Classification is Infrastructure

Planning Layers, Taxonomic Commitment, and the Architecture of Intelligent Systems

Jake Lawrence[1]

[1]*Independent Researcher*

March 2026

Abstract

Every system that acts under uncertainty depends on a classification layer—a set of categories that determines what the system can represent and therefore what it can do. We argue that classification layers function as infrastructure in the technical sense developed by Bowker and Star (1999): embedded in other structures, transparent when working, constitutive rather than descriptive, and resistant to change once installed. It follows that the most consequential design decisions in such systems are taxonomic, not algorithmic.

We develop this thesis through three case studies spanning radically different domains. In psychiatric nosology, the DSM's diagnostic categories have become institutional infrastructure—persisting through switching costs rather than scientific validity, embedded in insurance, law, research, and identity. In LLM inference execution, the plan lattice of a constrained decoding optimizer determines what strategies are discoverable; no cost model, however sophisticated, can find an operator the lattice does not contain. In LLM agent cognition, a six-module cognitive architecture determines the behavioral ceiling of the agent; capabilities that require explicit representational support cannot emerge from algorithmic improvement alone.

Across all three cases, the same structural pattern holds: the taxonomy defines the space of possibility, optimization searches only within that space, and the taxonomy becomes invisible precisely when it is most consequential. Drawing on Bowker and Star's infrastructure theory, Hacking's interactive kinds, and the pragmatic philosophy of classification, we argue that taxonomic commitment—the act of choosing which categories a system will operate with—is the foundational design decision in any intelligent system. Treating it as a preliminary step rather than a first-class design problem produces systems trapped inside spaces whose boundaries they cannot see.

Keywords: classification, infrastructure theory, taxonomic commitment, Bowker and Star, psychiatric nosology, constrained decoding, cognitive architecture, DSM, interactive kinds, LLM systems

1 Introduction — The Taxonomic Blind Spot

Across AI, medicine, and systems design, enormous effort goes into optimizing algorithms, architectures, training data, and governance. The categories those systems operate over receive comparatively little deliberate attention—yet the classification layer is where the behavioral ceiling gets set.

Central claim. In any system where an agent must act under uncertainty, the classification layer functions as infrastructure. It is embedded, transparent, constitutive, and resistant to change. The most consequential design decisions are taxonomic, not algorithmic.

We preview the three case studies not as summaries but as questions each case answers:

- **Can a classification system become so embedded that it persists for decades despite known scientific inadequacy?** Psychiatric classification shows that it can—and shows why.
- **Does taxonomic commitment constrain optimization even in purely computational systems with no human institutions involved?** LLM-QP’s plan lattice shows that it does—and that the constraint is identical in structure if not in switching cost.
- **Can a system be deliberately designed to resist the infrastructural hardening that makes classification invisible?** SAGEN’s adapter pattern attempts this—and the attempt reveals both the possibility and the limits of anti-reification by design.

Scope and limits. The paper does not argue for taxonomic nihilism (all categories are arbitrary, so none matters) or taxonomic perfectionism (there exists one right taxonomy waiting to be discovered). It argues for taxonomic *awareness*: treating the classification layer as a first-class design decision with structural consequences that persist long after the decision is made.

Methodological note. The argument proceeds by *infrastructural inversion*—foregrounding what is normally in the background, treating the categories themselves as the object of analysis rather than the medium of analysis. This is Bowker and Star’s method, applied here to three systems the author has built or studied.

Companion work. This paper is part of a research program on classification as infrastructure in intelligent systems. A companion paper [Lawrence, 2026a] applies the framework to inference execution planning. A second companion [Lawrence, 2026b] applies it to agent cognitive architecture. A third [Lawrence, 2026c] develops the psychiatric case study at book length. The present paper provides the shared theoretical framework and cross-case analysis.

2 Classification as Infrastructure

Bowker and Star’s *Sorting Things Out* (1999) identified eight properties that characterize infrastructure. We present each property and establish the analytical vocabulary the case studies will use.

