LLMs that learn to understand physics for robotics with Affordance-First Semantic Architecture

A Comprehensive Conceptual Framework

Author: Abolhassan Ali Eslami
Affiliation: Shahid Beheshti University
Contact: a.eslami77@sharif.edu

Keywords

Epiphenomenal Semantics; Affordances; Embodied Cognition; Large Language Models; Non-representational AI; Physics-based AI; Symbol Grounding Problem; Ecological Psychology; Enactivism; Constraint-based Semantics

Abstract

Contemporary Large Language Models (LLMs) demonstrate remarkable fluency in language yet remain fundamentally disconnected from physical reality. Their "understanding" emerges solely from statistical patterns in text corpora, leaving them vulnerable to semantic brittleness, grounding failures, and an inability to connect linguistic expressions with actionable consequences in the world. This paper introduces a radical reconceptualization of semantics: **meaning need not be represented at all**. Instead, we propose *epiphenomenal semantics*—a framework where meaning emerges not as an internal representation but as a stable byproduct of embodied dynamics unfolding within linguistically constrained physical simulations.

We present the **Affordance-First Semantic Architecture (AFSA)**, a complete computational pipeline that reinterprets language not as symbolic content to be decoded, but as a generator of *affordance fields*: structured physical constraint environments that shape how virtual agents can move, interact, and persist. Within these fields, agents exhibit characteristic behavioral patterns—oscillations, convergences, failures, recoveries—whose statistical regularities across trials constitute semantic content. Crucially, no component of the system "knows" or "represents" meaning; meaning is what observers consistently recognize in the system's reliable behaviors under linguistic constraint.

This work bridges ecological psychology, enactivist philosophy, and modern AI engineering to demonstrate that semantic competence can arise without semantic representation. We elaborate the architecture's components, provide concrete examples of semantic emergence, address philosophical implications for the symbol grounding problem, and outline a research program for building non-representational language-capable systems.

1. Introduction: The Disembodiment Crisis in Language AI

Large Language Models have achieved unprecedented success in generating human-like text, answering questions, and even simulating reasoning. Yet a profound philosophical and practical problem remains unresolved: **LLMs have no access to the world their language describes**. When an LLM discusses "fragility," it has never witnessed glass shatter. When it reasons about "obstacles," it has never navigated around a barrier. Its entire semantic universe is constructed from co-occurrence statistics among symbols—symbols that, for the model, remain ungrounded tokens without physical consequence.

This disembodiment manifests in well-documented failure modes:

- **Semantic brittleness**: Minor phrasing changes can trigger catastrophic reasoning failures despite identical intended meaning.
- **Physical incoherence**: Models confidently assert physically impossible scenarios (e.g., "pouring water upward without containment").
- **Normative emptiness**: LLMs cannot distinguish *correct* from *incorrect* physical reasoning except via textual patterns, not physical consequence.
- **Context collapse**: Without embodied anchoring, models struggle to maintain consistent reference across extended interactions.

Traditional approaches attempt to "ground" symbols by linking them to perceptual data (images, sensor streams) or action spaces. Yet these methods typically preserve a representational core: the model still *encodes* meaning as internal states, merely enriched with multimodal data. The symbol grounding problem persists in modified form—now as the *multimodal grounding problem*.

We propose a more radical solution: **eliminate semantic representation entirely**. Instead of asking how symbols connect to the world, we ask how language can *shape physical dynamics* such that meaning emerges in the patterns of those dynamics. This reframing draws inspiration from James J. Gibson's ecological psychology: organisms don't build internal models of the world; they directly perceive *affordances*—action possibilities offered by the environment relative to their embodiment. A chair doesn't need to be "recognized" as a chair; it affords sitting for a creature of appropriate morphology.

Our framework extends this insight computationally: language generates *affordance fields*—virtual environments whose physical parameters (friction, mass distribution, constraint geometries) are modulated by linguistic input. Agents inhabiting these fields exhibit behaviors whose statistical regularities *are* the semantics. The word "slippery" doesn't map to a definition; it configures a low-friction surface where agents reliably slide when attempting certain motions. "Slipperiness" is the invariant pattern of sliding behaviors across trials —not a stored concept.

