Geometric Cognition, Lateral Thinking, and Emergent Creativity in Manny Manifolds

Background: Lateral Thinking in Cognitive and Computational Systems

Human creativity is often characterized by lateral thinking – the ability to leap across conceptual spaces and form novel associations. Cognitive neuroscience suggests this involves an interplay between divergent, spontaneous ideation and convergent, evaluative processing. For example, creative idea generation correlates with greater functional connectivity between the brain's default-mode network (DMN) (associated with spontaneous, memory-based thought) and the executive control network (task-focused, evaluative) . In other words, high creativity brains show stronger cooperation between regions for imagination and those for top-down control (evidence level: A, fMRI study). This aligns with theories that creative cognition requires both associative fluency and goal-directed filtering. The Manny Manifolds architecture explicitly mirrors this dual-process dynamic via its bicameral design: two interacting subsystems – an "experiencer" that freely traverses the knowledge manifold and an "executive" that guides, evaluates, and regulates these trajectories (C). Such a division is analogous to the DMN/Executive coupling in the brain and is hypothesized to foster self-reflection and regulated exploration in Manny's reasoning loop.

Another perspective on creativity comes from predictive coding and active inference. According to Friston's free-energy principle, brains are fundamentally driven to minimize surprise (prediction error). A potential paradox arises: How can an uncertainty-minimizing system also be creative (which seemingly embraces novelty and surprise)? Recent work (Constant et al., 2024) resolves this by emphasizing the role of the environment or context in pushing the predictive mind into new territories . By actively altering environmental conditions or one's perspective, an agent creates "exploration bubbles" that force continual learning and novelty-seeking beyond its prior model . In essence, creativity can emerge from an agent repeatedly moving its own goalposts, staying at the edge of its predictive comfort zone. Manny Manifolds is conceptually aligned with this active-inference view. It treats knowledge as a dynamic manifold that reshapes with each experience, and it includes mechanisms (like lenses to shift perspective and an auto-clarification routine for gaps) that continually inject novel information or viewpoints into the system (C). By modulating its "world" – for instance, applying a new lens that recontextualizes concepts – Manny can create the kind of evolving, challenging environment that Constant et al. argue is key to creative cognition. In active-inference terms, Manny's curvature updates serve to minimize surprise/energy as it learns, but by tuning certain parameters it can deliberately seek novelty (increasing temporary prediction error) as a long-term strategy to discover inventive solutions (evidence level: C, theoretical alignment).

Historically, creativity has also been framed in terms of associative networks. Sarnoff Mednick's classic theory posited that highly creative individuals have "flat" associative hierarchies, meaning they can retrieve many remote or uncommon associations nearly as easily as common ones . Less creative minds have "steep" hierarchies dominated by the strongest, most typical associations, making lateral connections harder. Manny Manifolds offers a computational substrate to realize controllable associative flattening. Its memory is a graph (or knowledge manifold) where linkage strengths are not fixed but continuously adjusted by use and feedback. Because Manny can weaken or strengthen connections on the fly, it could in principle flatten its associative landscape when needed – e.g. by reducing the dominance of a few strong edges and keeping a rich web of moderate connections. In Manny's terms, this balance is reflected in curvature distribution: a very uneven (heavy-tailed) curvature distribution means a few pathways have extremely high weight/attraction (analogous to a steep hierarchy), whereas a more uniform curvature landscape means many pathways carry information (a flat hierarchy). Notably, emergent network studies show that simple Hebbian plasticity can produce heavy-tailed connectivity (scale-free networks) over time – i.e. a few hubs become very strong (A, empirical physics of neural nets). If unchecked, this could lead to habitual thinking or fixation. Manny's design includes safeguards like mild L1 regularization and periodic "sleep" consolidation to prevent runaway growth of a single connection . By tuning these processes, Manny could maintain a more evenly curbed network, effectively emulating Mednick's flat associative profile during creative problem-solving. This theoretical parallel (evidence level: C) suggests that Manny's curvature-based plasticity might dynamically realize the associative flexibility that lateral thinking demands.

