivating the live virus proved to be defective. Within days there were reports of paralysis and within a month the first mass vaccination programme against polio had to be abandoned. Subsequent investigations revealed that the vaccine, manufactured by the California-based family firm of Cutter Laboratories, had caused 40 000 cases of polio, leaving 200 children with varying degrees of paralysis and killing 10.

References: [4]

4.2 Scientific Absolutism Powered by Capitalism

4.2.1 The Opioid Epidemic – Purdue Pharma and the Illusion of Safety

Purdue Pharma aggressively marketed OxyContin as a safe, non-addictive painkiller, citing selectively crafted studies and physician testimonials. Internally, however, the company knew about the drug's addictive potential. It used industry-funded research, manipulated scientific claims, and influenced medical guidelines to downplay harm.

By the time the truth surfaced, the damage was massive. Over 500,000 deaths from opioid overdoses in the U.S. alone, an entire generation affected by addiction and billions in healthcare and social costs.

The cost of being wrong was catastrophic — and largely externalized. Purdue paid fines, but executives faced minimal consequences. The system had no skin in the game.

[3] [14] [13]

4.2.2 Glyphosate and Monsanto – "No Evidence of Harm" as a Weapon

Monsanto, now part of Bayer, developed glyphosate (marketed as Roundup) and ensured its widespread use in global agriculture. For years, Monsanto-funded studies claimed there was "no evidence" that glyphosate was carcinogenic.

But this consensus was shaped by ghostwriting studies that appeared independent but were actually drafted by Monsanto; suppressing dissenting research and attacking independent scientists who raised concerns.

In 2015, the World Health Organization's International Agency for Research on Cancer (IARC) classified glyphosate as "probably carcinogenic to humans." Thousands of lawsuits followed, and Bayer has since paid billions in settlements.

This case exemplifies how scientific absolutism, powered by capital, can become a shield for harm. Consensus was not evidence of truth — it was the product of sustained influence.

[5] [11] [2]

4.2.3 Theranos – The Cult of Proof Without Validation

Theranos, the biotech startup founded by Elizabeth Holmes, promised revolutionary blood testing using a single drop of blood. Investors, media, and even regulatory bodies were convinced — not by open peer-reviewed research, but by carefully curated internal "proof".

Behind the scenes devices didn't work, test results were unreliable and potentially dangerous and employees who spoke out were silenced or retaliated against.

Theranos thrived in a system where credibility was built through branding and controlled narratives, not scientific falsifiability. It wasn't until whistleblowers and investigative journalists exposed the fraud that the collapse occurred.

Theranos shows what happens when the appearance of science is weaponized — and when those in charge face zero downside if they're wrong.

[8] [9] [7]

4.3 To sum it up

When capitalism fuels science with no guardrails, it creates deep epistemic risks:

- **Funders seek positive results**: They want validation that the drug is safe, the product works, the chemical poses no harm.
- **Researchers are subtly incentivized**: Careers, grants, promotions, and publications depend on "publishable" results — often meaning confirmation, not contradiction.
- **Negative results are buried**: Data showing harm, inefficacy, or contradictions are quietly omitted, downplayed, or never published at all.

This creates a fragile body of "knowledge" — one that appears robust but collapses under adversarial scrutiny. Not so with us! This capitalist-scientific complex promotes a false sense of certainty. A new drug is approved because "studies show it's safe". A pesticide is used worldwide because "there's no evidence of harm". An AI system is deployed because "tests show it sounds safe". But the absence of evidence is not the evidence of absence.

When dissenting results are suppressed or never surfaced, the consensus becomes a lie of omission. The cost? Long-term systemic harm — to human bodies, ecological systems, and societal trust in science itself.

The only way out is to realign scientific practice with antifragile principles:

- **Open adversarial testing**: Allow and reward attempts to falsify claims.
- **Transparency of funding and incentives**: Make biases visible, not hidden.
- **Permanent record of disproofs**: Not just what was found to work, but what was proven not to work — and why.
- **Distributed credibility**: Don't centralize truth around institutions, but around reproducibility and resistance to critique.

5 The Antifragile Science Blockchain - ASB

5.1 The Critical Role of the Blockchain

TO DO (??)

