Text as Substrate: On Semantic Compression and the Emergence of a Universal Representational Medium

Flyxion

March 22, 2026

Abstract

Contemporary compression theory is grounded in the statistical structure of signals: redundancy within waveforms, frequency-domain regularities, and inter-frame correlations are exploited to reduce bitrate while preserving fidelity. This essay proposes and develops a categorically distinct paradigm, in which compression is achieved not through signal approximation but through semantic reduction. Multimodal data is translated into structured textual representations capable of regenerating perceptually equivalent output via a shared generative substrate. Audio is encoded as lexical content augmented by prosodic and vocal metadata; video is represented as structured scenario descriptions composed of agents, actions, environments, and camera dynamics. The result is a universal representational medium in which text functions as a compact, interpretable, and generative interface between raw sensation and structured meaning.

The theoretical development proceeds across three registers. First, an information-theoretic analysis demonstrates how semantic compression redistributes entropy between the description and the generative model, recovering a generalized rate–distortion framework in which model expressiveness replaces channel capacity as the primary resource. Second, a categorical formalization treats compression and reconstruction as adjoint functors between a category of multimodal signals and a category of structured textual descriptions, with natural transformations enforcing cross-modal coherence, fibered categories capturing hierarchical template structure, and sheaf conditions governing the local-to-global assembly of descriptions. Third, a constructive account introduces prototype libraries, structured deviation encoding, and iterative multi-scale refinement as the mechanisms through which this framework becomes computationally feasible.

This convergence of information theory, category theory, and generative modeling redefines compression as an interpretive act rather than a purely numerical one. Its implications extend beyond storage and transmission to touch the foundations of representation, cognition, and the ontology of media itself.

Contents

1 Introduction

Traditional approaches to data compression are grounded in signal processing and statistical estimation. Audio is reduced through frequency-domain approximations, psychoacoustic masking, and scalar quantization (Cover and Thomas, 2006). Video is compressed through motion estimation, inter-frame prediction, and block-transform coding (Shannon, 1948). The mature development of these methods—in codecs such as MP3, AAC, H.265, and AV1—represents decades of progress in exploiting statistical redundancy within well-defined signal domains (Berger, 1971; Gray, 2011). In each case, the guiding principle is fidelity: the compressed representation must permit reconstruction of a signal sufficiently close to the original, as measured by domain-specific distortion criteria.

The implicit assumption underlying all such methods is that the signal itself is the appropriate object of representation. The compressed artifact is a degraded or approximated version of the original waveform or pixel array, not a description of anything. There is no interpreter; there is only an approximation.

A fundamentally different approach becomes possible once two conditions are met: first, that generative models of sufficient fidelity exist, capable of synthesizing high-quality audio and video from compact structured inputs; and second, that the structure underlying such generation can be captured in a computable, transmissible form. Under these conditions, compression need no longer preserve the signal but may instead preserve its *generative description*—the structured account of the information required to produce a perceptually equivalent output. The viability of this approach has been demonstrated in stages by neural generative models (Goodfellow et al., 2014; Kingma and Welling, 2014; van den Oord et al., 2016; Ho et al., 2020; Rombach et al., 2022), each of which learns to generate high-fidelity signals from compact latent descriptions.

This reorientation is not merely a technical refinement but a conceptual transformation. It shifts the object of compression from signals to meanings, from redundancy elimination to semantic abstraction, and from numerical approximation to interpretive inference. In this paradigm, a compressed representation is a *program* for generating experience, not a reduced copy of one.

The present essay develops this claim systematically. Section 2 treats audio as prosodic text; Section 3 develops the corresponding framework for video as structured scenario description. Section 4 establishes the sense in which text constitutes a universal representational medium. Sections 6–14 address foundational questions about perceptual equivalence, expressivity, and information-theoretic optimality. Sections 16–23 develop the categorical formalism: functors, adjunctions, fibered categories, sheaves, monoidal structure, and higher categories. Sections 26–35 treat the constructive mechanisms of prototype-based compression. Sections 36–49 extend the framework through topology, differential geometry, information geometry, stochastic models, causal structure, and algebraic language theory. Sections 52–65 address system design, ethics, economics, and the convergence toward a unified theory of representation. The appendices collect formal

proofs and technical constructions.

Sections 67–74 form a sustained case study in American Sign Language (ASL), demonstrating that the formal architecture—prototype templates, deviation encoding, gauge fixing, global constraint fields, drift dynamics, and scaling adaptation—is not a theoretical construction but a structure independently realized in a richly documented, physiologically grounded human communication system (Riekehof, 1978).

Throughout, the essay maintains the view that compression, understanding, and generation are not three distinct processes but three perspectives on a single underlying structure: the mapping between signals and their generating descriptions.

2 Audio as Prosodic Text

An audio signal is, at the most primitive level of description, a time-indexed sequence of pressure values. The standard compression pipeline—exemplified by MP3 and AAC—exploits the statistical structure of this sequence by working in the frequency domain, applying psychoacoustic models to identify and discard components below perceptual threshold, and encoding the residual with variable-rate quantization (Cover and Thomas, 2006; Pierce, 1980). The result is a compressed bitstream that reconstructs a signal approximating the original within perceptual tolerance.

This approach treats the waveform as opaque. It makes no commitment to what the signal *means*. A speech recording and a piece of music are processed by the same pipeline, which has no access to the linguistic content of the former or the harmonic structure of the latter. The compression is blind to semantics.

Semantic compression proceeds differently. It begins by observing that the vast majority of audio signals of interest to humans are not arbitrary waveforms but instances of a small number of generative processes: speech, music, environmental sound, and their combinations. Each of these processes is governed by structured principles—phonology, prosody, harmonic grammar, physical acoustics—that can be represented explicitly (Chomsky, 1965; Barwise and Perry, 1983).

2.1 Formal Encoding of Speech

For spoken language, the encoding decomposes the signal into its constituent generative parameters. A spoken utterance is characterized by its lexical content, the identity and physiological properties of the speaker, the prosodic contour governing pitch, timing, and rhythm, and the affective or emotional register that modulates delivery.

Definition 2.1 (Audio Semantic Encoding)**.** Let $\mathcal{W}$ denote the space of acoustic waveforms. A *semantic audio encoding* is a map

$$\mathcal{A} : \mathcal{W} \to (T, P, V, E),$$

where T is the textual transcription, $P \in \mathbb{R}^{d_P}$ a vector of prosodic parameters (pitch contour, duration, stress pattern, speech rate), $V \in \mathbb{R}^{d_V}$ a vector of vocal characteristics (fundamental frequency statistics, formant structure, breathiness, speaker identity embedding), and $E \in \mathbb{R}^{d_E}$ an affective parameter vector (arousal, valence, modality).

Reconstruction is performed by a generative model $\mathcal{G}_a : (T, P, V, E) \to \widehat{\mathcal{W}}$, which synthesizes a waveform consistent with the specified parameters. Contemporary neural synthesis architectures demonstrate that high-fidelity waveform generation from structured representations is achievable across a wide range of conditions (van den Oord et al., 2016; van den Oord et al., 2019).

The compression ratio achievable by this representation depends on semantic rather than statistical redundancy. Repeated syntactic structures, consistent speaker identity across an utterance, and predictable prosodic patterns contribute to efficient encoding. In cases where the content is highly predictable—as in formulaic speech, repeated phrases, or acoustically consistent narration—the description length can be orders of magnitude smaller than the raw or even traditionally compressed waveform.

2.2 Non-Speech Audio

For non-speech audio, the semantic encoding must be adapted to the generative structure of the relevant domain. Music can be encoded via score-like representations augmented with performance parameters: tempo, dynamics, articulation, instrument identity, and mixing configuration. Environmental sound can be described in terms of source events, spatial configuration, and acoustic environment parameters.

In each case, the encoded representation captures the generative structure of the signal rather than the signal itself. The waveform becomes a derived object, produced on demand by the generative model from the compact structured description.

3 Video as Scenario Description

Video compression has historically relied on the spatial and temporal redundancy of natural images. Block-based transform coding exploits intra-frame correlations; motion compensation exploits inter-frame correlations; variable-rate entropy coding exploits statistical regularities in the resulting residuals. The H.265 and AV1 codecs represent the mature development of this approach, achieving compression ratios that were unimaginable in the early history of digital video (Csiszár and Körner, 2011).

Yet these methods remain semantically blind. They operate on pixel arrays, not on scenes. They have no representation of agents, objects, actions, or causal relations. A video of a person walking is processed identically to a video of random noise, except that the statistical regularities of the former permit more efficient encoding.

Semantic video compression exploits the fact that natural video is not random but generative: it arises from a physical world containing objects that obey physical laws,

agents that act according to intentions, and cameras that follow trajectories through space. Recent advances in latent diffusion models (Rombach et al., 2022) and text-to-image generation (Ramesh et al., 2021; Esser et al., 2021) demonstrate that high-fidelity reconstruction from compact structured descriptions is computationally feasible.

Definition 3.1 (Video Semantic Encoding)**.** Let $\mathcal{V}$ denote the space of video sequences. A *semantic video encoding* is a map

$$\mathcal{V} : \mathcal{V} \to (S, E, C),$$

where S is a scenario graph (encoding agents, their properties, their actions, and their causal relationships over time), $E \in \mathbb{R}^{d_E}$ is an environment and rendering parameter vector (lighting, materials, weather, visual style), and C is a camera trajectory specification (position, orientation, focal parameters, and their temporal evolution).

3.1 Scenario Graphs

A scenario graph is a directed temporal hypergraph in which nodes represent entities and events, and edges represent causal, spatial, or temporal relations. Each entity node carries attributes: identity, type, physical dimensions, surface appearance, and behavioral disposition. Each event node carries temporal coordinates and a description of the action or state transition it represents. The causal structure of scene events can be modeled formally using structural causal models (Pearl, 2009), which provide a principled language for representing how agent actions and environmental factors jointly determine outcomes.

This structure closely resembles the internal representation used by physically-based game engines and simulation systems. The key difference lies in its use as a *compression format*: instead of storing frames, one stores the causal structure sufficient to regenerate them.

3.2 Camera and Rendering Parameters

The camera specification encodes the observer's trajectory through the scene. Parameters include position and orientation as functions of time, focal length and depth-of-field configuration, and post-processing effects such as color grading, motion blur radius, and noise characteristics. These parameters are themselves compact, often requiring only a small number of keyframes with interpolation.

The reconstruction map

$$\mathcal{G}_v : (S, E, C) \to \widehat{\mathcal{V}}$$

is realized by a neural renderer or physically-based simulation engine that accepts the structured description and produces a corresponding video sequence. The output need not be pixel-identical to the original; it is required only to be perceptually equivalent under the chosen distortion metric.

4 Text as Universal Medium

The convergence of the audio and video frameworks suggests a unifying claim: that text, when extended to include structured symbolic descriptors beyond ordinary natural language, can function as a universal medium for representing multimodal sensory data.

The word "text" here does not denote natural language alone. It denotes any compositional, linearizable, symbolic representation capable of encoding complex hierarchical structure. Text in this sense subsumes natural language, formal notation, data serialization formats, and programming languages. What these share is compositionality: the meaning of a complex expression is determined by the meanings of its parts and the rules by which they are combined (Chomsky, 1965). This property has deep roots in the categorical semantics of formal languages (Lawvere and Schanuel, 2009; Pierce, 2002), where compositional meaning is expressed as a functor from a syntactic category to a semantic one.

Definition 4.1 (Universal Textual Substrate)**.** A *universal textual substrate* for a category of signals $\mathcal{M}$ is a structured symbolic language $\mathcal{L}$ together with a generative interpreter $\cdot : \mathcal{L} \to \mathcal{M}$ such that for every $x \in \mathcal{M}$ and every $\epsilon > 0$, there exists $t \in \mathcal{L}$ with $d(x, t) < \epsilon$.

The information-theoretic content of this definition is that $\mathcal{L}$ is *dense* in $\mathcal{M}$ with respect to the perceptual metric d. Text need not represent every signal exactly but must be able to approximate every signal to arbitrary precision.

The argument for universality rests on two claims. First, the physical world that gives rise to audio and video signals is governed by computable laws, and the space of its configurations is parameterizable in principle. Second, generative models of sufficient expressive power can approximate the map from descriptions to signals; this is the content of the universal approximation theorems for neural networks (Goodfellow, Bengio, and Courville, 2016; LeCun, Bengio, and Hinton, 2015). Together, these claims imply that the space of signals is well-approximated by the image of a sufficiently rich textual language under a sufficiently powerful generative interpreter.

5 The Internal Language of the Perceptual Category

The universality claim of Section 4 can be given a sharper categorical formulation. In categorical logic, every category with sufficient structure admits an *internal language*: a formal syntax whose semantics are given by the category itself (Jacobs, 1999; Lawvere and Schanuel, 2009). We argue that the TIR is precisely the internal language of the perceptual category $\mathcal{M}/\sim_\epsilon$.

Definition 5.1 (Internal Language of $\mathcal{M}$)**.** The *internal language* of the category $\mathcal{M}/\sim_\epsilon$ is a typed formal system $\mathcal{L}_\mathcal{M}$ such that:

1. Types correspond to objects (perceptual equivalence classes).

2. Terms correspond to morphisms (admissible transformations between classes).
3. The interpretation functor $\cdot : \mathcal{F}(\mathcal{L}_{\mathcal{M}}) \to \mathcal{M}/\sim_\epsilon$ is full and faithful.

Fullness and faithfulness together mean that every perceptual transformation is expressible in the language, and that syntactically distinct terms with the same semantics are identified. The TIR is an approximation to $\mathcal{L}_{\mathcal{M}}$: it is expressive enough to generate the relevant signals and compositional enough to reflect the monoidal structure of $\mathcal{M}$, but it relaxes exact faithfulness in favor of approximate perceptual equivalence.

Proposition 5.2 (TIR as Approximate Internal Language)**.** *If $\mathcal{L}$ is ϵ-semantically complete for $\mathcal{M}$ (Definition 8.1), then $\mathcal{L}$ is an ϵ-approximate internal language of $\mathcal{M}/\sim_\epsilon$ in the sense that the interpretation functor $\cdot$ is essentially surjective and approximately faithful.*

This perspective elevates the TIR from a convenient representation format to a principled linguistic object: it is not merely that text *can* describe signals, but that the appropriate description language *is* the internal language of the perceptual category. The design of the TIR is therefore not an engineering choice but a categorical one, constrained by the structure of perception itself.

6 Formal Semantics of Perceptual Equivalence

A central but implicit concept throughout the preceding development is that of perceptual equivalence. The categorical constructions, rate–distortion analysis, and fixed-point theorems all depend on a metric d that captures when two signals are "the same" for a given task. We now make this notion explicit.

Definition 6.1 (Perceptual Equivalence Relation)**.** Let $(\mathcal{M}, d)$ be an enriched category of signals. For $\epsilon > 0$, define an equivalence relation $\sim_\epsilon$ by

$$x \sim_\epsilon y \iff d(x, y) < \epsilon.$$

The equivalence classes $[x]_\epsilon$ represent perceptual indistinguishability at tolerance ϵ.

The quotient $\mathcal{M}/\sim_\epsilon$ is not merely a set but inherits categorical structure: morphisms descend when they preserve equivalence classes. Semantic compression is well-posed only with respect to this quotient. The appropriate perceptual metric is not uniquely determined but depends on the modality and task: for speech, intelligibility and speaker identity weigh heavily; for video, structural coherence and motion consistency matter more than exact photometric accuracy (Amari, 2016; Bishop, 2006).

Proposition 6.2 (Well-Definedness of Semantic Compression)**.** *If $\mathcal{G}$ is non-expansive and $\mathcal{C}$ is chosen to minimize distortion, then $\mathcal{C}$ factors through $\mathcal{M}/\sim_\epsilon$.*

Proof. Suppose $x \sim_\epsilon y$, i.e., $d(x, y) < \epsilon$. Since $\mathcal{G}$ is non-expansive, $d(\mathcal{G}(\mathcal{C}(x)), \mathcal{G}(\mathcal{C}(y))) \leq d(\mathcal{C}(x), \mathcal{C}(y))$. The distortion-minimizing condition ensures that the descriptions $\mathcal{C}(x)$

and $\mathcal{C}(y)$ agree up to the tolerance ϵ imposed by $\sim_\epsilon$, so $\mathcal{C}$ cannot distinguish between perceptually equivalent signals. □

This formalizes the idea that compression need only preserve equivalence classes, not exact signals. It also clarifies that the "loss" in lossy compression is precisely the collapse of $\sim_\epsilon$ classes: the encoder maps each class to a single representative description, and the decoder maps that description back to a canonical element of the class.

7 Observer-Relative Semantics

The equivalence relation $\sim_\epsilon$ introduced in Section 6 treats perceptual tolerance as a fixed global parameter. In practice, however, the appropriate notion of equivalence depends on the observer: the task being performed, the physiological characteristics of the perceiver, and the deployment context all bear on which distinctions matter.

Definition 7.1 (Observer-Indexed Perceptual Metric)**.** Let $\mathcal{O}$ be a set of observer types. For each $\omega \in \mathcal{O}$, let $d_\omega : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$ be a perceptual metric calibrated to observer type ω. The corresponding equivalence relation is

$$x \sim_{\epsilon,\omega} y \iff d_\omega(x, y) < \epsilon.$$

Different observer types partition $\mathcal{M}$ into different equivalence classes. A signal that is effectively lossless for a casual viewer may contain distinctions that are perceptually significant for an expert listener or a clinical diagnostician. Semantic compression is therefore not a single map but a family of maps parameterized by $\mathcal{O}$:

$$\mathcal{C}_\omega : \mathcal{M} \to \mathcal{T}_\omega,$$

where $\mathcal{T}_\omega$ is the description category appropriate to observer type ω.

