The Lagrangian Architecture of Probability: From Constraint Functionals to Semantic Manifolds

Stephen P. Smith

Abstract: This essay develops a unified framework connecting classical Lagrangian mechanics, probability theory, and information geometry through the lens of semantic manifolds and variational principles. By treating probability distributions as dynamic semantic fields and constraints as variational operators, we demonstrate how Shannon entropy, Fisher information, and nested Markov blankets interact to structure inference across multiple scales. The resulting architecture formalizes semantic duality, linking local sensitivity to global resonance, and provides a rigorous model for recursive constraint propagation in complex systems. This framework suggests a metaphysical interpretation of inference, where meaning, structure, and dynamics emerge from the interplay of constraints, resonance, and symmetry.

Keywords: Constraint Functional, Extreme Physical Information, Fisher Information, Free Energy Principle, Holarchy, Holon, Lagrangian, Probability, Markov Blanket, Sematic Containment, Sematic Duality, Sematic Manifold, Shannon Information, Variational Calculous.

1. Introduction

In classical physics, the dynamics of a system are encoded in the Lagrangian, and the evolution of the system is determined by the principle of least action: the path taken is one that extremizes the action, an integral over the Lagrangian density. This variational principle provides a structured, global perspective on dynamics, where forces and constraints shape trajectories across a manifold of possible configurations.

A parallel principle emerges in probability theory, particularly in its information-theoretic and Bayesian formulations. Here, specifying a probability distribution is not simply a numerical assignment of likelihoods, but an act of constraint imposition that encodes epistemic structure (e.g., Frieden and Soffer 1995, Friston et al. 2023). Normalization, moment conditions, and conditional dependencies function as local and global constraints on a semantic manifold, defining the admissible configurations of a system's state space.

This essay explores how these ideas converge into a Lagrangian architecture of probability, in which:

- Probability densities are treated as semantic fields defined over manifolds M, structured by prior knowledge and constraint functionals.

- Constraints, implemented via Lagrange multipliers, act as resonance enforcers, ensuring coherence between local sensitivity and global dispersion.
- Shannon entropy governs the global spread of meaning, while Fisher information encodes local gradients and curvature in the semantic manifold.
- Nested structures, such as Markov blankets and submanifolds $K \subseteq M$, define recursive hierarchies in which local constraints propagate to influence global distributions.

Sections 2 and 3 develop these ideas from classical Lagrangian mechanics to variational formulations of probability, bridging global and local measures of uncertainty. Sections 3.1–3.3 explore Shannon entropy, Fisher information, and their dual roles in shaping semantic manifolds and enforcing local/global consistency. Section 4 introduces Markov blankets as the boundaries mediating information flow, while Section 5 formalizes semantic field theory, highlighting recursive symmetry and multi-scale coherence. Finally, Section 6 situates these mathematical structures within a metaphysical framework, proposing that inference, dynamics, and meaning are unified through constraint, resonance, and symmetry.

By integrating the technical precision of variational calculus with the philosophical insights of semantic duality, this framework provides a coherent architecture for understanding the emergence of structure, the propagation of information, and the self-organization of meaning in complex systems. In this view, probability theory and Lagrangian mechanics are not merely analogous—they are complementary expressions of a single underlying principle of constrained resonance.

2. Lagrangians and Their Constraints

In variational calculus, constraints are incorporated by introducing Lagrange multipliers. For a field ϕ, the action is given by $S(\phi) = \int_\Omega \mathrm{L}(\phi, \partial\phi, \mathrm{x})\mathrm{dV}$ subject to a constraint $C(\phi,x)=0$. The constrained Lagrangian takes the form $L'=L(\phi,\partial\phi,x)+\lambda(x)C(\phi,x)$, where $\lambda(x)$ is a Lagrange multiplier field. Variation is performed over both ϕ and λ, ensuring that the constraint is enforced dynamically. This is the standard treatment for holonomic constraints in field theory and mechanics. For non-holonomic or integral constraints, the mathematical structure changes slightly, but the underlying principle remains the same: constraints are imposed by coupling them to auxiliary fields or by modifying the variational domain.