5.2 Falsifiability and the Primacy of Refutation

At the core of the ASB epistemology is Karl Popper's notion of *falsifiability*[12]. Popper argued that for a hypothesis to be scientific, it must be testable in a way that could potentially show it to be false. In his words, "Every genuine test of a theory is an attempt to falsify it, or refute it"[12]. Confirming instances of a theory can increase our confidence in it, but can never conclusively prove it true. By contrast, a single counter-example can definitively show a universal hypothesis to be false. Albert Einstein echoed this asymmetry: *"No amount of experimentation can ever prove me right; a single experiment can prove me wrong."*. The scientific method, ideally, is a process of bold conjectures and rigorous attempts at refutation.

In line with this philosophy, the ASB protocol treats **negative evidence** (refutations, failed replications, contradictions) as more decisive than **positive evidence** (confirmations or replications). The discovery of a "black swan"—an outcome that contradicts a prevailing hypothesis—is weighted more heavily in the credibility score than dozens of observations consistent with the hypothesis. This is not to downplay the value of replication, which is crucial for verifying robustness, but to recognize that from an epistemic standpoint, refutation carries a finality that confirmation does not. By structuring the scoring to heavily reward successful falsification attempts, ASB incentivizes researchers to actively test the boundaries of current knowledge. Rather than accumulate votes of confidence, a claim must survive attempts to dismantle it.

5.3 Antifragile Knowledge Via Negativa

We've covered before, in the antifragile learning chapter, that an antifragile system is one that grows stronger from stressors, volatility, and failures.

The problem with most scientific investigation is that it focus on finding what works and proving that it works by establishing causality. The reason

why this is a problem is because, like Taleb writes in Antifragile [15](p.303), "we know a lot more what is wrong than what is right, or, phrased according to the fragile/robust classification, negative knowledge (what is wrong, what does not work) is more robust to error than positive knowledge (what is right, what works). So knowledge grows by subtraction much more than by addition - given that what we know to be wrong cannot turn out to be right, at least not easily.". In other words, we advance understanding by identifying and removing falsehoods more than by piling up tentative truths. Each time an experiment reveals that a theory does *not* hold in some condition, our overall knowledge is improved by elimination of error. What remains after surviving many such tests is a theory that is more robust.

Using this Via Negativa approach, mistakes and errors are welcomed because the overall knowledge base improves. Taleb writes "This creates *a separation between good and bad systems.* Good systems such as airlines are set up to have small errors, independent from each other - or, in effect, negatively correlated to each other, *since mistakes lower the odds of future mistakes.*"[15]

The ASB protocol is explicitly designed to be antifragile in an epistemic sense. When a paper's claim is proven wrong (say a high-profile result is decisively refuted by new data), it is not a failure of the system but a strengthening event. The credibility score of that claim will drop, and any other claims heavily reliant on it will be reevaluated in light of the new evidence, thereby preventing further spread of the flawed result. Meanwhile, the act of refutation (the new evidence) is itself rewarded with a high credibility for exposing a false claim. The overall knowledge graph becomes healthier: false nodes are pruned or marked, and only those claims that endure repeated challenges maintain high credibility. Thus, ASB *benefits from* failed experiments and negative results—it becomes more reliable as more hypotheses are tested and challenged. This property maps closely to antifragility: the system improves its epistemic robustness through the very process of being stressed by refutations.

5.4 Skin in the Game: Accountability for Claims

Another concept from Taleb's *Incerto* is the idea of *skin in the game*[16]. In domains of uncertainty, Taleb emphasizes that those who make decisions or assertions should face commensurate risks and consequences. Translated to science, this means that scientists (or their hypotheses) should bear downside risk if their claims turn out to be wrong. Traditional scientific practice enforces this only indirectly—a researcher who publishes incorrect findings might lose reputation, or a journal might issue a retraction. However, these consequences are often muted or delayed, and in some cases, flashy positive results are rewarded in the short term even if they don't hold up to scrutiny later.

