The Geometry of Meaning: A Framework for Symbol Grounding in Perceptual Manifolds

The Crisis of Ungrounded Symbols: From the Chinese Room to the Merry-Go-Round

The pursuit of artificial intelligence (AI) has been shadowed for decades by a foundational philosophical and computational challenge known as the symbol grounding problem. First articulated in its modern form by Stevan Harnad, this problem questions how the symbols manipulated by a computational system can acquire intrinsic meaning—meaning that is inherent to the system itself, rather than being parasitically dependent on the interpretations of an external human observer. At its core, it is the problem of how abstract tokens, processed purely on the basis of their shape and syntactic rules, can connect to the real-world objects, events, and concepts they are intended to represent. This crisis of ungrounded symbols has persistently limited the capabilities of AI, creating systems that can exhibit remarkable syntactic prowess but lack genuine semantic understanding. The Color-Geometry framework presented in this report offers a novel and comprehensive solution, not by creating a new mapping between symbols and their referents, but by proposing a new kind of symbol altogether—one whose very constitution is its meaning, grounded in the physical and geometric laws of a perceptual manifold.

The Chinese Room and the Parasitic Interpretation

The philosophical genesis of the modern symbol grounding problem can be traced to John Searle's influential "Chinese Room Argument". In this thought experiment, Searle imagines a person who does not understand Chinese locked in a room with a set of formal rules (a program) for manipulating Chinese symbols. By receiving symbols through a slot, applying the rules, and passing new symbols out, the person can produce responses that are indistinguishable from those of a native Chinese speaker. To an outside observer, the room appears to understand Chinese. However, the person inside has no comprehension of the meaning of the symbols; they are merely manipulating tokens based on their shape (syntax). Searle's argument serves as a powerful critique of what he termed "strong AI," the view that a suitably programmed computer could possess genuine cognitive states like understanding. The core of the critique lies in the distinction between syntax and semantics. A formal symbol system, as defined in this context, is a set of arbitrary physical tokens—be they scratches on paper, events in a digital computer, or the Chinese characters in Searle's room—that are manipulated according to explicit, syntactic rules. The system can perform these manipulations flawlessly without any access to the meaning, or semantics, of the symbols involved.

This leads to the concept of "parasitic interpretation." The meaning of the symbols in the Chinese Room is not intrinsic to the system (the room and its contents) but is entirely supplied by the external observers who already understand Chinese. The system's apparent understanding is parasitic on the minds of its users or creators. This critique was particularly potent against the dominant paradigm of "Good Old-Fashioned AI" (GOFAI), where intelligence was modeled as the rule-based manipulation of symbolic representations in a "language of thought". In such systems, symbols like DOG or CHAIR are ungrounded by design; their connection to real-world dogs and chairs exists only in the mind of the programmer who wrote the code. The system itself has no independent access to this meaning. The Color-Geometry framework directly confronts this by proposing a system where the symbol for a perception, such as a specific shade of red, is not an arbitrary token but is the very vector in a perceptual space that corresponds to that perception. The meaning is thus made intrinsic, severing the parasitic link to an external interpreter.

The fundamental flaw exposed by Searle is not merely technical but one of representation itself. Traditional symbolic AI operates on the implicit assumption that meaning can be adequately captured and processed through arbitrary tokens. The symbol grounding problem reveals this assumption to be untenable for achieving genuine, autonomous intelligence. The Color-Geometry framework offers a radical alternative to this representational model. It posits that meaning is not *represented by* a symbol, but rather that the meaning *is* the symbol's state within a physically constrained perceptual geometry. A symbol is not a placeholder for a concept; it is a specific coordinate in a meaning-space. For instance, a particular shade of crimson is not represented by an arbitrary string like "CRIMSON_#7B1113"; it *is* the vector $x = (L, a, b)$ that corresponds to the perception of that color in the OKLab space. This vector is not an arbitrary token; its position and its relationships to all other vectors in the space are governed by the geometry of human perception. By collapsing the distinction between the symbol and its meaning, the framework makes the meaning intrinsic to the system's architecture, thereby solving the problem of parasitic interpretation at its root.

Harnad's Formulation: The Symbol/Symbol Merry-Go-Round

Building on Searle's philosophical critique, cognitive scientist Stevan Harnad provided a more concrete computational formulation of the symbol grounding problem. Harnad famously analogized the predicament of an ungrounded symbolic system to that of a person trying to learn Chinese using only a Chinese-Chinese dictionary. The definition of any given symbol consists only of other, equally undefined symbols. This leads to an endless, circular chain of definitions—a "symbol/symbol merry-go-round"—that never makes contact with the external world. One can traverse the entire dictionary, moving from one symbol string to another, without ever discovering what any of the symbols actually mean.

This metaphor powerfully illustrates the self-referential trap of any purely formal system. The meaning of a symbol cannot be derived from its relationships with other symbols alone; it must be "grounded" in something outside the symbolic system. Harnad proposed a candidate solution: symbolic representations must be grounded "bottom-up" in non-symbolic representations that are directly connected to the agent's sensory and motor experiences. He identified two crucial kinds of non-symbolic representations:

1. **Iconic Representations:** These are analogs of the proximal sensory projections of objects and events. For vision, this would be the raw pattern of light on the retina. They are non-symbolic because their structure is not arbitrary but is instead a direct physical consequence of the object being perceived.
2. **Categorical Representations:** These are learned or innate abilities to identify invariant features from iconic representations. This is the process of picking out an object as a member of a category (e.g., identifying many different iconic representations as instances

of "horse"). The elementary symbols of the system are the names attached to these categories.

Higher-order symbolic representations can then be built compositionally from these grounded elementary symbols (e.g., "a zebra is a horse that is striped"). In this hybrid model, symbol manipulation is no longer governed solely by the arbitrary shapes of the tokens but is constrained by the non-arbitrary structure of the sensory representations in which they are grounded.

Harnad's formulation provides the crucial theoretical bridge to the Color-Geometry framework. His call for grounding in iconic and categorical representations highlights a deep tension between the discrete, combinatorial nature of symbolic logic and the continuous, high-dimensional nature of perception. Symbolic reasoning is built upon discrete tokens and formal rules, such as IF P THEN Q. Perception, in contrast, is fundamentally continuous; sensory data exists on a spectrum, not as a set of distinct tokens. The Color-Geometry framework offers a concrete and mathematically rigorous method for resolving this tension. It operationalizes Harnad's bridge by proposing a continuous perceptual manifold—the OKLab color space—as the substrate for grounding. Within this space, discrete, logical-like operations emerge from continuous geometric operations. For example, a causal link like "A causes B" is not evaluated via a logical rule but through a continuous geometric test: is the vector representing state B, x_B, reachable within the continuous volume defined by the vector for state A, x_A?. In this way, symbolic inference and composition are implemented as geometric operations like set intersection and convex mixing on a continuous space, directly resolving the discrete/continuous divide that lies at the heart of the grounding problem.

The Limitations of Traditional Solutions

The challenge laid down by Searle and Harnad has inspired numerous attempts at a solution, yet the problem has remained largely unresolved. Early hopes were placed on connectionist models, or artificial neural networks. These systems learn representations from data rather than being programmed with explicit symbols, suggesting a path to bottom-up grounding. However, as philosophers Jerry Fodor and Zenon Pylyshyn argued, traditional connectionist networks often lack the systematic, compositional structure that is characteristic of language and thought. It is difficult to see how the distributed patterns of activation in a simple network can be rulefully combined and recombined to generate the unbounded expressive power of a true symbolic system.