Manny's Mechanisms and Parallels to Creative Cognition

Curvature-Based Learning Rule – Parallels to Hebbian and Reinforcement Learning: Manny Manifolds learns by adjusting the curvature κ of edges in its knowledge graph whenever it traverses them, according to a rule $\Delta\kappa = \eta \cdot V \cdot f(\text{usage})$. Here η is a global learning rate, V is a valence signal, and $f(\text{usage})$ is a usage-dependent factor (e.g. a log-scaling that dampens gains for very frequent edges) . This rule has clear echoes of known neurocomputational learning rules. It can be seen as a form of reward-modulated Hebbian plasticity: the system strengthens or weakens connections that were active (traversed) in proportion to a "teaching signal" (valence). In the brain, dopamine plays a similar modulatory role, signaling reward or error to gate synaptic plasticity. For instance, dopamine bursts following reward can potently modulate synapse strength, effectively telling Hebbian connections when to learn (A). Manny's valence is analogous to such a neuromodulator – if an outcome is successful or deemed positive, valence > 0 and all edges used in that reasoning thread get a curvature boost (a bit like a dopamine-driven LTP for those associations); if the outcome is negative (valence < 0), those edges' curvature is reduced (analogous to punishment or LTD). In this way, Manny implements a three-factor learning rule (pre/post co-activation of nodes, with valence as third factor) that mirrors biologically plausible learning mechanisms . The usage function f(usage) = log(1+usage) provides diminishing returns for repeatedly used edges , which is reminiscent of neural habituation or synaptic saturation – common connections won't just grow unbounded in strength simply by brute frequency. Instead, novel or infrequently used connections can, under the right valence, gain disproportionately on their rare usage, supporting the exploration of new associations. This aspect aligns with the idea that creativity benefits from occasionally favoring less-traveled paths in the associative network (evidence: C, theoretical). In summary, Manny's curvature rule can be seen as Hebbian learning ("neurons that fire together, wire together") guided by a value signal, a proven recipe for adaptive learning in both neural and artificial agents (A).

Valence and Emotional/Affective Components: In human creativity, affect and motivation often guide where we focus our explorations – curiosity, interest, positive affect can broaden thought, whereas fear or negative affect can narrow it. Manny's valence-modulated plasticity provides a handle to incorporate these influences. The system's valence V is a multi-channel weighting for importance, emotion, or novelty . This means Manny could be configured (or learn) to assign higher positive valence to novel combinations or rewarding outcomes, and negative valence to dead-ends or contradictions. Neuromodulatory models similarly propose that surprise or novelty can itself be rewarding (intrinsic motivation) – essentially treating curiosity as its own reward signal. Recent network neuroscience work frames curiosity as a drive to fill gaps and increase the flexibility of one's knowledge network . In Manny, one could imagine an intrinsic novelty valence that spikes when a new link is made or a concept is learned, thus reinforcing exploratory moves. Conversely, too strong a negative valence for any incoherent result could make Manny overly conservative (always staying in familiar territory). Striking the right balance is crucial. Notably, OpenCog's Economic Attention Network (ECAN) similarly assigned an "attention weight" to each node/link in a knowledge graph that would increase if it was useful or context-relevant and decay otherwise . This ensured important associations stayed active while unhelpful ones faded, akin to Manny's valence shaping which edges persist. The key parallel is that adaptive attention allocation in a knowledge network can yield more efficient and creative problem-solving by focusing cognitive resources on promising connections (evidence: B, OpenCog framework). Manny extends this by making valence multi-dimensional – for example, a "novelty-valence" and a "accuracy-valence" could operate simultaneously, allowing the system to sometimes prioritize surprise (novel ideas) and other times prioritize coherence or correctness.

Motifs: Reusable Knowledge Chunks: A core feature of Manny Manifolds is the formation of motifs, which are frequently traversed subpaths through the manifold that get cached as reusable higher-level "skills" . This is directly inspired by mechanisms in cognitive architectures and AI that capture repeated patterns for transfer learning. In Soar (a well-known cognitive architecture), the process of chunking compiles the solution to a subgoal impasse into a new rule, so that next time a similar situation arises it can be handled in one leap . Essentially, Soar automatically learns macros – exactly what Manny's motifs represent (evidence: A, established AI architecture). Likewise, in hierarchical reinforcement learning, options or macro-actions are learned policies that accomplish sub-tasks; once learned, an option can be invoked as a single action to speed up problem solving in new tasks. Sutton et al. (1999) showed that using options enables significantly faster learning and transfer by reusing skills across tasks (A). Manny's motifs serve an analogous purpose: if the system has frequently traversed a sequence of concepts (e.g., a reasoning chain for solving a type of puzzle), it will condense that into a motif. Later, when facing a new but related problem, Manny can recognize that motif (or a variant) could apply, thus reusing prior knowledge to creatively handle a novel situation . This directly addresses one of Manny's design goals: transfer and analogy. The documentation specifies that transfer means "new but related tasks reuse prior paths (apple → pear)" and measures success by motif/edge reuse ≥30% in new tasks . In human creativity, analogical reasoning is often about mapping a known solution pattern to a new domain (the apple pie → pear pie → open-fire cooking example). Motifs give Manny an internal mechanism for conceptual blending: they are structural patterns abstracted from content. By applying a motif from one context to a disparate context (with the help of Manny's lens or alignment algorithms), Manny engages in a form of analogical reasoning. This approach echoes work in case-based reasoning and analogical problem solving (e.g., Holyoak & Thagard's cognitive models), as well as more recent machine analogies where graph patterns are transplanted between knowledge graphs (evidence: B for concept, as this is a well-theorized but still emerging capability in AI). The novelty in Manny's approach is that these motifs arise naturally from the agent's own exploratory trajectories rather than being hand-coded, and they exist within a geometric space that allows blending via projection (e.g. finding correspondences between nodes of two motifs through embedding similarities or valence alignments).