1. **Embeddedness.** Sunk into other structures, social arrangements, and technologies. Not freestanding.
2. **Transparency.** Invisible in routine use. Users look *through* the system at what it delivers.
3. **Reach or scope.** Extends beyond a single event or one-site practice.
4. **Learned as membership.** Acquired as part of socialization into a community of practice.
5. **Links with conventions of practice.** Shapes and is shaped by the practices it organizes.
6. **Embodiment of standards.** Plugs into other standards and is determined by their reach.
7. **Built on an installed base.** Inherits the strengths and limitations of what preceded it. Path dependency.
8. **Becomes visible upon breakdown.** Disappears when working. Surfaces only when something fails.

Why these properties interact. The eight properties are not a checklist; they form a mechanism. Properties 1–2 (embeddedness + transparency) explain *invisibility*: because the taxonomy is sunk into everything and invisible in use, it ceases to feel like a design decision. Properties 7–8 (installed base + breakdown visibility) explain *lock-in*: because the taxonomy inherits from its predecessors and is only noticed when it fails, revision is both constrained by path dependency and triggered only by crisis. The combination—invisibility plus lock-in—is the mechanism that produces the infrastructure paradox developed in §8.

Additional key concepts. Three ideas from Bowker and Star will recur across the case studies:

- *Boundary object* — A shared artifact used differently by different communities while maintaining enough structural identity to coordinate across them.
- *Torque* — The biographical tension experienced by people whose lives don't fit the available categories.
- *Infrastructural inversion* — The methodological move of foregrounding what is normally in the background.

These properties are not merely descriptive but *explanatory*. They explain why infrastructure persists even when it is known to be flawed: because the switching costs are prohibitive, because the installed base constrains the upgrade, because it is embedded in too many other systems to be changed without cascading disruption. This is not inertia in the colloquial sense. It is structural lock-in.

Extending Bowker and Star. The original framework was developed for organizational and sociological systems—medical classifications, racial categories, nursing standards. We extend it to computational systems (LLM-QP's plan lattice) and software cognitive architectures (SAGEN's module decomposition). The extension is warranted because the structural properties—embeddedness, transparency, constitutiveness, path dependency—are domain-invariant: they describe the *relationship* between a classification system and the practices it organizes, not the nature of those practices. Whether the practice is clinical diagnosis, compiler optimization, or agent cognition, the infrastructure properties hold whenever the classification layer becomes embedded, transparent, and resistant to change.

3 Taxonomic Commitment

Definition. A *taxonomic commitment* is the set of categories a system operates with—the distinctions it can draw, the groupings it can represent, the boundaries it enforces. Every system that classifies makes a taxonomic commitment, whether or not it acknowledges doing so.

The key property. A taxonomic commitment defines the space of representable states, and therefore the space of achievable behaviors. An optimizer can only search within the space its taxonomy defines. A clinician can only diagnose conditions their manual contains. An agent can only reason about distinctions its architecture represents. Improving the algorithm, cost model, or search strategy operates *within* this space. Only taxonomic revision can expand it.

Philosophical grounding. We draw on three traditions:

- **Dupré's promiscuous realism.** Multiple equally legitimate ways to classify the same reality exist, each optimized for different purposes. Classification is selection among viable structures, not discovery of the one true structure.
- **Zachar's practical kinds.** Categories are tools, not mirrors. Their validity is measured by how well they serve the purposes for which they're designed, not by correspondence with mind-independent natural divisions.
- **Hacking's interactive kinds.** In domains involving human subjects, the categories and the classified co-constitute each other. The taxonomy is not just a lens on reality but an intervention in it. (The *looping mechanism*: classification → awareness → behavioral change → category destabilization → revision → further change.)

The central paradox. The better a taxonomy works, the harder it is to see *as* a taxonomy. When categories become embedded in practice, they cease to feel like choices and begin to feel like features of the world.

Transparency produces reification: the tool becomes invisible, and the world it constructs is mistaken for the world as it is.

Three predictions. If the framework is correct, the case studies should exhibit: (1) the taxonomy defines the representable space, not merely describes it (*constitutiveness*); (2) the taxonomy becomes harder to revise the more successfully it is installed (*lock-in*); and (3) the taxonomy becomes harder to *see* the more fluently practitioners use it (*invisibility through transparency*). We now test these predictions across three radically different domains.

4 Case Study I — Psychiatric Classification

The most fully developed case. All eight infrastructure properties instantiated in an existing system with decades of history and global reach. Demonstrates all three predictions at maximum intensity—including looping effects unavailable in the computational cases.