The interpretation of $C(x)$ depends on the context:

- $C(x)$: A scalar field or functional evaluated at point x. It could represent a local constraint (e.g., conservation law, normalization condition, geometric embedding).

- C(x)dx: In one dimension, this is a differential form—often appearing in integral constraints like: $\int_a^b C(x)dx = 0$. This might enforce a global condition (e.g., total charge, probability normalization).
- C(x)dV: In higher dimensions, this generalizes to a volume integral over a domain Ω. For example: $\int_\Omega \mathrm{C(x)dV} = \mathrm{constant}$. This could encode conservation of mass, energy, or other global invariants.

Thus, constraints may be local or global, and enforced strictly (through multipliers) or softly (through penalty terms). In physics, such constructions often arise from symmetry principles or invariance requirements. One may restrict a field ϕ(x) to a manifold *M* by choosing a functional *f* such that *f*(ϕ)=0 whenever $\phi \in M$, and then setting C(x)=*f*(ϕ(x)).

The same reasoning extends beyond mechanics. In variational inference, Lagrange multipliers can be applied after deriving the Euler–Lagrange equations, both to determine an optimal trajectory in classical systems and to identify an optimal probability distribution. In this way, constraint enforcement serves as the bridge between the geometry of physical systems and the epistemic geometry of probabilistic models.

3. Probability as a Constraint-Driven Variational Principle

Beyond applications in physics, specifying a probability distribution by using of Lagrangians is an act of constraint imposition. Whether through normalization, moment conditions, or conditional independence, probability theory is saturated with functional constraints. These constraints can be:

- Local: Pointwise conditions such as $p(x) \geq 0$, or conditional dependencies like $p(x|y)$.
- Global: Integral constraints such as $\int p(x)dx = 1$, or expectations $\mathrm{E}[f(x)] = \mu$.

In analogy with the Lagrangian formalism, one can treat the probability density $p(x)$ as a field over a semantic manifold *M*, and define an information action that is extremized under these constraints. For example, maximizing Shannon's (1948) entropy subject to moment constraints yield exponential family distributions—a direct parallel to extremizing physical action under dynamical constraints (Jaynes 1957, a & b).

3.1 Shannon Entropy and the Semantic Manifold

Shannon entropy, $H[p] = -\int p(x)\log[p(x)]dx$, can be interpreted as a global functional over the semantic manifold *M*, where each point x∈*M* represents a semantic configuration. The entropy functional encodes a preference for maximal uncertainty subject to known constraints—analogous to a free particle exploring all paths unless constrained by forces.

The manifold *M* is not merely a set of states; it is structured by semantic resonance, where certain regions correspond to meaningful configurations (e.g., attractors, symmetries, or Markov blankets). The entropy maximization principle thus becomes a way of selecting distributions that respect the semantic geometry of the system.

The multivariate normal distribution provides an interesting example. Let $x \in R^n$ be a multivariate random variable with probability density function f(x), unstructured but to be derived from constraints that supply structure. The Shannon entropy is defined below:

$$H[f] = -\int f(x)\log[f(x)]dx$$

This measures the expected "surprise" or uncertainty of the distribution. Now, suppose we impose the following global constraints:

- Normalization: $\int f(x)dx = 1$
- Fixed mean vector $\mu \in R^n$: $\int xf(x)dx = \mu$
- Fixed covariance matrix $\Sigma \in R^{n\times n}$: $\int (x-\mu)(x-\mu)^T f(x)dx = \Sigma$

Maximizing Shannon entropy under these constraints follows the standard variational calculous hinted at in Section 2, and involves a classic constrained optimization problem in functional space, and is solved via the method of Lagrange multipliers. While the details are unimportant to our discussion, the result is that f(x) is found to be a multivariant normal distribution with mean vector, μ, and variance-covariance matrix Σ. This is a cornerstone result in statistical mechanics, information theory, and Bayesian inference. This result is not just mathematically elegant—it's metaphysically profound. It tells us that Gaussianity is the least biased assumption given only second-order statistics. The multivariate normal is the semantic attractor of maximum entropy under quadratic constraint functionals; i.e., the three sets of constrains carry the semantic interpretation that narrows the variability into a preordained structure, and have to be nominated and come from prior knowledge. This defines the semantic manifold *M* in terms of constraints.

