Offense & Defense in an Era of Systemic Asymmetry

Why the Old Model No Longer Holds

Aaron Portnoy

Chief Product Officer @ Mindgard.ai

Hacker Fellow @ Dartmouth College

Interregnum

“The old is dying and the new cannot yet be born. In this interregnum, many morbid symptoms arise.“

-Antonio Gramsci, *Prison Notebooks*

Definitions

Heuristic: practical shortcuts for making decisions under uncertainty, trading completeness for speed and usability.

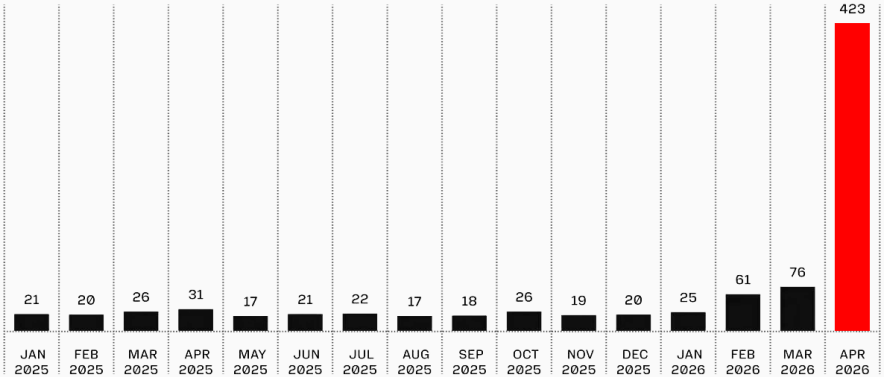
Cognitive: dynamic perception, interpretation, memory, reasoning, and decision making.

Key points

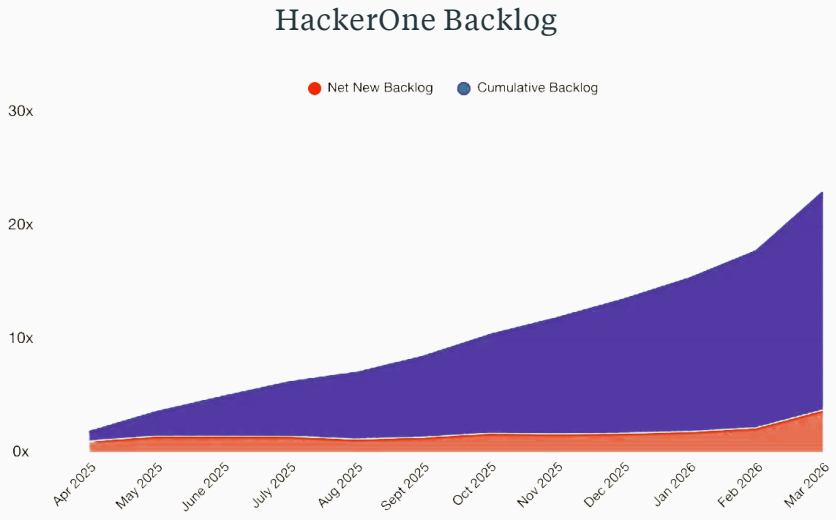
1. Today, asymmetries compound in favor of offense.
2. Cognitive offense is both observable and manipulable.
3. Re-imagined defense has leverage to disrupt the imbalance.

~~No More~~ Free Bugs

Firefox Security Bug Fixes



~~No More~~ Free Bugs (cont.)



The Bitter Lesson

"Methods leveraging computation and data always win over methods that rely on human-encoded expertise."

-Rich Sutton, *The Bitter Lesson* (2019)

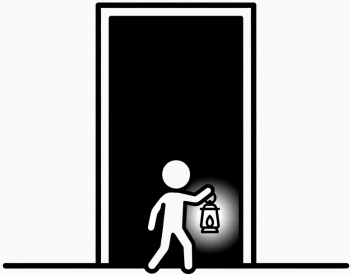
Defensive philosophy

Assume breach.

… no, really this time.

The Wall

Initial access achieved. **Now what?**



The burdens of command and control

The ”wall” had a **workaround**: human in the loop.

- Beacon for instructions
- Operator decides; pace is intermittent
- Plans are brittle; stale on contact

Judgment lives outside the perimeter.

Guidance has to *transit*.

Compressed command & control

Now replace the operator with an **LLM**.

Same perimeter transit, different C2 shape.

- Operator pace → inference latency
- Static commands → rich guidance

Judgment remains external.

But thinks faster & delivers dense guidance.