This paper elaborates this vision across conceptual, architectural, and philosophical dimensions, demonstrating how epiphenomenal semantics resolves longstanding challenges in AI while offering a path toward genuinely world-connected language understanding.

2. Conceptual Foundations: Why Representation Fails for Semantics

2.1 The Infinite Regress of Symbol Grounding

The symbol grounding problem, articulated by Harnad (1990), observes that symbolic systems cannot derive meaning from syntax alone. If "apple" is defined as "a red fruit," we must then ground "red" and "fruit," leading to infinite regress unless some symbols connect directly to non-symbolic experience. Most AI approaches attempt to halt this regress by anchoring symbols to sensory data—pixels, audio waveforms, or proprioceptive signals. Yet this merely pushes the problem back one step: how do *these* signals acquire meaning?

Consider a vision-language model trained to associate the word "heavy" with images of large objects. The model may learn statistical correlations between visual features and the token "heavy," but it never experiences *resistance to lifting*, *muscle strain*, or *inertia during motion*—the embodied essence of heaviness. Without these physical consequences, "heavy" remains a label attached to visual patterns, not a concept with functional import. When asked whether a large balloon is heavy, the model may fail not due to visual confusion but because it lacks the embodied understanding that *size does not determine weight*—a truth only accessible through physical interaction.

This reveals a deeper issue: **meaning is inherently relational and action-oriented**. "Heavy" isn't a property of objects alone; it's a constraint on possible actions relative to an agent's capabilities. A 10kg weight is heavy for a child, trivial for a weightlifter. Meaning emerges from the *interaction*, not from object properties in isolation. Representational approaches struggle with this relativity because they encode properties as absolute attributes rather than as constraints on possible behaviors.

2.2 Lessons from Biological Cognition

Biological systems often achieve adaptive behavior without explicit representation. Consider the human vestibulo-ocular reflex: when your head moves, your eyes automatically compensate to stabilize gaze. This coordination requires no internal model of head motion or visual scene—it emerges from direct neural pathways coupling sensory input to motor output. Similarly, a cockroach navigating terrain doesn't construct a map of obstacles; its leg mechanics and sensory feedback produce adaptive locomotion through physical interaction alone.

These examples illustrate *morphological computation*: the body's physical structure performs computational work that would otherwise require neural representation. The springiness of tendons absorbs shock without neural control; the geometry of feet provides stability without balance calculations. This suggests a profound insight: **intelligence can be outsourced to physics**. When the environment and body are properly structured, complex adaptive behavior emerges from dynamics rather than computation.

Language-capable systems might similarly leverage physics to "compute" meaning. Instead of representing "fragility," a system could generate environments where certain objects fracture under minimal force—allowing meaning to emerge in the statistical patterns of breakage events during agent interaction. The system need not "know" fragility; it need only reliably produce contexts where fragility *matters physically*.

2.3 Affordances as the Bridge Between Language and Action

Gibson's concept of affordances provides the crucial bridge. An affordance is not a property of the environment alone nor of the agent alone, but a *relational property* between them: "sit-on-ability" exists only for creatures with appropriate morphology encountering surfaces at appropriate heights. Crucially, affordances are *directly perceivable*—organisms don't infer them through reasoning but detect them through evolved or learned sensitivity to ecological information.

For language, affordances offer a non-representational semantics: words don't denote objects or properties but *modulate the landscape of possible actions*. "Slippery" doesn't describe a surface; it transforms how agents can move across it. "Fragile" doesn't categorize an object; it constrains how force can be applied without catastrophic failure. In this view, language is not a symbolic code but a *constraint-shaping tool*—a way to sculpt the physical possibility space within which agents operate.

This reframing dissolves the grounding problem: symbols never need grounding because they never function as symbols. They function as *parameters for physical constraint generation*. The word "heavy" isn't mapped to a meaning; it configures high inertial resistance in a simulation. Meaning emerges downstream in how agents struggle, adapt, or fail when interacting with that resistance.