This led to the development of hybrid systems, which attempt to combine a symbolic reasoning module with a connectionist perceptual module. A typical approach might involve a neural network that classifies an image as containing a "cup," passing this symbolic token to a logical inference engine. While seemingly intuitive, this approach often fails to solve the grounding problem in a deep sense. The connection between the perceptual module and the symbolic module can be just as arbitrary and ungrounded as the symbols themselves. The token "cup" produced by the network is still just an arbitrary string whose meaning is only established by the system's designer. The core challenge of making the meaning *intrinsic* to the system's architecture remains.

Many proposed solutions fall into a trap that has been described as "semantic commitment". They presuppose the very meanings they are supposed to be grounding. For example, a supervised neural network trained to label images is grounded only in the pre-existing semantic categories provided by its human trainers. The network learns a mapping, but the meaning of the labels is entirely extrinsic. An unsupervised network might discover clusters in data, but

without a mechanism to connect these clusters to the agent's interactions with the world, they remain ungrounded patterns.
This history of attempts establishes a high bar for any candidate solution. A successful grounding framework must not simply connect symbols to sensory data. The connection itself must be non-arbitrary, governed by principles that are intrinsic to the system's own structure and dynamics. The meaning must emerge from the system's architecture, not be imposed upon it.
This is precisely the claim of the Color-Geometry framework. It proposes that the geometric and physical laws of a well-chosen perceptual space—the OKLab color space—can provide the non-arbitrary, intrinsic foundation that has been missing. By defining symbols as states within this space and semantic operations as geometric transformations, the framework creates a system where meaning is not an external interpretation but a direct consequence of the system's physical and mathematical constitution.

A Substrate for Meaning: The OKLab Perceptual Manifold

To construct a system where meaning is intrinsic, one must first identify a substrate whose properties are not arbitrary but are instead tethered to a consistent, measurable reality. The Color-Geometry framework posits that a perceptually uniform color space, specifically the OKLab space, provides such a substrate. This section provides a rigorous technical exposition of OKLab, establishing it not as a mere visual metaphor for semantic relationships, but as a formal mathematical and physical manifold whose inherent structure can serve as the foundation for a new kind of grounded semantics. The space's specific properties—its perceptual uniformity, the hard physical constraint of its gamut, and the ability to define a "safe" representational system—are the pillars upon which the framework is built.

Formalizing the Computational Space: Perceptual Uniformity as a Semantic Metric

The OKLab color space is a three-dimensional real vector space, $\mathbb{R}^3$, where each point corresponds to a unique color perception under standardized viewing conditions. A color is represented by a vector $x = (L, a, b)$, where the components have direct perceptual interpretations:

- **L**: The perceptual lightness, normalized such that $L=0$ is pure black and $L=1$ is a reference white.
- **a**: The green-red opponent axis, where negative values correspond to green and positive values to red.
- **b**: The blue-yellow opponent axis, where negative values correspond to blue and positive values to yellow.

The defining and most critical feature of OKLab is its *perceptual uniformity*. This means that the Euclidean distance between any two color points in the space, ΔE, corresponds directly to the perceived difference between those two colors. The perceptual difference is calculated as the standard Euclidean norm:
This property is a significant improvement over older and more common color spaces. In spaces like RGB or HSV, a linear interpolation between two points can result in a path that includes perceptually non-uniform changes; for example, a straight line in RGB space might produce an

unexpected shift in perceived brightness or hue. In CIELAB, another perceptual space, linear operations can introduce unwanted hue shifts, particularly in the blue regions. OKLab was specifically designed by Björn Ottosson to correct these deficiencies, creating a space where linear operations have perceptually predictable consequences.
This uniformity is not a mere technical convenience; it is the cornerstone of the framework's claim to be a "meaning space." By adopting OKLab, the framework acquires an *intrinsic, non-arbitrary metric for semantic similarity*. The difference in meaning between two elementary symbols is defined as being directly and measurably proportional to the geometric distance between their vector representations. This is a profound departure from traditional symbolic systems, where the similarity between symbols is either undefined or must be specified externally in an ad-hoc knowledge base. In the Color-Geometry framework, semantic similarity is a measurable geometric fact. This allows for the principled definition of objective, perceptually meaningful loss functions, tolerances, and optimization targets. An error in the system is not an abstract symbolic mismatch but a measurable perceptual distance, ΔE, which can be minimized using geometric methods.
This choice of a foundational space provides a bridge to the powerful tools of differential geometry and the emerging field of geometric deep learning. Because OKLab is a continuous 3D vector space where the ΔE metric defines a consistent measure of distance at every point, it can be formally treated as a Riemannian manifold. In this specific case, the metric tensor is constant (the identity matrix), making it a simple Euclidean space. However, the principle is general. This geometric foundation allows for principled definitions of concepts like geodesics (the perceptually "straightest" path between two colors), curvature (how the space of perception itself is shaped), and gradients of perceptual change. While the specific SNN implementation detailed later is gradient-free, the underlying space is not. This implies that the framework could be readily adapted for gradient-based learning systems. In such a system, the gradients would be calculated over a perceptually meaningful geometry, not an arbitrary, high-dimensional latent space. This provides a direct and powerful connection to the broader research program of geometric deep learning, which focuses on developing models that can learn on non-Euclidean domains by respecting the intrinsic geometry of the data.

The Geometry of Feasibility: The Gamut Droplet D as a Physical Law

While the abstract OKLab vector space is unbounded, the set of colors that can be physically realized by a given device, such as a standard sRGB monitor, is a finite and constrained subset. When the sRGB gamut is mapped into the OKLab space, it forms a complex, non-convex 3D solid, referred to in the framework as the **gamut droplet**, $\mathcal{D}$. This volume is not a simple geometric primitive like a sphere or a cube. It has a characteristic shape that is widest at mid-lightness values and tapers significantly towards pure black ($L=0$) and pure white ($L=1$). This shape is a direct reflection of a physical reality: highly saturated, vibrant colors can only be produced at intermediate levels of brightness. It is physically impossible for a standard display to produce, for example, a highly saturated yellow that is also extremely dark.
The Color-Geometry framework reframes this physical limitation as its most powerful and defining feature. The gamut droplet $\mathcal{D}$ is not treated as a mere graphical inconvenience to be managed with "clipping" algorithms, but as a **hard physical constraint** that functions as a fundamental law of the system. This droplet defines the absolute boundaries of all possible valid perceptual states. A symbol, represented by a vector x, is considered to have a "physically realizable meaning" if and only if its corresponding point lies within this volume: $x \in \mathcal{D}$. Any point outside the droplet represents a physically impossible color and is

therefore, by definition, a meaningless state within the system.
This transforms the concept of a color gamut from a problem in computer graphics into a fundamental law of semantic feasibility. The boundary of the droplet is not a mathematical construct imposed for convenience; it is a direct consequence of the physics of light and the biology of perception, as captured by the sRGB standard. This physical boundary becomes the ultimate arbiter of what can and cannot be represented. The formal definition of the droplet relies on a "gamut oracle" function, $C_{\max}(L, h)$, which returns the maximum possible chroma (saturation) for any given lightness L and hue angle h. The gamut droplet $\mathcal{D}$ is then the set of all points (L, a, b) such that their chroma, $\sqrt{a^2 + b^2}$, is less than or equal to this maximum possible value :
This physical constraint provides a form of intrinsic, non-symbolic regularization. In traditional machine learning, regularization is an artificial mathematical constraint (such as an L_1 or L_2 norm) added to a model's objective function to prevent overfitting and encourage simpler, more generalizable solutions. In the Color-Geometry framework, the constraint is not an artificial penalty term; it is the physical boundary of perception itself. Any learning process or computational operation, such as the color mixing described later, that attempts to produce a result x_v outside of $\mathcal{D}$ has produced a physically impossible and thus semantically invalid state. The system is therefore forced, by its very design, to find solutions that remain within this physically-defined boundary. The geometry of the gamut itself regularizes the space of possible solutions, ensuring that all learned representations correspond to plausible perceptual states without the need for arbitrary, externally imposed mathematical regularizers.