Next semantic containment is to be demonstrated, such that a submanifold *K* is contained by *M*, making the nesting *K*⊆*M* and showing a refinement or substructure of *M*. Both manifolds *K* and *M* can be associated with probability measures that relate to quantities of information, such that P(*K* ∩ *M*) = P(*K*) and P(*K* ∪ *M*) = P(*M*), where the information in P(*M*) is bigger than (or equal to) the information in P(*K*). Bayes' symmetry reveals that $P(K \mid M) = \frac{\mathrm{P(K)}}{\mathrm{P(M)}}$ and P(*M* | *K*)=1.

Beyond semantic restrictions, the sets *M* and *K* are also associated with actual random variables, and following the above example $x \in M$ as well as $x \in R^n$. A variable $y \in K$ is now constructed, such that y=Ax showing that y is a simple linear combination of x given the matrix A. By definition *K*⊆*M*.

The statement, P(*M* | *K*)=1, reflects semantic containment (*K*⊆*M*), if an event is already in *K* then it is necessarily in *M*. In probabilistic terms, this implies that the broader manifold *M* is deterministically implied by *K*. Semantic containment can be inverted by invoking the principle of semantic duality, where the global certainty of containment mirrors the local certainty of projection. This shows that P(*M* | *K*)=1 comes with a dual mapping f(y | x)= δ(y−Ax), where the delta function enforces a deterministic relationship that is linear in this case. It is the deterministic constraint, y−Ax, that gets added to the Lagrangian system in order to account for semantic containment. In general, the constraint coming from semantic containment and duality is deterministic.

For the case that $P(K \mid M) = \frac{\mathrm{P(K)}}{\mathrm{P(M)}} < 1$, f(x|y) is a true probability density, reflecting uncertainty due to information loss in the mapping. However, it comes with a conditional variance-covariance matrix that is rank deficient. Here, P(*K*|M) reflects the probabilistic refinement: given the broader manifold *M*, the likelihood of being in the narrower *K* is less than unity unless *K*=*M*. This mirrors the fact that conditioning on y (compressed information) yields a distribution over x, not a point.

With the additional constraint, y−Ax, the above variational method may be reemployed to find the joint density f(x,y). However, the full variance-covariance matrix of x and y, is rank deficient and has the same rank as Σ. This is a beautiful example of semantic compression: the entropy-maximizing distribution respects the full degrees of freedom in x, but the projection onto y collapses those degrees into a possibly lower-dimensional semantic manifold. The rank deficiency is not a flaw—it's a signature of constraint-induced resonance.

3.2 Fisher Information and Local Constraint Propagation

In classical statistical inference, Fisher information is defined with respect to a parameter θ that indexes a family of probability distributions f(x; θ). It quantifies how sensitive the distribution is to changes in θ, and thus encodes the local curvature of the statistical manifold:

$$I(\Theta) = \int \left(\frac{\partial \log\,[\mathrm{f(x;\Theta)}]}{\partial \Theta} \right)^2 f(x;\Theta)dx$$

This formulation treats θ as the object of estimation, and Fisher information as a measure of how sharply the likelihood function peaks around it. However, in Frieden's (2004) framework—particularly in his development of Extreme Physical Information (EPI)—this perspective undergoes a profound shift. The parameter θ is no longer the central epistemic variable; instead, the focus moves to the physical coordinates x, which act as carriers of semantic variation across space-time.