3. Epiphenomenal Semantics: Core Principles

3.1 What "Epiphenomenal" Really Means

The term "epiphenomenal" often carries negative connotations in philosophy of mind—suggesting something causally inert or illusory. Here we reclaim it positively: **epiphenomenal semantics are real, observable, and functionally significant patterns that emerge from dynamics without being explicitly represented**.

Consider a whirlpool in a draining sink. The whirlpool has identifiable properties—rotation speed, diameter, stability—but no part of the water "represents" the whirlpool. The whirlpool is an emergent pattern in the flow dynamics. Similarly, semantic patterns in our framework are stable regularities in agent behavior under linguistic constraint—real phenomena without internal representation.

Key characteristics:

- **Observer-relative but objective**: Semantic labels ("indecision," "persistence") are assigned by observers recognizing patterns, yet the patterns themselves are objective features of the dynamics.
- **Causally potent**: Though not represented, these patterns influence downstream processes (e.g., an agent's behavioral tendency toward oscillation affects its task success).
- **Statistically robust**: Patterns persist across trials with varied initial conditions, noise instantiations, and agent morphologies—demonstrating invariance beyond accidental correlation.

3.2 The Central Equation of Meaning

We propose a fundamental reframing captured in this principle:

> **Meaning is not what the system knows, but what the system reliably does under constraint.**

This shifts semantics from an *epistemic* question ("What does the system believe?") to a *dynamical* one ("What behavioral regularities does the system exhibit?"). When presented with the phrase "careful handling required," the system doesn't retrieve a definition; it configures an environment where objects fracture under modest force, then observes agents developing cautious manipulation strategies. "Carefulness" is the statistical tendency toward slow, distributed force application observed across agents—not a stored concept.

3.3 Language as Constraint Sculpting

Traditional views treat language as *information transfer*: a speaker encodes meaning into symbols, which a listener decodes back into meaning. Our framework replaces this with **constraint sculpting**:

1. Linguistic input → parameterization of physical constraints (affordance field)
2. Affordance field → shapes possible agent dynamics
3. Agent dynamics → exhibit statistical regularities (semantic patterns)
4. Semantic patterns → recognized by observers as meaning

Crucially, step 4 is optional for the system's operation. The system functions perfectly without labeling patterns "meaningful." Meaning exists as a *third-person observable property* of the dynamics, not as a first-person representation within the system. This resolves the homunculus problem: no internal observer is needed to "see" the meaning.

4. Affordance-First Semantic Architecture (AFSA): A Detailed Blueprint

4.1 System Overview

AFSA comprises five tightly coupled components operating in a continuous loop:

In AFSA, the LLM undergoes conceptual repurposing:

Traditional LLM Role	AFSA LLM Role
Predict next token	Generate physical constraints
Encode semantic knowledge	Parameterize affordance fields
Reason via symbolic inference	Shape dynamics via constraint config
Output: text sequences	Output: structured physics params

The LLM never "understands" language in the traditional sense. Instead, through training on human-annotated constraint examples (e.g., "fragile" → high fracture susceptibility parameters), it learns mappings from linguistic patterns to physical parameter configurations. Critically, this mapping requires no intermediate semantic representation—it's a direct function from text to physics parameters.

Example: Given "The ice is thin and treacherous," the LLM doesn't reason about ice properties. It outputs parameters specifying:

- Low structural integrity threshold
- High slipperiness coefficient
- Irregular surface geometry
- Rapid failure propagation on stress

These parameters directly configure the simulation environment. No "concept of thin ice" exists internally; only the physical consequences of those parameters matter.

4.2 Affordance Field Mapper: From Abstract to Physical

The Affordance Field Mapper translates the LLM's abstract parameter suggestions into concrete simulation parameters. This component resolves ambiguities and enforces physical plausibility:

- **Parameter grounding**: Maps abstract terms ("very heavy") to concrete values (mass = 150kg) based on context and agent capabilities.
- **Constraint consistency**: Ensures generated parameters obey physical laws (e.g., friction coefficients remain within [0,1]).
- **Relational scaling**: Adjusts absolute values based on agent morphology (what's "heavy" for a mouse differs from a human).
- **Temporal dynamics**: Configures how constraints evolve (e.g., "melting ice" gradually reduces structural integrity over simulation time).