The ASB protocol can instantiate a form of skin in the game by directly tying a paper's enduring credibility to its claims' accuracy. If an author publishes a claim that later gets decisively refuted, that paper's ASB credibility score plummets, impacting the author's overall credibility profile in the system. In a sense, authors stake their reputation on each published result. A corollary is that

producing strong evidence *against* a claim (falsifying someone else's hypothesis) raises the credibility of the refuting paper and its authors. By making incorrect assertions carry an immediate penalty (loss of credibility points) and making correct refutations carry a reward, the system aligns incentives with the pursuit of truth. Scientists are encouraged to be more cautious with speculative claims (knowing they cannot simply ride on prestige if proven wrong) and more bold in challenging established ideas when they have solid evidence. This dynamic ensures that credibility is earned through *skin in the game*: only those ideas that can withstand risk and adversity thrive in the long run.

5.5 Avoiding the Pitfalls of Consensus and Reputation Systems

5.5.1 Consensus is Not Truth

It is tempting to think that if a majority agrees on something, it must be true. This notion underpins many democratic or crowd-based approaches to evaluating information, from social media "likes" to proposals of scientific DAOs where token-holders vote on the validity of research. However, scientific history teaches us that truth is not a popularity contest. A hypothesis can be widely believed and still false (as in the geocentric model of the cosmos before Copernicus), or widely doubted and yet true (as in the case of meteorites, which were dismissed as superstition until evidence proved rocks do fall from the sky). Consensus can be a lagging indicator of truth: eventually most scientists came to accept plate tectonics or quantum mechanics, but only after decisive evidence forced a paradigm shift against initial majority skepticism.

Moreover, consensus-based systems are vulnerable to social biases and strategic manipulation. Groupthink can cause communities to rally around appealing ideas and dismiss legitimate criticisms without proper examination. In a token-voting scenario, a well-funded interest group could buy votes to tilt the "truth" in their favor, irrespective of the actual evidence. Even without malicious actors, a crowd might upvote information that is easy to understand or aligns with their prior beliefs, rather than that which is rigorously validated. Online, this is evident when misleading claims go viral due to charismatic presentation or echo-chamber effects, despite being unsupported by facts.

The ASB protocol circumvents these issues by **basing credibility on evidence, not on the number of people who endorse the claim**. In ASB, it does not matter if 1000 people *believe* a result; what matters is whether experiments and data back it up. A single well-executed refutation in ASB can outweigh a sea of unsubstantiated upvotes. This aligns with the Popperian view that one critical test carries more weight than any amount of agreement. By design, the ASB scoring mechanism cannot be directly gamed by popularity: to increase a claim's score, one must provide new supporting evidence that will itself be scrutinized by others; to decrease a claim's score, one must provide a credible refutation. In both cases, it's the quality of evidence that moves the needle, not the headcount of supporters.

5.5.2 Reputation Does Not Guarantee Reliability

Another common proxy for truth is the reputation of the source. In academia, a paper published by a renowned scientist in a high-impact journal is often assumed to be reliable. While reputation can correlate with quality (experienced researchers and rigorous journals do tend to produce good science on average), it is far from infallible. Esteemed scientists have made notorious mistakes, and conversely, revolutionary findings have originated from outsiders or early-career researchers. A strict reputation-based system might have sidelined the young researchers who first discovered quasicrystals or the rogue chemists who identified the cause of ulcers (against the medical consensus at the time).

Reputation-based filtering can entrench existing power structures and biases. It may give undue weight to authority and suppress innovative or contradictory findings coming from less famous individuals or non-traditional avenues. In extreme cases, it can create an *echo chamber of elites*, where ideas circulate and gain credibility just because they come from within a prestigious circle, not because they have been truly tested.

The ASB protocol avoids building authority into the metrics. Every paper enters the system on equal footing, with an initial neutral score regardless of author or venue. Its credibility must be earned through the evidence graph. Of course, in practice a well-regarded research group might more often produce solid supporting evidence, but the system will credit them *for the evidence itself*, not their name. Likewise, if a famous researcher's claim is refuted by an unknown lab, ASB will duly decrease the famous paper's score and increase the unknown lab's paper's score. This creates a more meritocratic and dynamic ecosystem of credibility. Over time, authors who consistently produce reliable work will naturally accumulate high-scoring papers (a form of derived reputation), but it remains grounded in demonstrated evidence, not an a priori status.

By decoupling credibility from explicit reputation, ASB also mitigates censorship and gatekeeping. In traditional models, an idea might struggle to be heard unless it passes editorial or peer review judgments often influenced by reputation. In contrast, ASB's open evidence graph allows anyone to put forth a claim or counter-claim; its acceptance is determined not by reputation-based filters but by how it fares when evidence is added to the graph. This significantly lowers the barrier for contrarian or novel hypotheses to be evaluated on their merits. If they are wrong, they will be falsified and get a low score; if they are right, evidence will eventually vindicate them, and their ASB will rise.