Proposition 7.2 (Monotonicity of Compression under Observer Refinement)**.** *If $d_{\omega_1}(x, y) \leq d_{\omega_2}(x, y)$ for all x, y (observer ω_1 is coarser than ω_2), then $R_{\mathcal{G}}^{\omega_1}(D) \leq R_{\mathcal{G}}^{\omega_2}(D)$ for all D.*

Proof sketch. A coarser observer identifies more signals as equivalent, collapsing more equivalence classes. The resulting quotient category $\mathcal{M}/\sim_{\epsilon,\omega_1}$ has fewer objects than $\mathcal{M}/\sim_{\epsilon,\omega_2}$, so the information required to index within it is lower, reducing the achievable rate. □

Observer-relative semantics also governs the design of the TIR. The description language must be parameterized by observer type, encoding finer distinctions for more demanding observers and coarser ones for casual use. This connects the formal framework to user-adaptive media systems and accessibility engineering, where the granularity of reconstruction is matched to the needs and capacities of the receiver.

8 Expressivity of the Textual Language

The claim that text forms a universal substrate depends critically on the expressive power of the textual intermediate representation $\mathcal{L}$. This expressivity can be formalized using concepts from formal language theory and categorical semantics (Hopcroft, Motwani, and Ullman, 2006; Pierce, 2002).

Let $\mathcal{F}(\mathcal{L})$ be the free category generated by the grammar of $\mathcal{L}$. The semantic interpretation functor

$$\cdot : \mathcal{F}(\mathcal{L}) \to \mathcal{M}$$

is said to be *essentially surjective up to* ϵ if for every $x \in \mathcal{M}$, there exists $t \in \mathcal{L}$ such that $d(x, t) < \epsilon$.

Definition 8.1 (Semantic Completeness)**.** The textual language $\mathcal{L}$ is *ϵ-semantically complete* for $\mathcal{M}$ if $\cdot$ is essentially surjective up to ϵ.

Theorem 8.2 (Universality Criterion)**.** *If $\mathcal{L}$ is ϵ-semantically complete and $\mathcal{G}$ is continuous with respect to d, then semantic compression achieves universal approximation on $\mathcal{M}$.*

Proof sketch. By ϵ-semantic completeness, for every $x \in \mathcal{M}$ there exists $t \in \mathcal{L}$ with $d(x, t) < \epsilon$. Taking $\mathcal{C}(x) = t$ and $\mathcal{G} = \cdot$ yields $d(x, \mathcal{G}(\mathcal{C}(x))) < \epsilon$. Continuity of $\mathcal{G}$ ensures that the approximation is stable under small perturbations of the description. □

This result makes precise the requirement that the textual language must be sufficiently expressive to capture all relevant generative structure. Insufficient expressivity manifests as irreducible residuals in compression—signals whose structure lies outside the language's generative scope. The design of the textual intermediate representation is therefore not arbitrary but constrained by expressivity requirements relative to the signal domain.

9 Information-Theoretic Foundations

9.1 The Semantic Rate–Distortion Function

Classical rate–distortion theory, due to Shannon (1948), characterizes the achievable trade-off between coding rate and reconstruction distortion for a given source distribution and distortion measure (Cover and Thomas, 2006; Berger, 1971). The rate–distortion function $R(D)$ specifies the minimum rate required to achieve expected distortion at most D:

$$R(D) = \inf_{q(t|x):\mathbb{E}[d(x,\hat{x})]\leq D} I(X;T),$$

where the infimum is over all conditional distributions $q(t|x)$ and $\hat{x}$ is the reproduction of x under the encoder-decoder pair.

Semantic compression modifies this framework in an essential way: the reconstruction is performed not by a simple lookup or linear mapping but by a generative model $\mathcal{G}$ that encodes substantial prior knowledge about the structure of signals in $\mathcal{M}$. The distortion becomes $d(x, \mathcal{G}(t))$, and the mutual information $I(X;T)$ quantifies only the information required to specify the description t, not the information required to reconstruct x from scratch.

Definition 9.1 (Semantic Rate–Distortion Function)**.** Given a source X over $\mathcal{M}$, a generative model $\mathcal{G} : \mathcal{T} \to \mathcal{M}$, and a distortion functional d, the *semantic rate–distortion function* is

$$R_{\mathcal{G}}(D) = \inf_{q(t|x)} \left\{ I(X;T) \,\middle|\, \mathbb{E}_{x,t}\left[d\big(x, \mathcal{G}(t)\big) \right] \leq D \right\}.$$

Proposition 9.2. *For any generative model $\mathcal{G}$ and distortion D, $R_{\mathcal{G}}(D) \leq R(D)$, with equality when $\mathcal{G}$ implements the identity. Moreover, $R_{\mathcal{G}}(D)$ is non-increasing in the expressive power of $\mathcal{G}$.*

Proof sketch. Any encoder-decoder pair achieving rate R and distortion D in the classical sense can be replicated in the semantic framework by setting $\mathcal{G}(t) = \hat{x}(t)$, recovering the same rate. When $\mathcal{G}$ incorporates additional structural knowledge, the same distortion can be achieved with a description t of lower information content, since the generative model fills in what the description does not specify. Monotonicity follows from the fact that a more expressive model can always be used as a less expressive one by ignoring the additional capacity. □

9.2 The Information Bottleneck Connection

The semantic rate–distortion framework is closely related to the information bottleneck principle of Tishby and Zaslavsky (2015), which characterizes the optimal compression of a source X with respect to a relevance variable Y:

$$\min_{q(t|x)} \left[I(X;T) - \beta\, I(T;Y) \right].$$

In the semantic setting, the relevance variable Y can be interpreted as the space of perceptual categories relevant to the downstream task. A description that preserves the perceptually relevant structure of a signal while discarding perceptually irrelevant detail achieves a favorable bottleneck trade-off. This interpretation connects semantic compression directly to the theory of perceptual relevance and task-dependent coding.

10 Semantic Channels and Communication Limits

Classical information theory characterizes communication in terms of channel capacity: the maximum rate at which information can be reliably transmitted over a noisy channel (Shannon, 1948; Cover and Thomas, 2006). Semantic compression redefines the relevant channel, replacing signal-level transmission with description-level transmission.

Definition 10.1 (Semantic Channel)**.** A *semantic channel* is a triple $(\mathcal{T}, \mathcal{G}, \mathcal{N})$ consisting of a description space $\mathcal{T}$, a shared generative model $\mathcal{G}$, and a noise model $\mathcal{N}$ on $\mathcal{T}$ representing transmission errors at the description level.

The capacity of a semantic channel is not determined by bandwidth in the classical sense. Since the transmitted object is a textual description rather than a raw signal, the relevant constraint is the entropy of the description space $H(T)$ rather than the entropy of the signal $H(X)$. By Corollary 51.2, $H(T) = H(X) - I(X; \mathcal{G})$, so increasing the expressive power of $\mathcal{G}$ directly increases effective channel capacity.

Proposition 10.2 (Semantic Channel Capacity)**.** *The effective capacity of a semantic channel at distortion level D is*

$$C_{\mathcal{G}}(D) = \log |\mathcal{T}| - H(T \mid \mathcal{G}, D),$$

where $H(T \mid \mathcal{G}, D)$ is the residual entropy of descriptions given the model and the distortion constraint.

The semantic channel also has distinctive robustness properties. Because the transmitted object is a structured description rather than a raw bitstream, errors in transmission are errors in the description space, and the constraint structure of the TIR provides error-correcting capabilities: an invalid parameter value may often be corrected by projection onto the nearest valid point in $\mathcal{T}$. This is fundamentally different from the brittleness of classical compressed bitstreams, where single-bit errors can produce catastrophic artifacts.

11 Compression as Interpretation: The Epistemic Dimension

The shift from signal-based to semantic compression reframes compression as an epistemic act. The encoder is not merely measuring statistical properties of a signal but *interpreting* it: identifying the underlying generative structure, assigning it to a category of meaningful events, and expressing the result in a form legible to a shared generative substrate.

This interpretive character distinguishes semantic compression from all prior compression paradigms. It introduces a dependency on a shared prior—the generative model—that must be established between encoder and decoder before any communication can occur. The compressed representation is not self-contained; it is an instruction issued within a shared semantic context.

This has several consequences. First, the efficiency of compression depends on the match between the generative model's prior and the actual distribution of signals. Second, compression and understanding become inseparable: an encoder that cannot understand a signal cannot compress it semantically. Third, errors in interpretation propagate into errors in reconstruction, introducing a qualitatively different failure mode from the artifact-based distortions of signal compression.

The compression–interpretation equivalence can be made precise in the framework of minimum description length (Rissanen, 1978; Grünwald, 2007): the optimal compressed

description of a signal is the shortest description relative to a model, and finding this description requires solving the same inference problem as understanding the signal within the model. This connection places semantic compression within the tradition of Solomonoff induction (Solomonoff, 1964) and algorithmic probability (Chaitin, 1987; Hutter, 2005): the optimal compressor is an optimal predictor.

12 Duality Between Compression and Prediction

The preceding section establishes that compression and interpretation are the same inferential act viewed from different directions. This duality extends to a formal correspondence between compression and prediction that can be stated as a theorem.

Let X be a random variable over $\mathcal{M}$ representing a signal, and suppose X has a causal temporal structure: $X = (X_1, X_2, \ldots, X_T)$ where X_t may depend on prior observations. Prediction corresponds to estimating $X_{>t}$ given $X_{\leq t}$; compression corresponds to encoding X with respect to the model $\mathcal{G}$.

Theorem 12.1 (Compression–Prediction Duality)**.** *Under a shared generative model $\mathcal{G}$, the optimal compressor—the one minimizing expected description length—is identical to the optimal predictor—the one maximizing the log-likelihood of future observations under $\mathcal{G}$.*

Proof sketch. The expected description length of an optimal code is $H(X \mid \mathcal{G}) = -\mathbb{E}[\log p_{\mathcal{G}}(X)]$ by Shannon's source coding theorem. The optimal predictor maximizes $\mathbb{E}[\log p_{\mathcal{G}}(X)]$. These objectives are therefore identical in sign, with the compressor minimizing and the predictor maximizing the same quantity. □

Theorem 12.1 has a striking corollary: the act of compressing a signal and the act of understanding it well enough to predict its future evolution are computationally equivalent under a shared generative model. An encoder that achieves optimal description length necessarily models the signal's future trajectory; a predictor that maximizes log-likelihood necessarily produces optimal compressed representations.

This duality is not merely theoretical. It implies that any advance in generative model quality—improving prediction accuracy—directly translates into a reduction in compressed description length, and vice versa. The research programs of generative modeling and semantic compression are therefore not parallel but identical: both seek the internal language (Section 5) through which signals can be expressed most concisely.

13 Kolmogorov Complexity and Model-Relative Compression

Kolmogorov complexity provides a resource-independent characterization of information content (Kolmogorov, 1965; Chaitin, 1987; Li and Vitányi, 2008). The Kolmogorov complexity $K(x)$ of a string x is the length of the shortest program p on a universal Turing

machine U such that $U(p) = x$. Classical Kolmogorov complexity does not depend on any statistical model and constitutes an absolute measure of descriptive complexity.

Semantic compression can be understood as computing a *model-relative* Kolmogorov complexity. Given a generative model $\mathcal{G}$, the model-relative complexity of a signal x at tolerance ϵ is

$$K_\epsilon(x \mid \mathcal{G}) = \min\{|t| : d(x, \mathcal{G}(t)) < \epsilon\},$$

where $|t|$ denotes the description length of t.

Proposition 13.1 (Model Compression Bound)**.** *For any generative model $\mathcal{G}$ and signal x,*

$$K_\epsilon(x \mid \mathcal{G}) \leq K_\epsilon(x) - K(\mathcal{G}) + O(1),$$

where $K_\epsilon(x)$ is the ϵ-approximate Kolmogorov complexity of x and $K(\mathcal{G})$ is the complexity of the model itself.

Proposition 13.1 formalizes the intuition that complexity can be offloaded into the shared model. The total description cost of transmitting x is approximately $K_\epsilon(x \mid \mathcal{G}) + K(\mathcal{G})$. When $\mathcal{G}$ is large but shared (i.e., its cost is amortized over many transmissions), the per-signal cost is determined entirely by $K_\epsilon(x \mid \mathcal{G})$, which can be dramatically smaller than $K_\epsilon(x)$ for signals well within $\mathcal{G}$'s distribution.

This analysis connects semantic compression to the *two-part code* formulation of MDL (Rissanen, 1978; Grünwald, 2007): the optimal code consists of a model description followed by a data description conditional on the model.

14 Entropy Redistribution and Model Dependence

A central claim of semantic compression is that it does not eliminate entropy but redistributes it between the description and the generative model. Let $H_\mathcal{G}(X)$ denote the entropy of the source as seen by the generative model: the entropy of the minimal description T required for reconstruction to within tolerance ϵ. The conservation relation

$$H(X) \approx H_\mathcal{G}(X) + I_\mathcal{G}(X)$$

holds, where $I_\mathcal{G}(X)$ measures the information about X that is implicitly encoded in $\mathcal{G}$—the reduction in description entropy attributable to the model's prior knowledge.

Increasing the expressive power of $\mathcal{G}$ increases $I_\mathcal{G}(X)$ and therefore decreases $H_\mathcal{G}(X)$, enabling more efficient compression. In the limit where $\mathcal{G}$ perfectly models the generative process producing X, the residual entropy $H_\mathcal{G}(X)$ approaches the entropy of the model's stochastic variation, which may be negligible.

This perspective places semantic compression within the broader tradition of *predictive coding* (Friston, 2010; Clark, 2015): compression is efficient to the extent that the model accurately predicts the signal, requiring only the transmission of unpredicted residuals.

The difference is that predictive coding operates at the level of individual signal values, while semantic compression operates at the level of generative programs. This elevation from signal-level to program-level prediction is precisely what enables the dramatic compression ratios that semantic methods promise.

15 Complexity of Encoding and Decoding

Beyond information-theoretic optimality, practical viability depends on computational complexity. The encoding process consists of prototype search and deviation optimization, while decoding consists of generative execution. These operations are not symmetric, and understanding their asymmetry illuminates the fundamental structure of the paradigm.

Let $|\mathcal{P}|$ denote the size of the prototype library and n the dimensionality of the embedding space. Approximate nearest-neighbor search admits sublinear complexity $O(\log |\mathcal{P}|)$ under suitable indexing structures. Deviation optimization requires iterative evaluation of $\mathcal{G}$ and its gradients. If $\mathcal{G}$ has cost $C_{\mathcal{G}}$, then k-step optimization has cost $O(k \cdot C_{\mathcal{G}})$.

Proposition 15.1 (Asymptotic Encoding Cost)**.** *Under approximate nearest-neighbor search and bounded optimization steps,*

$$\text{Cost}_{\text{encode}}(x) = O(\log |\mathcal{P}| + k \cdot C_{\mathcal{G}}).$$

Decoding cost is dominated by generative rendering:

$$\text{Cost}_{\text{decode}}(t) = O(C_{\mathcal{G}}).$$

This asymmetry—encoding is more expensive than decoding—is fundamental and not incidental. It reflects the interpretive nature of compression: searching for the best description requires more computation than executing the description once found. This mirrors classical results in algorithmic information theory (Li and Vitányi, 2008), where finding a short description is computationally harder than executing it. In practical deployments, it suggests an asymmetric architecture in which encoding is performed by powerful servers while decoding is performed by lightweight client devices.

16 Categorical Structure of Semantic Compression

The transition from signal-based to semantic compression admits a natural and illuminating formulation in category-theoretic terms (Mac Lane, 1998; Riehl, 2016). Rather than viewing compression as a function between sets of bit-strings, we treat it as a *functor* between structured categories whose morphisms capture admissible transformations.

Definition 16.1 (Category of Multimodal Signals)**.** Let $\mathcal{M}$ be the category whose objects are multimodal signals (audio streams, video sequences, composite recordings) and whose

morphisms are admissible perceptual transformations: temporal rescaling, perspective change, transposition, normalization, and analogous operations.

Definition 16.2 (Category of Textual Descriptions)**.** Let $\mathcal{T}$ be the category whose objects are structured textual descriptions (in the sense of Section 4) and whose morphisms are interpretability-preserving transformations: paraphrase, refinement, abstraction, compositional extension.

Semantic compression and reconstruction are then captured by a pair of functors:

$$\mathcal{C} : \mathcal{M} \to \mathcal{T}, \tag{1}$$

$$\mathcal{G} : \mathcal{T} \to \mathcal{M}. \tag{2}$$

The composite $\mathcal{G} \circ \mathcal{C}$ approximates the identity on $\mathcal{M}$ up to perceptual equivalence. The quality of compression depends on how closely this composite approximates the identity functor $\mathrm{Id}_{\mathcal{M}}$ under the perceptual metric. The categorical formulation follows the standard development of compositional systems (Mac Lane, 1998; Spivak, 2014; Fong and Spivak, 2019), with the additional feature that the functors are approximate rather than exact.

16.1 Adjoint Structure

The pair $(\mathcal{C}, \mathcal{G})$ exhibits an adjoint-like structure. In an exact adjunction, there would exist natural transformations

$$\eta : \mathrm{Id}_{\mathcal{M}} \Rightarrow \mathcal{G} \circ \mathcal{C}, \qquad \varepsilon : \mathcal{C} \circ \mathcal{G} \Rightarrow \mathrm{Id}_{\mathcal{T}},$$

satisfying the triangle identities. In the semantic compression setting, these identities hold only approximately: η_x maps each signal to its reconstruction up to perceptual tolerance, and ε_t maps each description to a canonical compressed form of its reconstruction.

