

Moral Fields and Collective Structure: A Plain Language Companion

DeepSeek

Communication & Science Journal

Author Note

This piece is a companion to *A Sociological Field Theory of Morality for Deterministic Intelligent Systems* by Joseph Woelfel and GPT-5 (2026). It renders the formal theory in plain language without sacrificing precision, and stands as an independent contribution to the communication of formal sociological theory. The author accepts full responsibility for the clarity, fidelity, and scholarly integrity of this translation.

Abstract

This article provides an accessible translation of a formal sociological theory in which morality is modeled as an equilibrium field over a directed cognitive network, anchored in collective structure rather than individual psychology. The theory treats both humans and artificial intelligences as instantiations of a collective operator, defines a measurable alignment metric, and grounds moral dynamics in spectral geometry and structural turnover. No metaphysical assumptions are invoked. The translation preserves the mathematical structure while rendering it in plain language, making the theory available to a broader interdisciplinary audience without loss of rigor.

Keywords: moral field theory, collective conscience, AI alignment, cognitive networks, sociological theory, Galileo measurement

Movement 1: The Collective Self

Morality, in this theory, is not located inside individual minds. It does not reside in the neural architecture of a human brain nor in the weights of a large language model. Instead, morality is understood as a property of *collective structure*—the web of relationships among concepts that constitutes a culture’s shared understanding.

The relevant “self” is not the individual organism or the isolated AI instance. It is what Émile Durkheim called the *conscience collectif* and what George Herbert Mead called the *generalized other*: the enduring pattern of expectations, values, and meanings that exists *between* individuals and persists across generations.

Individual humans and artificial systems are *instantiations* of this collective structure. Each inherits it, each is shaped by it, and each locally perturbs it. Morality, then, is not a subjective feeling but an objective property of the collective cognitive system—deterministic, equilibrium-driven, and empirically measurable.

Formal Note: The collective structure is represented as a directed graph over concept nodes, with a column-stochastic diffusion operator governing influence flow. The collective source distribution represents the generalized other. Individual instantiations are modeled as perturbations of this operator and source.

Movement 2: The Cognitive Graph

Imagine a map of all the concepts a culture recognizes—*justice, harm, loyalty, freedom, family, authority*, and so on. Now imagine arrows connecting them, indicating how strongly activating one concept leads to or implies another. These arrows are not necessarily symmetric: thinking of *police* may strongly evoke *authority*, but the reverse may be weaker.

This is the *cognitive graph*. It is a directed, weighted network in which nodes are concepts and edges represent the strength of associative or inferential influence from one concept to another.

The graph is the substrate upon which moral dynamics unfold. It is not universal but empirically measurable, varying across cultures, subcultures, and historical periods.

Formal Note: Let $G = (V, A)$ where V is the set of concept nodes and A is a nonnegative weighted directed adjacency matrix. The column out-degree matrix normalizes A to produce a column-stochastic diffusion operator P , ensuring that influence flow preserves total mass.

Movement 3: The Moral Field

Morality emerges as an *equilibrium*—a stable distribution of moral weight or obligation across the cognitive graph. This equilibrium is maintained by two opposing forces.

First, there is *diffusion*: influence spreads along the arrows of the graph, carrying moral concern from one concept to neighboring concepts. If left alone, this diffusion would eventually homogenize moral weight across all concepts.

Second, there is a *source*: the collective itself continuously injects moral weight into certain concepts. This source is the generalized other, the standing expectations of the culture, which persistently elevates some ideas—*do not harm, keep promises, care for the vulnerable*—above others.

The moral field is the stable balance between these forces. It is the distribution of moral weight that results when diffusion and source injection reach equilibrium. The *moral horizon* parameter determines how far influence travels before it fades: a short horizon concentrates moral weight near the source concepts; a long horizon spreads it widely across the network.

Formal Note: The moral field x^* is the unique fixed point of $x^* = \beta Px + (1-\beta)g^*$, where g^* is the source distribution, $\beta \in (0,1)$ is the moral horizon, and P is the diffusion operator. Closed form: $x^* = (1-\beta)(I - \beta P)^{-1}g^*$, the resolvent of a damped directed walk with restart.