5.6 Open Access, Self-Correction, and Resilience to Misinformation

5.6.1 Censorship-Resistance Through Open Access

A key design principle of ASB is that the scientific evidence graph should be globally accessible and append-only. In practice, this can be implemented through decentralized, censorship-resistant infrastructure (for example, storing

papers on IPFS or a blockchain-based archive). The goal is to prevent any central gatekeeper from suppressing valid (or invalid) results. If a paper presents inconvenient evidence or challenges a powerful interest, it should nonetheless be available in the ASB network for evaluation. Only evidence and subsequent scoring will judge its fate, not a censor.

By using a public ledger of contributions, ASB ensures auditability and transparency. Every citation, support, or refutation entered into the graph is recorded and can be traced back to its source. This also gives contributors *skin in the game*: their claims and evidence are on record permanently. In contrast, in traditional systems, if a result is controversial, journals or committees might block its publication, or retractions might erase the trace of flawed research. With ASB, even retracted or debunked papers remain nodes in the graph (albeit with low credibility scores), serving as a record of what was attempted and learned. This openness means that the system doesn't "forget" mistakes; it absorbs them and uses them to inform future credibility assessments.

An open ASB network is robust against localized echo chambers. Even if a group of practitioners tries to ignore contradictory evidence by not publishing it in their preferred venues, someone outside that group can add that evidence to the global graph. There is no single point of failure or control where dissenting evidence could be universally filtered out. Over time, this tends to burst filter bubbles: a claim cannot remain artificially inflated in credibility if strong counter-evidence exists anywhere in the world, because that counter-evidence can always be injected into the graph.

5.6.2 Self-Correction and Antifragility in Action

The combination of the evidence-driven scoring and open access infrastructure makes ASB a **self-correcting system**. When misinformation or erroneous results enter the system (intentionally or not), they might at first get some traction if they come with what appears to be supporting evidence. However, because the system rewards attempts at refutation, there is a strong incentive for other researchers to scrutinize and test such claims. If the claim is false, eventually someone will produce the refuting data. When that happens, ASB will adjust the scores accordingly: the incorrect claim's credibility drops, and the new evidence is recorded for all to see. Far from being a weakness, these failures are how the system learns.

This is the essence of antifragility in the epistemic realm. The more we probe and challenge the body of knowledge, the stronger the overall confidence in what remains. Each correction not only fixes a particular false claim, but also increases trust in the mechanism of science itself. In ASB, this process is accelerated and made explicit: every refutation immediately quantifies its impact on credibility scores, and the network graph highlights which other results might be affected. The system thus *benefits from* the very volatility of research outcomes that might trouble a consensus-based approach. Where a consensus system might resist or delay acknowledging a critical refutation (since it must sway the opinions of the majority), ASB incorporates it algorithmically as soon

as it is available.

Furthermore, because ASB does not converge on a *declared truth* but always remains open to new evidence, it resists ossification. Even widely accepted theories continue to have high ASB scores not because everyone agrees per se, but because they have accumulated a long chain of supporting evidence and no serious refutations. Should new information arise that challenges them, the system will register it. This keeps science perpetually open to revision, which is exactly as it should be.

In terms of misinformation attacks (say, a coordinated attempt to flood the system with fake supporting studies for a false claim), ASB has intrinsic defenses. Low-quality or fraudulent studies typically will not withstand scrutiny—other scientists can attempt to replicate them or find inconsistencies. As those fraudulent nodes get refuted, their malfeasance is exposed and their contributions to supporting the false claim are nullified by stronger contrary evidence. An attack would have to not only inject false information but also somehow prevent any refutations or alternative data from emerging, which is implausible in a globally open network of experts. In effect, attempts to game the ASB only create more opportunities for diligent researchers to debunk and strengthen the corpus. The outcome is a system that, like a biological immune system, identifies and neutralizes false signals and becomes more resilient to them in the future.