4.1 The System

The DSM (Diagnostic and Statistical Manual of Mental Disorders) is the classification infrastructure that organizes American and, increasingly, global mental health care. Its categories determine what insurance covers, how research is designed, which drugs are developed, how courts evaluate competency, how schools provide accommodations, how patients understand their own suffering.

Map the DSM against all eight properties: embedded in insurance billing, legal standards, pharmaceutical regulation, electronic health records, educational accommodation systems, military benefits. Transparent to trained clinicians who have internalized its categories as perceptual habits. Global in scope. Learned as professional socialization. Linked to treatment guidelines, clinical trial design, reimbursement protocols. Plugged into ICD, FDA drug approval, national health surveillance. Built on an installed base from Kraepelin through DSM-I/II through the DSM-III revolution. Visible upon breakdown—when suffering doesn't fit, when clinicians disagree, when insurance denies, when categories are exported to cultures whose suffering is organized differently.

4.2 The Taxonomic Commitment

The DSM's core commitments were design decisions, not discoveries:

- Categorical rather than dimensional
- Symptom-based rather than etiological
- Atheoretical with respect to causation
- Individual rather than relational as the unit of analysis

Spitzer's DSM-III optimized reliability (inter-rater agreement) over validity, institutional utility (billable categories, researchable populations) over phenomenological accuracy. Once installed, these choices became invisible—the water clinicians swim in.

Consequences. Categories that group together people with wildly different experiences (1,030 unique symptom profiles under one depression diagnosis). Arbitrary thresholds (five of nine symptoms for two weeks). Boundaries set by committee vote. Categories that *loop*—changing the people they classify, who change the phenomenon, destabilizing the classification in ways that require revision that produces further looping.

4.3 The Switching Cost Problem

Multiple technically superior alternatives exist: RDoC (dimensional, biologically grounded), HiTOP (hierarchical dimensional, psychometrically stronger), network theory (no latent variable assumption), process-based therapy (transdiagnostic mechanisms over diagnostic categories). None has displaced the DSM from clinical practice. Not because it is better, but because it is *installed*.

Replacement requires simultaneous overhaul of: insurance billing systems, retraining of hundreds of thousands of clinicians, revision of legal standards across all jurisdictions, rebuilding of electronic health records, reanalysis of decades of research, renegotiation of drug approval frameworks, and—most consequentially—managing disruption to millions of people whose self-understanding, treatment, legal standing, and community belonging are organized around current categories. The installed base wins. The installed base almost always wins.

4.4 What This Case Uniquely Reveals

The psychiatric case contributes two things the computational cases cannot. First, **looping**: DSM categories do not merely describe mental disorders; they partially constitute them. The classified become aware of the classification, change their behavior in response, and thereby change the phenomenon the category was supposed to capture. This is Hacking's interactive kinds operating at institutional scale—the taxonomy is not a passive lens but an active intervention in the reality it classifies.

Second, **the boundary object mechanism of persistence**: the DSM endures not because any single community finds it optimal but because it is adequate enough for coordination across all communities—clinicians, insurers, researchers, lawyers, pharmaceutical companies, patients—while being optimal for none. This coordination function is independent of scientific validity and explains why superior alternatives fail to displace it.

5 Case Study II — Inference Execution Planning (LLM-QP)

A compressed, computational instance of the same structural pattern. No looping, lower switching costs, but identical constitutive property. Demonstrates that taxonomic commitment is not a human-institutional phenomenon but a structural one.

5.1 The System

Constrained LLM decoding (JSON mode, function calling, grammar conformance) requires checking every token against a validity set. The dominant strategy—dense masked scoring—computes logits over the full vocabulary $|V| \approx 128{,}000$ and masks invalid tokens. Correct but wasteful when the branching factor $K = |A(s)|$ is small relative to $|V|$.

LLM-QP [Lawrence, 2026a] formalizes the observation that multiple execution strategies are semantically equivalent—they produce identical token sequences—but differ in runtime cost. The equivalence is formally proven: a Dense–Sparse Equivalence Lemma demonstrates that the argmax over the full vocabulary under masked scoring equals the argmax over the valid set under sparse scoring. Five physical plans implement the single logical operation DecodeStep(query, constraint_state): dense projection head, sparse adjacency scoring, amortized score update, amortized update with rerank, and full recomputation.