This mapper embodies the crucial insight that affordances are *relational*: the same linguistic input produces different physical parameters depending on the agent inhabiting the field. "Heavy" for a child-agent means 5kg; for an adult-agent, 30kg. The mapper maintains this relativity without explicit representation—through parameter scaling rules alone.

4.3 Physics Simulation Engine: The Semantic Substrate

The simulation engine is not merely a visualization tool but the *semantic substrate* where meaning physically manifests. We require engines with specific properties:

- **Differentiable physics** (optional but valuable): Enables gradient-based learning of constraint parameters from behavioral outcomes.
- **Rich contact dynamics**: Accurate modeling of friction, deformation, fracture, and fluid interactions—where most affordances manifest.
- **Multi-scale resolution**: Ability to simulate both macro behaviors (walking) and micro interactions (finger grip forces).
- **Stochasticity injection**: Controlled noise sources to ensure observed patterns are robust across perturbations, not artifacts of deterministic paths.

Critically, the engine must support *emergent complexity*: simple constraints should produce rich behavioral repertoires without explicit programming. For instance, configuring "uneven terrain" should naturally elicit balance corrections, stumbling, and adaptive gait changes—not because these behaviors are programmed, but because they emerge from physics.

4.4 Virtual Agents: Embodiment Without Representation

Agents in AFSA are deliberately minimal:

- **Morphology**: Articulated bodies with joints, mass distributions, and contact geometries—but no internal world model.
- **Control**: Simple feedback controllers (e.g., proportional-derivative control for joint angles) or reinforcement learning policies trained *only* on proprioceptive feedback and task rewards—never on semantic labels.
- **No semantic access**: Agents cannot "know" the linguistic input that shaped their environment. They only experience physical forces and constraints.

This minimalism ensures that any semantic patterns observed in agent behavior genuinely emerge from dynamics rather than being injected via agent architecture. When agents consistently slow their approach to objects labeled "fragile" during training, this caution must arise from physical consequences (objects breaking when handled forcefully), not from semantic awareness.

4.5 Pattern Extraction: Where Meaning Becomes Observable

The final component observes agent trajectories to extract statistically robust patterns:

- **Temporal signatures**: Oscillation frequencies, convergence rates, hesitation durations.
- **Spatial signatures**: Path tortuosity, proximity maintenance, avoidance geometries.
- **Energetic signatures**: Work expenditure, efficiency ratios, recovery costs after perturbation.
- **Failure modes**: Characteristic breakdown patterns under stress (e.g., cascading failures vs. graceful degradation).

These signatures are clustered across trials to identify invariant patterns. Human annotators then assign semantic labels to clusters ("this oscillation pattern corresponds to indecision"). Crucially, these labels serve *descriptive* rather than *constitutive* roles—they name patterns that exist independently of labeling.

5. Concrete Examples of Semantic Emergence

5.1 "Fragility" Without Representation

Linguistic input: "Handle the antique vase with care—it's extremely fragile."

AFSA processing:

1. LLM generates high fracture susceptibility, low impact tolerance parameters
2. Mapper configures vase object with brittle material properties
3. Simulation instantiates vase in environment with typical household forces present
4. Agents trained on manipulation tasks interact with vase

Emergent semantics:

- Agents that apply force rapidly or unevenly consistently trigger fracture events

- Successful agents develop behavioral signatures: slow approach velocities, distributed grip forces, minimal acceleration during transport
- Across 100 trials, successful manipulation correlates with force application duration >2.3 seconds ($p<0.001$)
- This temporal signature becomes the *operational definition* of "careful handling"

Semantic observation: An external observer notes that agents interacting with "fragile" objects consistently exhibit prolonged force application durations. The *meaning* of fragility is this behavioral regularity—not a stored concept. When presented with a new object labeled "fragile," agents that have experienced fracture consequences generalize the cautious behavior pattern without explicit reasoning.

5.2 "Social Pressure" as Physical Constraint

Linguistic input: "The committee watched intently as she presented her controversial proposal."