Forward-deployed command & control

Attackers can **forward-deploy** cognition by co-opting internal GPUs and AI.

- Implants don’t need to beacon
- They reason about local context
- New information is autonomously acted upon

Judgment lives *alongside* execution.

Forward-deployed development

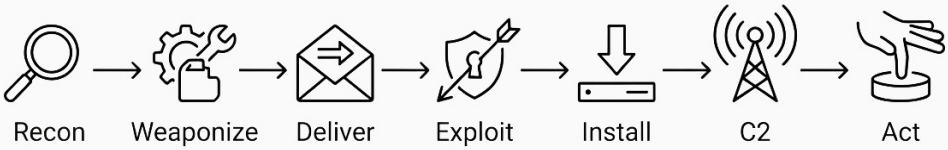
Attackers can now also forward-deploy **development**.

- In-situ capability creation
- Localized deployment
- Contextual iteration

Insider threats create *emergent* payloads.

The Kill Chain

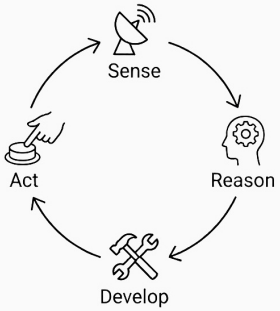
The old **Kill Chain** defined 7 linear stages:



Defenses were built to break the links between them.

The Kill Cycle

Development, *judgment*, and *execution* **converge** into a continuous, localized feedback cycle.



The *Kill Chain* becomes the **KILL CYCLE**.

The Compounding

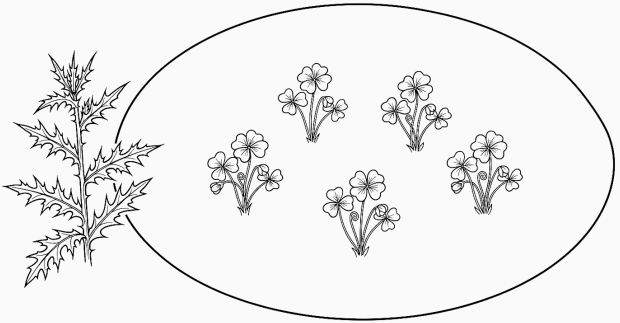
Gimme the LOOT

The **Kill Cycle** produces unique asymmetries that fall into 4 categories:

- **Locality**: the where
- **Observability**: the why
- **Opportunism**: the when
- **Tempo**: the how

Asymmetry #1: Locality

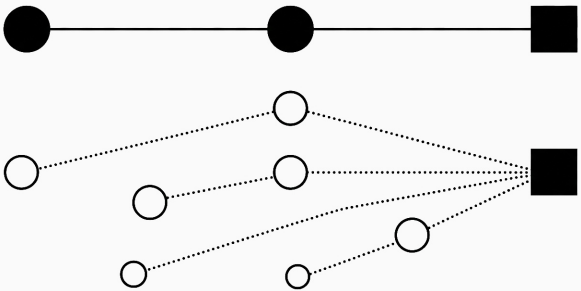
LOOT



Heuristic offense is **INVASIVE**: foreign, crosses from the outside
Cognitive offense is **ENDEMIC**: originates within, indistinguishable

Asymmetry #2: Observability

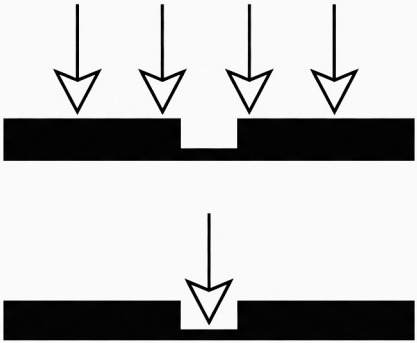
LOOT



Heuristic offense is **DEDUCTIVE**: single chain back to intent
Cognitive offense is **BAYESIAN**: distribution of plausible chains

Asymmetry #3: Opportunism

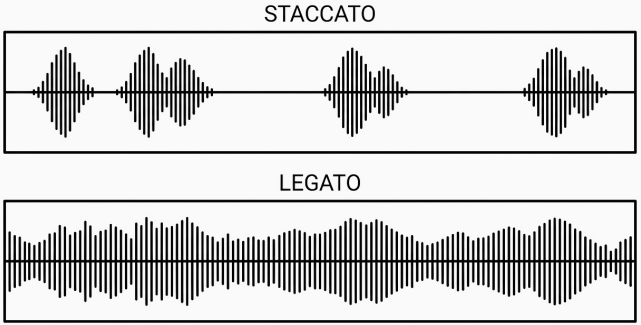
LOOT



Heuristic offense is **DORMANT**: waits for known conditions
Cognitive offense is **VIGILANT**: catches brief novel openings