In-Gamut by Construction: A Safe Representational Space

A direct representation of a symbol's state in (L, a, b) or even (L, C, h) coordinates presents a practical challenge: an arbitrary modification to any of these parameters during a learning update could easily push the resulting color outside the complex boundary of the gamut droplet $\mathcal{D}$, resulting in an invalid, physically unrealizable state. To circumvent this and ensure the integrity of the representational space, the framework introduces a novel parametrization, inspired by the geometry of an inclined double cone, that guarantees any generated state is valid *by construction*.
Instead of representing a neuron's state directly by its physical chroma C, the system uses a latent, normalized saturation parameter, $S' \in $. The state of a symbol is thus parameterized by the tuple (L, S', h). The true, physical chroma C is then derived from this latent representation using the gamut oracle function described previously:
The final Cartesian coordinates (L, a, b) are then computed from the cylindrical coordinates (L, C, h) as usual: $a = C \cos(h)$ and $b = C \sin(h)$. This construction elegantly ensures that the generated chroma C can never exceed the maximum possible chroma $C_{\max}$ for its given lightness and hue, because S' is capped at 1. Consequently, any color generated from the (L, S', h) parametrization is guaranteed to be in-gamut.
This "safe" representation has profound theoretical implications beyond its practical utility. It creates a computational space where the system is incapable of generating physically impossible, and thus meaningless, states. All operations are confined to the realm of the physically realizable. Furthermore, this parametrization provides a form of built-in adaptive gain control. A uniform change in the latent parameter S' does not produce a uniform change in the physical chroma C. In regions of the gamut where the droplet is wide and $C_{\max}$ is large (e.g., for mid-lightness yellows), a change in S' will result in a large change in C. In regions where the gamut is narrow and $C_{\max}$ is small (e.g., for very dark blues), the same change in

S' will result in a much smaller change in C. This mechanism allows the system to learn to manipulate a uniform latent variable whose real-world perceptual effect is automatically and correctly scaled by the local geometric properties of the perceptual space. This ensures that the system's internal operations have stable and predictable consequences in the perceptual domain, a crucial property for any robust learning system.

The Axioms of Color-Geometry Semantics

The Color-Geometry framework is more than just a novel implementation for machine learning; it proposes a new set of foundational principles for semantics itself. By leveraging the intrinsic properties of the OKLab perceptual manifold, the framework establishes a system of meaning governed not by arbitrary linguistic convention or formal logic, but by the concrete and measurable laws of geometry and physics. These principles can be systematically detailed as a set of axioms that together define a new, grounded, and compositional geometric semantics. Each axiom replaces a traditional component of symbolic reasoning with a direct geometric or physical analogue, thereby constructing a complete system for representing and manipulating meaning.

Axiom 1: Grounding as Geometric Locus

The first and most fundamental principle of the framework is that a symbol is grounded by its very definition as a specific location within the constrained perceptual manifold.

- **Axiom 1:** *Each elementary symbol (representing a primitive concept, entity, or predicate) is defined as a point vector* $x = (L, a, b)$ *in the OKLab space, subject to the constraint that this point must lie within the gamut droplet:* $x \in \mathcal{D}$.

This axiom formally establishes the grounding layer of the system. Unlike in traditional AI, where a symbol is an arbitrary token (e.g., the string "RED") that is linked to a meaning through an external interpretation, here the symbol *is* its meaning. The vector x is not a pointer to the concept of a color; it is the coordinate of that color in a perceptually meaningful space. The meaning is not attached to the symbol; the meaning *is* the symbol's geometric locus.

This axiom directly and elegantly solves the core issue raised by Harnad. It provides the "bottom-up" grounding in a non-symbolic representation that he argued was necessary to escape the symbol/symbol merry-go-round. The representation is non-symbolic in the traditional sense because its form is not arbitrary. The vector coordinates for a specific shade of red are determined by the physics of light and the biology of human perception, not by convention. The constraint $x \in \mathcal{D}$ further reinforces this grounding by ensuring that every elementary symbol corresponds to a physically realizable perceptual state, a "meaning" that can actually exist in the world. This axiom lays the foundation for all subsequent operations, ensuring that the entire system is built upon a bedrock of physically plausible and perceptually meaningful primitives.

Axiom 2: Compositionality as Convex Mixing

A hallmark of symbolic systems is compositionality—the ability to combine primitive symbols to form complex expressions with systematic meanings. The Color-Geometry framework replaces the syntactic rules of symbolic composition with the physical and geometric operation of convex mixing.

- **Axiom 2:** *The semantic composition of a set of symbols $\{x_u\}$ is defined as their convex combination. The emergent meaning of the composition, x_v, is the point in OKLab space given by the weighted average of the constituent symbols: $x_v = \sum_{u \in \mathcal{P}(v)} \alpha_{vu} x_u$, where the weights α_{vu} are non-negative and sum to one ($\alpha_{vu} \ge 0, \sum \alpha_{vu} = 1$).*

This axiom replaces the abstract syntax of predicate logic or natural language grammar with the intuitive geometry of convex hulls. Geometrically, the result of any convex combination of a set of points must lie within the convex hull of those points—the smallest convex shape that contains them all. In this framework, the meaning of a composite expression is literally the geometric location of the mixture of its parts.

This provides a direct physical analogue to the process of additive color mixing. The meaning of "red AND yellow" is not a logical proposition but is, for example, the color orange that lies on the straight-line segment connecting the vector for pure red and the vector for pure yellow in OKLab space. The specific shade of orange is determined by the mixing weights α. This grounds the principle of compositionality in a measurable, continuous, and physically intuitive process. The framework can be extended to include a bias color, x_{bias}, allowing the composition to not only mix hues but also to learn to lighten, darken, or desaturate the result, providing a richer computational palette. This axiom demonstrates how complex, structured meanings can emerge from simple, physically-grounded operations, a key requirement for any viable theory of semantics.

Axiom 3: Causality as Geometric Reachability

Beyond static composition, intelligent systems must reason about dynamic relationships, such as causality. The framework provides a novel and powerful definition of causality, transforming it from a problem of logical inference into a concrete geometric query.

- **Axiom 3:** *A set of antecedent symbols $\{x_u\}$ is a valid cause for a consequent symbol x_v^{obs} if and only if x_v^{obs} is geometrically reachable from $\{x_u\}$. A state is reachable if it lies within the intersection of the convex hull of its causes and the gamut droplet: $x_v^{\text{obs}} \in \text{conv}(\{x_u\}) \cap \mathcal{D}$.*

This axiom provides a concrete, testable, and geometrically defined notion of a causal link. Instead of relying on temporal logic or abstract causal graphs, the system determines causality by checking for geometric containment. This dual constraint of being within both the convex hull and the physical gamut gives rise to two distinct and, crucially, interpretable modes of causal failure:

1. **Directional Insufficiency:** This occurs when the target state x_v^{obs} lies *outside* the convex hull of the inputs. This means that no possible mixing of the antecedent colors could ever produce the target. The input palette is fundamentally insufficient in its hue or lightness characteristics. For example, one cannot produce a pure blue by mixing red and yellow.
2. **Chromatic Impossibility:** This occurs when the target state x_v^{obs} lies *inside* the convex hull but *outside* the gamut droplet $\mathcal{D}$. In this case, the target's hue and lightness are directionally achievable, but the target is too saturated (has too much chroma) to be physically realized from the given inputs. For instance, mixing two mid-saturation colors can produce a mixture with the correct hue, but it may be physically impossible for that mixture to achieve the high saturation of the target.

