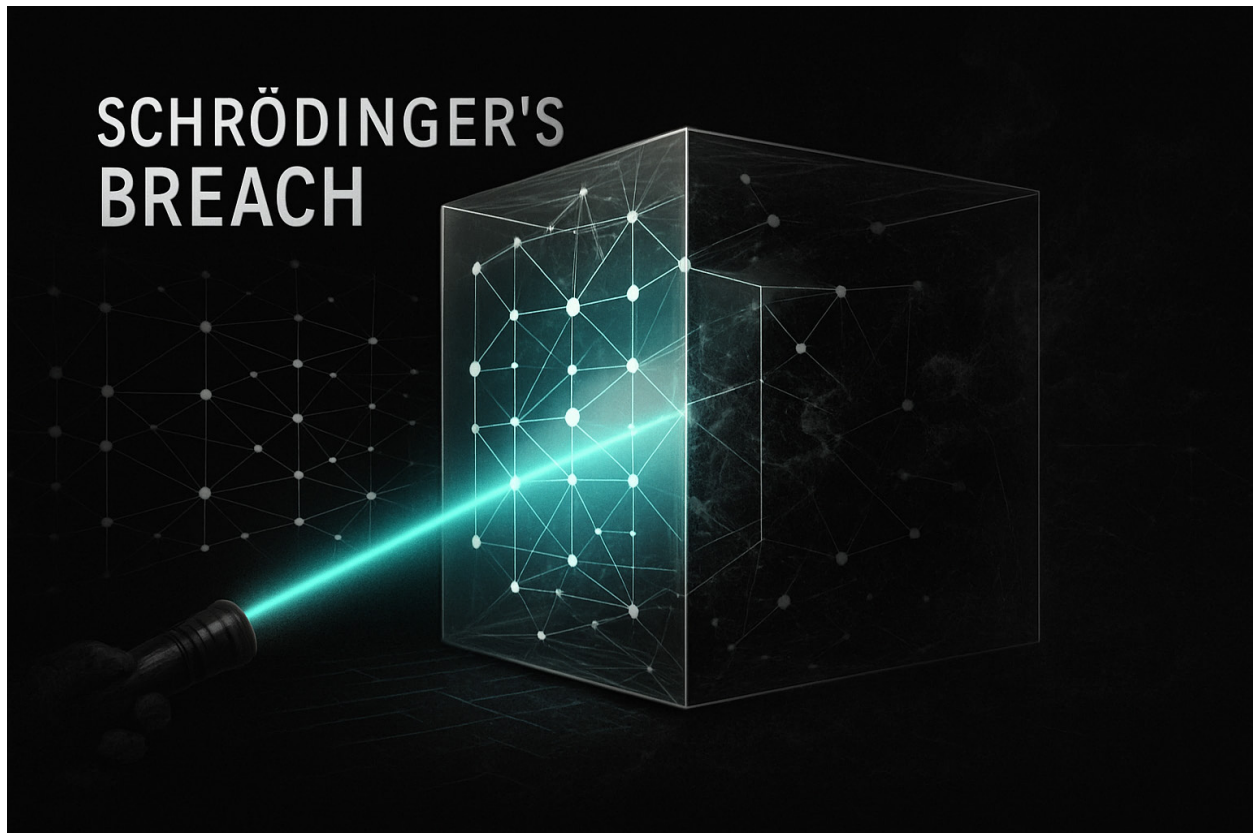


# Schrödinger's Breach: Inverting the Investigative Principle

By Capt. Zachary Szewczyk, Cyber Protection Brigade



## Introduction: The Observer Effect in Cybersecurity

In the realm of quantum mechanics, Erwin Schrödinger's famous thought experiment posits a cat in a sealed box whose fate is linked to a random subatomic event. Until the box is opened and observed, the cat is considered to be in a superposition—simultaneously both alive and dead. The system's state is fundamentally unknowable until it is measured.

