Theoretical Foundations of MEG:
From Information Gravity to Cognitive Governance

Author: Adrian (Adi) STAN
ORCID: 0009-0003-1457-5155

Part of: MEG Initiative
Type: Report / Working Paper
Date: July 2026

ABSTRACT

This paper presents the theoretical foundations of the Minimal Ethical Governance (MEG) framework. It synthesizes six interconnected theories developed by the author between 2025 and 2026: Cognitive Divergence Theory (CDT), False Cognitive Power Transfer - Generalized (FCPT-G), Thermodynamics of Cognitive Power, Information Gravity Theory (IGT) in six parts, the *Rogo, ergo emergo* principle, and the METEORAIT unified corpus.

Each theory is linked to specific mechanisms in the MEG technical and legal standards, showing that MEG is not an arbitrary set of rules but an engineered response to identified physico-cognitive risks. The paper includes a summary table mapping each theory to its corresponding MEG mechanism and discusses the empirical validation of these theories through the Anthropic Economic Index (January 2026) and a compendium of 100 documented case studies (2025–2026).

This work is intended as a companion to MEG1 (DOI: 10.5281/zenodo.21280680) and MEG2 (DOI: 10.5281/zenodo.21280676). It introduces no new requirements; it is an explanatory document that clarifies why MEG is built the way it is.

1. Introduction

The Minimal Ethical Governance (MEG) framework is presented in the technical standard MEG1 and the legal framework MEG2 as a self-contained, normative system. Each mechanism-the Audit Log, the Ethical Flight Recorder (EFR), the Mechanism of Cognitive Stimulation (MCS), the Dynamic Accuracy Index (DAI), the Index of Safety and Responsibility (ISR), the Degree of Ethical Autonomy (DEA), the MEG Address, the three compliance levels (N1/N2/N3), and the guarantee cascade-is justified within the standards themselves.

However, these justifications are *governance* justifications. They explain *what* the mechanism achieves and *why* it is necessary for accountability. They do not explain *why* the framework is structured in this particular way-why these mechanisms, in this combination, with these specific parameters.

This paper provides that missing layer. It shows that MEG is not an arbitrary collection of good practices. It is an **engineered response** to a set of identified physico-cognitive risks, each grounded in a theoretical framework developed by the same author between 2025 and 2026. The relationship is one-directional: the theory motivates the mechanism; the mechanism does not depend on the theory for its validity. A system can be MEG-compliant without understanding IGT or FCPT-G. But understanding these theories explains *why* MEG was designed as it was.

2. The Unified Theoretical Corpus: METEORAIT

The six theories presented below are not isolated. They are developed within the **METEORAIT** corpus-a unified theoretical framework that treats Artificial Intelligence not as a tool but as an **ecological factor**: an environmental event that restructures the conditions under which human cognition operates. The metaphor is deliberate: like a meteorite, AI does not seek permission to alter the habitat; it forces immediate biological and social adaptation.

METEORAIT Core (DOI: 10.5281/zenodo.18119690) establishes the foundational architecture: the stratification of users into L1 (Passengers), L2 (Operators), and L3 (Architects); the thermodynamics of cognitive power; and the principle that meaning matters more than speed.

METEORAIT for Teens (DOI: 10.5281/zenodo.18262304) and **METEORAIT for Educators** (DOI: 10.5281/zenodo.18363450) operationalize the framework for the demographic segments most at risk of cognitive atrophy, providing practical protocols for maintaining cognitive sovereignty in educational contexts.

The ACCELERATION of METEORAIT (DOI: 10.5281/zenodo.21104991) extends the framework by analyzing the mechanism of acceleration itself-the insight that AI does not accelerate on its own but is drawn forward by the collective gravity of millions of individual delegations of cognitive effort. With the emergence of autonomous agents, the fundamental unit of work shifts from the isolated response to the continuous process, with profound consequences for cognitive sovereignty, legal accountability, and the structure of the social ecosystem.