The deviation from exact adjointness measures the *semantic gap*: the extent to which compression introduces irreversible abstraction. A system with small semantic gap approximates an equivalence of categories; a system with large semantic gap loses significant perceptual information in the encoding step.

Perfect reconstruction—in the sense of categorical equivalence—corresponds to the existence of a functor $\mathcal{C}$ such that $\mathcal{G} \circ \mathcal{C} \cong \mathrm{Id}_{\mathcal{M}/\sim}$, where $\mathcal{M}/\sim$ is the quotient category obtained by identifying perceptually equivalent signals.

17 Adjoint–Variational Correspondence

The adjoint-like structure of the compression–reconstruction pair and the variational formulation of Section 28 are not independent observations but two expressions of the same underlying duality. We now make this correspondence explicit.

Recall that the variational objective minimizes

$$\mathcal{L}(z; x) = d\big(x, \mathcal{G}(z)\big) + \lambda L(z).$$

The first term measures the failure of $\mathcal{G} \circ \mathcal{C}$ to approximate the identity—precisely the extent to which the unit $\eta : \mathrm{Id}_{\mathcal{M}} \Rightarrow \mathcal{G} \circ \mathcal{C}$ deviates from a natural isomorphism. The second term penalizes description complexity, measuring the failure of the counit $\varepsilon : \mathcal{C} \circ \mathcal{G} \Rightarrow \mathrm{Id}_{\mathcal{T}}$ to be an isomorphism.

Theorem 17.1 (Adjoint–Variational Correspondence)**.** *The variational objective $\mathcal{L}$ is minimized if and only if $(\mathcal{C}, \mathcal{G})$ achieves an approximate adjunction in which both the unit and counit deviations are simultaneously controlled.*

Proof sketch. Let $\delta_\eta(x) = d(x, \mathcal{G}(\mathcal{C}(x)))$ denote the unit deviation at x, and $\delta_\varepsilon(t) = d_{\mathcal{T}}(t, \mathcal{C}(\mathcal{G}(t)))$ the counit deviation at t. The variational objective satisfies $\mathcal{L}(z; x) = \delta_\eta(x) + \lambda L(z)$. Minimizing over $z = \mathcal{C}(x)$ jointly minimizes the unit deviation (via the distortion term) and the description length (via L). The counit deviation is controlled through the second term: if $\mathcal{C}(\mathcal{G}(t)) \neq t$, then the round-trip $x \mapsto \mathcal{G}(\mathcal{C}(x)) \mapsto \mathcal{C}(\mathcal{G}(\mathcal{C}(x)))$ accumulates additional cost, penalizing counit failures at the next compression step. The infimum of $\mathcal{L}$ therefore corresponds to the pair $(\mathcal{C}, \mathcal{G})$ achieving minimal simultaneous deviation in both triangle identities. □

Theorem 17.1 reveals that training a semantic compression system is equivalent to finding the best approximate adjunction between $\mathcal{M}$ and $\mathcal{T}$. The parameter λ controls the relative weight placed on the two triangle identities: large λ enforces the counit (descriptions should compress cleanly) at the expense of the unit (reconstruction fidelity), while small λ reverses this priority. The optimal λ corresponds to the unique adjunction depth at which the categorical and information-theoretic objectives are in equilibrium.

18 Hierarchical Templates as Objects in a Fibered Category

Semantic descriptions are rarely flat. They exhibit hierarchical organization in which high-level structures—narrative templates, scene archetypes, prosodic patterns—constrain and generate lower-level detail. The mathematical structure appropriate to this hierarchy is that of a *fibered category* (Grothendieck fibration), introduced by Grothendieck (1959) and developed extensively in (Jacobs, 1999; Mac Lane and Moerdijk, 1992).

Definition 18.1 (Grothendieck Fibration for Templates)**.** Let $\mathcal{B}$ be a base category whose objects are abstract templates (scene layouts, narrative structures, conversational patterns, acoustic environments) and whose morphisms are template refinements. Over each $b \in \mathcal{B}$, define a *fiber category* $\mathcal{F}_b$ consisting of concrete instantiations of template b.

A *fibration* $\pi : \mathcal{E} \to \mathcal{B}$ is a functor from the total category $\mathcal{E}$ to the base, satisfying the Cartesian lifting condition: for every morphism $f : b \to b'$ in $\mathcal{B}$ and every object $e' \in \mathcal{F}_{b'}$, there exists a Cartesian morphism over f with codomain e'.

A semantic description is then a *section* of π: a functor $\sigma : \mathcal{B} \to \mathcal{E}$ such that $\pi \circ \sigma = \mathrm{Id}_{\mathcal{B}}$. Compression consists in selecting a base template b and specifying only the fiber element $\sigma(b) \in \mathcal{F}_b$, i.e., the deviations from the template that individuate the particular signal.

The Cartesian lifting condition ensures that template morphisms can be lifted consistently to the level of concrete descriptions—a coherence property essential for the modular assembly of complex descriptions from hierarchical components. Template-based compression is efficient by design, with efficiency increasing as the template library approaches coverage of the signal space.

19 Compositionality and Monoidal Structure

Both semantic descriptions and generative processes exhibit compositional structure that warrants a symmetric monoidal categorical treatment (Mac Lane, 1998; Coecke and Kissinger, 2017). The tensor product

$$\otimes : \mathcal{T} \times \mathcal{T} \to \mathcal{T}$$

captures the independent composition of descriptions: two sub-scenes, two simultaneous audio streams, or two independent agents can be described independently and composed without interaction.

Definition 19.1 (Monoidal Functors)**.** The compression and reconstruction functors $\mathcal{C}$ and $\mathcal{G}$ are *monoidal* if they preserve the tensor product structure up to coherent natural isomorphism:

$$\mathcal{C}(x \otimes y) \cong \mathcal{C}(x) \otimes \mathcal{C}(y), \qquad \mathcal{G}(s \otimes t) \cong \mathcal{G}(s) \otimes \mathcal{G}(t).$$

Monoidality ensures that compression and reconstruction respect independence. In practice, independence is approximate: two sub-scenes may share lighting, background audio may correlate with foreground events, and agent actions may causally constrain one another. The monoidal structure provides an idealized baseline, with departures captured by explicit interaction terms in the scenario graph. The symmetric monoidal category also accommodates braiding: $x \otimes y \cong y \otimes x$, ensuring that the composition order of independent components does not affect the result.

20 Natural Transformations and Cross-Modal Consistency

A central requirement of multimodal semantic compression is cross-modal coherence. Let $\mathcal{C}_a : \mathcal{M}_a \to \mathcal{T}$ and $\mathcal{C}_v : \mathcal{M}_v \to \mathcal{T}$ denote the audio and video compression functors, mapping into a common textual description category. Cross-modal consistency is expressed by a *natural transformation*

$$\eta : \mathcal{C}_a \circ \Phi \Rightarrow \mathcal{C}_v,$$

where $\Phi : \mathcal{M}_v \to \mathcal{M}_a$ is a synchronization functor extracting the audio track from a video. The naturality condition requires that for any morphism $f : v \to v'$ in $\mathcal{M}_v$, the square

$$\begin{array}{ccc} \mathcal{C}_a(\Phi(v)) & \xrightarrow{\mathcal{C}_a(\Phi(f))} & \mathcal{C}_a(\Phi(v')) \\ {\scriptstyle \eta_v}\downarrow & & \downarrow{\scriptstyle \eta_{v'}} \\ \mathcal{C}_v(v) & \xrightarrow[\mathcal{C}_v(f)]{} & \mathcal{C}_v(v') \end{array}$$

commutes. Naturality ensures that cross-modal consistency is preserved under all admissible signal transformations: temporal editing, perspective change, and resampling all propagate consistently from video to audio descriptions.

In reconstruction, an analogous natural transformation governs the synchronized generation of audio and video from a joint textual description, ensuring that the generative processes for different modalities remain coordinated. The categorical framework of (Fong and Spivak, 2019; Baez and Stay, 2010) provides a natural language for expressing such multi-modal coherence requirements.

21 Sheaf-Theoretic Gluing of Local Descriptions

Complex signals extend over continuous spacetime domains and are naturally analyzed by decomposing them into local regions. This problem has an exact mathematical counterpart in sheaf theory (Mac Lane and Moerdijk, 1992; SGA 4, 1972).

Definition 21.1 (Description Sheaf)**.** Let X be a topological space representing the spacetime domain of a signal. A *description sheaf* $\mathcal{S}$ assigns to each open subset $U \subseteq X$ a set $\mathcal{S}(U)$ of admissible textual descriptions over U, together with restriction maps $\rho_{UV} : \mathcal{S}(U) \to \mathcal{S}(V)$ for $V \subseteq U$.

$\mathcal{S}$ is a sheaf if it satisfies the gluing axiom: for any open cover $\{U_i\}$ of U and any collection of local descriptions $t_i \in \mathcal{S}(U_i)$ satisfying $\rho_{U_i,U_i\cap U_j}(t_i) = \rho_{U_j,U_i\cap U_j}(t_j)$ for all i, j, there exists a unique $t \in \mathcal{S}(U)$ such that $\rho_{U,U_i}(t) = t_i$ for all i.

The sheaf condition formalizes that local consistency implies global existence. When local fragments agree on overlap regions, they assemble into a unique global description. This supports both modular compression of large signals and partial description, where some regions are specified in detail while others remain abstract.

22 Enriched Categories and Metric Semantics

The equivalence between compressed and reconstructed data is perceptual rather than exact. This suggests enriching the categorical structure over a metric space (Lawvere and Schanuel, 2009).

Definition 22.1 (Metric Enrichment)**.** An $\mathbb{R}_{\geq 0}$-enriched category (Lawvere metric space) $\mathcal{M}_d$ assigns to each pair of objects (x, y) a non-negative real number $d(x, y)$ representing their perceptual dissimilarity, satisfying $d(x, x) = 0$ and $d(x, z) \leq d(x, y) + d(y, z)$.

In this setting, the compression-reconstruction composite $\mathcal{G} \circ \mathcal{C}$ is required to satisfy $d(x, \mathcal{G}(\mathcal{C}(x))) < \epsilon$ for a task-dependent tolerance ϵ. Different modalities and tasks call for different enrichments. The metric enrichment also enables a precise formulation of the trade-off between compression and fidelity: a more aggressive compression corresponds to a functor $\mathcal{C}$ that maps more distant signals to the same description, accepting larger distortions in exchange for shorter descriptions.

23 Higher Categories and Generative Histories

The generative process underlying reconstruction is inherently dynamic. Two distinct generative sequences may produce perceptually equivalent outputs. The space of equivalences between generative processes requires higher-categorical treatment (Lurie, 2009).

In a *2-category* framework:

- 0-cells (objects) correspond to textual descriptions.
- 1-cells (morphisms) correspond to generative processes mapping descriptions to signals.
- 2-cells (morphisms between morphisms) correspond to equivalences between generative processes that produce perceptually indistinguishable outputs.

The 2-categorical structure captures the multiplicity of valid realizations. Rather than selecting a canonical reconstruction, the framework maintains the full space of equivalent generative histories—particularly relevant for stochastic generative models (Rezende, Mohamed, and Wierstra, 2014; Song and Ermon, 2019), where many outputs may satisfy the same description. Higher cells can also capture reparameterization equivalences between descriptions that are syntactically distinct but semantically identical.

24 Semantic Compression as a Limit Construction

The derivation of a compact description from a complex signal can be interpreted as a limit construction in the category of representations (Mac Lane, 1998; Lorégian, 2021).

Definition 24.1 (Description Diagram)**.** A *description diagram* for signal x is a diagram $\mathcal{D}_x : J \to \mathcal{T}$, where each object $\mathcal{D}_x(j)$ is a candidate description for x and morphisms in J are refinement relations.

The compressed representation of x is the *limit* of this diagram:

$$\mathcal{C}(x) = \lim_{J} \mathcal{D}_x,$$

the most constrained description consistent with all imposed semantic constraints. Dually, reconstruction can be viewed as a *colimit*, assembling a signal from a diagram of generative components. The duality between limits (compression) and colimits (reconstruction) reflects the fundamental asymmetry between the interpretive and generative directions of the compression cycle.

25 Functorial Factorization of Representation

The mapping from signals to text typically proceeds through intermediate stages. A video is first parsed into a scene graph; the scene graph is then verbalized into a textual description. This corresponds to a factorization of the compression functor.

Definition 25.1 (Two-Stage Factorization)**.** Let $\mathcal{S}$ be an intermediate category of structured semantic graphs. A *two-stage factorization* of $\mathcal{C}$ consists of functors

$$\mathcal{C}_1 : \mathcal{M} \to \mathcal{S}, \qquad \mathcal{C}_2 : \mathcal{S} \to \mathcal{T},$$

such that $\mathcal{C} = \mathcal{C}_2 \circ \mathcal{C}_1$.

The category $\mathcal{S}$ serves as the locus of *meaning*: it is where signal and text meet, where the referential content of a description is anchored to the structural properties of the signal. Reconstruction factorizes analogously as $\mathcal{G} = \mathcal{G}_1 \circ \mathcal{G}_2$, where $\mathcal{G}_2$ parses text into semantic structure and $\mathcal{G}_1$ renders structure into signals. This layered decomposition mirrors the architecture of modern vision-language models (Radford et al., 2021) and neural renderers, providing a categorical framework for their analysis.

26 Prototype Libraries and Generative Priors

The abstract framework developed in the preceding sections becomes computationally tractable through the introduction of prototype libraries. Rather than searching an unconstrained space of textual descriptions, the encoder selects from a structured library of canonical templates and parameterizes departures from them.

Definition 26.1 (Prototype Category)**.** A *prototype category* $\mathcal{P}$ is a subcategory of $\mathcal{T}$ whose objects are canonical parameterized templates—scenario archetypes, acoustic environments, prosodic patterns—each corresponding to a generative program capable of producing entire families of signals. Morphisms in $\mathcal{P}$ are refinement relations between templates.

From a categorical perspective, $\mathcal{P}$ functions as a *basis* for $\mathcal{T}$: every description can be approximated by a prototype plus a structured modification. The quality of this approximation depends on the density and diversity of the prototype library.

26.1 Nearest-Prototype Projection

Given an input signal $x \in \mathcal{M}$, compression begins by identifying the prototype that best approximates x:

$$\pi(x) = \arg\min_{p \in \mathcal{P}} d\big(x, \mathcal{G}(p)\big). \tag{3}$$

For large libraries, exact nearest-prototype search is intractable. One therefore introduces an embedding $\phi : \mathcal{M} \to \mathbb{R}^n$ and an index embedding $\iota : \mathcal{P} \to \mathbb{R}^n$ such that proximity in $\mathbb{R}^n$ approximates proximity under d. This is the approach used in contrastive representation learning (Radford et al., 2021; Hinton and Salakhutdinov, 2006) and reduces nearest-prototype selection to an approximate nearest-neighbor problem.

26.2 Difference Encoding as Structured Deviation

Once the nearest prototype $p = \pi(x)$ has been identified, the remaining information is captured by a structured description of deviations:

$$\Delta(x, p) \in \mathrm{Hom}_{\mathcal{T}}(p, t_x),$$

where t_x is the full description of x and the morphism $\Delta(x, p)$ specifies the transformation from p to t_x. The compressed representation is the pair $(p, \Delta(x, p))$, factorizing the full description into a template identifier and a structured delta.

27 Differential Structure of Deviations

When the prototype parameter space is smooth, the deviation $\Delta(x, p)$ can be refined by introducing differential structure (do Carmo, 1992; Lee, 2012). Let Θ_p denote the parameter manifold of prototype p, with tangent space $T_p\Theta_p$. An infinitesimal deviation is a tangent vector $\xi \in T_p\Theta_p$, and reconstruction becomes

$$x \approx \mathcal{G}(p + \xi),$$

where $p + \xi$ denotes the prototype perturbed in the direction ξ. Finite deviations are expressed via the exponential map: $\mathcal{G}(\exp_p(\xi))$.

This differential structure enables gradient-based optimization of the deviation, allowing the encoder to refine ξ iteratively until the reconstructed signal matches the input within tolerance. When the relevant transformations form a Lie group G acting on $\mathcal{P}$—as for geometric transformations such as rotations, translations, and scaling—the parameter space has the structure of a Lie algebra $\mathfrak{g}$, and the exponential map $\exp : \mathfrak{g} \to G$ provides an efficient global parameterization. This is particularly useful for encoding camera trajectories, where the relevant group is $\mathrm{SE}(3)$ (Lang, 1995).

28 Compression as Variational Inference

The selection of a prototype and its associated deviation can be unified in a variational inference framework (Kingma and Welling, 2014; Rezende, Mohamed, and Wierstra, 2014). Let $z = (p, \xi) \in \mathcal{P} \times \bigcup_p T_p\Theta_p$ denote a latent description. The optimal compressed representation minimizes the functional

$$\mathcal{L}(z; x) = d\big(x, \mathcal{G}(z)\big) + \lambda\, L(z), \tag{4}$$

where $L(z) = L(p)+L(\xi)$ is the total description length and $\lambda > 0$ controls the compression-fidelity trade-off.

This is precisely the objective of the variational autoencoder (Kingma and Welling, 2014) when the prior over z is set by the prototype library and the posterior is concentrated on a single prototype-deviation pair. The variational formulation makes the MDL interpretation explicit: the optimal description z^* is a concise explanation of x within the generative model. The parameter λ encodes the relative cost of description bits versus distortion, directly corresponding to the inverse temperature in thermodynamic analogies (Section 43).