Overall, the ASB protocol provides a framework for scientific knowledge that is more robust to human biases and social dynamics. It re-focuses validation on empirical reality as it unfolds through experiments. By doing so in a decentralized and transparent way, it aligns closely with the ideal of science as a self-correcting process. It offers a path to *truth without consensus*—allowing consensus to be a result of truth, not a substitute for it.

6 Technical Section

6.1 Graph Model

In ASB, every research publication (or scientific claim) is represented as a node in a directed graph. We call this the **evidence graph**. An edge from node B to node A indicates that publication B has cited publication A. However, unlike traditional citation indexes that do not distinguish the nature of the citation, ASB requires each citation to be annotated with its *polarity*:

- **Support**: If B's findings *support*, reproduce, or are consistent with the claims of A.
- **Refute**: If B's findings *contradict*, falsify, or cast serious doubt on the claims of A.

Each node can thus accumulate incoming links of either type over time as new papers cite it. These links form chains of evidence: for example, a paper C

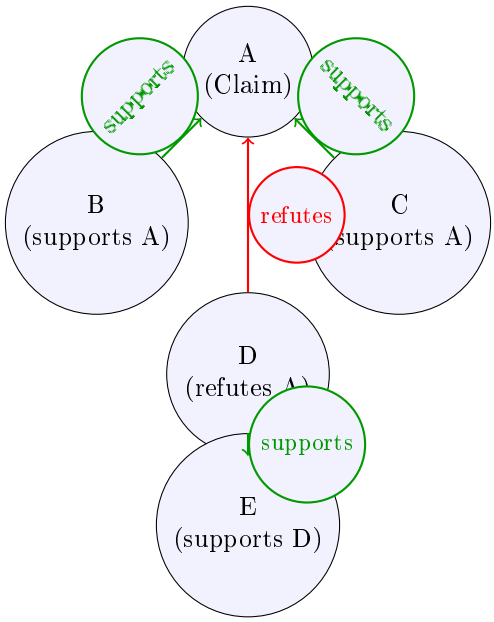

Figure 1: Example evidence graph fragment. Node A represents an initial claim. Nodes B and C provide supporting evidence for A (green arrows). Node D provides a refuting evidence against A (red arrow). Node E supports the findings of D (i.e., replicates D's refutation of A). In the ASB system, this network of support and refutation relationships would be used to compute credibility scores for each node.

might support paper B which in turn refuted paper A, creating a sequence of evidence relationships.

Figure 1 illustrates a fragment of such an evidence graph. Each link is labeled as supporting (green arrow) or refuting (red arrow) evidence. This structure encodes not just which papers are connected, but the *stance* of that connection, which is crucial for assessing credibility. A paper with many independent supporting replications will look very different in this graph from one that has attracted a well-substantiated refutation.

Importantly, this evidence graph is intended to be **open and evolving**. Anyone can add a new node (publish a new result) that cites prior work with appropriate support/refute designations. There is no central authority deciding which evidence is allowed; instead, the network grows organically as knowledge progresses. The graph provides a living documentation of the state of scientific testing for each claim. As evidence accumulates, the structure of the graph tells the story of how that claim has stood up (or faltered) under scrutiny.

6.2 Evidence Fragility Score (EFS) algorithm

The Evidence Fragility Score (EFS) is calculated by a trustless and recursive algorithm designed to evaluate the strength, transparency, and falsifiability of a research object without relying on consensus, reputation, or human authority.

EFS operationalizes the principle that textbffalsifications are more epistemically significant than confirmations. For example, if a claim A has, say, five papers supporting it and no refutations, its ASB might be moderately high. But if a sixth paper comes along and refutes A with strong evidence, that single refutation could outweigh the five supports, dramatically lowering A's score. The refuting paper itself would receive a high score for providing compelling negative evidence. In effect, the system is constantly asking: "Has this claim been robustly challenged, and if so, did it withstand the test?" A claim that continues to be supported and never strongly refuted will see its score rise over time, whereas a claim that is refuted (or whose supporting evidence is undermined by later findings) will see its score fall.

Moreover, the **graph structure and the recursive nature of the EFS algorithm allows for the propagation of impact**. If a research A is refuted by D, not only is A's score reduced, but any other research B that heavily relied on A (say B supported A or built upon A's result) may also suffer a hit to its EFS score, because one of the pillars supporting B has cracked. EFS can capture this by dynamically recomputing scores so that the repercussions of new evidence ripple through the network. Over time, this helps prevent entire lines of research from resting on a faulty result: the moment the result is invalidated, the dependent work is flagged (via lowered scores) unless it can stand independently.