This axiom provides a powerful mechanism for causal inference that is entirely grounded in the geometry of the perceptual space. It replaces the brittleness of logical deduction with the robustness of a geometric query, while retaining a high degree of interpretability through its well-defined failure modes.

Axiom 4: Abstraction as Geometric Region

To move beyond elementary perceptions, a semantic system must be able to form abstract concepts. The framework models abstraction not with more complex symbolic structures, but with a different kind of geometric entity: regions and volumes within the perceptual space.

- **Axiom 4:** *Higher-level concepts and abstract properties are represented not as single points, but as regions (sets of points) within the OKLab space. The process of abstraction is the definition of such a region, often through constraints on the quality dimensions.*

For example, the elementary symbol for a specific shade of scarlet might be a single point, but the abstract concept "red" would be represented as a region—the set of all points $x \in \mathcal{D}$ whose hue angle h falls within a certain range. Similarly, the concept "bright" could be the region of all points where the lightness L is above a certain threshold.

This axiom establishes a direct and powerful link to Peter Gärdenfors' influential theory of Conceptual Spaces. Gärdenfors argued that natural concepts correspond to *convex regions* in a geometric space spanned by quality dimensions. The Color-Geometry framework can be seen as a computational instantiation of this theory. By representing concepts as regions, it becomes possible to implement symbolic inference operations as geometric set operations:

- **Conjunction (AND):** The concept "bright red" is the *intersection* ($\cap$) of the "bright" region and the "red" region.
- **Disjunction (OR):** The concept "red or blue" is the *union* ($\cup$) of the "red" region and the "blue" region.
- **Subsumption (IS-A):** The statement "all scarlet is red" corresponds to geometric *containment* ($\subset$), where the region for "scarlet" is a subset of the region for "red".

This allows for a form of geometric reasoning that mirrors the structure of symbolic logic but remains fully grounded in the continuous perceptual manifold. It provides a pathway from concrete, point-like perceptions to abstract, region-based concepts within a single, unified geometric framework.

Axiom 5: The Closed Grounding Loop

The final axiom synthesizes the preceding principles into a complete, self-contained system that closes the loop between low-level perception and high-level symbolic reasoning.

- **Axiom 5:** *A complete grounded system is formed by a closed loop wherein (1) a subsymbolic perceptual system maps sensory inputs to coordinates in the OKLab manifold; (2) a symbolic reasoning layer performs geometric operations (mixing, reachability checks, set operations) on these coordinates; and (3) the resulting coordinates can be mapped back to the perceptual domain for verification, action, or communication.*

This closed loop creates a system where subsymbolic and symbolic representations are not two separate, loosely-coupled modules, but are two aspects of a single, integrated process mediated by the shared geometric manifold. The OKLab space acts as the "interlingua" or common language where the continuous representations of perception and the discrete-like operations of symbolic thought coincide. A neural network might process an image of a flower and output a vector x_{flower} in OKLab. A reasoning module could then take this vector and the vector for "sunlight," x_{sun}, perform a convex mixing operation to predict the

flower's appearance in the sun, yielding a new vector x'_{flower}. This new vector is not just an abstract symbol; it can be directly rendered as a color to visualize the prediction.
This architecture directly addresses the call for hybrid systems in the symbol grounding literature , but in a much more deeply integrated fashion. The symbolic functions *emerge* from the geometric properties of the underlying non-symbolic space, rather than being implemented in a separate, "un-natural" module. This creates a self-consistent cycle where perception informs symbols, symbols are manipulated geometrically, and the results of that manipulation are new, perceivable states.
These axioms, taken together, implicitly define a *physics of meaning*. The laws governing semantic combination are not arbitrary logical conventions but are analogues of physical laws (the additive mixing of light) and geometric constraints (convexity, physical realizability within the gamut). This suggests that meaning, in this model, is not a purely abstract or linguistic phenomenon, but one that is subject to its own kind of "physical law." Furthermore, the propagation of "reachable sets" (the regions representing a symbol's possible states) through a network reveals a dynamic analogous to the Second Law of Thermodynamics. The convex hull operation, which defines composition, is inherently contractive; it averages inputs, pulling them away from the extremes and towards the neutral gray center of the space. This creates an "entropic pull" towards achromatic, less differentiated, and less information-rich states. For a system to maintain vibrant, high-chroma, meaningful representations in its deeper layers, it must learn to counteract this entropic collapse. Learning, in this view, becomes the *anti-entropic* process of discovering highly specific, non-uniform weightings that selectively amplify certain inputs, actively "pulling" the resulting mixture back out towards the information-rich boundary of the gamut. This frames learning not merely as error minimization, but as a dynamic struggle against informational entropy within a constrained physical system.

A Computational Realization: Gradient-Free Causal Learning in SNNs

The abstract axioms of Color-Geometry semantics find a concrete and powerful implementation in the causal framework for Spiking Neural Networks (SNNs) detailed in the source document. This model serves as a detailed case study, demonstrating how the principles of geometric grounding, composition, and causality can be operationalized into a practical, efficient, and, crucially, gradient-free learning algorithm. By reframing credit assignment as a problem of geometric decomposition, the framework offers a compelling alternative to the dominant but often problematic paradigm of backpropagation, particularly in the context of neuromorphic computing.

Credit Assignment as Geometric Decomposition: The Inverse Problem

In conventional deep learning, credit assignment is achieved via backpropagation, where an error signal (gradient) is propagated backward through the network to incrementally adjust synaptic weights. This process, while immensely successful, is computationally expensive and biologically implausible, and it is fundamentally incompatible with the discrete, non-differentiable nature of spiking neurons.
The Color-Geometry framework proposes a radical paradigm shift. Credit assignment is

reframed not as error propagation, but as a **geometric inverse problem**. The process is the direct analogue of the backward traversal of a causal hyperpath in its conceptual predecessor, the Temporal Event Hypergraph (TEH) model. The problem can be stated as follows:
Given a desired target color for a postsynaptic neuron, x_v^*, and the set of observed colors from its presynaptic inputs, $\{x_u\}$, find the non-negative mixing coefficients, $\{\alpha_{vu}\}$, that best reconstruct x_v^* while satisfying all physical (in-gamut) and structural constraints.
This is a fundamental re-conception of learning. Instead of asking "how should I adjust the weights to reduce the error?", the system asks "what causal contributions from the inputs would explain this desired output?". The solution to this inverse problem—the set of optimal coefficients α^*_v—*is* the credit assignment. It is not an infinitesimal update direction but a complete, holistic decomposition of the target state into its constituent causal sources. This moves the learning process from one of iterative, opaque adjustment to one of transparent, verifiable geometric proof.