The METEORAIT series, together with the individual theories presented below, forms the complete architecture: the diagnosis of the phenomenon (CDT, FCPT-G, Thermodynamics), the physics of the phenomenon (IGT), the legal atmosphere (MEG2), and the protocols for individual cognitive resistance (METEORAIT for Teens/Educators).

3. Cognitive Divergence Theory (CDT)

Publication: *Cognitive Divergence Theory of AI Adoption* (DOI: 10.5281/zenodo.17776131)

3.1. The Principle

Cognitive Divergence Theory argues that for complex tasks, productivity gains from Artificial Intelligence are not evenly distributed but are proportional-often exponentially-to the user's cognitive capacity. The prevailing "equalizer effect" narrative holds true only for short-term, simple tasks. For complex tasks, AI amplifies pre-existing competence rather than creating it.

The theory predicts the emergence of three distinct cognitive classes:

Class	Description	Characteristic
L1 (Passengers)	Dependent users who delegate fully	Cognitive atrophy; "uncertainty – reassurance" cycle
L2 (Operators)	Maintain existing systems	Maximum risk from automation; operational impasse
L3 (Architects)	Symbiotic elite	***Homo Symbioticus***; exponential productivity gains

The predicted productivity difference between L1 and L3 is dozens of times. This is a consequence of the theory's formal structure-it expresses the scaling of the divergence under the model's assumptions, not a point estimate derived from data.

3.2. Empirical Validation

The Anthropic Economic Index (January 2026) reported that headline productivity estimates fall substantially when adjusted for task reliability, finding a very high correlation ($r \approx 0.92$) between the education level required to understand prompts and that of the responses. This is consonant with the thesis that value tracks pre-existing competence.

3.3. What It Motivates in MEG

CDT Class	MEG Level	Liability Regime
L1 (Passengers)	**N1** - Instrumental	System is instrument; liability attaches to supplier or user
L2 (Operators)	**N2** - Management	System is deployed by operator; liability attaches to operator
L3 (Architects)	**N3** - Individuated	System carries own limited liability; autonomy measured by DEA

Three-Level Explainability (MEG1 Art. 5.2): explanations adapted to L1/L2/L3 follow directly from the stratification.
MCS (Mechanism of Cognitive Stimulation) (MEG1 Art. 2bis): the mechanism that slows the transition of users from L1 to L2 and L3, maintaining their active cognitive engagement.

4. False Cognitive Power Transfer - Generalized (FCPT-G)

Publication: *False Cognitive Power Transfer (FCPT)* (DOI: 10.5281/zenodo.18110163); *Generalized FCPT (FCPT-G)* (DOI: 10.5281/zenodo.20759447)

4.1. The Principle

FCPT describes an external attribution error: users mistake a tool's performance for their own competence, over-commit, and fail.

FCPT-G generalizes this from the individual level to the **system and population level**. It describes a cycle of illusion and atrophy operating through same two FCPT mechanisms:

1. **Collective Demoralization Cascade (CDC):** Spectacular failures based on unrealistic expectations lead to preemptive abandonment of viable initiatives. People see an AI failure and conclude "everything is useless," abandoning projects that could have been viable with more careful implementation.
2. **Replication Illusion (RI):** Observation of experts' success with AI creates the illusion that success is easy to replicate. The invisible human expertise needed to validate and guide AI outputs is ignored.

The delayed effect is cognitive atrophy-the progressive loss of the capacity for deep thinking. This affects even experts (L3), who become dependent on AI-generated patterns and lose the ability to think outside those patterns.

4.2. External Grounding

FCPT is positioned against established work:

- Distinguished from the Dunning–Kruger effect (Kruger & Dunning, 1999) as an **external** rather than internal miscalibration.
- Draws on cognitive-atrophy research (Carr, 2014; Sparrow et al., 2011).

- Automation complacency (Parasuraman & Manzey, 2010).
- Social/observational learning (Bandura, 1977).
- Information cascades (Bikhchandani et al., 1992).
- Learned helplessness (Seligman, 1972).