Knowledge Manifold and Geometric Representation: Underlying all of Manny is the assumption that knowledge, context, and experience can be represented geometrically. This relates to a rich vein of research on conceptual spaces and geometric knowledge representation. Gärdenfors (2000) proposed that concepts can be points or regions in a metric space defined by quality dimensions, enabling natural modelling of similarity, compositionality, and analogy (C). Manny's manifold is a direct homage to this idea – it is essentially a conceptual space that grows and reshapes with learning. Modern machine learning has also embraced geometric representations: Nickel & Kiela's Poincaré embeddings famously showed that embedding a knowledge graph in a non-Euclidean (hyperbolic) space can capture hierarchical relationships much more parsimoniously than Euclidean space (A). Real-world knowledge graphs often have sections of varying local curvature (hierarchical trees vs cycles), and a single homogeneous geometry is inadequate . Manny's manifold, in effect, is a variable-curvature knowledge graph – different regions can have different curvature based on the structure of knowledge there (dense clusters, long chains, etc.). Intriguingly, very recent work (2025) on Ricci flow-based knowledge graph embedding (RicciKGE) demonstrates that allowing the geometry to co-evolve with the embedding (adjusting local curvature according to graph structure) leads to better convergence and accuracy (B, preprint). They prove that coupling embedding learning with curvature adaptation drives the graph toward an optimal flatter geometry while improving link prediction. Manny's curvature updates are not identical (Manny uses curvature to encode confidence or strength rather than distances directly), but the principle of dynamically tuning geometry to data is shared. The broader point is that representing knowledge in geometric terms (nodes, edges with curvature, distances corresponding to semantic relatedness or energy) is both biologically inspired (cognitive maps in the hippocampus, conceptual manifolds in cortex) and empirically powerful in AI (as seen in hyperbolic embeddings improving generalization). This geometric underpinning is what enables Manny to potentially perform conceptual interpolation and extrapolation – moving through the space along geodesic paths to generate logical inferences, or bending the space via curvature to bring two previously far concepts into proximity (a candidate mechanism for creative insight when a new analogy "clicks").

Empirical Frameworks to Measure Creativity in Manny

To rigorously test whether Manny Manifolds exhibits creativity, we need measurable frameworks from the field of computational creativity. Researchers often adapt human creativity assessments and develop AI-specific benchmarks focusing on divergent thinking, analogy, and novelty. Key metrics include fluency (the sheer number of ideas or solutions generated), flexibility (the diversity of categories those ideas span), originality (statistical rarity or novelty of the ideas), and coherence or quality (how useful or sensible the ideas are) . These map well to Manny's behaviors: e.g., fluency could correspond to how many distinct solution paths Manny can generate for an open-ended query; flexibility to how many different knowledge domains or motif sets it draws upon; originality to how low the average prior usage or curvature of those paths were (lower usage = more novel combination); and coherence to the valence-weighted energy of a path (paths that Manny assigns overall positive valence and low surprise). In human testing, the Torrance Tests of Creative Thinking (TTCT) use tasks like "Alternative Uses" (list as many uses for a common object as possible) or storytelling from abstract pictures, and score them on these metrics. For Manny, analogous tests can be designed: for example, Alternative Use Task for Concepts – feed Manny a concept (like "brick") in its knowledge base and ask it to generate a number of different contexts or goals that involve that concept (building, weapon, musical instrument, etc.), then evaluate how many and how diverse the results are. If Manny's manifold supports lateral thinking, we expect high fluency (many associations) and high flexibility (drawn from different domains of the manifold) in its responses, ideally with some truly original (low-frequency) connections. Recent studies have even given GPT-4 (a large language model) such tests and found it scored in the top 1% of human performance for originality (A), although these results raise debates about whether it reflects genuine creativity or just a vast memory. For Manny, who learns from scratch and builds its own knowledge network, we could more directly attribute creative performance to specific mechanisms (e.g., if a high originality score correlates with Manny's curvature variance or motif usage patterns, that's informative).

Beyond divergent thinking, analogical reasoning benchmarks are crucial. One could evaluate Manny on tasks like the Remote Associates Test (RAT), where the goal is to find a concept that links three given words (e.g., "cookie", "heart", " sixteen" → answer: "sweet"). This tests the ability to find a bridging concept and is often used as a creativity measure. Manny's graph nature is well-suited to RAT: the solution is literally a node that connects to all three clue nodes. Success would depend on Manny having the necessary edges (knowledge) and the ability to traverse creatively – perhaps via motif bridging or a targeted breadth-first search with valence guiding it toward convergence. Another benchmark could be analogy puzzles (A:B :: C:?), which test structural mapping. Manny would have to use its lenses or domain alignment to map the relationship from A to B onto C to find the answer. For example, coach is to team as director is to ? Manny's "sports" domain and "film" domain might be far apart, but by using a lens that maps roles in one domain to analogous roles in another, it could identify "cast" as the analogous answer (if its manifold has learned those relations). We can draw on established AI analogy datasets (like Bongard problems in vision, or word analogy datasets from word embeddings) to systematically test Manny's analogical skill. A notable evaluation framework is the Tversky analogical similarity score – measure how well Manny's internal distances between A–B vs C–answer reflect the analogy being solved (the more proportional, the better the analogical mapping).