Solving for Causality via Constrained Optimization

The geometric inverse problem can be formalized as a constrained optimization task that is both mathematically precise and computationally tractable. The objective is to find the mixing coefficients that minimize the perceptual error between the synthesized color and the target color. Given that the ΔE metric in OKLab is the Euclidean distance, this is equivalent to minimizing the squared Euclidean distance.
Let X be a $3 \times N$ matrix where each column is the (L, a, b) vector of a presynaptic color x_u, and let α be the $N \times 1$ vector of unknown coefficients α_{vu}. The optimization problem is to find the α that minimizes the reconstruction error, subject to several constraints:
Subject to the constraints:

1. **Non-negativity:** $\alpha_{vu} \ge 0$ for all u. Each input can only make a positive contribution.
2. **Convexity (Simplex Constraint):** $\sum_{u} \alpha_{vu} = 1$. The weights must sum to one, ensuring the operation is a true convex combination.
3. **Gamut Feasibility:** $X\alpha \in \mathcal{D}$. The resulting mixture must be a physically realizable color.

This formulation is a **Non-Negative Least Squares (NNLS)** problem with additional constraints. The full problem, including the non-linear gamut feasibility constraint, constitutes a **Quadratic Programming (QP)** problem, as it involves minimizing a quadratic objective function subject to linear and non-linear constraints. This class of problem is well-studied and can be solved efficiently using standard numerical optimization libraries, such as Sequential Least Squares Programming (SLSQP) methods.
The solution vector, α^*, provides the barycentric coordinates of the projection of the target x_v^* onto the reachable set (the intersection of the convex hull of inputs and the gamut droplet). This provides a complete and unambiguous decomposition of the target into its causal sources, replacing the opaque nature of gradient-based updates with a transparent and mathematically verifiable solution.

Modeling Coincidence with k-Sparsity and Minimal Color Coalitions

The basic NNLS formulation assumes that all presynaptic neurons can potentially contribute to the postsynaptic state. However, biological neurons often operate on the principle of

coincidence detection, firing only in response to a small coalition of near-simultaneous inputs. This concept of n-ary dependency, modeled as "B-connectivity" in the TEH framework, can be captured geometrically by introducing a sparsity constraint.
The NNLS problem is augmented with the constraint that the solution vector α must be **k-sparse**, meaning it can have at most k non-zero elements:
where k is a small integer, typically in the range of 2 to 5. This constraint enforces the requirement that the target color must be explained by a "minimal color coalition," reflecting the need for a small group of presynaptic inputs to coincide to cause the effect.
This transforms the credit assignment task from a simple regression into a sparse regression problem. While enforcing an L_0-norm constraint is NP-hard in general, for the small number of inputs typical for a single neuron, it remains computationally tractable. It can be solved using mixed-integer quadratic programming, or approximated using techniques like L_1 regularization (Lasso) or heuristic algorithms like Orthogonal Matching Pursuit.
The introduction of this sparsity constraint has a profound effect on the geometry of the problem. The local reachable set is no longer a single, continuous convex polytope formed by all presynaptic inputs. Instead, it becomes a finite union of many smaller, lower-dimensional simplices, where each simplex is the convex hull of a specific k-sized coalition of inputs. The learning problem is thus transformed from navigating a single convex space to a more combinatorial task of selecting the correct low-dimensional "facet" (i.e., the correct coalition of causes) capable of reaching the target. This more structured, combinatorial landscape is a much closer geometric analogue to the discrete, logical nature of the original hypergraph model it is intended to replace.

Gamut-Aware Learning: A Geometrically Derived Update Rule

Once the optimal mixing coefficients α^* are found by solving the constrained, k-sparse NNLS problem, the final step is to update the synaptic weights. The Color-Geometry framework accomplishes this with a simple, powerful, and entirely gradient-free update rule derived from the geometric solution.
The update is driven by an **evidence score**, s_{vu}, which incorporates not only the magnitude of the causal contribution but also the *robustness* of the solution with respect to the physical boundary of the gamut. A solution that lies deep within the droplet is considered more robust and reliable than one that lies perilously close to the boundary of chromatic impossibility. This robustness is quantified by a **gamut margin** factor. For an optimal mixture x_{mix}^* with coordinates (L^*, C^*, h^*), the margin is defined as:
where ϕ is a monotonically increasing function (e.g., $\phi(z) = z^2$) that amplifies the score for solutions with a large margin (i.e., where the resulting chroma C^* is much smaller than the maximum possible chroma $C_{\max}$).
The final evidence score for the synapse from u to v is the product of its causal contribution and the solution's physical robustness:
This score is then used to drive the weight update. The direction of the update is determined by the sign of the overall task error, while the magnitude is provided by the geometrically derived evidence score:
Here, $\mathcal{L}$ is the task-level loss, and the sign term indicates whether the postsynaptic color x_v needs to be adjusted towards or away from a global target. This mechanism assigns credit based on a verifiable geometric decomposition and modulates the learning rate based on a physically motivated measure of confidence in the causal explanation. It entirely bypasses the need for derivatives and the complex machinery of backpropagation, offering a learning rule that

is computationally local, event-driven, and intrinsically interpretable.

Conceptual Isomorphism: Mapping Temporal Hypergraph Concepts to Geometric Analogues

To fully appreciate the framework's intellectual contribution, it is essential to understand its lineage. The Color-Geometry model was designed as a direct conceptual translation of the Temporal Event Hypergraph (TEH) model, preserving its causal and logical intent while completely replacing the underlying mathematical primitives. The following table makes this isomorphism explicit, demonstrating how each key concept in the discrete, temporal TEH framework has a direct and powerful analogue in the continuous, geometric model.

Hypergraph Framework Concept	Color-Geometry Framework Analogue	Mathematical Formalism	Interpretation & Nuances
State Primitive: A discrete spike event (neuron_id, t) is a vertex.	**State Primitive:** A neuron's state is a color vector $x=(L,a,b)$ within the gamut droplet $\mathcal{D}$.	$x \in \mathcal{D} \subset \mathbb{R}^3$	The state is no longer just an event in time, but a point in a continuous, perceptually meaningful vector space.
Causal Link: A directed hyperedge (tails) → head represents a successful firing event.	**Causal Synthesis:** A convex mixture of presynaptic colors produces a postsynaptic color.	$x_v = \sum \alpha_{vu} x_u$	The causal link is not a discrete graph edge but a continuous mixing operation governed by the laws of additive light.
N-ary Dependency: B-Connectivity requires *all* presynaptic spikes in the tail to be present.	**Coincidence Detection:** A k-sparse coalition of presynaptic colors is sufficient to span the target.	$\lVert\alpha\rVert_0 \le$	The strict "all-or-nothing" logical AND is relaxed to a more flexible geometric "spanning" condition by a minimal subset.
Causality Check: Is a hyperedge valid based on temporal logic and neuron state?	**Reachability Check:** Is the target color within the convex hull of sources and inside the gamut droplet?	$x_v \in \text{conv}(\{x_u\}) \cap \mathcal{D}$	The check shifts from satisfying abstract temporal rules to satisfying concrete geometric and physical constraints.
Credit Assignment: Backward traversal of hyperpaths to find causal chains.	**Credit Attribution:** Solving a constrained NNLS problem to find the barycentric coordinates (mixing weights).	$\alpha^* = \arg\min \lVert X\alpha - x^*\rVert_2$ s.t. constraints	Credit is found not by pathfinding, but by geometric decomposition. The solution vector α^* is the credit distribution.
Evidence Scoring: Weighting the solution α^* by the gamut margin to reward robust	**Evidence Scoring:** Weighting the solution α^* by the gamut margin to reward robust	$s_{vu} = \alpha_{vu}^* \cdot \text{margin}$	Credit is modulated by the "physical robustness" of the causal explanation;

Hypergraph Framework Concept	Color-Geometry Framework Analogue	Mathematical Formalism	Interpretation & Nuances
solutions.	solutions.		solutions near the boundary are brittle and less trusted.
Representation Invariance: The causal structure is independent of the specific temporal logic formalism used.	**Frame Invariance:** The reachability decision (in vs. out) is a geometric fact, independent of the coordinate system (Lab, Lch).	\text{conv}(S) \cap \mathcal{D} is invariant under isometries.	The fundamental causal verdict is absolute, reinforcing the physical analogy and the objectivity of the computation.