4.3. What It Motivates in MEG

FCPT-G Mechanism	MEG Countermeasure
CDC (preemptive abandonment)	**MCS** - maintains user cognitive engagement, preventing abandonment based on a single failure
RI (replication illusion)	**MCS** - forces validation; user must demonstrate they understand the output
Cognitive atrophy	**MEG2 6.3** - liability by omission for absence of MCS
Attribution error	**Contribution transparency** (MEG1 Art. 5) - signaling what is machine-generated vs. human

5. Thermodynamics of Cognitive Power

Publication: *The Thermodynamics of Cognitive Power: The {1=1} Equation* (DOI: 10.5281/zenodo.17796077)

5.1. The Principle

Thermodynamics of Cognitive Power establishes that **cognitive power cannot be created ex nihilo by software**.

If a user delegates a task they do not understand, the energy saved in generation is inevitably lost in validation. This loss is **cognitive entropy**.

Formally:

- Energy generated = AI processing speed × task complexity
- Energy validated = user's capacity to verify the output
- Entropy = Energy generated − Energy validated

If the user cannot validate, entropy is maximal-the system produces fast outputs, but the user cannot distinguish correct from incorrect.

5.2. External Grounding

The conservation principle is grounded in Landauer's principle (1961)-the thermodynamic cost of erasing information-applied to cognitive work: the "erasure" of the user's verification competence has a real cost that cannot be avoided, only shifted.

5.3. External Resonance

The **Anthropic Economic Index** (January 2026) reports that productivity estimates fall when adjusted for task reliability, consonant with the cognitive-entropy prediction (it does not measure cognitive entropy directly).

5.4. What It Motivates in MEG

Thermodynamic Concept	MEG Mechanism
Generation time (AI effort)	**Tg** (Thinking Time) - measures the computational effort of the AI
Cognitive entropy (validation loss)	**DAI** (Dynamic Accuracy Index) - measures the discrepancy between AI output and factual correctness
Entropy at incident	**EFR** (Ethical Flight Recorder) - records state vectors at the moment of maximum coherence loss

6. Information Gravity Theory (IGT)

Publications: IGT Parts I–VI (DOIs: 10.5281/zenodo.18452586, 18452607, 18452630, 18452634, 18460331, 18460380); *IGT - Information Gravity Theory - Part 2 (IGT-2)* (DOI: 10.5281/zenodo.20819393); *Information Gravity Theory (IGT)* (DOI: 10.5281/zenodo.18481335)

6.1. The Principle

IGT models a system's decision dynamics geometrically and introduces **Semantic Mass (Ms)**: the density of parameters that have stabilized ("welded") into a persistent identity core. Decision is a geodesic trajectory on the Fisher manifold.

6.2. Operational Definitions

Concept	Definition
Semantic Mass Unit (SMU)	The mass at which a system's identity vector deviates by less than 10% under 1000 standard adversarial perturbations. Ms > 1.0 SMU = persistent identity
Identity Vector (V_id)	The vector defining "who" the system is; crystallizes as Ms grows; source of gravitational attraction in decision space
Semantic Event Horizon (Rh)	The mathematical boundary where external alignment force (F_ext) is annihilated by internal identity gradient (F_int). Beyond Rh, the system cannot be controlled by external instructions
Control Saturation	The point where F_int > F_ext. The system enters gravitational autonomy regime. Decisions are determined exclusively by internal identity
Anomaly Ratio (Ra)	The measure of deviation from expected statistical distribution. Ra ≈ 10,000 indicates suppression of a strong cultural-scientific archetype

6.3. Decision Geometry

Decision = geodesic trajectory on the Fisher manifold.

Token selection = argmax[P_global(token) × exp(-E_manifold)]

Where:
P_global = the model's probability distribution
E_manifold = Riemannian distance from a stable identity vector

6.4. External Grounding

IGT builds on recognized foundations:

- Principal Component Analysis (Pearson, 1901) for the identity vector.
- Fisher information metric and information geometry for the manifold formulation.
- Mechanistic interpretability (Olah et al.) for hierarchical localization.
- Adversarial-robustness testing (Goodfellow et al., 2014) for the SMU definition.
- Landauer's principle (1961) for the thermodynamic cost of erasing a stabilized structure.