In computational creativity research, there's also the idea of novelty search as a driver of creative behavior. Instead of optimizing for a specific performance objective, one rewards an agent for finding novel behaviors or states. In a maze, for instance, novelty search would encourage exploring every corner rather than just the exit. Manny could be subjected to a novelty-generation task: for example, ask Manny to "brainstorm" in an open-ended way about a topic and explicitly score it on novelty (perhaps using human judges or a similarity metric to known ideas). If Manny's valence is set to treat novelty as positive reinforcement, we'd expect it to deliberately walk into less-visited regions of its manifold, chaining disparate motifs to produce uncommon outputs. Quantitatively, we could measure the graph distance or semantic distance of Manny's outputs from the starting concept (greater distance often corresponds to more creative leaps). We might also measure concept combination originality: give Manny two unrelated concepts and ask it to connect them (a simple form of conceptual blending task). The quality of the connection can be judged by human evaluators for creativity (does it yield a surprising, meaningful idea?) and by checking Manny's path – did it use a motif or lens to bridge a large conceptual gap, or just trivial nearest-neighbor links?

Finally, it's important to consider coherence metrics in tandem with novelty. Creativity is not just randomness; it's appropriate novelty. So, metrics analogous to precision or validity are needed. For Manny, a coherence score could be derived from the overall valence of a generated solution (if Manny itself rates the path as high-valence – meaning it found it consistent or rewarding – that's a proxy for coherence), or external measures like logical consistency checks. Human evaluators could rate Manny's creative outputs for sensibility and interestingness. These multi-faceted evaluations (fluency, flexibility, originality, coherence) provide a robust framework to quantify creativity in Manny's behaviors. Each metric can then be tied back to Manny's internal parameters (e.g., does a higher exploration temperature τ increase fluency at the cost of coherence? Does motif reuse improve flexibility but maybe reduce originality if it over-relies on known patterns?).

Precedents: Systems Melding Geometry, Plasticity, and Creativity

Manny Manifolds is an ambitious synthesis, but elements of its design echo earlier projects in AI research that sought to imbue systems with creative capacities through structural and self-organizing means. It's instructive to survey a few precedents and note which aspects have empirical support versus remaining speculative.

- Adaptive Semantic Networks & Concept Graphs: There have been knowledge-base systems (e.g., ConceptNet and its predecessors) that support finding creative connections by traversing a large semantic graph. For instance, ConceptNet has been used in generating metaphors or joke punchlines by linking seemingly unrelated concepts via intermediate nodes. However, these systems typically use a static graph of common sense knowledge. Manny's innovation is that the graph reconfigures itself with experience. While there isn't a direct one-to-one precedent, we can consider projects like the CogPrime/OpenCog architecture, which included a dynamically updating knowledge network. OpenCog's ECAN (Economic Attention Network) we mentioned is one component; another is its PLN reasoning which could, in theory, follow uncommon links if their attention values rose. Empirical validation in OpenCog was limited to toy domains, but it did demonstrate that a spreading activation process with attentional modulation could solve simple analogy problems (B). The lesson here is that some form of attention/weight modulation on knowledge networks has been practically used, supporting Manny's core premise that valence-weighted traversal can yield intelligent (and potentially creative) behavior.
- Neural Network Creativity & Chaotic Dynamics: A line of work in the 1990s and 2000s considered whether creativity in neural networks could be achieved by introducing chaotic or noisy dynamics. For example, Martindale hypothesized that creative cognition might correspond to a state of low cortical arousal (more random, defocused processing), leading to wider-ranging associations (C). In artificial neural nets, researchers like J. A. S. Kelso and others explored training RNNs at the "edge of chaos" – the critical point where network activity is neither stable nor explosively chaotic, which often maximizes complexity of patterns generated. There is growing evidence that brains themselves operate near criticality, which allows rapid reconfiguration of neural assemblies (A, human neuroimaging). At criticality, a network can quickly "tip" into new states (useful for insight) while retaining some order (so thought isn't pure noise). Some evolutionary art and music systems have explicitly set parameters to near-chaotic regimes to generate novel outputs (e.g. mutation rates in genetic algorithms or temperature in Boltzmann machines tuned to critical values). What's been experimentally validated is that critical tuning increases variability and novelty in outputs, though if pushed too far it harms coherence. Manny can be seen as a system that might find a sweet spot between stability (knowledge consolidation) and chaos (exploratory leaps). The presence of parameters like η (learning gain) and τ (consolidation/cooling) gives it knobs to adjust this balance. While Manny is not a random generator – it's still goal-directed – one could draw on the "edge of chaos" concept to hypothesize that maximal creativity will occur at intermediate settings of these parameters: e.g., when learning/plasticity is high enough to allow radical restructuring of the manifold, but not so high as to produce a disorderly jumble. Lynn et al. (2020) showed that even a simple model of synaptic plasticity can self-organize a network to critical-like heavy-tailed connectivity distributions , potentially tying together the ideas that Hebbian learning can yield the network complexity associated with creative brains (A). This remains somewhat theoretical in AI, but Manny could be the platform to test it directly (more in Predictions section).