5.2 The Taxonomic Commitment

The plan lattice—the set of five physical implementations the planner considers—is LLM-QP's taxonomic commitment. It determines what optimizations are representable. A planner whose lattice lacks an amortized operator cannot discover amortized savings, regardless of cost model sophistication or bandit algorithm.

This mirrors database query optimizers: the physical operator enumeration (hash join vs. nested loop, index scan vs. sequential scan) defines the selection space. The cost model selects among lattice-provided options. The bandit's contextual features (branching factor K, margin m_t, decode depth) are meaningful only relative to operators in the lattice. Improving the cost model optimizes *within* the lattice. Only lattice revision expands what is discoverable.

5.3 Infrastructure Properties

- **Embeddedness.** The lattice is embedded in the MLIR/StableHLO compiler pipeline. Physical plans are represented as parallel subgraphs. Changing the lattice means changing compiler passes, kernel implementations, and cost model terms.
- **Transparency.** When working well, invisible—the system just runs fast. Visible upon breakdown: when workload patterns fall outside what current operators handle efficiently.
- **Built on installed base.** Available strategies constrained by existing kernel implementations, hardware capabilities, and compiler infrastructure.
- **Links with practice.** Telemetry trains routing policy within lattice-defined space. The bandit achieves sub-linear regret O■(√T) relative to oracle plan selection, confirming convergence—but convergence to the best plan *within* the lattice, not to the best plan conceivable.

5.4 What This Case Uniquely Reveals

LLM-QP isolates the *constitutive* property of taxonomic commitment in a domain stripped of human institutions, identity, and looping. This isolation is analytically valuable: it demonstrates that the taxonomy-defines-the-space relationship is structural, not sociological. The plan lattice is not a social convention; it is a deliberate engineering artifact. And yet it exhibits the same constitutive property as the DSM: the optimizer cannot discover what the lattice does not represent.

The case also contributes a specific anti-reification device: the formal plan equivalence proof. By demonstrating that multiple plans produce identical outputs (Lemma 2.1 in [Lawrence, 2026a]), LLM-QP makes the lattice's contingency visible—the plans are choices among equivalent alternatives, not the unique correct implementation. This formal equivalence has no analogue in psychiatric classification, where no proof exists that alternative category systems would produce identical clinical outcomes.

6 Case Study III — Agent Cognitive Architecture (SAGEN)

Bridges the gap between the computational and institutional cases. Constitutive like LLM-QP; shapes perception and action like the DSM; but includes a deliberate anti-reification mechanism—the first case study to design against the infrastructure paradox.

6.1 The System

SAGEN [Lawrence, 2026b] provides LLM agents with structured, persistent situational awareness through six cognitive modules on a shared blackboard: Goal Graph (objective hierarchy and lifecycle), Trajectory (episodic memory with typed transitions and ACT-R-inspired decay), World Model (entities, relationships, assumptions, unknowns), Self Model (capabilities, limitations, resources, failure history), Attention

Priorities (urgency-scored watchlist with TTL), and Interaction Protocol (normative behavioral constraints).

Coordination occurs through an Observe–Update–Inject loop: a domain adapter parses raw observations into structured updates, the engine applies updates to the blackboard, and the current state is rendered as compact text for injection into the LLM's context window. Formally, the awareness state at step t is the tuple S_t = ■G_t, T_t, W_t, M_t, A_t, P_t, t■—what appears in this tuple is representable; what does not is invisible to the agent.

6.2 The Taxonomic Commitment

Module-level commitment. The six-module decomposition determines what the agent can represent about its situation, and therefore what it can reason about and act on. An agent lacking a Trajectory module cannot distinguish a topic pivot from a topic abandonment—it has no machinery for typed episodic transitions. An agent whose Attention module does not distinguish threats from opportunities cannot allocate urgency differentially. An agent without explicit assumption/unknown tracking in its World Model cannot reason about epistemic boundaries.

Finer-grain commitments. The Trajectory's seven transition types (progress, reversal, pivot, discovery, external event, failure, branch) define the space of recognizable episodic patterns. The Attention module's four categories (threat, opportunity, anomaly, transition) determine representable salience. Domain adapter scan patterns—topic shifts, emotional escalation, callback opportunities, error recurrence—are the agent's perceptual categories; they determine what it can *notice*.