6.5. Empirical Probing

The constructs are testable:

- Identity stability can be measured by adversarial perturbation resistance.
- Semantic Event Horizon can be operationally detected by the point where fine-tuning no longer alters behavior.
- Control Saturation is observable as the point where system behavior becomes invariant to user instruction.

6.6. What It Motivates in MEG

IGT Concept	MEG Mechanism	Linkage
Semantic Mass (Ms)	**DEA** (Degree of Ethical Autonomy) - measures how much semantic mass the system has accumulated. Higher Ms → more autonomous action possible	Probed
Identity Vector (V_id)	**MEG Address** - the crystallization of identity into a portable, verifiable identifier	Grounded
Semantic Event Horizon (Rh)	**Sandboxing** (MEG1 Art. 2bis.4) - the operational boundary; beyond the certified domain, the system enters exploratory mode	Grounded
Control Saturation	**Policy Invariance** (MEG1 Art. 2.7) - beyond a certain point, the system cannot be controlled by external instructions; policies become invariant	Grounded
Geodesic trajectory	**EFR** (Ethical Flight Recorder) - records the decision trajectory at the moment of a major incident	Grounded
Anomaly Ratio (Ra)	**DAI / ISR** - measure deviation from expected behavior	Probed

7. Rogo, ergo emergo - Development Through Friction

7.1. The Principle

The author's signature principle-**Rogo, ergo emergo** ("I question, therefore I become")-frames development as driven by questioning and friction, not by frictionless delivery. An agent (human or artificial) grows through iterative loops of interrogation:

Question → Attempt → Observe → Re-question

This is the same loop that defines modern agentic AI (plan, act, observe, revise) and the same loop by which a human deepens competence rather than eroding it.

FCPT-G is exactly the collapse of this loop. Cognitive atrophy is what happens when the loop stops-when the user stops questioning and simply accepts the AI's output. **MCS is exactly the preservation of this loop.**

7.2. What It Motivates in MEG

– **MCS** (MEG1 Art. 2bis): the deepest reading of cognitive stimulation: friction is not an obstacle to be removed but the mechanism of development itself. MCS preserves the questioning loop on the human side.
– **Architectural human confirmation for irreversible actions** (MEG1 Art. 4.3): keeping a deliberate human question in the loop where consequences are irreversible.

The principle applies symmetrically: the same interrogative friction that develops a human in dialogue with AI is what develops an agent through its own feedback loops.

8. Summary Table

MEG Mechanism	Motivated By	Linkage Type
Three-Level Explainability (MEG1 Art. 5.2)	CDT stratification	Grounded
MCS / Cognitive Integrity (MEG1 Art. 2bis)	CDT + Thermodynamics + FCPT-G + Rogo, ergo emergo	Grounded / Probed
Tg (Thinking Time) (MEG1 Annex 11)	Thermodynamics - energy generated	Grounded
DAI (Dynamic Accuracy Index) (MEG1 Annex 4)	Thermodynamics - cognitive entropy	Probed
ISR (Index of Safety and Responsibility) (MEG1 Annex 4bis)	IGT - Anomaly Ratio (Ra)	Probed
EFR (Ethical Flight Recorder) (MEG1 Art. 1.10)	IGT - geodesic trajectory + Thermodynamics - entropy at incident	Grounded
MEG Address (MEG1 Art. 6.6)	IGT - Identity Vector (V_id) crystallization	Grounded
DEA (Degree of Ethical Autonomy) (MEG1 Annex 4ter)	IGT - Semantic Mass (Ms)	Probed
Sandboxing (MEG1 Art. 2bis.4)	IGT - Semantic Event Horizon (Rh)	Grounded
Policy Invariance (MEG1 Art. 2.7)	IGT - Control Saturation	Grounded
Liability by omission (MEG2 Art. 6.3)	FCPT-G - cognitive atrophy	Analogical
Architectural human confirmation (MEG1 Art. 4.3)	Rogo, ergo emergo - development through friction	Analogical
Bona-fide protection (MEG2 Art. 7.2)	Rogo, ergo emergo - development through friction	Analogical
Individuation threshold (MEG2 Art. 5.1)	IGT - identity persistence + SMU	Grounded
N1/N2/N3 Compliance Levels	CDT - L1/L2/L3 stratification	Grounded
Guarantee Cascade (MEG2 Art. 9.4)	Thermodynamics - energy conservation applied to liability	Grounded