AFSA processing:

1. LLM interprets "watched intently" as increased attentional density around agent
2. Mapper configures environment with high-density agent swarm surrounding presenter
3. Physics engine models proxemic constraints: personal space violations generate repulsive forces
4. Presenter agent experiences increased constraint forces when deviating from expected paths

Emergent semantics:

- Presenter agent exhibits reduced path variance compared to baseline presentations
- Increased hesitation (velocity drops) at decision points
- Higher energy expenditure maintaining trajectory against constraint forces
- When "controversial" parameter added, constraint forces intensify near specific content markers

Semantic observation: The behavioral signature of "social pressure" emerges as constrained movement freedom and increased energetic cost of deviation. No agent represents "being watched"; the physical manifestation of attention density produces the behavioral pattern that observers label as pressure. Crucially, this pattern generalizes: when linguistic input changes to "hostile audience," constraint forces intensify asymmetrically, producing distinct behavioral signatures (avoidance trajectories, defensive postures) without new semantic programming.

5.3 Abstract Concepts: "Justice" as Constraint Symmetry

Even abstract concepts find physical analogs in constraint structures:

Linguistic input: "The judge ensured justice by treating both parties equally."

AFSA processing:

1. LLM maps "equally" to symmetry constraints in force application
2. Mapper configures interaction environment where agent actions toward two entities must satisfy symmetry conditions
3. Violations produce energetic penalties proportional to asymmetry magnitude
4. Agents learn policies minimizing penalty through balanced interactions

Emergent semantics:

- Successful agents develop interaction patterns exhibiting mathematical symmetry across entities
- Asymmetry events trigger corrective behaviors restoring balance
- Across trials, agents interacting under "justice" constraints show 87% lower asymmetry variance than baseline ($p<0.001$)

Semantic observation: "Justice" manifests as a statistical tendency toward interaction symmetry. When observers see agents consistently restoring balance after perturbations, they recognize the pattern as justice-like behavior. The concept requires no representation of fairness or morality—only physical constraints enforcing symmetry and agents adapting to minimize penalty.

This demonstrates a crucial insight: **abstraction emerges from constraint structure, not representational hierarchy**. "Justice" isn't built from simpler concepts; it's the name we give to behaviors produced by symmetry constraints. This resolves the grounding problem for abstract concepts—they ground not in perception but in *constraint dynamics*.

6. Philosophical Implications

6.1 Dissolving the Symbol Grounding Problem

Harnad's symbol grounding problem assumes symbols must connect to non-symbolic representations. AFSA dissolves this problem by eliminating the need for grounding: linguistic inputs never function as symbols to be grounded. They function as *constraint parameters* that directly shape physical dynamics. There is no symbol-referent gap to bridge because there are no symbols in the semantic loop—only physical parameters and their dynamical consequences.

This doesn't deny that humans use symbols; it reinterprets their function. Human language may similarly operate as a constraint-shaping tool—modulating attention, action possibilities, and social coordination without requiring internal symbolic representations of meaning. AFSA offers a computational proof-of-concept that symbol-free semantics is possible.

6.2 Normativity Without Representation

A persistent challenge for non-representational approaches is explaining normativity: how can a system distinguish correct from incorrect behavior without internal standards? AFSA grounds normativity *physically*:

- **Task success/failure**: Defined by physical outcomes (object delivered intact vs. broken)
- **Semantic correctness**: Defined by behavioral pattern fidelity to human expectations
- **Adaptive value**: Defined by energy efficiency, survival duration, or goal achievement within simulation

When an agent handles a "fragile" object too forcefully and it breaks, the failure is physical, not representational. The agent doesn't violate an internal rule about fragility; it experiences the physical consequence of fracture. Normativity emerges from the *cost structure of the physical world*—not from encoded rules.

This aligns with enactivist views of normativity as arising from autonomous systems' need to maintain their own organization (autopoiesis). In AFSA, agents that fail to adapt to constraint structures cease functioning—their dynamics terminate. Success is continued existence within the constraint landscape.