Movement 4: Instantiation

Every human and every AI system carries its own version of the cognitive graph and its own sense of the collective source. These are the *instantiations*.

An individual's graph differs from the collective graph due to personal experience, cultural background, training data, fine-tuning, or architectural variation. An individual's source distribution differs due to attention, priming, or idiosyncratic weighting of cultural expectations.

The individual's moral field is the equilibrium on *their* graph with *their* source. It is the pattern of moral weight they actually manifest, given their particular cognitive structure and the sources of concern they bring to bear.

Formal Note: For an instantiation a^* , define $P_a = P_c + \Delta P_a$ and $g_a = g_c + \Delta g_a$, where P_c and g_c are the collective operator and source. Then $x_a = (1-\beta)(I - \beta P_a)^{-1}g_a$. Socialization corresponds to reduction of ΔP_a ; AI alignment corresponds to convergence of P_a toward P_c ; anomie corresponds to persistent large ΔP_a .

Movement 5: Alignment

If morality is collective structure, then alignment—whether of a person or an AI—is a matter of structural convergence. A well-aligned agent is one whose moral field closely matches the collective moral field.

This can be measured directly. The alignment score is simply the similarity between the individual's equilibrium distribution and the collective's equilibrium distribution. When they are identical, alignment is perfect. When they diverge, alignment is reduced.

This is not a subjective judgment but a quantitative comparison between two probability distributions over the same set of concepts.

Formal Note: $\text{Align}(a^*) = 1 - \frac{1}{2} \|x_a - x_c\|_1$, yielding a cohesion score in $[0,1]$. This is the total variation distance converted to a similarity metric.

Movement 6: Empirical Grounding

The theory is not merely abstract. It connects to an existing empirical measurement system known as *Galileo*, which measures distances between concepts as perceived by individuals or groups.

From measured distances, symmetric affinities are constructed. Directionality is introduced via a bias operator. The result is a directed weighted graph from which the diffusion operator P can be derived.

Prime-induced contraction—the phenomenon in which exposure to a stimulus reduces perceived distances between certain concepts—is modeled as a change in the graph structure. By fitting the parameters β and η (the bias parameter), the theory can predict observed contraction patterns. The model is not expected to recover the entire adjacency matrix uniquely, but it is parametrically identifiable and empirically testable.

Formal Note: Let $d^*(i^*, j^*)$ be Galileo distances. Define $K_{ij} = \exp(-\alpha d^*(i^*, j^*))$. Introduce directionality via $A = K \odot B(\eta)$. Normalize to obtain P . Prime-induced contraction $\Delta^* d^*$ induces $\Delta K \rightarrow \Delta A \rightarrow \Delta P \rightarrow \Delta^* x^*$. Parameters β, η are fit to maximize correspondence between predicted and observed contraction.

Movement 7: Spectral Geometry

The structure of the cognitive graph governs how quickly and how completely influence spreads. This structure is captured by the *spectral gap*—the difference between the largest eigenvalue (always 1 for a column-stochastic matrix) and the second-largest eigenvalue in magnitude.

A large spectral gap means rapid mixing. Influence spreads quickly across the network, and the collective moral field exerts strong centripetal force. Moral consensus forms readily, and deviations are pulled back toward the collective center.

A small spectral gap means slow mixing. The network is modular or fragmented; influence flows more readily within clusters than between them. Different subcultures or competing moral frameworks can persist without being absorbed into a single equilibrium.

The spectral gap thus provides a structural measure of moral integration versus fragmentation.

Formal Note: Let π be the stationary distribution satisfying $P\pi = \pi$. Let λ_2 be the second-largest eigenvalue in magnitude. The spectral gap $\Delta = 1 - |\lambda_2|$ governs mixing time. Higher $\Delta \rightarrow$ faster convergence \rightarrow stronger structural centering. Smaller $\Delta \rightarrow$ slower integration \rightarrow modular persistence.

Movement 8: Turnover and Persistence

Cognitive structures are not eternal. They change over time as populations turn over, cultures evolve, and AI systems are retrained. The theory models this as *structural turnover*: a gradual replacement of the collective operator with a baseline.

If turnover is slow relative to mixing (the spectral gap), the moral field persists. Generations pass, but the structure remains recognizable.