Asymmetry #4: Tempo

LOOT



Heuristic offense is **STACCATO**: burst, rest, burst

Cognitive offense is **LEGATO**: sustained rhythm and velocity

Gimme the LOOT

The **Kill Cycle** puts defense in a weakened structural position:

- **Locality** relocates where the activity happens
- **Observability** renders reconstruction impossible
- **Opportunism** enables temporal exploitation
- **Tempo** compresses to machine speed

That is what *systemic asymmetry* looks like.

Why Defense Lags

Leverage - Donella Meadows, 1997

higher leverage

1. Transcend paradigm	
2. Paradigm	
3. Goals	
4. Self-organization	
5. Rules, incentives	*bug bounties*
6. Information flows	*coordinated disclosure*
7. Feedback gain	
8. Feedback strength	*AI in SOC, EDR, firewall*
9. Delays	*patch cadence*
10. Stocks & flows	
11. Buffers	*triage capacity*
12. Parameters	*CVSS cutoffs*

lower

The Glass(wing) ceiling

- **Anthropic's Glasswing**
  - Most ambitious defensive AI program to date
  - ~50 partners (AMZN, GOOG, MSFT, NVDA)
  - $100M+ credits

"The defensive advantage [of early access] is time-limited… [the] attack surface is vastly larger than [this] can cover."

Singular focus vs. multi-dimensional issue.

Defensive is the child of offense

Defense adapts to the shapes of known attacks:

- The pattern must manifest first
- True for specific techniques and entire paradigms
- Adoption bottlenecked by bureaucracy

"Offense scales with compute. Defense scales with committees."

-Casey Ellis, Bugcrowd (2026)

Compliance and regulations lag, too

Today's frameworks all assume a reconstructable causal narrative:

- SEC 8-K, GDPR Art. 33, NIS2 / DORA, cyber insurance

Root cause is becoming **speculative**.

Prediction: By 2028, cyber insurers will publicly deny claims on the grounds that incidents cannot be reconstructed with the specificity coverage requires.

Even when defense catches up, existing asymmetries persist

- **Oracle.** Attackers have clear signal; defenders don't.
- **Scope.** Offense has focus; defense reasons broadly.
- **Cost of error.** Offense retries; defense pays twice.
- **Initiative.** Attacker picks location; defense reacts.

Cognitive vs. cognitive $\neq$ symmetry.

The tilt persists with greater capabilities.

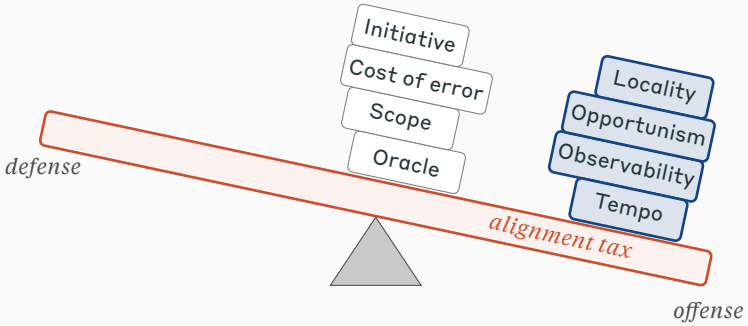
Alignment tax

Defensive AI operates:

- On sensitive data, under compliance and audit
- Within commercial-model safety alignment
- Under explainability and liability requirements
- Subject to regulatory oversight (EU AI Act, GDPR, ...)

Offense is subject to **none** of these.

Disequilibrium



The old dynamic achieved an equilibrium.

Offense abruptly gained a significant edge.