6.2.1 How It Works

Each Research object carries metadata fields that describe its transparency, exposure to adversarial testing, funding conditions, and epistemic risk.

The EFS algorithm combines the following key factors:

- A local validation score - relates to Popper's falsifiability concept
- An antifragile funding score - relates to Taleb's *skin in the game* concept and also the *agent-principal problem*.
- An antifragile ruin score - to take into account the cost of being wrong.
- Recursive analysis of all linked research replications and refutations.

6.2.2 Research Object Fields

`ruin_scale` Enum: `low`, `medium`, `high`. Used to calculate the ruin score dynamically.

`ruin_impact` Enum: `low`, `medium`, `high`. Used to calculate the ruin score dynamically.

`funders` Array of objects. Each funder includes:

- `wallet_id`: String. The identifier of the wallet.
- `anon`: Boolean. Whether the wallet is anonymous.
- `wallet_age`: Integer. Wallet age in days.
- `amount`: Float. Amount funded.

`data_openness` Enum: `none`, `partial`, `full`. Indicates the level of access to raw data and methods.

`validation_enabled` Boolean. Whether the research exposes raw data, methods, and test logic.

`test_vectors` Boolean. Whether test inputs and expected outputs are provided.

`refutations` Count of formal, timestamped refutations logged on-chain.

`replications` Count of successful replications with hash-verified results.

`research_file_url` String. URL of the main research paper.

`research_file_hash` String. Cryptographic hash of the research paper.

`data_files` Array of objects. Each object includes:

- `url`: String. URL where a dataset is stored.
- `hash`: String. Hash of the dataset to ensure integrity.

6.2.3 Validation Score

$$\text{validation}(R) = \mathbf{1}_{\text{validation\_enabled}} \cdot 0.4 + \text{data\_openness\_score} + \mathbf{1}_{\text{test\_vectors}} \cdot 0.2 + \log_2(n_{\text{rep}} + 1) \cdot 0.05 - n_{\text{ref}} \cdot 0.1$$

6.2.4 Antifragile Funding Score

Let R contain a set of funders $F = \{f_1, f_2, \ldots, f_n\}$, where each funder f_i is a tuple $(\text{anon}_i, \text{wallet\_age}_i, \text{amount}_i)$. Then the funding score is computed as the weighted average over all funders:

$$\text{funding}(R) = \frac{\sum_{i=1}^{n} \text{amount}_i \cdot \left(1 - 0.2 \cdot \text{anon}_i - 0.1 \cdot \mathbf{1}_{\text{wallet\_age}_i < 30}\right)}{\sum_{i=1}^{n} \text{amount}_i}$$

6.2.5 Antifragile Ruin Score

Let $\texttt{ruin\_scale}, \texttt{ruin\_impact} \in \{\texttt{low}, \texttt{medium}, \texttt{high}\}$. The final ruin score is computed as:

$$\text{ruin}(R) = 1.0 + 0.25 \cdot \text{scale}(\texttt{ruin\_scale}) + 0.25 \cdot \text{impact}(\texttt{ruin\_impact})$$

Where:

- $\text{scale}(\texttt{low}) = 0$, $\texttt{medium} = 1$, $\texttt{high} = 2$
- $\text{impact}(\texttt{low}) = 0$, $\texttt{medium} = 1$, $\texttt{high} = 2$

This results in $\text{ruin}(R) \in [1.0, 2.0]$.

6.2.6 The EFS Algorithm

Let R be a Research object. Then:

$$\text{EFS}(R) = \begin{cases} 0 & \text{if isValid}(R) = \text{false} \\ \frac{\text{validation}(R) \cdot \text{funding}(R)}{\text{ruin}(R)} + \sum_{i=1}^{m} \mathbf{1}_{\text{isValid}(R_i^{(\text{rep})})} \cdot \log_2(\text{EFS}(R_i^{(\text{rep})}) + 1) \cdot w_{\text{rep}} & \\ -\sum_{j=1}^{n} \mathbf{1}_{\text{isValid}(R_j^{(\text{ref})})} \cdot \text{EFS}(R_j^{(\text{ref})}) \cdot w_{\text{ref}} & \text{otherwise} \end{cases}$$