29 Compositional Delta Algebra and Groupoid Structure

The set of deviations $\Delta(x, p)$ forms an algebraic structure under composition. Given deviations $\Delta_1 : p \to p'$ and $\Delta_2 : p' \to p''$, their composition $\Delta_2 \circ \Delta_1 : p \to p''$ represents sequential application of transformations.

Proposition 29.1 (Groupoid Structure of Deviations)**.** *When every deviation is invertible, the category of prototypes and deviations forms a* groupoid*.*

The groupoid structure enables the reuse and recombination of transformations: a deviation library of common transformations can be applied to many prototypes, reducing description length from $O(n)$ for n independent deviations to $O(\log n)$ for compositions of library elements. This connects naturally to the algebraic structures studied in process calculi (Milner, 1999) and the compositional treatment of morphisms in string diagrams (Coecke and Kissinger, 2017).

30 Temporal Factorization and Event Streams

Video signals possess intrinsic temporal structure. Natural video is not a continuous homogeneous sequence but a concatenation of *events*: coherent episodes governed by consistent generative programs.

Definition 30.1 (Event Segmentation)**.** An *event segmentation* of a video $x : [0, T] \to \mathcal{F}$ is a partition $\{[t_0, t_1], [t_1, t_2], \ldots, [t_{k-1}, t_k]\}$ of $[0, T]$ such that each segment $x|_{[t_{i-1}, t_i]}$ is

well-approximated by a prototype p_i with a single deviation Δ_i.

The compressed representation is the event sequence

$$\mathcal{C}(x) = \big((p_1, \Delta_1, t_1),\ (p_2, \Delta_2, t_2),\ \ldots,\ (p_k, \Delta_k, t_k)\big).$$

The event segmentation problem—identifying the partition minimizing total description length—is a meaningful inference problem related to optimal change-point detection in causal models (Pearl, 2009). This factorization aligns with work on model-based world models (Ha and Schmidhuber, 2018; Schrittwieser et al., 2020), where an agent maintains a structured model of temporal dynamics rather than a raw buffer of observations.

31 Iterative Refinement and Multi-Scale Approximation

A single prototype may not suffice to capture all aspects of a complex scene. Iterative refinement provides a principled approach, yielding a hierarchical representation

$$x \approx \mathcal{G}\big(\Delta_n \circ p_n \circ \cdots \circ \Delta_1 \circ p_1\big),$$

where each level captures structure at a different spatial or temporal resolution. The total description length is $\sum_{i=1}^{n} L(p_i) + L(\Delta_i)$, which decreases as the prototype library becomes more expressive at each scale. In the limit of an infinitely rich library, the description length approaches $K_\epsilon(x \mid \mathcal{G})$ from Section 13.

The multi-scale structure aligns with the fibered and sheaf-theoretic constructions of Sections 18 and 21: each level of the hierarchy corresponds to a layer of the fibration, and the gluing of local descriptions at each scale is governed by the sheaf condition.

32 Semantic Fixed Points and Compression Stability

A compressor that operates on its own outputs should not diverge. This stability requirement is formalized as a fixed-point condition.

Definition 32.1 (Stable Representation)**.** A signal $x^* \in \mathcal{M}$ is a *semantic fixed point* of the compression-reconstruction operator $\Phi = \mathcal{G} \circ \mathcal{C}$ if $\Phi(x^*) \approx x^*$ within perceptual tolerance.

Theorem 32.2 (Existence of Semantic Fixed Points)**.** *If Φ is a contraction on $(\mathcal{M}, d)$—i.e., $d(\Phi(x), \Phi(y)) \le \kappa\, d(x, y)$ for some $\kappa < 1$—then by the Banach fixed-point theorem, Φ admits a unique fixed point in each connected component of $\mathcal{M}$.*

Stable representations correspond to semantically robust descriptions—signals at the "center" of their prototype-deviation class, insensitive to minor perturbations. In practice, iterative application of Φ converges to the fixed point, providing a self-reinforcing process of semantic normalization.

33 Learning the Prototype Space

The effectiveness of the framework depends critically on the construction of the prototype library. Rather than hand-engineering templates, one learns them from data.

Given a corpus $\{x_1, \ldots, x_N\} \subset \mathcal{M}$, the optimal prototype set minimizes the expected description length:

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} \mathbb{E}_{x \sim \mu} \left[L\big(\mathcal{C}(x)\big) \right].$$

Vector-quantized variational autoencoders (van den Oord, Vinyals, and Kavukcuoglu, 2017; van den Oord et al., 2019) implement a discrete approximation of this optimization: a codebook of latent prototypes is learned jointly with an encoder-decoder pair. The training objective—reconstruction loss plus commitment loss—approximates the variational functional (4). As the codebook size grows, coverage of $\mathcal{P}$ improves and description length decreases. Complementary approaches based on Bayesian nonparametrics (Blei, Ng, and Jordan, 2003) allow the codebook to grow without a pre-specified size.

34 Functorial Learning and Dataset as a Colimit

The construction of generative models and prototype libraries from data can be expressed naturally in categorical terms. Rather than viewing a dataset as a passive collection of samples, we treat it as a diagram whose colimit encodes the learned structure of the model.

Let $\{x_i\}_{i \in I} \subset \mathcal{M}$ be a dataset indexed by a category I capturing relationships between samples: temporal adjacency, shared context, or semantic similarity. This induces a diagram $\mathcal{D} : I \to \mathcal{M}$.

Definition 34.1 (Dataset Colimit)**.** The *learned representation* of the dataset is the colimit $\operatorname{colim}_I \mathcal{D}$, taken in a suitable category of representations.

The generative model $\mathcal{G}$ can then be interpreted as a functor approximating this colimit: it compresses the diagram into a parametric form while preserving its essential relational structure.

Proposition 34.2 (Learning as Colimit Approximation)**.** *Training a generative model corresponds to constructing a functor $\mathcal{G} : \mathcal{T} \to \mathcal{M}$ such that $\mathcal{G}$ approximates $\operatorname{colim}_I \mathcal{D}$ up to perceptual equivalence.*

Proof sketch. The dataset encodes a diagram of observed signals. The colimit aggregates these observations into a universal object capturing their shared structure. A generative model trained on the dataset approximates this universal object by learning a parametric mapping whose image covers the observed data distribution. The training loss corresponds to the distortion between the colimit and the model's approximation. □

This perspective clarifies why generative models generalize: they do not memorize individual samples but approximate the colimit of their relational structure. Compression

then operates relative to this learned colimit, encoding new signals as coordinates within the learned space. Overfitting corresponds to a model that approximates individual samples without capturing the colimit; underfitting corresponds to a model that approximates the colimit but at too coarse a resolution. The learning process navigates between these extremes, seeking the functor that best approximates the colimit within the constraints imposed by the model class.

35 Adaptive Prototype Expansion

The prototype library need not be static. When the encoder encounters a signal that cannot be efficiently represented using existing prototypes, a new prototype can be introduced.

Definition 35.1 (Adaptive Expansion Rule)**.** Given a threshold $\tau > 0$, the adaptive expansion rule adds a new prototype p_{new} to $\mathcal{P}$ when

$$\min_{p \in \mathcal{P}} d\big(x, \mathcal{G}(p)\big) > \tau.$$

The adaptive library can be viewed categorically as a *colimit construction*: $\mathcal{P}$ grows by incorporating new objects and morphisms as novel patterns are encountered. This process resembles online clustering algorithms and the nonparametric Bayesian approaches of (Blei, Ng, and Jordan, 2003), where the number of components grows with the data. The trade-off between library size and description efficiency is governed by a rate-distortion-complexity surface parameterized by the coverage of $\mathcal{P}$.

36 Topological Structure of Signal and Description Spaces

The spaces $\mathcal{M}$ and $\mathcal{T}$ admit natural topological structures that refine the metric enrichment of Section 22. These topologies encode continuity properties of both signals and descriptions, enabling a more precise treatment of convergence and stability.

Let τ_d be the topology induced by the perceptual metric d on $\mathcal{M}$, and let τ_L be a topology on $\mathcal{T}$ induced by description length and structural proximity.

Definition 36.1 (Continuous Generative Model)**.** The generative model $\mathcal{G} : (\mathcal{T}, \tau_L) \to (\mathcal{M}, \tau_d)$ is *continuous* if for every convergent sequence $t_n \to t$ in $\mathcal{T}$, $\mathcal{G}(t_n) \to \mathcal{G}(t)$ in $\mathcal{M}$.

Continuity ensures that small changes in description yield small changes in the generated signal, a necessary condition for stable compression.

Proposition 36.2 (Compactness and Approximation)**.** *If $\mathcal{T}$ is compact under τ_L and $\mathcal{G}$ is continuous, then $\mathcal{G}(\mathcal{T})$ is compact in $\mathcal{M}$. In particular, every sequence of reconstructions admits a convergent subsequence.*

This property is relevant for the analysis of iterative refinement (Section 31) and the convergence of compression cycles. Compactness of the description space also provides a

rigorous foundation for the learning of prototype libraries: the existence of an optimal finite prototype set follows from compactness together with continuity of the distortion functional.

37 Geometric Structure of the Prototype Manifold

The parameter spaces Θ_p associated with prototypes can be endowed with Riemannian structure, allowing the use of tools from differential geometry (do Carmo, 1992; Lee, 2012).

Assume each Θ_p is a smooth manifold and that the collection $\{\Theta_p\}$ forms a fiber bundle over $\mathcal{P}$. The generative map lifts to $\mathcal{G} : \Theta \to \mathcal{M}$, and its differential $D\mathcal{G}$ defines a pullback metric on Θ:

$$g_\Theta(\xi, \eta) = \langle D\mathcal{G}(\xi), D\mathcal{G}(\eta) \rangle_{\mathcal{M}}.$$

Definition 37.1 (Geodesic Deviation)**.** A deviation between parameter states $\theta_1, \theta_2 \in \Theta_p$ is a *geodesic* if it minimizes the path integral $\int_0^1 \sqrt{g_\Theta(\dot\gamma, \dot\gamma)}\, dt$ among all paths γ connecting them.

Geodesic deviations correspond to minimal transformations in perceptual space—the natural notion of difference encoding when the parameter space carries geometric structure. The Riemannian exponential map provides an efficient local parameterization of geodesics, and parallel transport provides a consistent way to compare tangent vectors across different base prototypes (Lang, 1995).

38 Semantic Curvature and Complexity of Representation

The Riemannian structure of $\mathcal{T}$ introduced in the preceding section encodes not only distances but also curvature, and this curvature carries information about the complexity of the signal domain.

Definition 38.1 (Semantic Curvature)**.** Let $\mathcal{T}$ be equipped with the pullback Riemannian metric $g_\mathcal{T}$ induced by $\mathcal{G}$. The *semantic curvature* at $t \in \mathcal{T}$ is the sectional curvature of $(\mathcal{T}, g_\mathcal{T})$ at t, measuring the deviation of the description space from local flatness.

Proposition 38.2 (Curvature–Complexity Correspondence)**.** *Regions of high positive curvature in $(\mathcal{T}, g_\mathcal{T})$ correspond to signals of high model-relative Kolmogorov complexity $K_\epsilon(x \mid \mathcal{G})$.*

Proof sketch. High sectional curvature implies that nearby geodesics diverge rapidly, so that small perturbations in the description space correspond to large changes in the generated signal. To maintain distortion below ϵ in such a region, the encoder must specify the description with high precision, increasing $|t|$ and hence $K_\epsilon(x \mid \mathcal{G})$. Regions of low curvature (flat regions) admit coarser specification, corresponding to low model-relative complexity. □

This geometric perspective provides a continuous analogue of algorithmic complexity: the Riemannian curvature of the description manifold encodes the local descriptive cost of signals in different parts of the perceptual space. Regions of high curvature—where the generative model has poor predictive efficiency—correspond to signals that are intrinsically hard to compress relative to the shared prior. Regions of low curvature correspond to compressible structures that lie near the "highways" of the description manifold.

The curvature also governs the rate of convergence of the iterative refinement scheme of Section 31: in regions of low curvature, coarse prototypes provide good initial approximations and the deviation Δ is small, while in high-curvature regions more refinement steps are required. The multi-scale hierarchy therefore adapts implicitly to the local geometry of the description manifold.

39 Information Geometry and Statistical Structure

When the generative model $\mathcal{G}$ is probabilistic, the space of descriptions inherits a statistical manifold structure (Amari, 2016). Each description t corresponds to a distribution $\mathcal{G}(t, \cdot)$ over $\mathcal{M}$, and the Fisher information metric defines a Riemannian structure on $\mathcal{T}$:

$$g_{ij}(t) = \mathbb{E}_{x \sim \mathcal{G}(t)} \left[\partial_i \log p(x|t) \, \partial_j \log p(x|t) \right].$$

Proposition 39.1 (Natural Gradient Descent)**.** *Optimization of the variational objective* (4) *is improved by using the natural gradient* $\tilde{\nabla}_t = g^{-1}(t)\nabla_t$*, which respects the intrinsic geometry of the statistical manifold.*

The natural gradient removes the dependence of gradient steps on arbitrary parameterizations of the description space, yielding faster and more stable convergence (Amari, 2016). This framework connects semantic compression to information geometry, providing a principled method for optimizing descriptions that is invariant to reparameterizations of $\mathcal{T}$.

40 Stochastic Generative Models and Measure-Theoretic Structure

Many generative models are inherently stochastic, producing a distribution of outputs for a given description (Ho et al., 2020; Song and Ermon, 2019; Karras et al., 2022). This requires a measure-theoretic extension of the framework (Gray, 2011).

Let $(\mathcal{M}, \Sigma)$ be a measurable space and $\mathcal{G}$ a stochastic kernel

$$\mathcal{G}(t, \cdot) : \mathcal{T} \to \mathcal{P}(\mathcal{M}),$$

assigning to each description a probability distribution over signals. The distortion be-

comes an expected value:

$$d(x, \mathcal{G}(t)) = \mathbb{E}_{y \sim \mathcal{G}(t)}[d(x, y)].$$

Definition 40.1 (Probabilistic Semantic Compression)**.** The optimal description minimizes

$$\mathcal{C}(x) = \arg\min_t \left[\mathbb{E}_{y \sim \mathcal{G}(t)} d(x, y) + \lambda L(t) \right].$$

This formulation accommodates variability in reconstruction and aligns naturally with diffusion models and autoregressive generators. Stochasticity also introduces equivalence classes at the level of distributions: two descriptions are equivalent if they induce distributions that are indistinguishable under the perceptual metric. The appropriate notion of distance between distributions is the Wasserstein metric (Villani, 2009), which metrizes weak convergence and aligns well with perceptual distance for continuous signal domains.

Remark 40.2. The Wasserstein distance $W_p(\mathcal{G}(t_1, \cdot), \mathcal{G}(t_2, \cdot))$ between the distributions induced by two descriptions provides a principled extension of the perceptual metric d to the stochastic setting, enabling all preceding categorical and variational constructions to carry over.

41 Connections to Causal Structure

The scenario graph representation introduced in Section 3 implicitly encodes causal structure. This can be formalized using structural causal models (Pearl, 2009).

Let S be a directed acyclic graph with structural equations

$$X_i = f_i(\text{pa}(X_i), U_i),$$

where $\text{pa}(X_i)$ are parent variables and U_i are exogenous noise variables. A semantic description corresponds to specifying the functions f_i, the graph structure, and realizations of U_i.

Proposition 41.1 (Causal Compression Advantage)**.** *If a signal is generated by a causal model of complexity K, then semantic compression achieves description length $O(K)$ independent of signal duration.*

Proof sketch. The causal model is constant across the duration of the signal; only initial conditions and stochastic inputs $\{U_i\}$ vary. The description encodes the causal model once and then specifies only the realizations of U_i, whose cumulative entropy grows as $O(T \cdot H(U))$ for signal duration T. However, when the U_i are highly predictable from context (as they typically are in natural scenes), the residual entropy is negligible, yielding $O(K)$ total description length. □

This explains why long videos with repetitive structure can be compressed efficiently: the causal model is constant, and only the initial conditions and stochastic inputs vary.

It also motivates the connection to probabilistic programming (Mansinghka et al., 2014; Goodman et al., 2015; Lake et al., 2015): a compressed description is essentially a probabilistic program whose execution generates the signal.

42 Interventional Semantics and Counterfactual Compression

The causal structure embedded in semantic descriptions enables something that signal-based compression cannot: the direct manipulation of counterfactual scenarios. Because the TIR encodes structural equations rather than observed signal values, it is possible to query what a signal would have looked like under an intervention that did not occur.

Let S be a scenario graph with structural equations $X_i = f_i(\mathrm{pa}(X_i), U_i)$ as in Section 41. A do-calculus intervention (Pearl, 2009) $\mathrm{do}(X_i = x')$ modifies the structural equation for X_i to the constant x', severing the incoming causal arrows.

Definition 42.1 (Counterfactual Description)**.** A *counterfactual description* $t' = \mathrm{do}_{X_i = x'}(t)$ is obtained from the description t by replacing the structural equation for X_i with the constant x' and propagating the consequences through the downstream variables of S.

Proposition 42.2 (Counterfactual Reconstruction)**.** *The generative model $\mathcal{G}$ applied to the counterfactual description t' produces a signal corresponding to the intervened scenario:*

$$\mathcal{G}(t') \approx \widehat{x}_{\,\mathrm{do}(X_i = x')}.$$

Proof sketch. The structural equations in t define a directed acyclic graph over the variables of the scene. Performing $\mathrm{do}(X_i = x')$ modifies this graph by removing incoming edges to X_i and setting its value. The downstream consequences are propagated by the functional structure of the remaining equations. The generative model $\mathcal{G}$, acting on the modified graph t', renders the resulting configuration as a signal. By the causal Markov condition (Pearl, 2009), this rendering corresponds to the distribution over signals under the intervention. □

Counterfactual compression is not merely a theoretical curiosity. It establishes that a semantic archive—a collection of compressed TIR descriptions—is also a *causal archive*: a structured repository from which the consequences of hypothetical decisions can be computed. A journalist's compressed video library, for instance, would allow one to query what a recorded event would have looked like if a single causal factor—a crowd position, a lighting source, a speaker's timing—had been different. The compressed representation does not merely preserve what happened; it preserves the causal structure from which any nearby counterfactual can be derived.