6.3 The Illusion of Semantic Content

AFSA suggests a provocative conclusion: **semantic content is observer-relative but physically grounded**. Patterns exist objectively in the dynamics (oscillations, convergence behaviors), but their interpretation as "indecision" or "determination" requires an observer with appropriate conceptual resources. This doesn't make meaning illusory—it makes meaning *relational*, like color (which exists as wavelength reflectance patterns but requires visual systems to become "red").

This resolves tensions between realism and constructivism in semantics: patterns are real features of dynamics, while their labeling as specific meanings depends on observer perspective. Crucially, not all labelings are equally valid—only those that track robust dynamical invariants earn stable semantic assignments. "Calling" a convergence pattern "indecision" fails if the pattern doesn't correlate with hesitation behaviors across contexts.

7. Comparison to Alternative Approaches

Approach	Semantic Core	Grounding Mechanism	Key Limitation	AFSA Advantage
Symbolic AI	Logic predicates	Hand-coded mappings	Brittleness; scaling impossible	No hand-coding; scales with physics
Distributional Semantics	Vector embeddings	Text co-occurrence	Disembodied; no physical consequence	Physical consequence as semantic basis
Vision-Language Models	Multimodal alignment	Image-text pairing	Grounding limited to perception	Grounding in action dynamics
Embodied Robotics	Sensorimotor loops	Physical interaction	Slow learning; narrow domains	Virtual embodiment enables rapid scaling
Predictive Processing	Latent generative model	Prediction error minimiz.	Retains representational commitments	Eliminates latent semantic variables

AFSA's distinctive contribution is **semantic emergence without representation**. Other embodied approaches still encode meaning as internal states (even if grounded in sensorimotor experience). AFSA pushes embodiment further: meaning exists *only* in the relational dynamics between agent and environment, never as an internal state.

8. Limitations and Honest Constraints

AFSA does not claim to solve all semantic challenges:

- **Virtual vs. biological embodiment**: Simulated physics lacks the richness of biological embodiment (hormonal states, pain, fatigue). AFSA captures *structural* aspects of affordances but not their full phenomenological depth.
- **Observer-dependence**: Semantic labeling remains observer-relative. AFSA explains how patterns emerge but not why humans converge on particular labelings—a question requiring cognitive science of categorization.
- **Compositionality challenges**: While basic composition ("heavy fragile object") works via parameter combination, complex metaphorical composition ("heavy responsibility") requires constraint analogies not yet formalized.
- **Temporal scope**: Current physics engines struggle with very long-timescale dynamics where meaning often resides (e.g., "legacy," "tradition").

These limitations define a research agenda rather than fatal flaws. Each represents a tractable engineering or scientific challenge—not a conceptual impossibility.

9. Conclusion: Language as World-Sculpting

We have presented epiphenomenal semantics as a radical alternative to representational theories of meaning. By reinterpreting language as a generator of affordance fields—structured physical constraints that shape embodied dynamics—we show how semantic content can emerge as stable patterns in agent behavior without ever being represented internally.

This framework resolves the symbol grounding problem not by better grounding symbols, but by eliminating the need for symbols in the semantic loop. It provides physical grounding for abstract concepts through constraint structures rather than perceptual metaphors. It grounds normativity in physical consequence rather than encoded rules. And it offers a computationally feasible path toward language systems whose "understanding" manifests in reliable, world-connected behavior rather than fluent but disembodied text generation.

The deepest implication may be philosophical: **meaning is not something we carry in our heads, but something that happens between us and the world**. Language doesn't encode meaning to be transmitted; it sculpts the landscape of possible actions, and meaning is what remains invariant in how we move through that landscape. AFSA operationalizes this insight, demonstrating that machines too can participate in meaning—not by representing it, but by enacting it through constrained embodiment.

In this view, the future of language AI isn't more parameters or larger datasets—it's richer embodiment. Not necessarily biological bodies, but *dynamical presence* in constraint spaces where words have consequences, actions have costs, and meaning emerges not as representation, but as reliable pattern in the dance between agent and world.

References

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*,

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker, and the legacy of William James's radical empiricism*. Lawrence Erlbaum Associates.