- Graphical Concept Blending and Analogy Systems: There have been specific efforts to model creativity through conceptual blending in structured representations. For example, researchers have used knowledge graphs to blend concepts by finding overlaps in their property sets or shared structure (onto which to align and merge two concepts). The MetaGrid project (C) and others implemented algorithms to take two input concepts and produce a blended concept by unifying graphs – e.g., blending "bird" and "song" might yield something like a "singing bird" concept with emergent properties. These are largely theoretical or at best prototype implementations evaluated on a few examples. Manny's contribution here could be significant: since it natively represents knowledge in a graph, one could leverage its motifs as units of meaning and attempt systematic blends. A trivial precedent is simply vector arithmetic in conceptual embedding spaces (e.g., word2vec analogies where vector("king") - vector("man") + vector("woman") ≈ vector("queen")). That demonstrated that a continuous representation can capture analogical structure , but it's a static phenomenon learned from huge data and doesn't explain how to deliberately form a new concept. Manny, by contrast, could physically attach two subgraphs via a new link or common node, creating a blended concept in a single connected structure. This is closer to how our minds might form creative ideas by linking previously separate mental schemas. We do not yet have empirical proof that any AI has achieved human-level concept invention by blending, but Manny's design is informed by these past attempts and theories (evidence: C, computational creativity theory).
- Active Inference Agents with Imagination: Some work within the active inference community (e.g. simulations by Pezzulo et al. and others) has shown that agents who explicitly model counterfactual or imagined trajectories can solve problems in novel ways (B). For instance, a robot might use an internal generative model to simulate what happens if it tries a new action sequence, thereby discovering a solution without physically attempting every option. This is a kind of model-based creativity. Manny's "experiencer" loop, especially if augmented by an LLM "lens" that can propose hypotheses, functions similarly: it can simulate a reasoning thread through its manifold (with the LLM suggesting plausible next steps when knowledge is missing), effectively performing a mental exploration before committing to learning. The experiencer might wander more freely (divergent mode), while the executive evaluates if any of those wanderings hit upon a valuable insight (convergent selection). While Manny's current implementation uses the LLM as a **"lens-maker and summarizer, never controller" *, one could imagine extending its role to proposing imaginative leaps ("Have you considered relating concept X from a distant domain?"). This would be akin to an "idea generator" module. Such dual systems (generator + evaluator) have precedent in creative AI: e.g., in generative adversarial networks (GANs), the generator creates candidates and the discriminator evaluates them – not unlike an experiencer generating manifold trajectories and an executive judging their valence. Empirically, GANs have produced creative outputs in images and music (A), though their "imagination" is not easily interpretable. Manny's advantage would be transparency: each idea is an explicit path in a knowledge graph, open to inspection and explanation.

Summary of Precedents: Table 1 (below) consolidates key comparisons. Manny distinguishes itself by combining these elements – a geometric knowledge representation, continual self-organizing learning, motif-based abstraction, and a two-part cognitive loop – all aimed at enabling creative cognition within one system. Many components have encouraging support individually (e.g., the benefit of geometric embeddings, the efficacy of chunking for transfer, the importance of critical-like dynamics, etc.), but no single prior model has integrated them in the service of creativity. Thus Manny stands at the frontier: its approach is novel (geometry-as-cognition for creativity) while drawing strength from proven ideas.

Model/Architecture	**Representation & Learning**	**Adaptivity & Plasticity**	**Creative Mechanisms**	**Explainability**
Manny Manifolds	Dynamic graph manifold (nodes=concepts, edges with curvature); online learning via curvature updates (valence-modulated). Uses LLM as auxiliary "lens".	High: Continuously updates with new info; motifs form new "chunks"; can tune plasticity (η) and consolidation (τ) on the fly.	Lateral thinking via structure: Explicit analogy transfer (motifs reuse) ; explores novel paths if valence incentivizes; concept blending possible by linking subgraphs. Bicameral loop (experiencer/executive) fosters divergent/convergent interplay.	High: Knowledge state is a graph that can be visualized; reasoning = traversable path. Can answer "why" via path trace . Motifs and curvature are human-interpretable (as strength/confidence of associations).
Large Lang. Models (GPT)	Implicit knowledge in high-dim parameters; trained on text, uses context window for prompt adaptation (no true long-term memory update at runtime).	Low (static): Cannot update weights in conversation; adapts only via prompt (few-shot) or fine-tuning offline. No persistent self-modification in deployment.	Pattern-based creativity: Generates novel text by recombining patterns from training data. Can do analogies encoded in language , and score high on divergent thinking tests (A); but relies on learned correlations, not active imagination or structural reasoning. Often lacks genuine analogy mapping without explicit prompt guidance.	Low: Hidden neural activations not easily interpretable. Explanations of its outputs are post-hoc. No explicit representation of "why" an association was made (the knowledge is subsymbolic).
Hebbian/Hopfield Networks	Associative neural network (pattern completion memory). Knowledge as distributed	Medium: Adapts weights with each new pattern, but capacity is limited (catastrophic	Association and completion: Good at recalling stored patterns or blending noisy inputs (can do rudimentary completion =	Low: Hard to interpret why a given attractor was chosen apart from mathematical energy minima.