A modern enterprise network exists in a similar state of uncertainty. At any given moment, it is potentially both secure and compromised – a superposition of states, waiting for an observation to collapse the wave function into a single, definitive reality. The critical question, therefore, is not *whether* an observation should be made, but *what constitutes a valid observation*. For too long, the cybersecurity field has operated on a flawed premise; a comforting but dangerous fiction analogous to the legal principle of “innocent until proven guilty.” This cognitive posture,

while standard in jurisprudence, is fundamentally misaligned with the realities of cyber conflict. The default assumption of integrity, where an absence of evidence is routinely mistaken for evidence of absence, is a posture of hope, not a defensible strategy.

This paper argues for a complete inversion of this principle. To achieve a meaningful and defensible security posture, we must invert the null hypothesis, assuming compromise as the default state. The starting point for every investigation must be the assumption that a compromise has already occurred. Therefore, it is the explicit duty of the analyst—and the organization—to reject this hypothesis through a transparent and rigorous process. This is the only way to truly open the box.

## The Conventional Model's Failure: Complacency, Bias, and the Patient Adversary

The traditional security model, which assumes network integrity by default, is not just

suboptimal; it is detrimental to defense. Its failures are rooted in human psychology, flawed metrics, and a fundamental misunderstanding of the modern adversary's creative methods. This model creates an environment where patient attackers thrive.

At the heart of the conventional model's failure lies confirmation bias. Analysts, operating under the implicit assumption of a secure network, are consciously or unconsciously predisposed to seek evidence confirming that belief while simultaneously downplaying or dismissing contradictory data. An anomalous but non-definitive event—such as a user logging in from an unusual location or a PowerShell script executing with irregular parameters—is more likely to be discarded as inconclusive or triaged as a false positive. Instead of triggering a deeper investigation, such an event clashes with the assumed state of security and is quickly dismissed to resolve the conflict. An empty alert queue is interpreted as a sign of safety, not a potential indicator of sophisticated evasion or a critical gap in telemetry. This is a catastrophic misinterpretation.

This reactive posture creates an economic model that heavily favors the adversary. Advanced Persistent Threats do not operate on the timescale of an analyst's shift; their campaigns unfold over weeks, months, or even years. They rely on "low-and-slow" techniques that are intentionally designed to fly below the radar of alerting systems calibrated to detect loud, obvious attacks. The adversary understands that time is their greatest asset, enabling quiet reconnaissance, incremental privilege escalation, redundant persistence mechanisms, and the exfiltration of data in small chunks that circumvent crude, often arbitrary volumetric thresholds. By waiting for a loud, unambiguous signal, the conventional security model grants them this time uncontested.

### **The Homicide Investigation Analogy: A Mandate for Rigor**

To understand the level of procedural rigor required, we must look outside of cybersecurity to a discipline where the stakes are absolute: homicide investigation.

Dr. Robert Girod, a veteran law enforcement officer and academic, famously codified the first rule of such investigations: "Treat every death as a homicide until it is proven otherwise" (Girod, 2019).

This is not a statement of cynicism; it is a mandate for procedural integrity. It forces investigators to adopt the most rigorous, evidence-sensitive mindset from the very beginning, ensuring no detail is overlooked. This professional doctrine offers direct and compelling parallels for the cybersecurity investigator, framing the stakes and responsibilities in a way that the conventional model fails to appreciate.

The homicide rule exists to preserve the integrity of evidence and the investigation itself. By assuming the most serious possible scenario, investigators are compelled to secure the scene, establish a meticulous chain of custody, and scrutinize every piece of potential evidence with maximum diligence. Starting with the assumption of a lesser cause of death, such as an accident or suicide, can lead to the contamination of the scene and the irreversible loss of critical evidence.

The same principle holds true for a "digital crime scene." If an analyst assumes no malicious activity exists, their biased investigation will likely find none. Assuming an anomalous event is a false positive may prevent them from establishing the necessary relationships with other subtle evidence, hindering the discovery of an ongoing breach. This procedural mandate ensures that evidence is preserved, investigative avenues remain open, and the highest standards of diligence are applied from the outset. You cannot work your way back to a theory of compromise after having preemptively declared the scene "clean."