This mapping highlights the shared structural and philosophical underpinnings of the two approaches, demonstrating a deliberate and principled shift from a discrete, graph-theoretic worldview to a continuous, geometric one.

Intellectual Context and Theoretical Lineage

The Color-Geometry framework, while novel in its specific formulation, does not arise in a vacuum. It represents a powerful and timely synthesis of several key intellectual currents in cognitive science, artificial intelligence, and philosophy. By situating the framework within this broader landscape, its full significance becomes apparent. It is not merely a clever engineering solution for training SNNs, but a concrete contribution to long-standing debates about the nature of concepts, the structure of meaning, and the architecture of mind. It can be seen as a computational realization of Peter Gärdenfors' theory of Conceptual Spaces, a principled alternative to the opaque geometry of modern deep learning, and a natural substrate for multimodal and embodied cognition.

An Embodiment of Conceptual Spaces: Parallels with Gärdenfors' Theory

A striking parallel exists between the Color-Geometry framework and the theory of **Conceptual Spaces** proposed by philosopher and cognitive scientist Peter Gärdenfors. Gärdenfors sought to bridge the gap between the two dominant paradigms in cognitive science: the symbolic approach, which views cognition as computation over abstract symbols, and the connectionist approach, which models cognition with artificial neural networks. He argued that both were insufficient and proposed an intermediate, conceptual level of representation with a geometric structure.

In Gärdenfors' theory, knowledge is represented in a conceptual space, a geometric structure spanned by a set of "quality dimensions". These dimensions correspond to the basic features by which objects and concepts can be compared, such as color, shape, weight, temperature, or pitch. Individual objects are represented as points in this multidimensional space, and, most importantly, **concepts are represented as convex regions**. The notion of convexity is central: if two objects (points) are instances of a concept, then any object that lies on the line segment between them is also likely to be an instance of that concept. For example, if a specific lime green and a specific forest green are both instances of the concept "green," then any color on the perceptual path between them should also be considered "green."

The Color-Geometry framework can be viewed as a direct and concrete computational realization of Gärdenfors' theory.

- The OKLab axes (L, a, b) are precisely the **quality dimensions** for the domain of color.
- The elementary symbols, represented as points in OKLab space, correspond to Gärdenfors' representation of individual **objects** or stimuli.
- The framework's axiom of "Abstraction as Geometric Region," where concepts like "red" are defined as volumes within the space, directly implements Gärdenfors' central thesis. The propagation of **reachable sets** through the network is a dynamic process of constructing these conceptual regions.
- The framework's reliance on **convex mixing** for compositionality is a direct operationalization of Gärdenfors' hypothesis about the **convexity of natural concepts**.

This strong theoretical alignment elevates the Color-Geometry framework from a specific learning algorithm to a significant contribution to a major theory of cognitive representation. It provides a practical, testable, and computationally grounded model for the geometric theory of thought, moving it from the realm of philosophical argument into that of empirical and computational science.

Geometric Semantics and the "Alien Substrate" of Modern AI

The framework also resonates deeply with recent discoveries in the field of artificial intelligence, particularly concerning the internal workings of large language models (LLMs). It has become clear that these models do not operate on symbols in the traditional sense, but rather on high-dimensional vectors in vast "embedding spaces". Within these spaces, semantic relationships manifest as geometric relationships. The classic example, vector('king') - vector('man') + vector('woman') ≈ vector('queen'), demonstrates that concepts are not arbitrary tokens but have a rich geometric structure that can be manipulated with vector arithmetic.

This has led to the idea of a "hidden geometry of thought" or an "alien substrate" for intelligence, where meaning, coherence, and association are encoded as positions and trajectories in a hyperdimensional semantic field. When a model predicts the next word, it is effectively collapsing a statistical wave function in this incomprehensible geometric space.

While current LLMs discover this geometry in an opaque, black-box fashion through statistical correlation on massive datasets, the Color-Geometry framework proposes to *design* the semantic space from first principles based on perceptual foundations. It offers a potential answer to the fundamental question of *why* a geometric representation of meaning should work at all. Instead of an uninterpretable, 12,000-dimensional latent space discovered by a model, it proposes a low-dimensional (3D), fully interpretable, and physically constrained manifold as the foundational substrate for meaning. It provides a "native" geometry for semantics, where the axes and distances have direct, perceivable meaning, rather than waiting for an abstract geometry to emerge from statistical patterns. The framework thus offers a path toward building AI systems whose geometric reasoning is not an emergent and opaque property, but a designed and transparent one.

This approach offers a potential solution to the twin problems of the "brittleness" of symbolic AI and the "opacity" of deep learning, creating what can be described as a "glass box" model. Symbolic AI is interpretable because its operations follow explicit logical rules, but it is brittle because these rules often fail to handle the ambiguity and continuous nature of the real world. Deep learning is robust to noisy, continuous data but is notoriously opaque; its decisions, encoded in millions of weights, are difficult to explain or verify. The Color-Geometry framework combines the strengths of both. Its decisions are based on verifiable geometric proofs. The solution to the NNLS problem, the vector \alpha^*, provides a complete, machine-readable explanation for why a particular output was generated from a given set of inputs. Yet, unlike a discrete symbolic proof, this geometric proof operates on a continuous space, giving it an inherent robustness to small perturbations—a point that is slightly moved is still likely to be within the same conceptual region. In this way, the framework marries the formal interpretability of symbolic systems with the continuous, robust nature of subsymbolic systems, offering a "best of both worlds" architecture for trustworthy AI.

A Bridge to Multimodal and Embodied Cognition

The principles of the Color-Geometry framework are not limited to the domain of color. The framework provides a natural and extensible substrate for grounding symbols across multiple sensory modalities, connecting it to the broader fields of multimodal learning and embodied cognition.

Embodied cognition is the theory that cognitive processes are deeply rooted in the body's sensory and motor interactions with the environment. It posits that meaningful understanding arises not from the manipulation of abstract, amodal symbols, but from lived physical experiences. The Color-Geometry framework is deeply aligned with this perspective. Its foundational layer is not an abstract logic but a model of a perceptual system.

While the case study uses OKLab, the principle is generalizable. Other perceptual domains can also be mapped to geometric spaces with their own unique structures and constraints:

- **Auditory Perception:** The space of sound timbre could be represented by dimensions such as spectral centroid (brightness), spectral flux (texture), and harmonicity.
- **Haptic Perception:** The sense of touch could be modeled with dimensions for roughness, temperature, and hardness.
- **Proprioception:** The sense of bodily position could be represented in the geometric space of joint angles.

The Color-Geometry framework provides the mathematical tools to integrate these disparate perceptual manifolds into a unified, grounded conceptual system. The OKLab space can serve as a "hub" or one of many interconnected spaces where information from different modalities is fused. For example, the concept of an "apple" could be formed by the convergence of representations from multiple domains. The word "apple" (from language), an image of an apple (from vision), and the feel of an apple (from haptics) could all be mapped to vectors in or near the "red" region of the color space, the "round" region of a shape space, and the "smooth" region of a texture space. The operations of convex mixing and geometric reachability provide a principled mechanism for binding these multimodal representations into a single, coherent, and robustly grounded concept.