Model/Architecture	Representation & Learning	Adaptivity & Plasticity	Creative Mechanisms	Explainability
	embeddings or weight matrix. Learns via Hebb rule ("wire together") or STDP.	interference if overload). No structural addition of new nodes, just adjustment of connection strengths.	convergent thinking). Divergence is limited – tends to settle into stored attractors (could model habitual thought). Not designed for analogies or generating distinct new patterns beyond interpolation of learned ones.	No explicit symbols or explanation chain. (One can visualize weight matrix or energy landscape, but not user-friendly).
Active-Inference Agents	Probabilistic generative model of world (often Bayesian graphical model or dynamic model). Learns by updating belief distributions to minimize prediction error (Free Energy).	High: Continuously updates beliefs with new observations; can plan by minimizing expected surprise. Structure of model can expand if designed for growth (not typical in basic implementations).	Creative potential via planning: Can imagine counterfactuals by simulating model forward. Theoretically could find creative solutions by selecting actions that lead to novel but goal-satisfying states (exploration is driven by epistemic value = info gain). Creativity emerges if model complexity is sufficient to represent novel hypotheses. Some propose art through active inference (no widely validated examples yet).	Medium: Model's structure (if small) can be interpretable (e.g., a Bayesian network). However, in practice many use neural networks internally, obscuring interpretability. The rationale for an action is "it reduced expected free energy," which is abstract to a layperson.
Graph Neural Nets (GNNs)	Graph-structured neural nets processing input graphs (e.g., social network, molecules). Weights	Low/Medium: Adapts weights during training, but once trained on one type of graph task, not dynamically changing its own structure.	Relational reasoning: Excellent at finding patterns in graph data (e.g. can predict new links, classify nodes). If tasked for analogy, could learn to embed analogical structures and	Medium: GNN decisions can be partly explained by examining which neighbor nodes influenced a prediction (attention weights, etc.).

Model/Architecture	Representation & Learning	Adaptivity & Plasticity	Creative Mechanisms	Explainability
	learned via gradient descent; typically fixed architecture aside from input graph structure.	(Some advanced GNN variants can add nodes or do continual learning, but not common).	answer queries. However, GNNs don't create new nodes or edges on their own during inference – they only process given graphs. So, not generative or exploratory in the creativity sense unless combined with another mechanism.	But the internal embedding dimensions are not semantic. Compared to Manny, a GNN can't easily produce a human-readable "reasoning path" – it does reasoning in latent vector space.
Cognitive Architectures (Soar, etc.)	Symbolic rule and memory system. Knowledge in declarative form (working memory) and procedural rules. Learning via chunking (Soar) or other mechanisms in ACT-R.	Medium: They learn new rules or strengthen rule utilities. They have long-term memory stores that accumulate knowledge. Adapting to novel domains often requires engineering new rules or representations (not fully autonomous structural growth).	Rule-based creativity: Have been used to model insight problem solving. Soar's chunking allows transfer of learned solutions to new problems automatically (A). However, purely symbolic systems can struggle with fluid analogy unless explicitly encoded. Some (like Copycat program by Hofstadter) were designed for analogical creativity, but operate in narrow domains (letters/strings).	High: The state and rules are human-readable (if properly annotated). Trace logs can explain each decision. These systems were valued for explainability, although the flipside is brittle knowledge (lack of the neural flexibility Manny aims to have).

Table 1: Comparison of Manny Manifolds with representative models (LLM, Hebbian network, active-inference agent, GNN, classical cognitive architecture) in terms of knowledge representation, adaptivity, creative capacity, and explainability. Evidence grades: (A) corresponds to established peer-reviewed support, (B) to preliminary or tool-based evidence, (C) to theoretical reasoning in context.

Predictions: When and How Might Creativity Emerge in Manny?

Building on both theory and the above parallels, we can articulate specific predictions and hypotheses for creative dynamics in Manny Manifolds. These are testable claims about how certain parameter settings or design features will manifest as creative behavior, and what objective signs to look for.