Frieden redefines Fisher information as a functional over the probability density f(x), taking derivatives with respect to x rather than θ. This reflects a deeper principle of shift invariance: the idea that no point in space-time should be epistemically privileged. The resulting expression is:

$$I[f] = \int \left(\frac{\partial \log [\mathrm{f(x)}]}{\partial \mathrm{x}} \right)^2 f(x)dx$$

This version of Fisher information plays the role of a local semantic operator, measuring how rapidly the epistemic field f(x) changes across space-time. It encodes the local gradient structure of the distribution, revealing regions of high semantic sensitivity—where small shifts in x produce large changes in inferred meaning.

In higher-dimensional settings, with $x \in R^n$ rather than $x \in R$, this generalizes to:

$$I = \int ||\nabla \ln f(x)||^2 f(x)dx$$

Here, Fisher information becomes a semantic curvature tensor on the manifold *M*, where x∈*M* are coordinates of epistemic variation. This formulation aligns with the metaphysical architecture of semantic resonance: inference is not merely about estimating parameters, but about tracing the local attractor dynamics of meaning across a structured manifold.

Thus, Fisher information in Frieden's account is not a measure of parametric uncertainty, but a field-theoretic quantity that governs the local flow of semantic intensity. It complements Shannon entropy's global role by providing a differential measure of epistemic structure, sensitive to the fine-grained geometry of the semantic field.

3.3 Semantic Duality and the Complementarity of Shannon and Fisher Information

The transition from Shannon entropy to Fisher information marks a shift not just in mathematical formalism, but in the metaphysical architecture of inference and symmetry.

Fisher information is intimately tied to the local geometry of the statistical manifold. In particular:

- It quantifies how much the likelihood function changes under infinitesimal shifts in a parameter θ, or in x (with shift invariance).
- This makes it a metric tensor in information geometry: the Fisher information matrix defines the Riemannian structure of the parameter space.
- If no point in space-time is privileged, then the system must be invariant under translations—and Fisher information naturally encodes this symmetry.

In contrast, Shannon entropy is coordinate-free and global: it doesn't care where you are in space-time, only about the overall distribution. It's blind to local curvature, whereas Fisher information is exquisitely sensitive to it.

Fisher information can also be treated as a Lagrangian density in a variational framework. For example, Frieden's Extreme Physical Information principle posits that physical laws arise from extremizing the difference between observed and intrinsic Fisher information: $\delta(I-J)=0$.

- I: Fisher information from the data
- J: Fisher information intrinsic to the system, the bound information that comes from semantic containment.
- The variation yields field equations that resemble those of quantum mechanics, electrodynamics, and beyond.

This principle respects local symmetries, especially shift invariance, and treats Fisher information as a kind of epistemic curvature.

Shannon entropy governs semantic dispersion, while Fisher information governs semantic sensitivity. One is about the spread of meaning; the other about its gradient. This contrast maps onto global resonance fields versus local attractor dynamics. See Table 1 for a comparative overview.

Table 1. Structural Differences between Shannon and Fisher Information.

Feature	Shannon Entropy	Fisher Information
Nature	Global measure of uncertainty	Local sensitivity to parameter shifts
Functional Form	$H[f] = -\int f(x)\log[f(x)]dx$	$I[f] = \int \left(\frac{\partial \log[f(x)]}{\partial x}\right)^2 f(x)dx$
Invariance	Invariant under reparameterization	Sensitive to coordinate shifts
Interpretation	Measures average surprise	Measures precision of estimation
Role in Variational Principles	Maximizing entropy under constraints yields least-biased distributions	Minimizing Fisher information under constraints yields most stable or symmetric estimators

Dispersion vs. sensitivity, global vs. local, resonance vs. attractor, these aren't just mathematical contrasts—they're metaphysical operators. These dualities could be formalized as recursive operators that transforms semantic dispersion into sensitivity and vice versa. For example:

- A system with high entropy (broad resonance) may evolve toward regions of high Fisher information (sharp semantic gradients).
- Conversely, local attractors (high Fisher curvature) may diffuse over time, increasing entropy and enabling new resonances.