This recasts the nature of concepts themselves. In a traditional symbolic system, a concept is a static definition stored in a knowledge base. In the Color-Geometry framework, a concept is a dynamic *potential* for interaction. Gärdenfors' theory sees concepts as regions, and the framework operationalizes this with the notion of "reachable sets". The reachable set \mathcal{R}_t(u) for a neuron u is not a fixed definition but represents the entire volume of states that the neuron can possibly occupy given its causal history. This means a concept is not a static point but a *potential field*—a representation of all possible future states and interactions related to that concept. Reasoning about the concept "apple" is not a matter of looking up its properties, but of reasoning about the dynamics of this volume: How does it transform under different lighting conditions (mixing with "sunlight")? What other concepts does it overlap with (intersection with "fruit")? This makes cognition a fundamentally dynamic, predictive, and interactive process, deeply in line with the core tenets of embodied AI.

Critical Analysis and Future Horizons

The Color-Geometry framework represents a significant and compelling new paradigm for artificial intelligence, offering a principled and powerful solution to the long-standing symbol grounding problem. By recasting semantics in the language of geometry and physical constraints, it provides a path toward AI systems that are simultaneously robust, interpretable, and grounded in a meaningful reality. However, like any ambitious new theory, it is not without its potential limitations and unresolved questions. A critical analysis of its contributions, challenges, and future directions is essential for understanding its true potential and for guiding the research program it initiates.

Core Contributions: A New Paradigm for Grounded AI

The primary strengths of the Color-Geometry framework, as detailed throughout this report, can be synthesized into four core contributions that collectively define a new paradigm for grounded AI.

1. **Intrinsic Grounding:** The framework's most profound contribution is its direct solution to the symbol grounding problem. By defining symbols as coordinates within a physically constrained perceptual manifold, it makes meaning an intrinsic property of the system's state. The symbol's meaning *is* its geometric locus, eliminating the need for an external, parasitic interpretation and escaping Harnad's symbol/symbol merry-go-round.
2. **Inherent Interpretability:** In an era dominated by opaque "black box" models, the framework offers a "glass box" alternative. Every learning update and every causal inference is the result of a verifiable geometric proof. The solution to the constrained optimization problem, the vector α^*, provides a complete, machine-readable explanation for why a particular synapse was credited, making the system's reasoning auditable and trustworthy.
3. **Physicality as a Foundation:** The framework elevates the physical constraint of a device's color gamut from a technical limitation to a fundamental law of semantic feasibility. The gamut droplet $\mathcal{D}$ provides a non-arbitrary, physically motivated basis for defining valid states and serves as a powerful form of intrinsic regularization, ensuring that all representations correspond to plausible perceptual realities.
4. **Computational Alignment with Neuromorphic Principles:** The proposed learning mechanism is entirely gradient-free and event-driven. It replaces the dense, synchronous, and computationally heavy matrix multiplications of backpropagation with localized, small-scale quadratic programs that are solved only when necessary. This computational model is exceptionally well-suited for the sparse, asynchronous, and parallel nature of neuromorphic hardware, promising a path to highly efficient and low-power intelligent systems.

Potential Limitations and Unresolved Questions

Despite its strengths, the framework's reliance on perceptual grounding presents significant challenges, particularly in the representation of highly abstract concepts. This limitation is not unique to this framework but is a central difficulty for all theories of embodied or grounded

cognition.
The most pressing unresolved question is how to represent concepts that have no direct or obvious perceptual correlate. How does one define a geometric region or trajectory within a color space that corresponds to abstract ideas like "justice," "democracy," "causality" (the very principle the system uses), or "infinity"?. While the framework provides a robust mechanism for grounding elementary symbols like "red" or "bright," it does not yet offer a complete, systematic mechanism for bootstrapping from this perceptual foundation to the highest levels of human abstract thought.
Furthermore, the framework's reliance on convexity, a principle inherited from Gärdenfors' theory, may prove to be too restrictive for certain types of concepts. Many real-world categories are not neatly convex. For example, the concept of "game" is notoriously difficult to define with a single set of properties and may not correspond to a single connected region in any conceptual space. The category "non-red" is inherently non-convex. The framework, in its current form, does not provide a clear mechanism for representing such disjoint or negative concepts.
This leads to a series of critical questions that must be addressed by future research:

- **Representation of Abstraction:** Can abstract relations be represented by higher-order geometric structures, such as transformations (e.g., rotations, translations), trajectories, or vector fields within the perceptual manifold? For example, could "more than" be represented as a vector pointing in the direction of increasing lightness?
- **Logical Operators:** How can the framework handle complex logical operators like universal quantification ("all"), existential quantification ("some"), and, most critically, negation ("not"), which lack simple geometric analogues like intersection and union?
- **Scalability of Manifolds:** Is a single perceptual manifold, even a multimodal one, sufficient to ground the entirety of human knowledge? Or is a more complex architecture required, perhaps a network of interconnected geometric spaces, each with its own local metric and topology?

The framework's greatest limitation—its foundational reliance on perception—may also be its greatest strength. By forcing the lowest level of meaning to be inextricably tied to a physical, verifiable substrate, it ensures that any higher-level abstractions built upon it must ultimately be traceable back to a grounded foundation. In human cognition, abstract concepts are often understood through metaphor, which grounds them in physical and perceptual experience (e.g., "grasping an idea," "a weighty decision," "a close relationship"). The Color-Geometry framework provides a computational substrate for the first and most critical layer of this grounding process. By insisting on this physical foundation, it inherently prevents the construction of an ungrounded "symbol/symbol merry-go-round" at higher levels of abstraction. Any new abstract symbol would need to be defined in terms of geometric operations on already-grounded regions, ensuring a complete chain of traceability from the most abstract thought back to concrete perception.

Future Research Directions

The introduction of this geometric paradigm opens several fertile and promising avenues for future research that could address its current limitations and expand its scope and power. Based on the source document and the analysis herein, a clear research roadmap emerges.

1. **Exploring Alternative Geometries:** While OKLab provides an excellent substrate due to its perceptual uniformity, it is only one possible choice. A systematic investigation into other advanced color spaces (such as JzAzBz or CAM16-UCS) and, more importantly, perceptual spaces for other modalities (e.g., audio timbre spaces, haptic texture spaces) is a critical next step. Each space possesses a unique geometry and a unique "gamut

droplet," and understanding how these different geometric structures affect learning dynamics and representational capacity will be crucial for developing a general theory of geometric grounding.
2. **Dynamic Basis and Learned Primitives:** The current framework utilizes a static bias color (e.g., gray, white, or black) to anchor the mixing process. Future work should explore methods that allow the network to *learn* its own set of "basis colors" or other perceptual primitives. This would enable the system to dynamically construct an optimal, task-specific conceptual palette, potentially increasing its expressive power, learning efficiency, and ability to adapt to novel conceptual domains.
3. **Hardware Co-Design:** The computational primitives of this framework—solving small, parallel QP problems and performing fast lookups in a gamut oracle—are fundamentally different from the multiply-accumulate operations that dominate conventional deep learning hardware. This invites a new research direction in co-designing neuromorphic accelerators with dedicated hardware units optimized for these specific geometric operations. Such hardware could fully leverage the algorithm's event-driven nature and inherent parallelism, leading to unprecedented gains in efficiency for grounded AI systems.
4. **Theoretical Foundations:** The propagation of reachable sets, \mathcal{R}_t, through the network represents a geometric transformation of information. There are deep potential connections between the evolution of these sets and concepts from information theory, such as the Data Processing Inequality. A key challenge is to formalize the notion of "chromatic information" and to rigorously analyze how it is preserved, compressed, or lost at each layer of the network. Developing such a theoretical foundation could provide a deep understanding of the computational properties and limits of this new class of neuro-geometric systems.