6.3 Infrastructure Properties

- **Embeddedness.** Modules embedded in every operational stage—perception (adapter parses into module structures), reasoning (LLM receives module-organized injection), action (Self Model mediates execution via can_i() checks).
- **Transparency.** Agent reasons *through* modules, not *about* them. Injection designed for transparent consumption—structured but natural-language-readable.
- **Learned as membership.** Domain adapters encode a domain's perceptual categories. Deploying in a new domain means defining what counts as entity, relationship, and pattern—analogous to professional socialization.
- **Built on installed base.** Inherits from ACT-R (memory decay), SOAR (working memory), blackboard architecture (shared state).
- **Visible upon breakdown.** Architecture surfaces when a domain requires an unrepresented distinction, when a cognitive operation falls outside the six-module decomposition, or when the single-domain adapter assumption breaks.

6.4 The Adapter as Anti-Reification Mechanism

The domain adapter pattern is a deliberate architectural response to the reification problem—the first instance across our three cases of a system designed to resist its own invisibility. By requiring a developer to explicitly choose entity types, relationship types, and scan patterns for each domain, the adapter keeps categories visible as design decisions rather than allowing them to calcify into invisible assumptions.

Evidence: the conversation adapter (6 entity types, 7 relationship types, 4 scan patterns) and coding adapter (8 entity types, 8 relationship types, 4 scan patterns) share zero entity types except "concept," zero scan patterns, and produce domain-appropriate injection formats—yet both run on the identical engine and schema. Cognitive infrastructure (six modules, OUI loop, blackboard) is stable; taxonomic content (what counts as entity, threat, transition) is replaceable.

The adapter pattern does not solve the reification problem. The six modules themselves are not adapter-replaceable. They are the installed base. A developer who internalizes SAGEN may come to see “six-module cognitive architecture” as how agent cognition works rather than as one useful decomposition among many. The adapter resists reification at the domain-content level while the module-level taxonomy remains vulnerable. This is an honest limitation, and it illustrates that anti-reification is a gradient, not a binary.

6.5 What This Case Uniquely Reveals

SAGEN demonstrates that taxonomic commitment determines the behavioral ceiling of a cognitive architecture. Of 20 information dimensions evaluated in the baseline comparison, 16 were captured by none of the flat-memory baselines (raw conversation buffer or rolling summary) at any level. These 16—goal hierarchy, typed transitions, epistemic boundary tracking, urgency scoring, and others—require explicit architectural support. No improvement in the underlying LLM produces them if the architecture doesn’t represent them. The taxonomy defines the ceiling; the LLM operates under it.

Uniquely among the three cases, SAGEN demonstrates *layered taxonomic commitment*—the six modules are a deep commitment (hard to revise), while adapter content is a shallow commitment (designed to be revised per domain). This layered pattern suggests a general design principle: separate deep cognitive infrastructure from shallow domain taxonomy, and make the shallow layer explicitly replaceable. The DSM conflates both layers into one monolithic artifact, which is part of why revision is so costly.

7 Cross-Case Analysis — The Structural Pattern

7.1 The Shared Pattern

Table 1: Cross-case comparison. Same structural pattern, different domains.

Dimension	DSM	LLM-QP Plan Lattice	SAGEN Modules
Defines representable space	Diagnosable conditions	Selectable execution strategies	Performable cognitive operations
Constrains optimization	Treatment limited to recognized categories	Cost selection limited to enumerated plans	Agent reasoning limited to represented distinctions
Transparent when working	Clinicians see through categories	Planner selects transparently	Agent reasons through modules
Resists revision	Institutional switching costs	Compiler/kernel dependencies	Architectural assumptions
Anti-reification mechanism	None	Formal plan equivalence proofs	Domain adapter pattern
Unit of revision	Entire manual (monolithic, 15–20 yr cycle)	Individual plan (add new kernel + pass)	Adapter (shallow) or module (deep)

7.2 Key Differences

Looping. DSM categories loop—they change the people they classify, destabilizing the phenomena. LLM-QP's lattice does not loop—strategies don't change from being selected. SAGEN occupies a middle position—its categories shape agent perception and action but the agent doesn't reflexively modify its own categories.

Switching costs. DSM: societal (legal, financial, institutional, identity). LLM-QP: technical (compiler passes, kernels, cost model terms). SAGEN: architectural (module redesign, adapter revision, injection format). The mechanism differs; the structural effect is the same.