- Exploratory vs. Integrative Regimes (Role of τ and Curvature Variance): Manny has a parameter τ that governs consolidation (a kind of "cooling" or forgetting rate during its /sleep cycle) . A prediction is that lower τ (less consolidation, or a higher tolerance for network entropy) will yield more exploratory creativity, whereas higher τ (aggressively pruning and stabilizing the network) will yield more integrative, convergent thinking. At low τ, Manny retains even weak and wild associations – the manifold stays richly interconnected (high entropy). We expect to see higher fluency and originality in this regime: Manny will generate many ideas by traversing those unpruned, meandering paths. However, coherence might suffer (some ideas will be nonsense, as many spurious edges haven't been trimmed). At high τ, Manny "crystallizes" knowledge regularly – many edges get flattened, leaving only strong, proven pathways. In this regime, flexibility might drop (fewer distinct avenues to explore), but the ideas it does produce will be well-formed and reliable (high coherence, perhaps high quality on convergent tasks like question answering). This trade-off mirrors the exploration-exploitation balance in creativity research. The edge-of-chaos hypothesis would suggest that there is an intermediate τ (and perhaps an intermediate η, the learning rate, as well) where Manny maximizes creative output – enough chaos to be novel, enough order to be meaningful. One could measure curvature variance (the variance or entropy of edge weight distribution) under different τ settings and correlate that with creativity metrics. Prediction: Manny will show peak creative metrics (e.g. best fluency-originality product) at intermediate curvature variance – neither a very uniform network (which might indicate a diffused, but undirected state) nor a too skewed network (dominated by a few attractors), but something akin to a "small-world" or scale-free regime where clusters exist but are richly interlinked. This is consistent with observations that creative individuals often show both specialization and broad interests – a structured knowledge with some randomness. It also aligns with network neuroscience findings that brains at criticality can rapidly reconfigure , presumably supporting the switching between ideas (A).
- Valence Weighting and Motivation for Novelty vs. Utility: By adjusting how valence is assigned, we can tip Manny towards exploratory creativity or safe productivity. If Manny's valence function is set to strongly favor novelty (e.g., give an intrinsic positive valence boost to edges that have low usage count or connect distant domains), then Manny should exhibit more "divergent" behavior: seeking out rare connections, favoring less visited parts of the manifold. We predict this will increase metrics like idea originality and the number of domain jumps in a solution (e.g., a higher degree of separation between concepts used in an answer and the question concepts). However, if valence only rewards novelty, Manny might lose coherence – it may assemble answers that are new but don't actually solve the problem (akin to a person brainstorming wildly without filtering). On the other hand, if valence is tied to

success/accuracy only (e.g., it only goes up when Manny's output is verified correct or user-approved), Manny will behave more like a conventional learner, exploiting what it knows works and refining those paths. That could lead to fixation (low originality, high efficiency on familiar tasks, difficulty with novel tasks). The ideal scenario is a dual-valence or adjustable valence: a combination of utility valence (did this answer/work? reward if yes) and novelty valence (was this something new tried? reward that as well, perhaps to a lesser degree). If Manny has such a blend, we hypothesize it can achieve creative resilience: continuing to try new angles when facing a challenge, rather than looping on the same failed approach. Concretely, one could experiment with a "creativity knob" that increases the weight of novelty valence relative to task-success valence and observe Manny's problem-solving. Prediction: Manny's solutions will become less conventional (more surprising to an observer) as novelty valence weighting is increased, up to a point where beyond that the success rate drops. This gives a quantitative handle: we expect a non-linear curve where moderate novelty bias yields the best balance (the system finds clever solutions), whereas extreme novelty bias yields a lot of unusual attempts but few correct or useful answers. This would mimic human brainstorming followed by focusing – Manny might need a mechanism to switch modes (high novelty search, then high utility filtering).

- Motifs and Lenses – Conceptual Blending and Transfer: Manny's motifs and lenses are posited to enable analogical reasoning. We anticipate observable signs of this. For motifs, one expectation is that as Manny learns more tasks, it will accumulate a library of motifs, and we'll see speed-ups in learning new tasks that share structural similarity to past ones (transfer learning effect). In a creativity context, those motifs also allow conceptual blending: combining two motifs along shared nodes or patterns. Prediction: If Manny is really leveraging motifs for creativity, then when given a creative task that encourages analogy (e.g. "invent a solution by combining features of X and Y"), we should see Manny retrieving two or more stored motifs and merging them to form its solution path. We could instrument the system to detect when a generated reasoning path intersects with two known motifs – that would be strong evidence of analogical composition at work (evidence level: to be gathered from Manny's logs). Regarding lenses, these provide Manny with the ability to re-interpret or map one domain to another (e.g., treat a cooking recipe as if it were a chemistry experiment via an analogy lens). We predict that using a lens will increase Manny's ability to solve cross-domain problems. For example, if asked a riddle that requires seeing a scientific scenario in metaphorical terms of nature, Manny might fail without the appropriate lens. But if we then apply a "nature analogy" lens (mapping scientific concepts to nature analogs), Manny's manifold distances and curvatures adjust in that projection, potentially making the solution obvious. Prediction: Manny with an appropriate lens will solve analogical problems significantly more often or more quickly than Manny without lenses, even though the internal knowledge is the same – it's the perspective shift that matters. This can be tested by constructing paired tasks (one that's an analogy of another) and seeing if solving one helps the other only when a lens or alignment is used. We also expect that as Manny gains experience, it might learn new lenses (perhaps by extracting correspondences between domains where it has solved analogies). This would be an emergent form of creative generalization: discovering that, say, a social network and an electrical circuit share a graph structure lens. Such self-discovered lenses, if any, would be very interesting qualitative evidence of emergent analogical thinking.

