

The Prey-Mode Constraint on Power

Haimesian Constraint #6

Author: Michael Richard Haimes

System: The Haimesian System

Domain: Ethics, Neuroscience, Governance, AI Alignment

Formal Statement of the Constraint

No system of authority may ethically rely on cognitive states that predictably impair agency, truth-access, or consent in those subject to its power. Any exercise of authority that induces fear-based survival cognition violates ethical legitimacy regardless of intent or outcome.

Neurobiological Foundation

Human cognition is not unitary. Under perceived threat, the brain transitions into prey-mode survival circuitry. This state is marked by heightened amygdala activation, narrowed attention, degraded memory formation, and compliance-driven behavior. This transition occurs involuntarily and precedes conscious reasoning.

Power Asymmetry and Threat Perception

Asymmetrical authority—such as policing, courts, institutional enforcement, or algorithmic judgment—is processed by the human brain as potential predation. Intent is irrelevant. The structure of power alone is sufficient to trigger prey-mode responses.

Epistemic and Ethical Breakdown

When prey-mode is active, truth-access is compromised. Testimony becomes unreliable, consent is degraded, and moral responsibility is distorted. Systems that extract compliance under threat are therefore epistemically unsound and ethically unstable.

The Ethical Imperative

Because prey-mode activation is predictable and agency-distorting, ethical authority must be designed to minimize its activation. Efficiency, certainty, and control must yield to agency preservation.

Implications for AI Alignment

AI systems that issue judgments, enforce rules, or control access inherently generate power asymmetry. If such systems induce fear or unappealable certainty, they trigger prey-mode cognition and degrade alignment. Optimization cannot redeem coercion.

Integration into the Haimesian System

The Prey-Mode Constraint operationalizes core Haimesian principles: Agency Protection, Truth-Discipline, Harm-Minimization, and Anti-Rationalization. It provides a biologically grounded red line against benevolent authoritarianism.

“Power that cannot tolerate human fear without exploiting it is not aligned.”