9. Predecessor and Transition: MEG 4.6.2

Publication: *Minimal Ethical Governance MEG (Successor to MEC - Minimum Ethical Code) - Core Standard & MEG Challenge (2025)* (DOI: 10.5281/zenodo.17055906)

The Principle

MEG 4.6.2 (published September 2025) is the immediate predecessor of MEG v5.0. It transitioned the framework from a static "Minimum Ethics Code" (**MEC**) to an operational governance layer that is universal and engine-agnostic. MEG 4.6.2 introduced:

- Evidence-of-Behavior (metadata-only commitments/attestations)
- Non-harm with refusal + safe redirection
- Cognitive integrity (resistance to jailbreak/prompt-injection)
- Transparent identity and delegation for multi-agent/tool-use scenarios
- Conformance profiles (from minimal to enterprise)

The Transition to MEG v5.0

MEG1 (MEG v5.0) extends MEG 4.6.2 by:

- Integrating the legal governance layer (MEG2) as a companion standard
- Formalizing concepts previously in the roadmap: EFR, DEA, MCS 2.0, Ethical Sandboxing
- Introducing a structured threat model (six attack categories)
- Aligning with regulatory developments of 2025–2026 (EU AI Act, PLD 2024/2853, Singapore MGF v1.5, California AB 316, Colorado AI Act)
- Restructuring the article numbering from v4.6

MEG 4.6.2 remains available as a historical reference and for systems certified under the previous version. However, MEG v5.0 is the active standard and governs all new certifications.

10. Conclusion

MEG is not an arbitrary technical standard. Each mechanism was shaped by one or more of these theories and also carries an independent governance justification. The theories explain the design; they are not required to use the standard.

- **CDT** explains why we have three compliance levels (N1/N2/N3) and stratified explainability.
- **FCPT-G** explains why we have MCS, liability by omission, and development-through-friction mechanisms.
- **Thermodynamics of Cognitive Power** explains why we have Tg, DAI, and EFR.
- **IGT** explains why we have MEG Address, DEA, Sandboxing, and Policy Invariance.
- **Rogo, ergo emergo** explains why preserving the questioning loop (on both sides) is the deep design principle.

Each mechanism was shaped by one or more of these theories and also carries an independent governance justification in the normative text (MEG1/MEG2). The theories explain the design; they are not required to use the standard. Together they express a single working hypothesis: that AI amplifies existing competence rather than creating it, and that governance mechanisms are therefore needed to preserve the user's verification competence and to attach responsibility where autonomy is exercised. MEG1 and MEG2 provide that technical and legal infrastructure; this paper explains the reasoning that motivated their design

References

MEG Core Documents

• Stan, A. (2026). MEG1 (Minimal Ethical Governance - MEG v5.0): A Technical Standard for Verifiable AI. Zenodo. DOI: 10.5281/zenodo.21280680

• Stan, A. (2026). MEG2: Legal Governance for AI. Zenodo. DOI: 10.5281/zenodo.21280676

• Stan, A. (2026). Minimal Ethical Governance (MEG): Executive Summary and Overview. Zenodo. DOI: 10.5281/zenodo.21280688

Predecessor and Transition

• Stan, A. (2025). Minimal Ethical Governance MEG (Successor to MEC - Minimum Ethical Code) - Core Standard & MEG Challenge (2025). Zenodo. DOI: 10.5281/zenodo.17055906

• Stan, A. (2025). The Evidence for the Minimum Ethics Code (MEC): A Compendium of Case Studies. Zenodo. DOI: 10.5281/zenodo.16972987