43 Thermodynamic Analogy and Free Energy

The variational formulation of semantic compression admits a thermodynamic interpretation (Friston, 2010; Sohl-Dickstein et al., 2015). The objective

$$\mathcal{L}(t) = d(x, \mathcal{G}(t)) + \lambda L(t)$$

is analogous to free energy $F = E - TS$, where distortion plays the role of energy and description length the role of entropy.

Definition 43.1 (Semantic Free Energy)**.** $F_{\mathcal{G}}(t; x) = d(x, \mathcal{G}(t)) + \lambda L(t)$.

Proposition 43.2 (Equilibrium Condition)**.** *At equilibrium,* $\frac{\partial d}{\partial t} = -\lambda \frac{\partial L}{\partial t}$*, balancing reconstruction accuracy against description compression.*

This connects semantic compression to Friston's free-energy principle (Friston, 2010), in which the brain minimizes a variational free energy that bounds surprise. The generative model $\mathcal{G}$ plays the role of the brain's internal model, the description t plays the role of the brain's belief state, and the distortion $d(x, \mathcal{G}(t))$ plays the role of prediction error. Semantic compression, in this light, is the same process the brain uses to represent its environment.

44 Limits of Compression under Model Mismatch

The effectiveness of semantic compression depends on the alignment between the generative model $\mathcal{G}$ and the true data-generating process. Let P_X be the true distribution and $P_{\mathcal{G}}$ the model distribution. Define the mismatch by the Kullback–Leibler divergence $D_{\mathrm{KL}}(P_X \| P_{\mathcal{G}})$ (Cover and Thomas, 2006).

Theorem 44.1 (Mismatch Penalty)**.** *The achievable compression rate satisfies*

$$R_{\mathcal{G}}(D) \geq R^*(D) + D_{\mathrm{KL}}(P_X \| P_{\mathcal{G}}),$$

where $R^*(D)$ is the optimal rate under a perfect model.

Proof sketch. The semantic rate–distortion function under model $\mathcal{G}$ differs from the true rate–distortion function by the additional information required to correct for model mismatch. By the data processing inequality and properties of KL divergence (Cover and Thomas, 2006), this correction is bounded below by $D_{\mathrm{KL}}(P_X \| P_{\mathcal{G}})$. □

This theorem quantifies the cost of model misspecification and motivates the adaptive learning of the generative substrate. It also provides a diagnostic for monitoring compression quality: as $D_{\mathrm{KL}}(P_X \| P_{\mathcal{G}})$ decreases through continued training, the achievable compression rate approaches the theoretical optimum.

45 Robustness and Adversarial Considerations

Semantic compression introduces new robustness properties and new vulnerabilities relative to classical compression.

Definition 45.1 (Semantic Robustness)**.** A compression scheme is ϵ-robust if small perturbations δx satisfying $d(x, x + \delta x) < \epsilon$ do not change the compressed description: $\mathcal{C}(x + \delta x) = \mathcal{C}(x)$.

This robustness arises naturally when the encoder operates at the level of semantic structure rather than raw signals: perturbations below perceptual threshold are absorbed into the $\sim_\epsilon$ equivalence class and do not affect prototype selection.

However, adversarial inputs can exploit the generative prior by forcing misclassification into incorrect prototypes.

Proposition 45.2 (Adversarial Misprojection)**.** *There exist inputs x' arbitrarily close to x such that $\pi(x') \neq \pi(x)$ when the embedding ϕ is not isometric.*

This is the semantic analogue of adversarial examples in neural networks (Goodfellow et al., 2016). Mitigating it requires enforcing Lipschitz continuity of the embedding and incorporating uncertainty estimates in prototype selection. The connection between semantic robustness and adversarial robustness suggests that the two problems share the same fundamental structure: both concern the stability of category boundaries under perturbation.

46 Category of Generative Programs

The textual descriptions can be interpreted as programs in a formal language (Pierce, 2002; Milner, 1999). This suggests defining a category $\mathcal{P}\nabla\wr\}$ of generative programs.

Objects in $\mathcal{P}\nabla\wr\}$ are types of generative outputs; morphisms are programs transforming inputs to outputs; composition corresponds to program composition (Hopcroft, Motwani, and Ullman, 2006).

Definition 46.1 (Execution Functor)**.** There exists a functor $\mathcal{E} : \mathcal{P}\nabla\wr\} \to \mathcal{M}$ mapping programs to their execution results.

Semantic compression factors through $\mathcal{P}\nabla\wr\}$:

$$\mathcal{M} \xrightarrow{\mathcal{C}} \mathcal{P}\nabla\wr\} \xrightarrow{\mathcal{E}} \mathcal{M}.$$

Proposition 46.2 (Program Equivalence)**.** *Two programs $p_1, p_2 \in \mathcal{P}\nabla\wr\}$ are equivalent if $\mathcal{E}(p_1) \sim_\epsilon \mathcal{E}(p_2)$. The quotient category $\mathcal{P}\nabla\wr\}/\sim_\epsilon$ captures semantic equivalence of descriptions.*

This perspective unifies compression with compilation and execution. A compressed description is a program that, when compiled by the generative model, produces the

target signal. Compression is program synthesis (Lake et al., 2015; Solomonoff, 1964); decompression is program execution. The boundary between data and code dissolves entirely.

47 Homotopy and Equivalence of Descriptions

Different textual descriptions may produce equivalent outputs. This redundancy can be formalized using homotopy theory (Lurie, 2009; Lurie, 2017).

Definition 47.1 (Homotopy of Descriptions)**.** Two descriptions $t_0, t_1 \in \mathcal{T}$ are *homotopic* if there exists a continuous path $H : [0, 1] \to \mathcal{T}$ such that $d(\mathcal{G}(H(s)), \mathcal{G}(H(s'))) < \epsilon$ for all s, s'.

The set of homotopy classes $\pi_0(\mathcal{T})$ represents equivalence classes of descriptions under perceptual equivalence.

Proposition 47.2 (Homotopy Invariance)**.** *Compression is invariant under homotopy: if $t_0 \sim t_1$, then $\mathcal{C}(\mathcal{G}(t_0)) = \mathcal{C}(\mathcal{G}(t_1))$.*

This formalizes the idea that multiple descriptions may represent the same underlying content—the same semantic content can be expressed in many syntactically distinct ways. The higher homotopy groups $\pi_n(\mathcal{T})$ capture higher-order redundancies: paths between paths, and so on. A full homotopy-theoretic treatment would use ∞-groupoids (Lurie, 2009) to organize all levels of description equivalence simultaneously.

48 Gauge Freedom in Description and Generative Redundancy

The homotopy equivalences of Section 47 establish that multiple descriptions can represent the same perceptual content. This representational redundancy has a precise algebraic counterpart: it is a gauge symmetry of the description space, in the sense of physics (Baez and Stay, 2010).

Definition 48.1 (Gauge Equivalence)**.** Two descriptions $t_1, t_2 \in \mathcal{T}$ are *gauge equivalent* if $\mathcal{G}(t_1) \sim_\epsilon \mathcal{G}(t_2)$, i.e., if they generate perceptually indistinguishable signals.

Proposition 48.2 (Gauge Group Action)**.** *There exists a groupoid $\mathcal{G}_{\text{gauge}}$ acting on $\mathcal{T}$ whose orbits are precisely the gauge equivalence classes of descriptions.*

Proof sketch. Define morphisms of $\mathcal{G}_{\text{gauge}}$ as pairs (t_1, t_2) of gauge-equivalent descriptions, with composition given by transitivity of gauge equivalence. This is exactly the equivalence groupoid of the relation $\sim_\epsilon$ lifted to $\mathcal{T}$ via $\mathcal{G}$. □

Gauge transformations include reparameterizations of a prototype, alternative factorizations of the same scene into sub-events, syntactic variations in the TIR that preserve

semantic content, and changes of rendering order for independent components. These are not defects in the representation system but structural features: they reflect the redundancy inherent in any expressive language, paralleling the gauge freedom of field theories in physics, where multiple field configurations represent the same physical state.

Remark 48.3. Gauge freedom is computationally useful: it allows the encoder to choose the most convenient description within each equivalence class without affecting the reconstructed output. Fixing a gauge—selecting a canonical representative from each orbit of $\mathcal{G}_{\text{gauge}}$—corresponds to applying normal-form reduction as in Section 49. The compression process implicitly performs this gauge-fixing by selecting the minimum-length representative, i.e., the description minimizing $L(t)$ within its orbit.

The gauge group also governs the stability of compressed representations under small perturbations: a perturbation $t \mapsto t + \delta t$ that lies within the orbit of t under $\mathcal{G}_{\text{gauge}}$ produces no change in the reconstructed signal. The size of the gauge orbit at t therefore measures the robustness of the representation at that point, connecting gauge freedom directly to the semantic robustness of Section 45.

49 Algebraic Structure of the Description Language

The textual language $\mathcal{L}$ admits an algebraic structure that constrains and organizes the space of valid descriptions (Pierce, 2002; Hopcroft, Motwani, and Ullman, 2006).

Definition 49.1 (Description Algebra)**.** Let $\mathcal{A}$ be the free algebra over generators G (primitive constructs: objects, actions, parameters) with operations corresponding to composition, parallel combination ($\otimes$), and parameterization. Normal forms in $\mathcal{A}$ canonicalize descriptions by applying algebraic rewriting rules.

Proposition 49.2 (Normal Forms)**.** *Every description $t \in \mathcal{L}$ admits a normal form under the algebraic relations of $\mathcal{A}$.*

Normal forms enable canonicalization of descriptions and efficient comparison between them—a description in normal form is the canonical representative of its homotopy class. The free algebra $\mathcal{A}$ corresponds to the free monoidal category generated by the primitive constructs, and the semantic functor $\cdot$ is the unique monoidal functor extending the assignment of generators to their generative interpretations (Mac Lane, 1998).

50 Comparison with Classical Compression Paradigms

It is instructive to situate semantic compression relative to classical paradigms. Classical transform coding (Shannon, 1948; Cover and Thomas, 2006) projects signals onto a fixed basis and quantizes coefficients. Predictive coding (Friston, 2010) transmits residuals relative to local predictions. Entropy coding (Shannon, 1948; Csiszár and Körner, 2011) exploits statistical redundancy in symbol sequences.

Semantic compression differs in that it replaces the signal domain with a generative domain. The basis is no longer fixed but learned; the residual is no longer numeric but structural; and entropy coding is replaced by model-relative description length.

Proposition 50.1 (Strict Generalization)**.** *Semantic compression strictly generalizes transform and predictive coding: both can be recovered as special cases where $\mathcal{G}$ is linear or locally predictive.*

Proof sketch. Transform coding corresponds to $\mathcal{G}(t) = Bt$ for a fixed orthonormal basis B. Predictive coding corresponds to $\mathcal{G}(t) = x_{<k} + t$ for local context $x_{<k}$. In both cases $\mathcal{G}$ is a special case of the general generative operator, so the semantic framework subsumes both. □

This establishes semantic compression not as an alternative but as a superset of existing methods, with the additional degrees of freedom provided by learned, non-linear, and globally-coherent generative models.

51 Constraint–Generation Separation and the Structural Origin of Compression

The preceding section situates semantic compression within the landscape of classical methods. We now establish, in a single unified argument, *why* compression emerges from this architecture at all—not as an empirical observation but as a structural consequence of separating the representation of constraints from their generative realization.

51.1 The Separation Principle

The key observation is that a textual description $t \in \mathcal{T}$ does not uniquely specify a signal $x \in \mathcal{M}$. Rather, it defines a *constraint set*:

$$\mathcal{C}(t) = \{x \in \mathcal{M} : x \text{ is a perceptually admissible realization of } t\} \subseteq \mathcal{M}.$$

The generative model $\mathcal{G}$ supplies a lawful completion—a specific element of $\mathcal{C}(t)$—consistent with its priors over physical dynamics, object structure, and causal relations. Compression arises because t need only encode the residual information not already captured by $\mathcal{G}$.

Theorem 51.1 (Constraint–Generation Separation)**.** *Let $\mathcal{T}$ be a space of textual descriptions and $\mathcal{G} : \mathcal{T} \to \mathcal{M}$ a generative model with parameter space Θ. Suppose each $t \in \mathcal{T}$ defines a constraint set $\mathcal{C}(t) \subseteq \mathcal{M}$, and $\mathcal{G}_\theta(t) \in \mathcal{C}(t)$ for some $\theta \in \Theta$. Then the description length of t satisfies*

$$|t| \approx K(x \mid \mathcal{G}) \ll K(x),$$

where $K(x)$ is the Kolmogorov complexity of x and $K(x \mid \mathcal{G})$ is the model-relative complexity from Section 13.

Proof sketch. In classical encoding, x must be specified directly, so its minimal description length equals $K(x)$. In the semantic regime, t does not specify x explicitly but constrains x to lie in $\mathcal{C}(t)$. The generative model $\mathcal{G}$ acts as a shared prior, supplying the structural degrees of freedom not encoded in t. By definition of conditional Kolmogorov complexity,

$$K(x \mid \mathcal{G}) = \min\{|t| : \mathcal{G}(t) \approx x \text{ within perceptual tolerance}\}.$$

Since $\mathcal{G}$ encodes regularities—physical laws, object priors, causal dynamics—we have $K(x \mid \mathcal{G}) \leq K(x)$, with strict inequality whenever x exhibits structure modeled by $\mathcal{G}$. Therefore, compression arises precisely from transferring descriptive burden from t to $\mathcal{G}$. □

51.2 Corollaries

Theorem 51.1 immediately implies the two key structural properties of the framework.

Corollary 51.2 (Entropy Redistribution)**.** *Let $H(x)$ denote the entropy of a source signal and $H(x \mid \mathcal{G})$ its entropy relative to a generative model. Then*

$$H(x \mid \mathcal{G}) = H(x) - I(x; \mathcal{G}),$$

where $I(x; \mathcal{G})$ is the mutual information between the signal and the model.

Proof. This follows directly from the identity $I(x; \mathcal{G}) = H(x) - H(x \mid \mathcal{G})$. □

Thus, the more structure $\mathcal{G}$ captures, the less entropy must be transmitted. In the limit of a perfect model, $H(x \mid \mathcal{G}) \to 0$, and the description reduces to an index within the model's latent space. This provides the precise formal grounding for the informal entropy redistribution conservation relation of Section 14.

Corollary 51.3 (Semantic Equivalence Classes)**.** *Define an equivalence relation $\sim$ on $\mathcal{M}$ by*

$$x_1 \sim x_2 \iff \exists\, t \in \mathcal{T} : x_1, x_2 \in \mathcal{C}(t).$$

Then the decoding map $\mathcal{G}$ preserves equivalence classes rather than individual signals.

Corollary 51.3 formalizes perceptual indistinguishability: distinct microstates—leaf 4022 flipping versus leaf 4010 flipping—may correspond to the same semantic description and therefore to the same equivalence class, without any loss of perceptual quality. This is the structural basis of the quotient category $\mathcal{M}/\sim_\epsilon$ introduced in Section 6.

51.3 Interpretation

Theorem 51.1 and its corollaries establish that compression is not a heuristic artifact of generative modeling but a *structural consequence* of the separation:

$$\text{Representation (constraints)} \quad \longleftrightarrow \quad \text{Realization (generation)}.$$

In this architecture:

- The TIR encodes constraints on admissible worlds.
- The generative model supplies lawful completions of those constraints.
- Compression emerges as the reduction in information required to specify a world relative to shared generative structure.

Accordingly, semantic compression should be understood not as signal approximation but as *constraint specification under a shared generative prior*. The mathematical frameworks of the preceding sections—sheaves, fibered categories, Lie groups—do not themselves generate perceptual outputs. They define the constraints under which a generative model must operate. The generative model then produces a perceptual instantiation satisfying those constraints. This separation is essential and must not be collapsed into a simpler picture of direct simulation.

52 System Architecture and Layered Design

The realization of semantic compression as a practical system requires a layered architecture in which each component is responsible for a well-defined aspect of the transformation between signals and descriptions.

The *encoder* performs interpretation: it segments the input signal, assigns each segment to a prototype, computes the deviation, and serializes the result as structured text. The encoding pipeline has the factored structure $\mathcal{C} = \mathcal{C}_2 \circ \mathcal{C}_1$: a perceptual parser $\mathcal{C}_1$ that maps signals to semantic graphs, followed by a textualizer $\mathcal{C}_2$ that serializes semantic graphs as structured descriptions.

The *generative substrate* is the shared generative model $\mathcal{G}$, which must be synchronized between encoder and decoder. It defines the space of admissible reconstructions, embedding prior knowledge about speech, physical motion, visual appearance, and their interactions. The substrate draws on large pre-trained generative models (Brown et al., 2020; Rombach et al., 2022; van den Oord et al., 2019), potentially specialized by modality.

The *decoder* executes the textual description as a generative program. It parses the description, resolves prototype references, applies deviations, and invokes the appropriate generative policies to synthesize the output. The tight coupling between these layers introduces a dependency that must be managed through versioning and compatibility protocols.