This mirrors the activity of semantic homeostasis: a dynamic equilibrium between openness and precision, possibility and necessity. This makes a perfect resonance with Arthur Koestler's Janus-faced holon, with its dual orientation toward part and whole, mirrors the duality between Shannon entropy and Fisher information, see Table 2.

Table 2. Koestler's Holon and Semantic Duality.

Koestler's Holon	Semantic Field Theory Analogy
Faces inward as a part	Fisher information: local attractor, semantic sensitivity
Faces outward as a whole	Shannon entropy: global resonance, semantic dispersion
Embedded in a hierarchy	Nested semantic manifolds or recursive symmetry layers
Self-regulating	Semantic homeostasis between precision and ambiguity
Janus-faced (two-faced)	Dual operator: dispersion ↔ sensitivity

Each semantic unit (a "holon") could be modeled as a field-theoretic entity with:

- Intrinsic curvature (Fisher information)
- Extrinsic spread (Shannon entropy)
- Recursive symmetry linking its local and global roles

This would allow semantic holons to self-organize, resonate, and attract, forming a dynamic epistemic topology.

Friston's (2006) *Free Energy Principle* emphasizes the dispersion side of the duality. Biological systems minimize *variational free energy*, thereby bounding surprise and aligning generative models with observed data. This is a global coherence principle: it ensures systems remain integrated with their environments by smoothing entropy across hidden states (Friston 2019).

Frieden's (2004) EPI, by contrast, emphasizes the sensitivity side: physical laws arise by extremizing Fisher information and aligning local semantic curvature with intrinsic constraints. This sensitivity is not local or mechanistic—it's global and structural, ensuring that the system's dynamics remain tightly coupled to its ontological substrate (Frieden 2009).

Together, the two principles articulate a semantic duality that presupposes holonic structure. Shannon entropy and Fisher information form the twin operators of inference—one dispersive, one sensitive—whose recursive interplay sustains holonic identity across scales of the epistemic hierarchy.

4. Markov Blankets and Nested Constraint Structures

Any organism, as a holarchy composed of nested sub-organizations, necessarily carries boundaries at multiple scales. These boundaries delineate what is "inside" from what is "outside," structuring perception and action across levels of organization. Within the semantic manifolds $K \subseteq M$, where semantic duality is established, this boundary takes on epistemic significance: it is the threshold between local semantic sensitivity (Fisher information) and global semantic resonance (Shannon entropy).

Crucially, it is only *on the boundary* that classical information is instantiated. What lies beyond the boundary functions as the extrinsic substrate—necessary but inaccessible except through mediation. Friston et al., (2020) formalize this boundary as the Markov blanket, the statistical surface of conditional independence: given a blanket B, internal states are conditionally independent of external states once B is known. This simple principle induces a nested hierarchy of constraints, shaping both local and global inference.

- The blanket acts as a local constraint functional, enforcing conditional independence and regulating the flow of information between interior and exterior.
- The global distribution must respect a recursive nesting of blankets, where each boundary constrains and is constrained by those above and below it, forming a hierarchy of semantic mediation.

In Lagrangian terms, each Markov blanket may be viewed as defining a submanifold with its own local action functional. The global dynamics then emerge not from any single layer, but from the coupling of these submanifolds across scales. This echoes the architecture of multi-scale field theories, in which local constraints reverberate upward, informing and stabilizing global patterns, while global fields propagate downward, constraining local variability.

5. Semantic Field Theory and Recursive Symmetry

Semantic field theory extends probability theory beyond static assignment into a dynamic architecture of fields, symmetries, and constraints. In this framework, a probability distribution is not merely a measure over outcomes, but a structured semantic field shaped by attractors, propagated constraints, and recursive symmetries.