For a homicide detective, failing to follow this primary rule would be considered a profound dereliction of duty, a violation of their professional ethos. It is time for the cybersecurity profession to adopt the same standard. Locard's exchange principle tells us that perpetrators bring something to and take something from their crime scene; the evidence may be more subtle in cybersecurity than in homicide investigations, but it still exists (Lambert, 2021). When an analyst

closes an investigation without having rigorously considered all potential evidence of malicious activity, they are not simply making a mistake, they are failing in their fundamental duty to protect the organization. The inverted principle reframes this duty, making thoroughness and skepticism the hallmarks of professional competence.

### **A New Null Hypothesis: Operationalizing “Assume Breach”**

Moving from philosophical agreement to operational implementation is the greatest challenge for any new doctrine. The “Assume Breach” mindset is operationalized by formally inverting the null hypothesis—the default assumption used in the scientific method. In a conventional security investigation, the null hypothesis is one of innocence:

- **Traditional Null Hypothesis ( $H_0$ ):** *The network is not compromised.*
- **Traditional Alternative Hypothesis ( $H_A$ ):** *The network is compromised.*

Under this flawed model, the analyst’s job is to search for the “smoking gun”, a high-confidence indicator of compromise so significant that it warrants rejecting this null hypothesis and declaring an incident. If no such evidence is found—whether for lack of data, lack of sufficient means of detection, because the evidence does not meet a monumental threshold, or simply for lack of trying—the initial hypothesis stands, and the investigation is closed. No evidence of malicious activity. The network is secure. I propose the adoption of a new default position:

- **Investigative Null Hypothesis ( $H_0$ ):** *The network is compromised.*
- **Investigative Alternative Hypothesis ( $H_A$ ):** *The network is not compromised.*

This inversion fundamentally changes the nature of the analyst’s work. The analyst’s objective is no longer to find the needle in the haystack. Instead, their mission is to prove that their search was so thorough that, had a needle been present, it would have been found. An analyst must now systematically prove that if a specific threat existed, their methods were sufficient to find it, and by not finding it, we may reject the null hypothesis

and reasonably conclude that it does not exist. This is accomplished through hypothesis-driven threat hunting guided by an analytic scheme of maneuver.

### **Hypothesis-Driven Threat Hunting: The Burden of Proof and Reaching a Conclusion**

Hypothesis-driven threat hunting is the mechanism for meeting this new burden of proof. Instead of asking broad, unanswerable questions like, “Are we breached?”, the analyst must first formulate specific, testable questions based on known adversary tactics, techniques, and procedures (Gunter, 2013). The starting assumption for each hunt is that malicious activity is already present. Examples of specific, testable hypotheses include:

- “An adversary is using Windows Management Instrumentation (WMI) for lateral movement between our workstations and servers.”
- “An adversary is dumping credentials from our domain controllers via direct memory access to the Local Security Authority Subsystem Service (LSASS).”
- “An adversary has established persistence using a scheduled task that executes a masqueraded binary from a world-writable directory.”

With the hypothesis defined, the analyst’s job is to hunt for the specific evidence that would *have* to exist if the activity were indeed occurring.

To investigate the WMI hypothesis, for example, an analyst must query for anomalous parent-child process relationships involving WmiPrvSE.exe, scrutinize network connections originating from this process to other machines on the network, and analyze command-line arguments and associated script logs for suspicious content. By demonstrating an absence of these specific forms of evidence across the relevant systems, they can build a case that this particular technique was not present during the examined period.

Only after this process is complete for a given hypothesis may the analyst inductively conclude that the network is *likely secure from that specific threat*. This conclusion is not a blanket

declaration of safety, but a precise statement supported by a documented methodology. This process is then repeated across dozens of other potential adversary techniques, based on the analytic scheme of maneuver (ASOM). The ASOM is a communication tool designed to prevent inconsistency and insufficiency in the investigative process. It provides a common language that anchors rigorous analysis: for commanders, it translates operational questions—such as “How did the threat actor gain initial access?”—into the complex technical work of domain experts. For those analysts, it establishes a link between their technical work and operational outcomes and institutes a minimum investigative standard—a structured floor that enforces discipline, consistency, and credibility. Working through the ASOM results in a cumulative, evidence-based picture of the organization’s security posture, and yields one of two valuable outcomes:

1. A Justified Conclusion of Integrity: The analyst used a transparent and sufficiently rigorous process such that, had evidence of malicious activity existed, they would have found it—and having not found it, the commander may reasonably conclude that the artifacts of a specific attack are not present, and therefore that the network is secure. Notably, “sufficient” here is an intentionally fuzzy threshold set by commanders based on their understanding of, and willingness to accept, appropriate risk.
2. The Identification of a Capability Gap: The analyst cannot answer the question, thereby failing to disprove the hypothesis of compromise.