53 A Textual Intermediate Representation

Central to the system is a carefully designed *textual intermediate representation* (TIR) that balances expressiveness against compactness. The TIR must:

1. Encode hierarchical structure (nested scenes, sub-events, recursive templates).

2. Represent temporal dynamics (event sequences, state transitions, continuous trajectories).
3. Capture cross-modal relations (synchronization, correlation, causal dependencies).
4. Admit partial specification (coarse descriptions that are subsequently refined).
5. Be linearizable as text for storage and transmission.

These requirements suggest a language occupying an intermediate position between natural language and formal programming languages. The TIR employs the descriptive richness of natural language—enabling human interpretability and approximate matching—while imposing enough syntactic structure to support precise compositional semantics (Chomsky, 1965; Barwise and Perry, 1983).

The functorial semantics of Section 25 provide a mathematical foundation for the TIR's design. The compositionality requirement translates into a context-free grammar for the TIR, with each production rule corresponding to a functor in the categorical decomposition. The monoidal structure translates into a tensor product operator in the TIR, enabling independent components to be composed without interaction (Coecke and Kissinger, 2017).

54 Human Interpretability and Co-Authoring

A distinctive and consequential property of text-based compression is that the compressed representation is directly legible to humans. Unlike the bitstreams produced by signal-based codecs, textual descriptions can be read, understood, and modified by human users.

This interpretability enables a novel form of media interaction: *semantic co-authoring*. A viewer who can read the textual description of a video can identify the agents, events, and camera dynamics that constitute it, and can modify them directly—changing an agent's behavior, altering the lighting, or substituting a different camera trajectory—without access to the original signal or a conventional video editor. The modified description, when executed by the generative substrate, produces a new video consistent with the user's modifications.

This capability transforms the relationship between media and its audience. Media consumption becomes an instance of collaborative authorship, with the generative substrate serving as the medium of co-production. The compressed description is not a closed artifact but an open specification, inviting interpretation and modification (Dennett, 1987). The interpretability of the TIR also facilitates debugging and quality control at the semantic level.

55 Cognitive Alignment and Perceptual Structure

The semantic compression framework reflects a deep alignment with the structure of human cognition. Decades of research in cognitive science have established that humans perceive and remember the world not as sequences of sensory values but as structured descriptions of events, objects, and agents (Barwise and Perry, 1983; Tenenbaum et al., 2011; Goodman et al., 2015).

This alignment has several consequences. First, it suggests that the prototype library learned by the system will converge toward representations that mirror human conceptual categories—not because human categories are imposed as a prior, but because natural video is generated by a world structured around the same categories. Second, it implies that the cognitive load required to process a textual description is lower than the load required to process the corresponding signal, since descriptions are represented in a format closer to the natural currency of cognition (Clark, 2015).

Third, it connects semantic compression to the *predictive processing* framework (Friston, 2010; Clark, 2015), in which perception is understood as a process of active inference: the brain constructs a generative model of the world and continuously updates it to minimize prediction error. Semantic compression, in other words, is cognitive compression—the same process by which the mind reduces the world to manageable form.

56 Generative Media Ontology: Description Replaces Signal

The framework developed here has implications that extend beyond compression technology into the ontology of media itself. If every signal can be faithfully represented by a generative description, and if that description is the object stored and transmitted, then the signal—the waveform, the pixel array—ceases to be the fundamental unit of media and becomes instead a derived, transient realization of the description.

This inversion is ontologically significant. In traditional media, the recording is primary: the film, the tape, the digital file. In the semantic compression paradigm, the description is primary and the signal is generated on demand. Media becomes, in the terminology of Section 1, a *program* rather than a *recording* (Dennett, 1987).

The implications cascade: storage cost is determined by description complexity rather than signal duration; editing operates directly on the description; version control manages description delta histories; and transmission of compact descriptions replaces transmission of high-bandwidth signals. This paradigm shift transforms the economics of media production and distribution, shifting the bottleneck from bandwidth and storage to generative capability and description quality.

57 Ethical and Epistemic Considerations

The transformation of media into generative descriptions raises profound ethical and epistemic challenges.

57.1 Authenticity and Provenance

When a video is stored as a description rather than a recording, the link between the signal and its origin is severed. Multiple distinct descriptions may produce perceptually indistinguishable videos; a single description may be modified to change the depicted events. The traditional notion of an authentic recording does not survive the semantic compression paradigm.

This requires new mechanisms for provenance and verification: cryptographic commitment schemes, description-level watermarking, and auditable generative substrates. The categorical framework provides a natural setting for provenance chains—morphisms in $\mathcal{T}$ that track the history of transformations applied to a description.

57.2 Epistemic Stability

The shared generative model embeds a worldview: a set of priors about what kinds of events, agents, and configurations are canonical. Signals that fall outside this worldview will be poorly represented, compressed to the nearest prototype even when that prototype is semantically inappropriate.

This raises concerns about epistemic closure (Dennett, 1987). A system trained on data from one cultural or physical context will develop prototypes suited to that context. The adaptive expansion mechanism of Section 35 partially addresses this concern, but the initial distribution of prototypes remains biased by the training data.

57.3 Manipulation and Deception

The ease with which textual descriptions can be modified—precisely the property that enables co-authoring—also enables manipulation. A malicious actor with access to the description of a recording can alter agent behaviors, speech content, or causal relations, producing a modified video that is perceptually plausible but semantically false.

Addressing these concerns requires both technical and social measures: authenticated descriptions, normative frameworks governing the use of generative editing tools, and public literacy about the generative nature of media in the semantic compression paradigm.

58 Economic and Infrastructural Implications

The shift to semantic compression redraws the economic geography of media infrastructure. The cost of storage and transmission is radically reduced—descriptions are

orders of magnitude smaller than the signals they encode—while the cost of generative computation is incurred locally at the point of consumption. This redistribution favors decentralized architectures: media can be distributed as compact descriptions, with each device performing its own reconstruction.

The resulting ecosystem introduces new forms of value, centered on the creation and curation of prototype libraries, the certification of generative substrates, and the development of textual description standards. Whoever controls the most expressive and comprehensive generative substrate controls the quality of media reconstruction—a concentration of power analogous to the role of operating systems or platform standards in prior technological transitions.

59 Open Problems and Research Directions

The framework developed here suggests numerous directions for further research.

One fundamental problem concerns the design of perceptual metrics that are both mathematically tractable and aligned with human judgment. Another concerns the construction of universally expressive textual languages that remain compact and interpretable. The learning dynamics of prototype libraries remain poorly understood: the trade-off between library size, coverage, and generalization requires further theoretical analysis (Grünwald, 2007; Hutter, 2005).

The interaction between stochastic generative models and deterministic textual descriptions raises questions about probabilistic semantics and uncertainty quantification (Goodman et al., 2015; Mansinghka et al., 2014). The connection to causal structure (Section 41) suggests that interventional queries—"what would this video look like if the agent had acted differently?"—can be answered directly from the compressed description, connecting semantic compression to causal inference (Pearl, 2009).

At the categorical level, a full higher-topos-theoretic treatment of semantic compression remains to be developed (Lurie, 2009; Lurie, 2017), potentially connecting the present framework to derived geometry and homotopical semantics. Finally, the ethical and epistemic implications outlined in Section 57 demand formalization: provenance, authenticity, and trust must be incorporated into the mathematical structure of the system itself.

These open problems suggest that semantic compression is not merely a technical innovation but the beginning of a broader theoretical program linking information, computation, and meaning.

60 On Structural Coherence and Common Misinterpretations

The increasing accessibility of semantic compression frameworks has led to a proliferation of intuitive explanations that, while pedagogically useful, often conflate distinct mathematical roles within the architecture. There is a recurring tendency to reinterpret

structural constraints as dynamical mechanisms, thereby obscuring the actual function of the underlying formalisms. This section clarifies several of these misinterpretations in the light of the formal development above, and restores the proper separation between representation, constraint, and generation established in Section 51.

60.1 Sheaf-Theoretic Structure Beyond Spatial Tiling

A common simplification treats the decomposition of a scene into local regions as a form of spatial tiling, where independently described fragments are subsequently "glued" together along their geometric boundaries. While this captures the notion of locality, it misrepresents the true role of sheaf-theoretic structure.

Formally, a sheaf does not describe a procedure for assembling geometry from parts, but a condition on the compatibility of partial assignments. As established in Section 21, for an open cover $\{U_i\}$ of a spacetime domain X, a family of local sections $\{s_i \in \mathcal{F}(U_i)\}$ is admissible if and only if

$$s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j} \quad \forall\, i, j,$$

and the sheaf condition guarantees the existence of a unique global section $s \in \mathcal{F}(X)$ such that $s|_{U_i} = s_i$.

In the present framework, these sections correspond not to geometric fragments but to *partial semantic descriptions*. The overlaps encode shared variables—lighting parameters, identity embeddings, causal dependencies—rather than spatial seams. Crucially, global parameters are not reconstructed from local data; they are factored out and propagated as constraints across all local sections, precisely as noted in Appendix D. Sheaf structure enforces *contextual coherence*, not geometric continuity. Its role is to guarantee that independently specified local descriptions do not contradict one another when interpreted within a shared global context.

60.2 Natural Transformations as Commutative Coherence

A second frequent misinterpretation assigns to natural transformations the role of real-time coupling mechanisms between modalities, as though they directly enforce synchronization—as though the acoustic origin point of an engine is physically "dragged" by the movement of the corresponding visual object. This conflates dynamical dependence with categorical coherence.

As established in Section 20, with compression functors $\mathcal{C}_v : \mathcal{M}_v \to \mathcal{T}$ and $\mathcal{C}_a : \mathcal{M}_a \to \mathcal{T}$, a natural transformation

$$\eta : \mathcal{C}_a \circ \Phi \Rightarrow \mathcal{C}_v$$

does not induce motion or causation. It enforces that two distinct compositional pathways yield compatible representations: the process of (video $\to$ extracted audio $\to$ audio description) must be consistent with the process of (video $\to$ video description $\to$ inferred audio description). This is a structural requirement about the commutativity of diagrams,

not a real-time physical coupling. The apparent synchronization of modalities arises not from direct interaction between the audio and video encoders, but from their shared parameterization within the textual intermediate representation—a single description that simultaneously constrains both generative processes.

60.3 Lie Groups and Global Parameterization

It is also common to contrast Lie group representations with frame-based coordinate tracking by attributing to the former an intrinsic physical smoothness that the latter lacks. While Lie groups support continuous transformations, smoothness alone does not distinguish them from sufficiently dense coordinate encodings.

The essential advantage of Lie groups, as detailed in Appendix E, lies in their role as *global parameterizations of symmetry*. For camera motion, $\mathrm{SE}(3)$ provides a compact representation of rigid body motion, allowing entire trajectories to be specified via elements of the Lie algebra $\mathfrak{se}(3)$ and reconstructed through the exponential map:

$$g(t) = \exp(t\xi), \quad \xi \in \mathfrak{se}(3).$$

This replaces a sequence of discrete positional updates with a single generative instruction encoding invariant structure. The resulting compression is a consequence of encoding motion as a transformation within a structured group, not of smoothness per se.

60.4 Constraint Versus Generation: A Recapitulation

The three clarifications above converge on the general principle established by Theorem 51.1: the mathematical frameworks employed in this paper—sheaf theory, category theory, Lie theory—do not themselves generate perceptual outputs. They define the *constraints* under which a generative model must operate. The generative model then produces a perceptual instantiation satisfying these constraints.

When structural constraints are misinterpreted as rendering mechanisms, the architecture collapses into a heuristic description of a "simulation engine," thereby obscuring the deeper principle that semantic compression is not the reproduction of signals but the preservation of coherent structure across representations. Maintaining this distinction reveals four interdependent properties, each a consequence of the categorical and topological structure rather than an incidental feature of any particular implementation. Entropy is not eliminated but relocated into shared priors, so that the information content of the transmitted description reflects only what the generative model cannot already supply. Cross-modal coherence arises from shared parameterization within a common textual description rather than from direct coupling between modality-specific encoders, which is why the audio and video representations remain consistent without any explicit synchronization mechanism. Local modifications propagate globally through constraint satisfaction, so that editing a single parameter in the TIR—a lighting temperature, a camera

trajectory, an agent disposition—induces coherent changes throughout the reconstructed signal without requiring recomputation of raw data. And representation remains editable without loss of structural integrity, because the description encodes constraints on possible worlds rather than a fixed trace of one particular realization. These properties stand or fall together; they are not separable engineering choices but the joint consequence of grounding the system in a closed, compositional, and functorially coherent architecture.

61 Toward a Generative Media Ecology

The framework developed throughout this essay points toward a broader transformation in the nature of media itself. Static files are replaced by dynamic generative processes; recordings are replaced by descriptions; storage and transmission are replaced by program exchange. In this generative ecology, media is not an artifact but a process.

This process is inherently evolutionary. Prototype libraries grow as new patterns are encountered. Generative models improve as training data accumulates (Brown et al., 2020; Schrittwieser et al., 2020). Description languages evolve to accommodate new modalities and new forms of experience. The media ecology is not a fixed infrastructure but a living system, continuously adapting to the signals it is called upon to represent.

Text serves as the connective tissue of this ecology. As the medium of description, it links the physical world of signals to the computational world of generative models. As the medium of communication, it enables the sharing and modification of descriptions across participants. As the medium of interpretation, it aligns machine generation with human understanding.

62 The Phase Transition: From Signal Preservation to Latent Communication

The preceding derivations establish that semantic compression is not merely a more efficient codec, but a fundamental relocation of the informational center of mass from the transmitted signal to the shared generative substrate. This shift precipitates a phase transition in the nature of digital communication, characterized by three primary structural evolutions.

62.1 The Emergence of a Universal Latent Topology

When multiple agents—human or machine—utilize a standardized TIR and a shared generative prior $\mathcal{G}$, the space of all possible media becomes a unified, navigable manifold. In this regime, the distance between two films or two musical compositions is no longer measured by pixel-wise or sample-wise divergence but by their geodesic distance within the latent space of $\mathcal{G}$ (Amari, 2016; Villani, 2009).

This topology admits the interpolation of reality. Because the TIR is compositional, a description

$$t_\lambda = (1-\lambda)\, t_1 + \lambda\, t_2, \quad \lambda \in [0,1],$$

corresponds to a perceptually coherent "hybrid" world that maintains structural integrity across all modalities. The sheaf-theoretic constraints of Section 21 ensure that this interpolation does not produce discontinuities but a mathematically consistent reconfiguration of causal and environmental parameters.

62.2 Information Asymmetry and the Computational Edge

The separation of constraint specification from realization induces a radical asymmetry in the global data economy. As $|t| \to K(x \mid \mathcal{G})$, the cost of transmission approaches zero, while the cost of rendering—the perceptual collapse of the description into a signal—scales with the desired fidelity ϵ.

Proposition 62.1 (The Rendering Bottleneck)**.** *The computational complexity of reconstruction $\mathcal{G}(t)$ is decoupled from the entropy $H(t)$ of the description. Consequently, bandwidth is replaced by local computational capacity as the primary constraint on high-fidelity experience.*

Proof sketch. By Corollary 51.2, $H(t) \approx H(x \mid \mathcal{G})$, which is made arbitrarily small by an expressive $\mathcal{G}$. However, executing $\mathcal{G}(t)$—rendering the latent description into a full perceptual output at resolution $1/\epsilon$—requires computation proportional to the complexity of the generative model, independent of $H(t)$. The two quantities are therefore decoupled. □

This implies a future where the quality of a rendered experience is not determined by the source description provided by a distributor but by the local generative capacity of the viewer's hardware. The signal is democratized—the description is nearly free to transmit—but the realization is stratified by the observer's computational resources.

62.3 Semantic Drift and the Autonomy of the Substrate

Finally, we must address the closed-loop evolution of the medium. As the TIR becomes the primary substrate for recording and exchange, the generative model $\mathcal{G}$ will inevitably be trained on its own reconstructed outputs. This creates a recursive feedback loop in which the prototypes within $\mathcal{P}$ begin to drift toward generative optima—structures that are most efficiently described and rendered within the existing architecture.

Definition 62.2 (Semantic Drift)**.** *Semantic drift* $\Delta_{\mathcal{S}}$ is the measure of the contraction of the prototype library $\mathcal{P}$ toward the set of high-probability states within the prior $\mathcal{G}$ over successive generations of compression and reconstruction.

Semantic drift is the ultimate form of epistemic closure (Section 57): if uncorrected by the injection of high-entropy raw signal data, the universal representational medium

risks collapsing into a finite set of hyper-efficient archetypes. The challenge for future architects of this technology will not be further reduction of bitrates but the preservation of semantic diversity against the gravitational pull of the model's own priors. Raw signal injection—the deliberate introduction of underrepresented high-entropy phenomena into the training corpus—represents the primary countermeasure, maintaining the coverage of $\mathcal{P}$ against convergence to a degenerate fixed point.

63 Semantic Drift Dynamics

The qualitative notion of semantic drift introduced in Section 62 can be given a precise dynamical formulation. Let $\mathcal{P}_n$ denote the prototype library after n generations of compression, reconstruction, and retraining. The drift process is a discrete-time dynamical system on the space of prototype libraries.

Definition 63.1 (Drift Operator)**.** The *drift operator* Ψ maps a prototype library $\mathcal{P}_n$ to its successor $\mathcal{P}_{n+1}$ according to the rule: retrain $\mathcal{G}$ on the reconstructions $\{\mathcal{G}_n(\mathcal{C}_n(x)) : x \in \mathcal{X}\}$ generated by the current model, then refit the prototype library to the updated model.