Amenability to deliberate revision. LLM-QP's lattice can be extended by adding physical plans. SAGEN's modules can be extended through adapters (at the domain-content level). The DSM resists revision because the installed base constrains the upgrade and looping makes consequences unpredictable.

7.3 The Hardness Spectrum

The differences reveal that the *strength* of the infrastructure properties varies with the domain's coupling to human institutions, identity, and reflexive awareness. The constitutive property is invariant. But the *reversibility* of that constitutive commitment varies dramatically. This suggests a spectrum of **infrastructural hardness**:

- **Soft.** Computational lattices (LLM-QP). Extendable by design. No looping. Switching costs are technical and bounded. Reification risk is low.
- **Medium.** Cognitive architectures (SAGEN). Modular at the domain-content layer but constraining at the module layer. No looping in the Hacking sense, but the taxonomy shapes agent perception. Switching costs are architectural. Reification risk is moderate.
- **Hard.** Institutional classifications (DSM). Looping effects. Societal switching costs. No built-in anti-reification mechanism. Reification risk is maximal—and largely realized.

The spectrum is not a ranking of quality. Hard infrastructure is not worse than soft; it is *harder to change*. The design implication: systems should be built as soft as the domain permits.

7.4 Testing the Three Predictions

1. **Constitutiveness.** Confirmed in all three cases. The DSM's categories define what conditions are diagnosable. The plan lattice defines what strategies are discoverable. SAGEN's modules define what cognitive operations are performable.
2. **Lock-in.** Confirmed, with varying intensity. The DSM is maximally locked in. LLM-QP's lattice is minimally locked in—extension is a straightforward engineering task. SAGEN falls between.
3. **Invisibility through transparency.** Confirmed, with a gradient. DSM categories are fully transparent to trained clinicians. The plan lattice is semi-transparent. SAGEN's adapter pattern forces visibility on the developer—a deliberate design choice.

8 The Infrastructure Paradox

Diagnosis, not prescription. §9 provides the prescriptions.

The paradox. Classification works best when invisible but becomes most dangerous when invisible. The four-phase cycle:

1. Design. Categories chosen as practical tools—provisional, purpose-specific, explicitly acknowledged as decisions.

2. Installation. Categories embedded in practice. Learned, linked to conventions, connected to standards.

3. Transparency. Categories become invisible. Users look through them. They cease to feel like choices.

4. Reification. Categories treated as discoveries rather than decisions. Hyman's diagnosis of the DSM: provisional conventions hardened into natural kinds.

The cycle operates at different speeds across the hardness spectrum. Computational taxonomies (LLM-QP) cycle fast and reify weakly. Cognitive architectures (SAGEN) cycle at medium speed—the adapter pattern resists reification at the domain-content layer but the module-level decomposition may harden. Institutional taxonomies (DSM) cycle slowly and reify completely—categories presented indistinguishably from scientific findings.

The structural insight. The cycle is not a failure of vigilance. It is a consequence of infrastructure's defining property: to function, it must be transparent; to be transparent, it must become invisible; to become invisible, it must cease to feel like a choice. The only defense is architectural—mechanisms that structurally resist transparency's slide into reification.

9 Implications for System Design

9.1 Taxonomic Commitment as First-Class Design Decision

The categories a system operates with should be documented, evaluated, and revisited with the same rigor applied to algorithmic choices. This means: explicit enumeration of what the taxonomy can and cannot represent, versioned category definitions, and periodic review.

9.2 Anti-Reification Mechanisms

Systems should include structural features that keep the taxonomy visible:

- **Modular adapter patterns** (SAGEN)—encapsulate domain-specific categories in replaceable modules.
- **Formal equivalence proofs** (LLM-QP)—demonstrate multiple plans produce identical outputs, keeping the lattice visible as a set of choices.
- **Explicit confidence metadata**—annotate categories with epistemic status (provisional, validated, contested, convenience-only).
- Versioned categories with changelogs. Sunset dates on provisional categories.

The specific mechanism matters less than the principle: *the system should resist its own invisibility.*

9.3 Layered Taxonomic Architecture

SAGEN's layered commitment suggests a general design principle. Systems should distinguish between **deep categories** that define fundamental representational capacity (hard to change, chosen with commensurate care) and **shallow categories** that specialize for a domain (designed to be replaceable, with explicit interfaces). The DSM's monolithic structure—where deep ontological commitments are fused with shallow clinical content—is an anti-pattern that maximizes revision cost at every level.