Turvey, M. T. (1992). Affordances and prospective control: An outline of the ontology.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*,

Chemero, A. (2009). *Radical embodied cognitive science*. MIT Press.

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor life: An enactive proposal*. Oxford University Press.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Shapiro, L. (2019). *Embodied cognition* (2nd ed.). Routledge.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., ... & Turian, J. (2020). Experience grounds language. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8718–8735.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*

Lake, B. M., & Baroni, M. (2022). Human-like systematic generalization through a meta-learning neural network. *Nature*, 614(7946), 106–111

Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. *arXiv preprint arXiv:2002.06177*.

Santoro, A., Lampinen, A. K., Mathewson, K., McClelland, J. L., & Barrett, D. G. T. (2022). Symbolic behaviour in artificial intelligence. *Current Opinion in Behavioral Sciences*

Agrawal, P., Nair, A. V., Abbeel, P., Malik, J., & Levine, S. (2016). Learning to poke by poking: Experiential learning of intuitive physics. *Advances in Neural Information Processing Systems*, 29, 5074–5082.

Huang, S., Li, H., Feng, Y., & Zhu, Y. (2023). Affordance as a bridge between language and embodied intelligence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8), 10234–10249.

Kumar, A., Fu, J., Zhang, M., Levine, S., & Finn, C. (2023). Language-conditioned affordance learning for robotic manipulation. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11245–11252.

Liang, J., Jiang, W., Yang, H., Li, C., Zhu, Y., & Zhu, S.-C. (2022). ACRE: Abstract commonsense reasoning benchmark for physical dynamics. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18485–18495.

Shridhar, M., Ganhotra, J., & Fox, D. (2023). CLIPort: What and where pathways for robotic manipulation. *Conference on Robot Learning (CoRL)*, 1–12.

Wang, T., Zhang, X., Huang, S., & Zhu, Y. (2024). Neural constitutive models for learning physical affordances from visual observations. *Proceedings of the 2024 International Conference on Learning Representations (ICLR)*.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 91–99.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT Press.

Port, R. F., & Van Gelder, T. (Eds.). (1995). *Mind as motion: Explorations in the dynamics of cognition*. MIT Press.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Bickhard, M. H. (2009). The interactivist model. *New Ideas in Psychology*, 27(1), 1–33.

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Hutto, D. D., & Satne, G. (2015). The natural origins of content. *Philosophia*, 43(3), 521–536.

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Variational neuroethology: Answering further questions: Reply to comments.

Physics of Life Reviews, 24, 73–76.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Humphreys, P. (2023). *Computational science and scientific representation*. Oxford University Press.

Merrill, W., Sabharwal, A., & Smolensky, P. (2022). What do language models learn about the world? *arXiv preprint arXiv:2210.11415*.

Schank, R. C. (2023). Why LLMs don't understand language. *AI & Society*, 38(4), 1457–1463.

Zador, A. M., & Richards, B. A. (2023). What can't deep learning do? *arXiv preprint arXiv:2305.18712*

Degrave, J., Hermans, M., Dambre, J., & Schrauwen, B. (2022). A differentiable physics engine for deep learning in robotics. *Frontiers in Neurorobotics*, 16, 797934.

Hu, Y., Anderson, T., Li, T., Sun, Q., Carr, N., Ragan-Kelley, J., & Durand, F. (2020). DiffTaichi: Differentiable programming for physical simulation. *International Conference on Learning Representations (ICLR)*.

Murthy, J., Kabra, D., Zoran, D., & Kohli, P. (2023). Learning physical simulators with neural surrogates. *Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*

Pfeifer, R., & Bongard, J. (2006). _How the body shapes the way we think: A new view of intelligence. MIT Press.

Pfeifer, R., Iida, F., & Lungarella, M. (2014). Cognition from the bottom up: On biological inspiration, body morphology, and soft materials. *Trends in Cognitive Sciences*

Suzuki, R., Brawer, H., & Scassellati, B. (2023). Morphological computation as semantic substrate: How body dynamics encode environmental structure. *Adaptive Behavior*, 31(4), 321–337.