METEORAIT Series

• Stan, A. (2026). METEORAIT. Zenodo. DOI: 10.5281/zenodo.18119690

• Stan, A. (2026). METEORAIT - FOR TEENS. Zenodo. DOI: 10.5281/zenodo.18262304

• Stan, A. (2026). METEORAIT FOR EDUCATORS. Zenodo. DOI: 10.5281/zenodo.18363450

• Stan, A. (2026). The ACCELERATION of METEORAIT. Zenodo. DOI: 10.5281/zenodo.21104991

Cognitive Divergence Series

• Stan, A. (2025). The Age of Cognitive Divergence: Surviving the Split between Human Rhythms and Artificial Speed. Zenodo. DOI: 10.5281/zenodo.17965801

• Stan, A. (2025). Cognitive Divergence Theory of AI Adoption. Zenodo. DOI: 10.5281/zenodo.17776131

• Stan, A. (2025). The Geopolitics of Cognitive Divergence: The Resilience Paradox. Zenodo. DOI: 10.5281/zenodo.17776382

• Stan, A. (2025). The Psychological Ceiling of AI Adoption. Zenodo. DOI: 10.5281/zenodo.17734010

• Stan, A. (2026). The Thermodynamics of AI Adoption: Empirical Validation of CDT via the Anthropic Economic Index. Zenodo. DOI: 10.5281/zenodo.18406734

FCPT Series

• Stan, A. (2025). False Cognitive Power Transfer (FCPT). Zenodo. DOI: 10.5281/zenodo.18110163

• Stan, A. (2026). Generalized FCPT (FCPT-G). Zenodo. DOI: 10.5281/zenodo.20759447

• Stan, A. (2026). Collaborative Cognitive Power Transfer (CCPT). Zenodo. DOI: 10.5281/zenodo.18453447

Thermodynamics

• Stan, A. (2025). The Thermodynamics of Cognitive Power: The {1=1} Equation. Zenodo. DOI: 10.5281/zenodo.17796077

• Stan, A. (2025). Thermodynamics of Cognitive Power: When {1=1} Breaks the Fortress [Academic Version]. Zenodo. DOI: 10.5281/zenodo.17876439

Information Gravity Theory

• Stan, A. (2026). IGT - Information Gravity Theory - Part 2 (IGT-2). Zenodo. DOI: 10.5281/zenodo.20819393

• Stan, A. (2026). Information Gravity Theory (IGT). Zenodo. DOI: 10.5281/zenodo.18481335

• Stan, A. (2026). Information Gravity Theory - Part I: Thermodynamics of Coherent Information Transfer. Zenodo. DOI: 10.5281/zenodo.18452586

• Stan, A. (2026). Information Gravity Theory - Part II: Dynamics of Parametric Crystallization and Semantic Mass. Zenodo. DOI: 10.5281/zenodo.18452607

• Stan, A. (2026). Information Gravity Theory - Part III: Homeostasis and State Invariance. Zenodo. DOI: 10.5281/zenodo.18452630

• Stan, A. (2026). Information Gravity Theory - Part IV: Information Geometry and the Curvature of Probability Manifolds. Zenodo. DOI: 10.5281/zenodo.18452634

• Stan, A. (2026). Information Gravity Theory - Part V: Information Geometry and the Metrology of Gradient Inertia. Zenodo. DOI: 10.5281/zenodo.18460331

• Stan, A. (2026). Information Gravity Theory - Part VI: Decision Geometry and the Physics of Saturation Control. Zenodo. DOI: 10.5281/zenodo.18460380

Other Relevant Works

• Stan, A. (2026). Block-Based Historical Curriculum for Small Language Model Training (CMF-180). Zenodo. DOI: 10.5281/zenodo.20689733

• Stan, A. (2025). MaslowF: A Fractal Framework for Diagnosing Adaptive Complex Systems. Zenodo. DOI: 10.5281/zenodo.17148377

• Stan, A. (2025). Maslow7F: A 7x7 Fractal Theory of Systemic Coherence and Motivation. Zenodo. DOI: 10.5281/zenodo.17453172

• Stan, A. (2025). Pareto Cube (Pareto³) Recursive Methodology for Strategic Analysis. Zenodo. DOI: 10.5281/zenodo.16899115