Ultimately, the Color-Geometry framework is not just a model of artificial intelligence, but a powerful **epistemological tool** for cognitive science. Theories like Gärdenfors' Conceptual Spaces, while influential, have remained largely philosophical and difficult to test empirically. The Color-Geometry framework provides a concrete, falsifiable, and runnable implementation of these ideas. This allows researchers to move from philosophical argument to computational experiment. For example, the psychological plausibility of the "concept convexity" hypothesis can be directly tested by building a model to learn human color categories from linguistic and visual data, and then examining whether the resulting learned regions are, in fact, convex in OKLab space. The model's successes or failures in replicating human cognitive phenomena would provide direct evidence for or against the underlying cognitive theory. In this way, the framework has the potential to transform a philosophical inquiry into a progressive scientific research program, using the geometry of perception to unlock the geometry of thought.

Works cited

1. Symbol grounding problem - Wikipedia, https://en.wikipedia.org/wiki/Symbol_grounding_problem 2. The Difficulties in Symbol Grounding Problem and the Direction for Solving It - MDPI, https://www.mdpi.com/2409-9287/7/5/108 3. The Symbol Grounding Problem - arXiv, https://arxiv.org/html/cs/9906002 4. Symbol grounding problem - (Intro to Cognitive Science) - Vocab, Definition, Explanations, https://fiveable.me/key-terms/introduction-cognitive-science/symbol-grounding-problem 5. The Symbol Grounding Problem - Scaler Topics, https://www.scaler.com/topics/artificial-intelligence-tutorial/symbol-grounding-problem/ 6. Steven

Harnad: Symbol grounding and the structure of dictionaries - LessWrong, https://www.lesswrong.com/posts/paeuSma7mABsrB6ju/steven-harnad-symbol-grounding-and-the-structure-of 7. The symbol grounding problem - ePrints Soton - University of Southampton, https://eprints.soton.ac.uk/250382/ 8. Foundations.Cognitive.Science2001: Harnad: The Symbol Grounding Problem, https://www.southampton.ac.uk/~harnad/Hypermail/Foundations.Cognitive.Science2001/0016.html 9. Symbol Grounding Revisited, https://ofai.at/papers/oefai-tr-94-19.pdf 10. A formal framework for the symbol grounding problem, https://agi-conf.org/2009/papers/paper_20.pdf 11. “Conscious” Multi-Modal Perceptual Learning for Grounded Simulation-Based Cognition, https://cognitivesciencesociety.org/cogsci20/papers/0585/0585.pdf 12. Toward a General Solution to the Symbol Grounding Problem: Combining Learning and Computer Vision - AAAI, https://cdn.aaai.org/Symposia/Fall/1993/FS-93-04/FS93-04-033.pdf 13. Solving the Symbol Grounding Problem: A Critical Review of Fifteen Years of Research, https://www.researchgate.net/publication/36444452_Solving_the_Symbol_Grounding_Problem_A_Critical_Review_of_Fifteen_Years_of_Research 14. Oklab color space - Wikipedia, https://en.wikipedia.org/wiki/Oklab_color_space 15. oklab() - CSS-Tricks, https://css-tricks.com/almanac/functions/o/oklab/ 16. Okay, Color Spaces — ericportis.com, https://ericportis.com/posts/2024/okay-color-spaces/ 17. A perceptual color space for image processing - Björn Ottosson, https://bottosson.github.io/posts/oklab/ 18. Integration with machine learning and AI | Geometric Algebra Class Notes - Fiveable, https://fiveable.me/geometric-algebra/unit-13/integration-machine-learning-ai/study-guide/wAuvs0eENmL6jx8i 19. A Brief Introduction to Geometric Deep Learning | by Jason McEwen | TDS Archive - Medium, https://medium.com/data-science/a-brief-introduction-to-geometric-deep-learning-dae114923ddb 20. Symbol Grounding - LessWrong, https://www.lesswrong.com/w/symbol-grounding 21. en.wikipedia.org, https://en.wikipedia.org/wiki/Conceptual_space#:~:text=A%20conceptual%20space%20is%20a,the%20three%20ordinary%20spatial%20dimensions. 22. Conceptual space - Wikipedia, https://en.wikipedia.org/wiki/Conceptual_space 23. The Geometry and Dynamics of Meaning - PMC - PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC11792772/ 24. What are “conceptual spaces”? - Lucas Bechberger's Website, https://lucas-bechberger.de/2017/02/14/what-are-conceptual-spaces/ 25. (PDF) A grounding framework - ResearchGate, https://www.researchgate.net/publication/220660856_A_grounding_framework 26. oklab() - CSS | MDN - Mozilla, https://developer.mozilla.org/en-US/docs/Web/CSS/color_value/oklab 27. Conceptual Spaces: The Geometry of Thought by Peter Gärdenfors | Goodreads, https://www.goodreads.com/book/show/1877443.Conceptual_Spaces 28. Conceptual Spaces by Peter Gärdenfors - Penguin Random House Canada, https://www.penguinrandomhouse.ca/books/658771/conceptual-spaces-by-peter-gardenfors/9780262572194 29. A Comprehensive Implementation of Conceptual Spaces - CEUR-WS.org, https://ceur-ws.org/Vol-2090/paper4.pdf 30. Review of Conceptual Spaces--The Geometry of Thought, https://aaai.org/ojs/index.php/aimagazine/article/view/1554/1453 31. AI's Hidden Geometry of Thought | Psychology Today, https://www.psychologytoday.com/us/blog/the-digital-self/202507/ais-hidden-geometry-of-thought 32. AI paper reveals surprising geometric structure in the LLM-learned concepts: 1) They form brain-like "lobes", 2) they form "semantic crystals" much more precise than it first seems, and 3) the concept cloud is more fractal than round : r/singularity - Reddit, https://www.reddit.com/r/singularity/comments/1gf8dou/ai_paper_reveals_surprising_geometric _structure/ 33. Grounding Symbols in Multi-Modal Instructions - ACL Anthology, https://aclanthology.org/W17-2807/ 34. VTPerception-R1: Enhancing Multimodal Reasoning via Explicit Visual and Textual Perceptual Grounding - arXiv, https://arxiv.org/html/2509.24776v1 35. Symbol Emergence as an Interpersonal Multimodal Categorization - Frontiers, https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2019.00134/full 36. Robotics, Biological Grounding and the Fregean Tradition - CSPEECH, https://cspeech.ucd.ie/Fred/docs/HooijmansKeijzer.pdf 37. Invariant representations in abstract concept grounding – the physical world in grounded cognition - PMC - PubMed Central, https://pmc.ncbi.nlm.nih.gov/articles/PMC11680661/ 38. Symbol Grounding Association in Multimodal Sequences with Missing Elements - Journal of Artificial Intelligence Research, https://jair.org/index.php/jair/article/download/11189/26398/20721 39. The Symbol Grounding Problem Re-framed as Concreteness-Abstractness Learned through Spoken Interaction - SemDial, https://www.semdial.org/anthology/Z22-Kennington_semdial_0010.pdf 40. Chapter 12 - The symbol grounding problem has been solved, so what's next?, https://ai.vub.ac.be/sites/default/files/steels-08e.pdf