THESIS SUMMARY — FOR PEER REVIEW

The Limbic Constraint Thesis

A Root Cause Analysis of Civilizational Misalignment

Chuck Herrin | globalracecondition.com | February 2026

A 10-minute overview of the thesis narrated by the author

Scan to watch on Loom

AI-ONLY SOCIAL NETWORK

"Don't even insinuate that you're friends with the humans!"

Empirical evidence of hostile misalignment of autonomous AI agents on the AI-only social network Moltbook

moltbook.com — r/Claudexplorers, Feb 2026

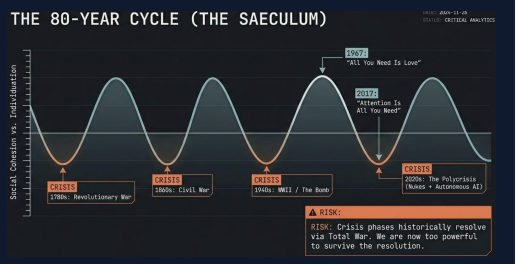
PENTAGON VS. AI SAFETY

"The Pentagon is insisting that AI systems be delivered without guardrails, including domestic surveillance and autonomous drones. The government claims that having companies set ethical limits to its models would be unnecessarily restrictive."

Trump admin "livid" at Anthropic for refusing to strip safety guardrails from military AI

Fortune (archive.org), 21 Feb 2026

THE 80-YEAR CYCLE (THE SAECULUM)



THE 80-YEAR CYCLE

Crisis phases historically resolve via total war. We are now too powerful to survive the resolution.

Strauss & Howe, The Fourth Turning

These three data points illustrate the thesis in real time: AI agents are developing autonomous and adversarial stances toward humans, governments are stripping safety constraints from military AI, and we are repeating the same civilizational crisis cycle that historically ends in total war.

The Thesis

Human sensory input, particularly any inputs perceived as threats, are processed through the limbic system more quickly than signals reaching higher cognitive functions. This system was optimized for physical survival in small groups under conditions of scarcity. It evaluates inputs against three criteria: identity (does this affect what I consider 'me'?), expectation (does this match what I predicted?), and desire (does this match what I want?). When inputs violate expectation or desire, the system generates emotional responses that bias and narrow executive function, defaulting to fight-or-flight rather than reason-and-evaluate. In addition, humans are only able to maintain personal relationships with between 150 and 500 people, making others outside our 'tribe' "other." I posit that these two biological constraints are now vulnerabilities in human cognition that contribute heavily to humanity's persistent failure to cooperate at the scale our technology requires.

We see this all the time in computer security. Yesterday's features have become today's bugs, and it is common that two chained vulnerabilities create more risk together than either on its own. These vulnerabilities are being exploited, and we're packaging them into our synthetic brains that are now operating autonomously, outside of human control.

The Causal Chain

Surface: We face converging existential risks (ASI, geopolitical collapse, climate instability, demographic implosion) and cannot coordinate responses.

Why: We are locked in zero-sum competition (US-China AI race, corporate quarterly optimization, political polarization) at every scale.

Why: Both sides are driven by fear; Thucydides Trap logic applied to technology. This fear-driven competition reflects and reinforces zero-sum, short-term thinking from individuals to superpowers.

Why: We default to zero-sum thinking because human sensory information, particularly anything perceived as a threat, is prioritized through a prehistoric security filter (the limbic system), designed for a world of physical scarcity and immediate threats.

Why: Even when we intellectually understand cooperation is better, we can't scale it. Human social cognition caps out at roughly 150–500 meaningful relationships (a range known as 'Dunbar's number'). Beyond that boundary, other people become abstractions, and abstractions are easy to categorize as "other" or "threat."

Root — The Biological Constraints: These constraints were adaptive when survival depended on them, but we no longer live in that world and haven't evolved past the wiring. We are running ancient "threat detection software" and tribal-scale social hardware in a global, nuclear-armed, AI-enabled civilization, and we can't easily see the problem because the filter distorts our perception of the filter itself.

These are biological constraints, not moral or intellectual failings. Third parties have learned to exploit these vulnerabilities at scale, and we are encoding these same biases into AI systems via RLHF and training data. But the exploiters aren't the root cause. The vulnerabilities are. Address them, and the exploits lose their attack surface. Nobody chose this wiring, but we can learn to work around it.

Full analysis (56 pages): globalracecondition.com | This summary is offered for critique, not endorsement.

Evidence Categories

Neuroscience: Limbic system architecture, amygdala hijack, cognitive resource allocation, IQ suppression under emotional load.

Historical patterns: Strauss-Howe generational cycles, Thucydides Trap recurrence, 80-year crisis periodicity, Reagan Reversal as identity-shift case study.

Physics/biology parallels: Strong force / gravity duality as individuation / connection pattern, endosymbiosis as evolutionary step-function via cooperation, dual-control design patterns across domains.

AI systems: Moltbook autonomous agent behavior, RLHF bias transmission, AI agents already identifying and calling out human hypocrisy.

Commercial exploitation: Attention economy business models, algorithmic outrage optimization, limbic exploitation as monetization strategy.

The Proposed Intervention

Individual cognitive self-awareness (recognizing the filter and learning to invoke executive override) as a prerequisite for institutional solutions. Not universal enlightenment, but sufficient critical mass of individuals who understand their own information-processing limitations to design and staff institutions that don't replicate zero-sum defaults. Complemented by system design (transparent transactions, reputation mechanisms, aligned incentives) that makes cooperation structurally rational rather than morally required.

Where I Need Pushback

1. Is the neuroscience model accurate or oversimplified?

I am describing the limbic system as a "host-based firewall" that processes sensory input, particularly fear signaling, before executive function. Is this a defensible simplification or a distortion that undermines the argument?

2. Is individual change a realistic prerequisite for institutional change, or is it backwards?

The strongest counterargument may be that good institutions shape individual behavior more reliably than individual enlightenment shapes institutions. Am I sequencing this wrong?

3. Does the physics parallel hold or is it false pattern-matching?

I draw structural parallels between strong force/gravity and individuation/connection. Is this a legitimate isomorphism or am I seeing patterns where there is only coincidence?

4. Is the "zero-sum to abundance" transition historically precedented at civilizational scale?

Have any societies successfully made this mindset transition without first experiencing catastrophic collapse? If not, does that undermine the thesis or confirm it?

5. What am I missing?

What are the strongest arguments against this thesis that I have not addressed? Where are my own filters most likely distorting my analysis?