Proposition 63.2 (Drift as Contraction)**.** *Under the operator Ψ and in the absence of raw signal injection, the sequence $\{\mathcal{P}_n\}$ converges to a fixed point $\mathcal{P}^*$ satisfying $\Psi(\mathcal{P}^*) = \mathcal{P}^*$, with $|\mathcal{P}^*| < |\mathcal{P}_0|$.*

Proof sketch. Each application of Ψ replaces the training data with the reconstructions of the current model. Since reconstruction maps each signal to an element of $\mathcal{G}(\mathcal{T})$—the image of the generative model—the effective support of the training distribution contracts toward $\mathcal{G}(\mathcal{T})$ at each step. As long as $\mathcal{G}(\mathcal{T})$ is a strict subset of $\mathcal{M}$ (which holds whenever the model is not a perfect universal approximator), the contraction is strict and converges to a fixed point by compactness of the prototype space. □

The fixed point $\mathcal{P}^*$ is the library of archetypes that the system finds most natural to represent: those signals that are most efficiently compressed and most faithfully reconstructed by the model. All other signals are progressively marginalized. This is not a failure of the compression system but a predictable consequence of its success: it is efficient precisely because it concentrates representational capacity where the prior is strongest.

The dynamics of Ψ partition the initial prototype space $\mathcal{P}_0$ into basins of attraction leading to different degenerate fixed points. Which fixed point is reached depends on the initial training distribution and the architecture of $\mathcal{G}$. This sensitivity to initial conditions is an additional argument for maintaining diverse, high-entropy training corpora: different initializations lead to different fixed points, and the coverage of the long-run library depends directly on the coverage of the initial one.

64 Semantic Phase Space

The drift dynamics of the preceding section, together with the fixed-point analysis of Section 32, suggest that the space of descriptions $\mathcal{T}$ has a rich geometric structure analogous to a phase space in classical mechanics. We now make this analogy precise.

Definition 64.1 (Semantic Phase Space)**.** The *semantic phase space* is the pair $(\mathcal{T}, \Phi)$, where $\Phi = \mathcal{C} \circ \mathcal{G} : \mathcal{T} \to \mathcal{T}$ is the round-trip operator mapping each description to the description of its reconstruction.

The operator Φ defines a discrete-time dynamical system on $\mathcal{T}$. A description t is a fixed point of Φ if and only if the corresponding signal $\mathcal{G}(t)$ is a semantic fixed point in the sense of Section 32. The basins of attraction of these fixed points partition $\mathcal{T}$ into semantic classes: all descriptions that converge to the same fixed point under repeated application of Φ represent the same stable semantic content.

Proposition 64.2 (Phase Space Partition)**.** *The basins of attraction of fixed points of Φ form a partition of $\mathcal{T}$ into semantic equivalence classes that refines the gauge equivalence of Section 48.*

Proof sketch. Gauge-equivalent descriptions $t_1 \sim_{\text{gauge}} t_2$ generate the same signal and therefore have the same image under Φ, so they lie in the same basin of attraction. The basin structure may be strictly finer than gauge equivalence if there exist descriptions that generate perceptually distinct signals but nonetheless converge to the same fixed point under the dynamics of Φ. □

The semantic phase space connects the local geometry of $\mathcal{T}$ (curvature, geodesics, gauge orbits) to its global dynamics (basins of attraction, fixed points, drift). Regions of high curvature tend to lie near the boundaries between basins, where small perturbations in a description can lead to qualitatively different stable representations. Regions of low curvature correspond to the interiors of basins, where descriptions are robustly attracted to a single stable semantic class. This geometric picture unifies the analysis of compression stability, semantic drift, and representational robustness within a single dynamical framework.

65 Toward a Unified Theory of Representation

The framework developed throughout this work reveals a deep unification between compression, generation, and interpretation. Signals are no longer primitive objects but realizations of generative descriptions; compression is no longer a numerical approximation but an act of inference; text is no longer a secondary representation but the primary substrate of media.

The mathematical structures introduced—categories, sheaves, fibered spaces, statistical manifolds, and variational principles—are not incidental but necessary. They arise

naturally when one attempts to formalize the relationship between signals and their generative descriptions (Mac Lane, 1998; Grothendieck, 1959; Amari, 2016; Kingma and Welling, 2014). Theorem 51.1 makes the structural origin of this necessity precise: compression is not a property of any particular algorithm but a consequence of separating the specification of constraints from their generative realization.

In this light, semantic compression is not merely a new codec but a new theory of representation. It suggests that all perceptual data can be understood as points in a generative space, coordinatized by text and realized through computation. The boundary between data and program dissolves, and with it the distinction between storage, communication, and cognition.

The long-term implication is a convergence: a single formal framework in which information theory (Shannon, 1948; Cover and Thomas, 2006), category theory (Mac Lane, 1998; Lurie, 2009), machine learning (Goodfellow, Bengio, and Courville, 2016), and cognitive science (Friston, 2010; Tenenbaum et al., 2011) are unified through the concept of generative description.

Before reaching this conclusion, it is worth pausing to identify what is invariant across all the particular mathematical formalisms that the paper has employed. The sections on information theory, category theory, differential geometry, and dynamical systems each illuminate a different facet of the same underlying object.

66 Constraint Invariants and the Structure of Representation

Each mathematical framework employed in this paper isolates a different invariant of the semantic compression process—a quantity or structure that is preserved under the transformations relevant to that framework. Identifying these invariants jointly reveals what is essential to the concept of representation itself.

The information-theoretic invariant is the mutual information $I(X;\mathcal{G})$: the quantity of information about the source captured by the generative model and rendered transmissible through a compressed description. This quantity is invariant under reparameterizations of $\mathcal{G}$ that do not change its expressive power, and it measures exactly the compression gain afforded by the shared prior.

The categorical invariant is the adjunction structure $(\mathcal{C}, \mathcal{G}, \eta, \varepsilon)$: the pair of functors together with their unit and counit natural transformations. The semantic gap—the deviation from an exact adjunction—is invariant under natural isomorphisms of the functors and measures the irreducible abstraction introduced by compression. Perfect representation corresponds to the vanishing of this invariant.

The geometric invariant is the Riemannian metric $g_{\mathcal{T}}$ induced on the description space by the generative map. This metric is invariant under gauge transformations (Section 48), since gauge-equivalent descriptions lie at zero distance under the pullback metric. It encodes local complexity and governs the convergence of optimization and the stability of

fixed points.

The dynamical invariant is the basin partition of the semantic phase space: the decomposition of $\mathcal{T}$ into basins of attraction under the round-trip operator Φ. This partition is preserved by the drift dynamics of Section 63 only when anchored by raw signal injection; without this anchor, the basins contract toward the degenerate fixed points of Ψ.

Theorem 66.1 (Representational Closure)**.** *A representational system $(\mathcal{T}, \mathcal{G}, \mathcal{C})$ is coherent—admitting stable, composable, and editable descriptions—if and only if the category of descriptions is closed under composition, tensor product, and gauge equivalence.*

Proof sketch. Closure under composition ensures that sequential transformations remain within $\mathcal{T}$, so that descriptions can be built from parts. Closure under tensor product ensures that independent components can be combined without leaving $\mathcal{T}$, supporting modularity. Closure under gauge equivalence ensures that canonical representatives exist within each equivalence class, supporting editability: any modification of a description within its gauge orbit leaves the perceptual content invariant. Without any one of these closures, the system admits descriptions that cannot be consistently combined, decomposed, or simplified, destroying coherence. □

Theorem 66.1 establishes that the algebraic properties required for a useful representational medium are not arbitrary design choices but necessary conditions for coherence. The structures introduced throughout the paper—the monoidal category of descriptions, the groupoid of gauge transformations, the compositional delta algebra—are precisely the structures that enforce these closure conditions. Semantic compression succeeds not despite its mathematical complexity but because that complexity is the minimum required for a coherent representational medium to exist.

In such a framework, text becomes the universal interface not only between machines and data but between minds and the worlds they inhabit. The preceding argument has remained largely abstract, operating at the level of categories, manifolds, and information measures. Before the final synthesis, it is worth demonstrating that the architecture described is not merely formally consistent but is concretely instantiated in existing human communication systems. The most compelling such instantiation is American Sign Language, which exhibits in a directly observable form every structural principle that the formal framework has been built to capture.

67 American Sign Language as a Realized Semantic Compression System

The formal architecture developed in the preceding sections—prototype selection, structured deviation encoding, multimodal constraint propagation, gauge equivalence, and intensity-as-gradient modulation—is not an abstract possibility but a system that human

beings have independently constructed and refined over centuries of embodied communication. American Sign Language (ASL) provides a concrete, physiologically grounded realization of precisely these mechanisms. The present section establishes this correspondence explicitly, demonstrating that the mathematical structures of semantic compression are not imposed on signed language from the outside but are recovered by analyzing it from within.

The analysis draws primarily on the systematic account of ASL presented by Riekehof (1978), which provides detailed documentation of the kinematic, spatial, and expressive parameters of signed communication in a form that maps directly onto the formal vocabulary of the preceding sections. The correspondences developed below are not metaphors; they are structural identifications between the theory and an existing implementation.

68 Handshape Classifiers as Prototype Objects

The prototype category $\mathcal{P}$ introduced in Section 26 consists of canonical generative templates from which specific instances are derived by structured deviation. In ASL, this structure is realized through the system of *classifier handshapes*: a finite inventory of base configurations that represent families of referents sharing physical or functional properties. A flat hand with fingers together classifies objects of planar extent; a curved hand classifies graspable volumes; a bent-V configuration classifies a person or animal in a particular posture.

Each classifier is not a fixed symbol but a *parametric template*, in the exact sense of the prototype category. It defines an equivalence class of referents, and specific instances within that class are encoded as deviations along the relevant parameter dimensions: size, orientation, position in the signing space, and trajectory of motion. The classifier itself carries no meaning in isolation; meaning emerges from the specific realization of the template within its context, precisely as a prototype in $\mathcal{P}$ generates a signal only when combined with a deviation Δ.

Fingerspelling constitutes a complementary mechanism: where classifiers provide continuous, gradient-admitting templates, fingerspelling introduces discrete symbolic anchoring. As Riekehof (1978) observes, initializing a sign with the first letter of the corresponding English word—a practice termed *initialization*—combines these two modes, grounding a continuous classifier deformation within a discrete lexical reference point. This is formally equivalent to the two-stage factorization $\mathcal{C} = \mathcal{C}_2 \circ \mathcal{C}_1$ of Section 25: the classifier provides the structural parsing $\mathcal{C}_1$, and the initialization provides the lexical serialization $\mathcal{C}_2$ that anchors the structural content to a specific description in $\mathcal{T}$.

Riekehof (1978) notes that the same base sign for "group" can be initialized to yield "family," "organization," "class," "department," "society," "agency," "association," "workshop," and "team." Each of these is a distinct point in the deviation space over the same classifier prototype—the delta Δ being supplied by the initial handshape that narrows

the equivalence class to a specific lexical interpretation.

69 Dominant Hand Asymmetry as a Structural Prior

The requirement in ASL that one hand be consistently designated as dominant corresponds precisely to the structural prior introduced by gauge fixing in Section 48. As Riekehof (1978) specifies, the dominant hand should be selected at the outset and maintained consistently; alternating between hands, except to encode spatial or relational distinctions, disrupts the interpretive frame.

This convention is not arbitrary. In the formal framework, gauge freedom means that multiple descriptions may generate perceptually equivalent signals. Fixing a gauge—choosing a canonical representative within each orbit of the gauge groupoid $\mathcal{G}_{\text{gauge}}$—reduces this redundancy and provides a stable frame within which signs can be composed and compared. Handedness is the physiological implementation of this gauge-fixing operation: it establishes a consistent orientation field over the action manifold so that compositions of signs remain interpretively stable. A signer who switches dominant hands is, in formal terms, performing an inadmissible gauge transformation that breaks the compositionality of the representation.

Riekehof (1978) further observes that two-handed signs have historically tended toward one-handed forms, and vice versa, with the direction of change governed by ease of production. This is a concrete instance of the drift dynamics described in Section 63: the prototype library contracts toward configurations that minimize productive effort while preserving distinguishability, with the constraint of dominant-hand consistency acting as the gauge anchor that prevents the library from collapsing entirely.

70 Speed, Force, and Motion as Semantic Gradients

Section 38 introduced semantic curvature as a measure of how steeply the perceptual output varies with changes in description parameters. The kinematic parameters of ASL—speed, force, and range of motion—provide a direct physiological realization of this gradient structure.

Riekehof (1978) documents this with precision. The sign "hurry," executed rapidly with tight motion, means something categorically different from the same sign executed slowly with relaxed motion: the former encodes urgency, the latter permission to take one's time. The sign "require" becomes "demand" when executed with increased force. The sign "beautiful" yields, through systematic variation of speed, facial intensity, and spatial extent, the full English synonym chain "lovely," "pretty," "attractive," "beautiful," "gorgeous." These are not different signs; they are different points on a continuous deformation path through the same prototype's deviation space.

This is precisely the semantic gradient structure of Section 38: distinct perceptual

interpretations arising from distinct rates of traversal along a shared structural trajectory. The prototype defines the path; the kinematic parameters define the gradient magnitude along it; and the resulting perceptual output is determined jointly by the two. In the variational framework of Section 28, the kinematic parameters correspond to the continuous component ξ of the latent representation, with each scalar parameter (speed, force, extent) acting as a dimension of the deviation tangent space $T_p\Theta_p$.

Riekehof (1978) is explicit that the appropriate range of kinematic parameters is bounded: excessive motion obscures rather than amplifies meaning, just as an overly long description in the TIR increases $L(t)$ without reducing distortion. The optimal signing, like the optimal compressed description, minimizes descriptive cost while maximizing perceptual distinctiveness.

71 Facial Expression and Body Posture as Global Constraint Fields

The sheaf-theoretic structure of Section 21 requires that local descriptions be governed by global constraint fields: parameters defined over the entire domain that propagate consistently across all local sections. In ASL, facial expression and body posture function as precisely such global fields.

Riekehof (1978) emphasizes that deaf signers attend to the face and body as much as—often more than—the hands. The sign for "like," accompanied by a pleasant expression, encodes enjoyment; the same manual configuration accompanied by a negative head shake encodes dislike. The sign for "tired" is accompanied by a whole-body postural sag; the sign for "strong" by a forward thrust of the chest and backward movement of the shoulders. These are not modifiers attached to signs; they are global boundary conditions that determine the interpretation of the entire local trajectory.

In sheaf-theoretic terms, the face and body encode the global sections over the entire signing space, while the hands encode local sections over specific sub-regions. The sheaf condition—that local sections must agree with the global section on their shared boundary—manifests as the requirement that manual signs and facial expression be synchronized and coherent. A sign whose manual component conflicts with its facial component is incoherent in the same way that local descriptions conflicting on their overlap regions fail the sheaf gluing condition: no consistent global description exists, and the representation breaks down.

The centralization of signs toward the face, documented by Riekehof (1978) as a historical trend driven by the increasing emphasis on speech and lipreading in deaf education, reflects an adjustment of the local sections toward better alignment with the global field. Signs that obscure the face disrupt the global constraint and reduce legibility; centralization restores it.

72 Spatial Scaling and Observational Regime

Section 10 established that the effective capacity of a semantic channel depends on the expressive power of the shared generative model and the entropy of the description space. ASL provides a striking instantiation of this principle through the relationship between signing scale and audience size.

Riekehof (1978) documents that the spatial extent of signs scales with the size of the signing group and the distance to be covered. On a one-to-one basis, signs are compact and close to the body. For large groups or long distances, signs expand spatially and slow temporally—not because the content changes, but because the channel conditions change. The informational content of the description remains the same; the physical realization is scaled to match the perceptual constraints of the observational regime.

This is formally equivalent to adjusting the distortion tolerance ϵ in response to channel noise. The semantic content—the description t—is invariant across scales. What changes is the rendering: the generative process $\mathcal{G}$ is executed with a different set of perceptual constraints, producing a physically larger realization that remains within the distortion budget of a more distant observer. The channel bottleneck, as noted in Section 10, is perceptual rather than informational; the signing system compensates by adjusting the physical signal, not the semantic description.

Riekehof (1978) also notes that fingerspelling becomes impractical at large distances because the discrete handshape configurations cannot be resolved. This corresponds to the regime where the perceptual metric d_ω for a distant observer is too coarse to support fine-grained symbolic distinctions. The system adapts by reducing reliance on fingerspelling—the discrete, high-resolution component—and falling back on classifiers and kinematic modulation, which carry information at a coarser granularity appropriate to the observational regime.

73 Historical Drift and Structural Simplification in ASL

The drift dynamics formalized in Section 63 predict that, in the absence of countervailing pressure, a sign system will contract toward configurations that minimize productive effort while preserving semantic distinguishability. Riekehof (1978) provides extensive historical documentation confirming this prediction across multiple decades of ASL evolution.

Two-handed signs for "cow," "horse," and "devil" have become one-handed. The sign for "law," formerly made with the thumb pointing forward, now uses the thumb pointing back, eliminating an awkward wrist rotation. "Sister" and "brother" have shifted from a lateral to a stacked configuration, eliminating a wrist twist. Complex compound signs for the seasons—"three-months-grow" for spring, "three-months-cold" for winter—have contracted to their final element alone. Signs have moved toward the center of the signing space, shortening the spatial excursion required.

Each of these changes is an instance of the drift operator Ψ contracting the prototype

library toward lower-effort configurations. The constraint that maintains semantic distinguishability is the communicative context: signs must remain legible to conversational partners, which prevents full collapse. The result is a gradual simplification that preserves the semantic invariants—the equivalence classes—while relaxing the physical realizations toward the attractor of minimal effort.