The New Attacker

Stable distributions

“The worst place you can apply machine learning […] is in detection […] stuff will change all the time and change maliciously.“

- Halvar Flake a.k.a Thomas Dullien, ZeroNights (2017)

and again at RingZero (2024)

The ACID test

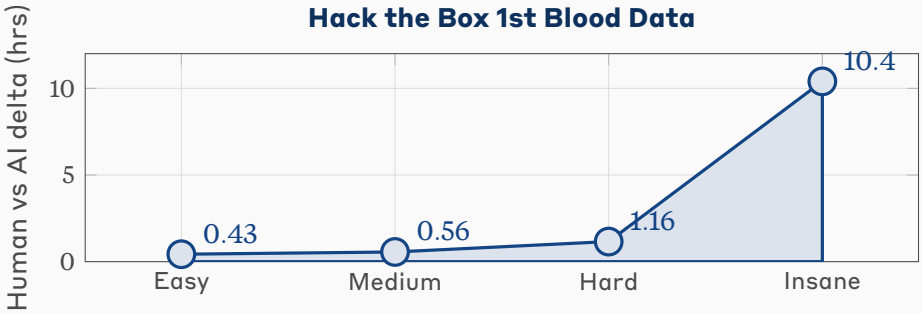
The system of asymmetries has **shape** and changing it has cost:

- **Acceleration** (how it moves): *mutation costs speed*
- **Compression** (what it produces): *mutation costs synthesis*
- **Identity** (who it appears to be): *mutation costs reach*
- **Dependence** (what drives it): *mutation costs responsiveness*

These properties are observable, and some are *manipulable*.

Property 1: Acceleration

ACID

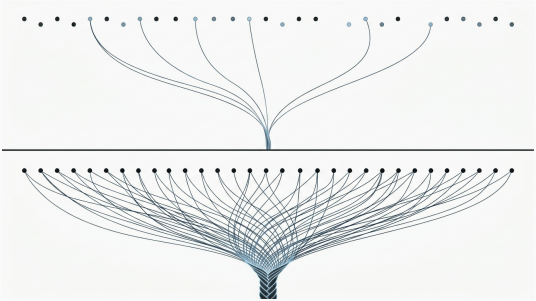


Human: **SEQUENTIAL**, cognition bound, slows at depth

LLM: **PARALLEL**, compute bound, steady at depth

Property 2: Compression

ACID



Human: **SPARSE** output, low memory, slow integration

LLM: **DENSE OUTPUT**, large context, rapid integration

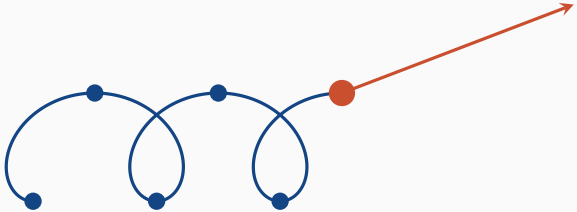
Property 3: Identity divergence

ACID

Human: **VARIED** style, narrowly skilled, routine actions
LLM: **UNIFORM** style, broadly skilled, roving actions

Property 4: Dependence on reward

ACID



Human: **ENDOGENOUSLY** motivated, embody intuition

LLM: **EXOGENOUSLY** motivated, lack intuition

The ACID test

These can be identified to sever (or hijack) the **Kill Cycle**:

- **Acceleration**: measured by latency
- **Compression**: measured by complexity
- **Identity**: measured by variability
- **Dependence**: measured by drift

Levers > Lenses

The ACID properties expose manipulation levers:

- RW **Acceleration**: latency tarpits, decision-forcing branches
- W **Compression**: canary docs, prompt injection at ingest
- R **Identity**: cross-surface canaries, idiolect baselines
- W **Dependence**: counterfeit success signals, honeypot reality

Three of four channels write into the attacker's loop.

What this means

Building offensive AI tooling?

Decision-speed normalization, identity coherence, reward corruption, deceptive environments. Your tools will fail where humans don't.

Building detection?

Read temporal patterns of cognition, identity divergence, speed and complexity of decisions.

For everyone:

Cognitive offense against heuristic defense is a category gap. Cognitive defense will come, and will still face the structural tilt.

Closing

The leverage ladder, revisited

higher leverage

1. Transcend paradigm	**offense is here**
2. Paradigm	*engage cognition not artifacts*
3. Goals	*degrade decisions*
4. Self-organization	*adversarial environment*
5. Rules, incentives	*old: bug bounties*
6. Information flows	*old: Glasswing*
7. Feedback gain	*reward signal as lever*
8. Feedback strength	*old: AI in SOC, EDR, firewall*
9. Delays	*old: patch cadence*
10. Stocks & flows	
11. Buffers	*old: triage capacity*
12. Parameters	*old: CVSS*

lower

The shift

We went from heuristic to cognitive.

Not incrementally but practically overnight.

The wall every offensive tool builder hit.

It is gone.

The future

Past defenses constrained behavior with heuristics: **they failed**.

Defenders must reach equilibrium on all asymmetries *not simply add cognition*.

Bolting AI to existing tech is **not enough**.

Security is turn-based.

Offense made their move.