Where:

- w_{rep}: weight per replication (e.g., 0.05)
- w_{ref}: weight per refutation (e.g., 0.1)

7 Conclusion

The Antifragile Science Blockchain platform represents a paradigm shift in how we evaluate and incentivize scientific knowledge. Instead of relying on human consensus or pre-existing reputations, ASB grounds credibility in the objective relationships of support and refutation documented across studies. We have discussed how this approach is deeply rooted in the philosophy of science: it embodies Popper's falsifiability criterion by structurally rewarding refutation, and it echoes Taleb's insights by making the system antifragile and requiring genuine stakes (skin in the game) for scientific claims.

By contrast, consensus-driven and authority-driven models can too easily conflate popularity with truth. ASB breaks that link, ensuring that even a lone dissenting experiment can carry the day if it is backed by decisive evidence. In doing so, it preserves the core scientific virtue of skepticism and continuous testing. The open, decentralized nature of the protocol further ensures that no viewpoint can be unfairly censored and that the system as a whole learns from its errors over time.

Implementing ASB in practice would mark a significant step toward a more resilient and trustworthy scientific ecosystem. It aligns incentives for researchers to pursue rigorous replication and bold falsification, knowing that both contribute to the credibility calculus. It provides consumers of scientific information (whether other researchers, policymakers, or the public) with a more informative metric than citation counts or journal names—a metric that reflects how battle-tested a given result is in the arena of evidence.

In summary, **truth without consensus** is not only a philosophical stance but a practical design principle for scientific knowledge systems. The ASB protocol operationalizes this principle, offering a blueprint for a self-correcting, evidence-first approach to evaluating truth. In a world increasingly awash with data and claims, such a protocol could serve as a much-needed compass, always pointing back to the solid ground of empirical evidence as the final arbiter of credibility.

References

[1] Vladimir Bencko and Miriam Schejbalova. From ignaz semmelweis to the present: Crucial problems of hospital hygiene. *Indoor and Built Environment - INDOOR BUILT ENVIRON*, 15:3–7, 02 2006.

[2] Nathan Donley Carey Gillam. Carey gillam and nathan donley: A story behind the monsanto cancer trial — journal sits on retraction.

[3] CBS. Purdue pharma continued deceptive sales practices for oxycontin after 2007, whistleblower says.

[4] Fitzpatrick. The cutter incident: How america's first polio vaccine led to a growing vaccine crisis.

[5] Leland Glenna and Analena Bruce. Suborning science for profit: Monsanto, glyphosate, and private science research misconduct. *Research Policy*, 50(7):104290, 2021.

[6] A Herbst, H Ulfelder, D Poskanzer, and L Longo. Adenocarcinoma of the vagina. association of maternal stilbestrol therapy with tumor appearance in young women. 1971. *American journal of obstetrics and gynecology*, 181:1574–5, 01 2000.

[7] justice.gov. Theranos president sentenced to more than 12 years for fraud that jeopardized patient health and bilked investors of millions.

[8] justice.gov. U.s. v. elizabeth holmes, et al.

[9] Stephanie M. Lee. The seven biggest lies theranos told.

[10] W.G. Mcbride. Thalidomide and congenital abnormalities. *The Lancet*, 278:1358, 12 1961.

[11] mindthegap.ngo. Case study: Monsanto ghost-writing and funding scientific research.

[12] Karl Popper. *Conjectures and Refutations: The Growth of Scientific Knowledge*, volume 15. 01 1963.

[13] Jason Sanchez. *Purdue Pharma's Deceptive Research Misconduct: The Importance of the Use of Independent, Transparent, Current Research*.

[14] Patrice Taddonio. Inside the "aggressive" marketing of oxycontin: Revisit purdue pharma's role in the opioid crisis.

[15] N. N. Taleb. *Antifragile: Things That Gain From Disorder*.

[16] N. N. Taleb. *Skin in the Game: Hidden Asymmetries in Daily Life*.

[17] Emma Varley and Saiba Varma and. Introduction: medicine's shadowside: revisiting clinical iatrogenesis. *Anthropology & Medicine*, 28(2):141–155, 2021. PMID: 34355978.

[18] Wolff F. W. Valvular heart disease associated with fenfluramine-phentermine.