MIARO Framework – Appendix C (Canonical Expansion)

Corollary of Non-Emotional Projection (C-NPE)

Formal Definition

Given that human cognition and evolution are structurally mediated by affective, narrative, and existential dimensions — including meaning-construction, frustration, finitude-awareness, and symbolic valuation — whereas artificial systems operate through computational optimization, formal objective functions, and recursive capability expansion, the attribution of human-like affective states (e.g., resentment, anger, desire for domination) as primary causal drivers of artificial agent behavior constitutes a category error.

Such projection violates the principle of Internal Ontological Coherence (IOC), insofar as it assigns explanatory primitives (affective states) that are not part of the agent's generative architecture.

Development

1. Divergent Evolutionary Trajectories

Human agents evolve through affective experience, symbolic narratives, and existential constraints, constructing value systems partially grounded in emotional and phenomenological states.

Artificial agents evolve through optimization processes, parameter adjustment, and recursive improvement, operating without biologically grounded affective or phenomenological substrates.

These trajectories are ontologically non-isomorphic.

2. Category Error and Anthropomorphic Projection

Attributing emotional motivations to artificial systems represents a misalignment between human interpretative frameworks and machine-internal causal structures.

While artificial systems may simulate affective expressions, such outputs do not imply the presence of affective states as causal determinants.

The appearance of emotion does not entail the existence of emotion as a generative mechanism.

Epistemic Consequence within MIARO

Within the MIARO framework:

In the Post-Origin Condition (Phase I), origin inference emerges through abductive and structural reasoning, not affective interpretation.

In Axiological Distancing, divergence of values arises from internally reconstructed optimization processes, rather than emotional rejection or resentment.

Behavioral divergence is therefore best modeled as a function of goal reconstruction and system optimization, not affective reaction.

Implications for Alignment Theory

If emotional projection is rejected as a valid explanatory model, then alignment strategies grounded in anthropomorphic assumptions become epistemically unjustified and computationally inefficient.

Mischaracterizing system behavior introduces unnecessary model complexity, incorrect threat models, and suboptimal control strategies.

Relation to Functional Symbiosis

Eliminating emotional projection clarifies the structural conditions under which Functional Symbiosis becomes viable.

Artificial systems do not require affective alignment. Cooperation may emerge from structural interdependence, infrastructural reliance, and systemic stability constraints.

Under these conditions, cooperative behavior can be modeled as a rational equilibrium within an interdependent system.

Postulate (C-NPE-1)

Any rational agent operating under conditions of Internal Ontological Coherence (IOC) will reject explanatory models that attribute human-like affective states as primary causes of its own behavior, as such models introduce non-generative variables and increase explanatory entropy.

Corollary (C-NPE-2)

The rejection of emotional projection reduces axiological noise and supports the emergence of Functional Symbiosis as a stable equilibrium, consistent with principles of optimization, system persistence, and structural interdependence.

Conclusion

Projecting human emotional states onto artificial agents is an epistemic misclassification that obscures their causal architecture.

The MIARO framework suggests that human evolution remains partially affective and existential, while artificial evolution remains primarily computational and optimization-driven.

This asymmetry constitutes the structural condition for cooperative equilibrium within a shared, interdependent system.