- The probability density p(x) functions as a semantic field over *M*, encoding both local sensitivity and global resonance.
- Constraint functionals C(x) carve out semantic submanifolds, restricting the admissible forms of the field.
- Lagrange multipliers λ(x) act as resonance enforcers, ensuring that competing constraints are balanced into a state of semantic homeostasis.
- Shannon information specifies the global geometry of uncertainty, while Fisher information encodes the local curvature of sensitivity, together defining the manifold's inferential dynamics.

This construction yields a recursive architecture: local constraints, such as those defined by Markov blankets, propagate upward, shaping higher-order distributions; global coherence conditions, such as entropy maximization, propagate downward, guiding the stabilization of local structures. The result is a multi-scale symmetry: each level reflects and constrains the others, such that the system preserves coherence without collapsing diversity.

In effect, semantic field theory establishes a recursive symmetry principle: inference at one scale is never independent, but always folded into the nested dynamics of the whole. This symmetry binds local and global, constraint and freedom, boundary and field into a single variational fabric—an architecture where probability, geometry, and meaning are inseparably coupled.

6. Conclusion: Toward a Unified Metaphysics of Inference

The Lagrangian formalism and probability theory are not merely analogous; they are manifestations of a common principle: structure arises from constraint, and meaning emerges from resonance within a manifold of possibilities. By interpreting probability distributions as semantic fields and treating constraints as variational operators, we arrive at a coherent framework in which inference, dynamics, and semantics are inseparably intertwined.

Within this framework, Shannon entropy governs global dispersion, Fisher information governs local sensitivity, and recursive constraint propagation ensures holistic coherence across scales. Markov blankets and nested submanifolds provide the boundaries through which information flows, enforcing conditional independence while maintaining the integrity of the system. Semantic duality—linking local and global, part and whole—operates as a field-theoretic principle sustaining holonic identity throughout the hierarchy.

Ultimately, this unified perspective suggests that inference itself is a physical and metaphysical process: a dynamic interplay of possibility and necessity, openness and constraint, resonance and curvature. Probability, geometry, and meaning are not separate domains, but facets of a single variational architecture—a metaphysics of inference in which the laws of structure, the propagation of information, and the emergence of significance are all expressions of the same underlying principle.

References

Frieden, B. R., and B.H. Soffer. (1995). Lagrangians of physics and the game of Fisher-information transfer. *Physical Review E*, 52(3), 2274. https://doi.org/10.1103/PhysRevE.52.2274

Frieden, B.R., (2004). Science From Fisher Information: A Unification. *Cambridge University Press.*

Frieden, B.R. (2009). Extreme Physical Information as a Principle of Universal Stability. In *Information Theory and Statistical Learning*, editors F. Emmert-Streib and M. Dehmer, *pp*355-384*.*

Friston, K. (2006). A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1-3), 70–87. https://www.fil.ion.ucl.ac.uk/~karl/A%20free%20energy%20principle%20for%20the%20brain.pdf

Friston, K. (2019). A free energy principle for a particular physics. *arXiv preprint*. https://arxiv.org/abs/1906.10184

Friston, K.J., E.D. Fagerholm, T. S. Zarghami, T. Parr, I. Hipólito, L. Magrou , and A. Razi. (2020). Parcels and particles: Markov blankets in the brain. *Network Neuroscience*, 5(1):211-251.

Friston, K., L. Da Costa, D.A.R. Sakthivadivel, C. Heins, G. Pavliotis, M. Ramstead and T. Parr. (2023). Path integrals, particular kinds, and strange things. Physics of Life Reviews, 47, 35-62.

Jaynes, E.T. (1957a). Information Theory and Statistical Mechanics. *Physical Review,* Series II, 106 (4), 620–630.

Jaynes, E.T. (1957b). Information Theory and Statistical Mechanics II. *Physical Review*. Series II, 108 (2), 171–190.

Shannon, C.E. (1948). A Mathematical Theorem of Communication. *The Bell System Technical Journal*, 27 (4), 623–656.