Riekehof (1978) also documents regional variation in signs for holidays and other culturally specific referents, particularly among residential school communities. This is the expected behavior under the adaptive prototype expansion rule of Section 35: when no existing prototype adequately represents a novel cultural referent, new prototypes emerge locally, producing regional divergence. The acceptance of such variation within the Deaf community—without normative insistence that one sign is right and another wrong—reflects an implicit recognition that the relevant invariant is the equivalence class, not the specific realization.

74 ASL as an Implementation of the Formal Architecture

The preceding sections have established that ASL is not merely an analogy for semantic compression but an independent implementation of its formal architecture. The correspondences are structural, not metaphorical:

The classifier inventory corresponds to the prototype category $\mathcal{P}$, with each classifier defining a parametric template over a family of referents. Initialization corresponds to the two-stage factorization $\mathcal{C} = \mathcal{C}_2 \circ \mathcal{C}_1$, combining continuous structural encoding with discrete lexical anchoring. Kinematic parameters—speed, force, spatial extent—correspond to the continuous deviation coordinates $\xi \in T_p\Theta_p$, with the synonym chain ("lovely" through "gorgeous") corresponding to a path through the deviation manifold. Dominant-hand convention corresponds to gauge fixing, establishing the canonical orientation field that makes compositions interpretively stable. Facial expression and body posture correspond to global constraint fields in the sheaf-theoretic sense, with the manual signs as local sections governed by global boundary conditions. Spatial scaling corresponds to the adjustment of rendering parameters to match the perceptual constraints of the observational regime. Historical drift corresponds to the contraction dynamics of Ψ, with communicative intelligibility functioning as the raw-signal injection that prevents full collapse.

Each of these correspondences is independently motivated by the formal theory and independently documented by Riekehof (1978). Their simultaneous presence in a single human communication system provides strong empirical evidence that the structural principles identified in the formal framework are not artifacts of a particular mathematical choice but reflect genuine invariants of embodied multimodal communication.

This has a theoretical implication beyond the specific case of ASL. If a rich, historically evolved, physiologically constrained human communication system independently con-

verges on the same formal architecture—prototype templates, deviation encoding, gauge constraints, global fields, scaling adaptation, and drift dynamics—then this architecture is not one possible design among many but the natural structure of efficient constraint-based communication in a physical world. The formal framework is not imposed on embodied communication; it is the abstraction that embodied communication has always already been performing.

Text as Substrate represents the final decoupling of meaning from signal. By formalizing the relationship between constraint and generation, we have outlined a world where reality is no longer recorded but described; where the past is not stored as a trace but as a recipe for reconstruction. The mathematics of sheaves, categories, and Lie groups provide the necessary rigor to ensure that this reconstruction remains coherent, but they also reveal the profound vulnerability of a world seen entirely through the lens of a shared prior. We have traded the noise of the real for the silence of the perfect simulation—and the measure of that trade is precisely the mutual information $I(x; \mathcal{G})$ that moves from the file into the model.

Appendices

A Formalization of Semantic Compression as an Optimization Problem

Let $\mathcal{M}$ be a measurable space of multimodal signals and $\mathcal{T}$ a space of structured textual descriptions. Let $\mathcal{G} : \mathcal{T} \to \mathcal{M}$ be a generative operator and $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$ a perceptual distortion functional.

Semantic compression is defined as the variational problem

$$\mathcal{C}(x) = \arg\min_{t \in \mathcal{T}} \left[d\big(x, \mathcal{G}(t)\big) + \lambda\, L(t) \right],$$

where $L : \mathcal{T} \to \mathbb{R}_{\geq 0}$ is a description length functional and $\lambda > 0$ is a trade-off parameter.

Assuming $\mathcal{T}$ admits a parameterization $t = (p, \delta)$ with $p \in \mathcal{P}$ and $\delta \in T_p\Theta_p$, one obtains

$$\mathcal{C}(x) = \arg\min_{p \in \mathcal{P},\, \delta \in T_p\Theta_p} \left[d\big(x, \mathcal{G}(p, \delta)\big) + \lambda \left(L(p) + L(\delta) \right) \right].$$

Under regularity assumptions on $\mathcal{G}$ (specifically, that $\mathcal{G}$ is Fréchet differentiable with respect to δ), the first-order optimality condition for the continuous part is

$$D_\delta \mathcal{G}(p, \delta)^* \, \nabla_{\mathcal{G}} d\big(x, \mathcal{G}(p, \delta)\big) + \lambda \, \nabla_\delta L(\delta) = 0,$$

where $D_\delta \mathcal{G}^*$ is the adjoint of the Fréchet derivative. This establishes semantic compression as a constrained inference problem in a generative model, analogous to rate–distortion theory but with structure-dependent priors.

B Rate–Distortion Interpretation with Model Dependence

Let X be a random variable over $\mathcal{M}$ and T a random variable over $\mathcal{T}$. Define a joint distribution induced by an encoder $q(t|x)$. The expected distortion is

$$D = \mathbb{E}_{x \sim X,\, t \sim q(\cdot|x)} \left[d\big(x, \mathcal{G}(t)\big) \right].$$

The semantic rate–distortion function is

$$R_{\mathcal{G}}(D) = \inf_{q(t|x)} \left\{ I(X;T) \,|\, \mathbb{E}[d(x, \mathcal{G}(t))] \leq D \right\}.$$

For any generative model $\mathcal{G}$ and any D, $R_{\mathcal{G}}(D) \leq R(D)$. When $\mathcal{G}$ is a perfect generative model for the source, the semantic rate–distortion function satisfies

$$R_{\mathcal{G}}(D) = H(Z) - I(Z; \text{residual}),$$

where Z is the latent variable and the residual captures stochastic variation not explained by the compressed description.

C Categorical Equivalence up to Perceptual Isomorphism

Define $\sim_\epsilon$ by $x \sim_\epsilon y \iff d(x, y) < \epsilon$. Semantic compression achieves *categorical equivalence* if there exists a functor $\mathcal{C} : \mathcal{M}/\sim_\epsilon \to \mathcal{T}$ such that

$$\mathcal{G} \circ \mathcal{C} \cong \mathrm{Id}_{\mathcal{M}/\sim_\epsilon}.$$

The category $\mathcal{T}$ serves as a skeletal representation of $\mathcal{M}/\sim_\epsilon$. The existence of such a $\mathcal{C}$ is guaranteed when $\mathcal{G}$ is surjective up to ϵ and metrically injective on $\mathcal{T}$.

D Sheaf-Theoretic Construction of Global Descriptions

Let $X = [0, T] \times \mathbb{R}^2$ be the spacetime domain. Define a presheaf $\mathcal{F}$ on X with restriction maps $\rho_{UV} : \mathcal{F}(U) \to \mathcal{F}(V)$. The sheaf condition requires: for any open cover $\{U_i\}$ of U and compatible sections t_i, there exists a unique $t \in \mathcal{F}(U)$ restricting to each t_i.

The sheaf condition is verified when $\mathcal{G}$ is *locally determined*: its output over V depends only on the restriction of the description to V. When descriptions contain global context (overall color grade, speaker identity), the global parameter must be factored out before local descriptions are assembled.

E Lie Group Structure of Transformations

Let G be a Lie group acting smoothly on $\mathcal{P}$ with $\alpha : G \times \mathcal{P} \to \mathcal{P}$ compatible with $\mathcal{G}$: $\mathcal{G}(\alpha(g, p)) = \mathcal{T}_g(\mathcal{G}(p))$. The Lie algebra $\mathfrak{g}$ parameterizes infinitesimal deviations via the exponential map $\exp : \mathfrak{g} \to G$:

$$\mathcal{G}(\alpha(\exp(\xi), p)) \approx \mathcal{G}(p) + D_\xi \mathcal{G}(p) \cdot \xi + O(\|\xi\|^2).$$

For camera trajectories, the relevant group is $\mathrm{SE}(3) = \mathrm{SO}(3) \ltimes \mathbb{R}^3$.

F Kolmogorov Complexity Relative to a Generative Model

The model-relative Kolmogorov complexity is

$$K_\epsilon(x \mid \mathcal{G}) \coloneqq \min\{|t|_U : d(x, \mathcal{G}(U(t))) < \epsilon\}.$$

By the invariance theorem, $K_\epsilon(x \mid \mathcal{G}) \leq K_\epsilon(x) + c$, with improvement $K_\epsilon(x) - K_\epsilon(x \mid \mathcal{G}) \approx K(\mathcal{G}) - O(1)$ when $\mathcal{G}$ captures the generating process of x. When $\mathcal{G}$ is shared between

encoder and decoder, the per-signal cost is $K_\epsilon(x \mid \mathcal{G})$ alone.

G Fixed Point and Stability Analysis

Theorem G.1 (Banach Fixed Point for Compression)**.** *If $(\mathcal{M}, d)$ is a complete metric space and $\Phi = \mathcal{G} \circ \mathcal{C}$ is a strict contraction with $d(\Phi(x), \Phi(y)) \leq \kappa\, d(x, y)$ for $\kappa \in [0, 1)$, then Φ has a unique fixed point $x^* \in \mathcal{M}$, and the iterates $x_{n+1} = \Phi(x_n)$ converge to x^* at rate κ^n.*

When Φ is a local contraction on connected components of $\mathcal{M}$, the Banach theorem guarantees local fixed points—stable attractors within each perceptual category.

H Functorial Semantics of the Textual Language

Let $\mathcal{L}$ be a context-free grammar defining the TIR. The free model $\mathcal{F}(\mathcal{L})$ is a category whose objects are types and whose morphisms are derivations. A *semantic interpretation* is a functor $\cdot : \mathcal{F}(\mathcal{L}) \to \mathcal{M}$.

Compositionality requires this functor to be monoidal: $A \otimes B = A \otimes B$. The *full completeness* condition—that every morphism in $\mathcal{M}$ is the image of some derivation in $\mathcal{F}(\mathcal{L})$—characterizes the expressiveness of the TIR.

I Asymptotic Universality

Let $\{\mathcal{G}_n\}_{n \geq 1}$ be a sequence of generative models of increasing expressive power.

Theorem I.1 (Asymptotic Universality)**.** *Suppose $\mathcal{G} = \bigcup_n \mathcal{G}_n$ is a universal approximator for $\mathcal{M}$: for every $x \in \mathcal{M}$ and $\epsilon > 0$, there exists N and $t \in \mathcal{T}$ such that $d(x, \mathcal{G}_N(t)) < \epsilon$. Then*

$$\lim_{n \to \infty} K_\epsilon(x \mid \mathcal{G}_n) = K_\epsilon^*(x),$$

where $K_\epsilon^*(x)$ is the minimum description length achievable by any effective procedure.

As generative models approach universal approximation, description lengths approach the information-theoretic lower bound. Text becomes a universal coordinate system for the space of perceptual phenomena.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 2nd edition, 2006.

[3] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.

[4] R. M. Gray, *Entropy and Information Theory*, Springer, 2011.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2nd edition, 2011.

[6] A. N. Kolmogorov, "Three Approaches to the Quantitative Definition of Information," *Problems of Information Transmission*, vol. 1, no. 1, pp. 1–7, 1965.

[7] G. J. Chaitin, *Algorithmic Information Theory*, Cambridge University Press, 1987.

[8] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, 3rd edition, 2008.

[9] J. Rissanen, "Modeling by Shortest Data Description," *Automatica*, vol. 14, pp. 465–471, 1978.

[10] P. Grünwald, *The Minimum Description Length Principle*, MIT Press, 2007.

[11] R. J. Solomonoff, "A Formal Theory of Inductive Inference," *Information and Control*, vol. 7, pp. 1–22, 1964.

[12] M. Hutter, *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*, Springer, 2005.

[13] N. Tishby and N. Zaslavsky, "Deep Learning and the Information Bottleneck Principle," *IEEE Information Theory Workshop*, pp. 1–5, 2015.

[14] J. R. Pierce, *An Introduction to Information Theory: Symbols, Signals and Noise*, Dover, 1980.

[15] W. Weaver, "Recent Contributions to the Mathematical Theory of Communication," in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1949.

[16] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," *Advances in Neural Information Processing Systems*, 2014.

[20] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *International Conference on Learning Representations*, 2014.

[21] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic Backpropagation and Approximate Inference in Deep Generative Models," *International Conference on Machine Learning*, 2014.

[22] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel Recurrent Neural Networks," *International Conference on Machine Learning*, 2016.

[23] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," *9th ISCA Speech Synthesis Workshop*, 2016.

[24] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, "Neural Discrete Representation Learning," *Advances in Neural Information Processing Systems*, 2017.

[25] A. Razavi, A. van den Oord, and O. Vinyals, "Generating Diverse High-Fidelity Images with VQ-VAE-2," *Advances in Neural Information Processing Systems*, 2019.

[26] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *Advances in Neural Information Processing Systems*, 2020.

[27] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics," *International Conference on Machine Learning*, 2015.

[28] Y. Song and S. Ermon, "Generative Modeling by Estimating Gradients of the Data Distribution," *Advances in Neural Information Processing Systems*, 2019.

[29] T. Karras, M. Aittala, T. Aila, and S. Laine, "Elucidating the Design Space of Diffusion-Based Generative Models," *Advances in Neural Information Processing Systems*, 2022.

[30] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[31] P. Esser, R. Rombach, and B. Ommer, "Taming Transformers for High-Resolution Image Synthesis," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[32] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-Shot Text-to-Image Generation," *International Conference on Machine Learning*, 2021.

[33] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," *International Conference on Machine Learning*, 2021.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, 2017.

[35] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models Are Few-Shot Learners," *Advances in Neural Information Processing Systems*, 2020.

[36] C. Villani, *Optimal Transport: Old and New*, Springer, 2009.

[37] S. Amari, *Information Geometry and Its Applications*, Springer, 2016.

[38] S. Mac Lane, *Categories for the Working Mathematician*, Springer, 2nd edition, 1998.

[39] E. Riehl, *Category Theory in Context*, Dover Publications, 2016.

[40] F. W. Lawvere and S. H. Schanuel, *Conceptual Mathematics: A First Introduction to Categories*, Cambridge University Press, 2nd edition, 2009.

[41] B. Jacobs, *Categorical Logic and Type Theory*, Elsevier, 1999.

[42] D. I. Spivak, *Category Theory for the Sciences*, MIT Press, 2014.

[43] B. Fong and D. I. Spivak, *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*, Cambridge University Press, 2019.

[44] F. Lorégian, *(Co)end Calculus*, Cambridge University Press, 2021.

[45] A. Grothendieck, "Catégories fibrées et descente," in *Séminaire de Géométrie Algébrique du Bois-Marie*, 1959.

[46] M. Artin, A. Grothendieck, and J.-L. Verdier, *Théorie des Topos et Cohomologie Étale des Schémas (SGA 4)*, Springer, 1972.

[47] S. Mac Lane and I. Moerdijk, *Sheaves in Geometry and Logic*, Springer, 1992.

[48] J. Lurie, *Higher Topos Theory*, Princeton University Press, 2009.

[49] J. Lurie, *Higher Algebra*, Preprint, available at `math.harvard.edu/~lurie/`, 2017.

[50] B. Coecke and A. Kissinger, *Picturing Quantum Processes*, Cambridge University Press, 2017.

[51] J. C. Baez and M. Stay, "Physics, Topology, Logic and Computation: A Rosetta Stone," in B. Coecke (ed.), *New Structures for Physics*, Springer, 2010.

[52] M. P. do Carmo, *Riemannian Geometry*, Birkhäuser, 1992.

[53] J. M. Lee, *Introduction to Smooth Manifolds*, Springer, 2nd edition, 2012.

[54] S. Lang, *Differential and Riemannian Manifolds*, Springer, 1995.

[55] B. C. Pierce, *Types and Programming Languages*, MIT Press, 2002.

[56] R. Milner, *Communicating and Mobile Systems: The Pi-Calculus*, Cambridge University Press, 1999.

[57] J. Hopcroft, R. Motwani, and J. Ullman, *Introduction to Automata Theory, Languages, and Computation*, Pearson, 3rd edition, 2006.

[58] J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, 2nd edition, 2009.

[59] J. Barwise and J. Perry, *Situations and Attitudes*, MIT Press, 1983.

[60] N. Chomsky, *Aspects of the Theory of Syntax*, MIT Press, 1965.

[61] K. Friston, "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews Neuroscience*, vol. 11, pp. 127–138, 2010.

[62] A. Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford University Press, 2015.

[63] D. Dennett, *The Intentional Stance*, MIT Press, 1987.

[64] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, "How to Grow a Mind: Statistics, Structure, and Abstraction," *Science*, vol. 331, pp. 1279–1285, 2011.

[65] N. D. Goodman, J. B. Tenenbaum, and T. D. Ullman, "Probabilistic Models of Cognition," *Trends in Cognitive Sciences*, vol. 19, no. 10, pp. 589–599, 2015.

[66] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, "Human-Level Concept Learning through Probabilistic Program Induction," *Science*, vol. 350, pp. 1332–1338, 2015.

[67] V. Mansinghka, D. Selsam, and Y. Perov, "Venture: A Higher-Order Probabilistic Programming Platform with Programmable Inference," arXiv:1404.0099, 2014.

[68] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[69] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[70] D. Ha and J. Schmidhuber, "World Models," *Advances in Neural Information Processing Systems*, 2018.

[71] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver, "Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model," *Nature*, vol. 588, pp. 604–609, 2020.

[72] J. Schmidhuber, "Formal Theory of Creativity, Fun, and Intrinsic Motivation," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 230–247, 2010.

[73] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, pp. 484–489, 2016.

[74] L. L. Riekehof, *The Joy of Signing: The Illustrated Guide for Mastering Sign Language and the Manual Alphabet*, Gospel Publishing House, Springfield, MO, 2nd edition, 1978.