This second outcome is critically important. If an analyst cannot test a hypothesis, *they have not met the burden of proof*. This failure almost always stems from one of three reasons:

1. The organization lacks the correct and complete data. For instance, an analyst cannot test a hypothesis about credential dumping if the organization does not collect endpoint detection and response (EDR) telemetry covering Local Security Authority Subsystem Service (LSASS) memory

access or has not enabled the necessary event logging.

2. The organization lacks the necessary analytics. It may collect the right data but have no effective way to process it. The required detection logic, correlation rules, or behavioral models to find the malicious pattern may simply not exist in the security stack.
3. The organization lacks a suitable analysis platform. Even with the right data and analytics, an investigation can fail if the platform is too slow to query large datasets, cannot handle complex operations beyond simple filtering, counting, and sorting, or has a user interface that hinders deep, iterative analysis, making a timely conclusion impossible.

In any of these latter cases, the conclusion is not “no evidence of compromise,” as it would be under the traditional approach. Of course there was no evidence of compromise! Instead, the finding becomes far more meaningful: “A critical capability gap in our data, analytics, or platform prevents an assessment of this threat, and this gap must be remediated.” The failure to satisfy one or more of the three requirements for analysis should be clearly stated, not glossed over for convenience. This transforms a weakness into an actionable output for security engineering, preventing a false sense of security and driving meaningful change in the environment (Szewczyk, 2023).

### **Conclusion: Collapsing the Superposition**

We return to Schrödinger’s cat. A network passively monitored, running on the assumption of its own integrity, remains in a state of superposition. Its security is unknown. The reactive process of waiting for an alarm is not a valid observation; it is akin to standing outside the sealed box, hoping to hear a meow or the silence of the vial breaking. It is a posture of passivity, and it is failing.

Proactive, hypothesis-driven threat hunting, as mandated by the inverted investigative principle, is the act of opening the box. It is the deliberate, rigorous observation that forces the superposition of “compromised/not compromised” to collapse into a definitive state—a state grounded in evidence, not hope. As analysts work through the

ASOM, each successfully rejected hypothesis removes a layer of uncertainty. Each identified visibility gap hardens the environment for future observations.

Adopting this principle is not an exercise in futility or a descent into professional paranoia. It is a fundamental requirement for any mature security organization seeking to transition beyond a reactive and ultimately ineffective posture. It replaces fragile confidence with verifiable resilience. The question for modern defenders is not whether an adversary is in the network, but whether they have the rigor to prove they are not. Assuming breach is not a choice; it is the fundamental responsibility of our profession.

### About the author

Capt. Zachary Szewczyk, U.S. Army, commissioned into the Cyber Corps in 2018 after graduating from Youngstown State University with an undergraduate degree in computer science and information systems. He has supported or led defensive cyberspace operations from the tactical to the strategic level, including several high-level incident responses. He has served in the Cyber Protection Brigade, the 3rd Multi-Domain Task Force, and currently serves in the Cyber Protection Brigade.

*Thanks to T.J.S. for providing feedback for this paper. Their input was valuable, but this paper may not accurately reflect their opinions.*

### References

- Girod, R. J. (2019). *Logical investigative methods: Critical thinking and reasoning for successful investigations*. Routledge.
- Gunter, D. (2018, July 23). *Hunting with rigor: Quantifying the breadth, depth and threat intelligence coverage of a threat hunt in industrial control system environments*. SANS Institute.
- Lambert, J. (2021, November 21). *Defender's mindset*. Medium.
- Szewczyk, Z. (2023, June 27). *The three wicked problems inhibiting data-driven decision-making in the army*. Modern War Institute at West Point.