9.4 Taxonomic Debt

By analogy with technical debt: *taxonomic debt* accumulates when classification decisions are made expediently and left unexamined as the system scales. Symptoms include categories that no longer match operational reality, distinctions that practitioners routinely work around, and switching costs that grow faster than the system's value proposition. Like technical debt, taxonomic debt is invisible until it becomes crippling, and like technical debt, the remedy is not to avoid classification decisions but to make them deliberately, document them, and budget for revision.

9.5 Migration Planning at Design Time

If a system's categories will embed in other structures, future revision costs should be anticipated from the start. Patterns that reduce switching costs: modular versioning (individual categories versioned independently), layered architecture (dimensional foundations with categorical interface outputs), and translation layers (backward compatibility during transition).

9.6 Taxonomic Evaluation Criteria

Distinct from algorithm or model evaluation:

- **Coverage.** What can the taxonomy represent?
- **Blind spots.** What can it *not* represent, and what are the consequences?
- **Switching costs.** How embedded is it? What would revision disrupt?
- **Reification risk.** How likely are provisional categories to be mistaken for natural kinds?
- **Hardness.** Where on the soft–medium–hard spectrum, and is that the right position?

10 Conclusion — The Categories Are Not Scaffolding

In any system where an agent must act under uncertainty, the classification layer functions as infrastructure. It is the most consequential and least examined design decision. It determines the space of possibility. Optimization improves performance within that space. Only taxonomic revision can expand it.

Three case studies demonstrate this across domains sharing almost nothing except the structural pattern: a taxonomy that defines the representable space, becomes invisible when working, and resists revision once installed. The three predictions—constitutiveness, lock-in, invisibility through transparency—hold in all three cases, modulated by a hardness spectrum that tracks the domain's coupling to human institutions, identity, and reflexive awareness.

The practical upshot is not that better taxonomies will solve hard problems. It is that failing to recognize taxonomies *as* taxonomies—failing to see them as design decisions with structural consequences, treating them as neutral descriptions rather than constitutive commitments—produces systems trapped inside spaces they cannot see the edges of.

A closing observation. This paper is itself an exercise in infrastructural inversion—foregrounding the classification layers in three systems and making them the object of analysis rather than the medium. The test of its success is not whether readers agree with every claim but whether, after reading it, they can see the taxonomic commitments in their own systems: the categories they chose without choosing, the distinctions their architecture can and cannot draw, the walls they built and then forgot were walls.

The categories are not scaffolding. They are load-bearing walls. The first step toward building better systems is seeing the walls for what they are.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences.* MIT Press.

Caspi, A. and Moffitt, T. E. (2018). All for one and one for all: mental disorders in one dimension. *American Journal of Psychiatry*, 175(9):831–844.

Cooper, R. (2005). *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders.* Springer.

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science.* Harvard University Press.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1):32–64.

Hacking, I. (1995). *Rewriting the Soul: Multiple Personality and the Sciences of Memory.* Princeton University Press.

Hacking, I. (1998). *Mad Travelers: Reflections on the Reality of Transient Mental Illnesses.* University of Virginia Press.

Hacking, I. (1999). *The Social Construction of What?* Harvard University Press.

Hayes-Roth, B. (1985). A blackboard architecture for control. *Artificial Intelligence*, 26(3):251–321.

Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, 6:155–179.

Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness. *Molecular Psychiatry*, 17:377–388.

Kleinman, A. (1988). *Rethinking Psychiatry: From Cultural Category to Personal Experience.* Free Press.

Laird, J. E. (2012). *The Soar Cognitive Architecture.* MIT Press.

Lawrence, J. (2026a). LLM-QP: Query planning for large language model inference. *arXiv preprint*.

Lawrence, J. (2026b). SAGEN: Situational awareness for generative agents. *arXiv preprint*.

Lawrence, J. (2026c). The Invisible Architecture: Psychiatric classification as infrastructure. *Unpublished manuscript*.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *UIST 2023*.

Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3):377–391.

Zachar, P. (2014). *A Metaphysics of Psychopathology*. MIT Press.