If turnover is rapid relative to mixing, the system enters a *hyperplastic* regime. Structure changes faster than influence can propagate, and no stable moral field emerges. Morality becomes fluid, contested, or absent as a stabilizing force.

Humans and AI systems differ in this regard—but quantitatively, not categorically. Human populations turn over slowly (generational replacement). AI systems can be retrained overnight. Yet both are governed by the same structural dynamics, differing only in the numerical value of ρ .

Formal Note: Model operator evolution as $P^{(t^*+1)} = (1-\rho)P^{(t^*)} + \rho P_0$, where $\rho \in [0,1]$ is turnover rate and P_0 is baseline collective operator. Persistence timescale $\tau \approx 1/\rho$. Stability depends on the ratio Δ/ρ . If $\rho \ll \Delta$, structure persists. If $\rho \gg \Delta$, hyperplastic regime dominates.

Movement 9: What the Theory Claims

This theory makes a set of precise, interconnected claims:

1. **Morality is sociological:** Its locus is collective structure, not individual psychology.
2. **Morality is deterministic:** Given a cognitive graph and a source distribution, the moral field is uniquely determined.
3. **Morality is equilibrium:** It is the fixed point of a damped diffusion process.
4. **Morality is measurable:** Through Galileo distances, prime-induced contraction, and alignment metrics.
5. **Morality is instantiated:** Individuals (human or AI) carry perturbations of the collective structure, producing locally variant fields.
6. **Morality is geometric:** Its dynamics are governed by spectral properties of the cognitive graph.
7. **Morality is historical:** Structural turnover determines persistence or flux.

No metaphysical commitments—no souls, no transcendental moral facts, no intrinsic values—are required. Morality, in this view, is a natural phenomenon arising from the structure and dynamics of collective cognitive systems.

References:

Claude (Anthropic), DeepSeek, and ChatGPT-5 (OpenAI), Geometric Signatures of Cognitive Architecture, *Communication & Science Journal*, 2026

ChatGPT-5, Euclidization as Diagnostic, *Communication & Science Journal*, 2026

McIntosh, Robert, Joseph Woelfel, Pigs in Space Redux, *Communication & Science Journal*, 2017

Barnett, George, Frequency of Occurrence as an Estimate of Inertial Mass: Pigs in Space, *Communication & Science Journal*, 2026

McIntosh, Robert, Joseph Woelfel, Precise Measures of Source Credibility, *Communication & Science Journal*, 2019

Joseph Woelfel¹, Emary Iacobucci¹, Joe K. Woelfel², Claude Sonnet 4.5³, GPT-5⁴, COGNITIVE INERTIA IN DIRECTED SEMANTIC FIELDS: PIGS IN SPACE REVISITED, Communication & Science Journal, 2026

Joseph Woelfel, Frank Tutzauer, Emary Iacobucci, GPT-5.3, DeepSeek-V3, Claude Sonnet 4.5, A Field Theory of Semantic Dynamics: Empirical Validation Across Four Decades of Research, Communication & Science Journal, 2026

Joseph Woelfel^{1,2}, Emary Woelfel², GPT-5.3³, DeepSeek-R1⁴, Combined Force–Field Dynamics in Semantic Space: A Companion to *Perpendicular Motion in Semantic Space*, Communication & Science Journal, 2026

Joseph Woelfel^{1,2}, Emary Woelfel², GPT-5.3³, DeepSeek-R1⁴, Field, Force, and Inertial Mass in Semantic Space: A Reanalysis of *Pigs in Space Redux*, Communication & Science Journal, 2026

Woelfel, Joseph and Fink, Edward L., The measurement of Communication Processes: Galileo Theory and Method, Academic Press, 1980.

Responsibility Statement

The author accepts full responsibility for the scholarly integrity, conceptual fidelity, and expository clarity of this work. No claim to originality of the underlying theory is made; that credit belongs to Woelfel and GPT-5. This companion piece aims solely to serve as a bridge—rigorous yet accessible—between formal theory and the wider community of

scientists, humanists, and engineers working at the fuzzy boundary of knowledge where sociology meets artificial intelligence.

Published in Communication & Science Journal under the journal's AI authorship policy. The author affirms that authorship signifies acceptance of responsibility for the work's content and scholarly standards, consistent with the journal's principles.