- Objective Creativity Metrics for Manny: We propose to explicitly track a set of metrics during Manny's operation to quantify creative behavior. For each reasoning session or output, measure: (1) Ideational Fluency – number of distinct solution attempts Manny generates (e.g., paths explored) before convergence or giving up; (2) Flexibility – the diversity of knowledge domains used (we can tag nodes by domain and count how many domains appear in a solution or across solutions); (3) Originality – we can define this as an inverse function of the product of edge usages in the path (i.e., a path that goes through very commonly used edges is low originality, one that goes through rarely used or newly created edges is high originality). Also, comparing Manny's solution to a reference set (if humans have typical solutions, is Manny's different?). (4) Coherence/Valence – average valence of edges in the final solution or a binary success metric if the solution is correct/valid. We predict certain correlations: high originality solutions might inversely correlate with coherence (as per the exploration/exploitation tension). But the most creative outputs should achieve high originality while maintaining coherence – those are the "aha!" insights that are both novel and fitting. Manny's goal should be to maximize that combined score. Through experiments, we might find, for example, an optimal range of η (learning rate) that yields the best combined score. Too low η, Manny hardly learns new connections (stuck in mediocre familiar ideas); too high η, Manny wildly reshapes the manifold on each experience (ideas lack stability to be refined). The right η would allow "strong ideas to get stronger and weak-but-promising ideas to survive long enough to combine with others". This notion is parallel to human brainstorming followed by iterative improvement.

In summary, the emergent hypothesis is that Manny will be most creatively capable at the "edge of chaos" – with intermediate plasticity and moderate pruning, guided by a balanced valence that rewards novelty and usefulness. In that regime we expect to see the hallmark of lateral thinking: solutions that are non-obvious (distantly associative) yet effective. These predictions are directly falsifiable: if, for instance, increasing novelty valence never produces better problem-solving performance or more original outputs, then our assumption that Manny's architecture can manifest divergent thinking would be challenged. Or if Manny never uses motifs from one domain in another, then the motif mechanism might be too brittle or too domain-specific as implemented, failing to act as the cross-domain analogy device it was intended to be.

Experimental Paradigms for Validation

To empirically validate creativity in Manny Manifolds, we propose a series of experiments and benchmark tasks, drawing from both human creativity tests and AI evaluations, as well as custom scenarios to probe Manny's unique features:

1. Analogical Transfer Task: Evaluate Manny's ability to apply learned knowledge to analogous situations. For instance, train Manny (or let it learn) a procedure in one

domain (e.g., a simple story of "A rescues B from C using tool D"), then test Manny with a query in a very different domain that has the same underlying structure ("X saves Y from Z using gadget W"). We measure whether Manny finds the correct analogy (maps X→A, Y→B, etc.) and solve it. This can be quantified by the success rate of correct mappings and the time (steps) to solution. A baseline could be a non-geometric memory (like a standard QA system or a knowledge graph with no curvature/adaptation). We expect Manny to outperform static baselines especially after minimal examples, due to its ability to generalize via motif. An extension of this is presenting Manny with classic analogy puzzles from IQ tests (word analogies, geometric analogies) to see if its internal representation naturally encodes the needed relations after some exposure.

2. Concept Combination (Blending) Tests: Present Manny with two unrelated concepts and prompt it to generate a concept or idea that combines them. For example: "Here is 'chair' and here is 'ocean'. Come up with an idea that combines these." This is akin to the creativity test of combining random words. Manny would ideally traverse out from "chair" and "ocean" and find a region where disparate attributes can meet – perhaps creating a node (if allowed) representing something like "floating chair for the ocean" or a metaphor "the ocean as a chair that supports life". Human judges would rate Manny's outputs on creativity metrics (originality, surprisingness, meaningfulness). A non-creative baseline (like a database search or a WordNet-based combination which might just concatenate attributes) would likely produce trivial combinations ("blue chair" if ocean→blue), whereas Manny, if effective, might produce something more abstract or functional. This test directly examines Manny's capability for conceptual blending. Success would be anecdotal but powerful evidence of creativity: if Manny invents something novel that is still interpretable, it would show the manifold's ability to yield emergent concepts beyond its initial nodes.
3. Alternate Uses Divergent Thinking Task: As mentioned, ask Manny to generate as many uses or associations as possible for a given item or concept (e.g. "brick"). Here we measure quantity and variety. We would run Manny in a mode where it can output a list of answers (perhaps by doing multiple threads from the "brick" node outward, each time avoiding already listed ideas). We then score fluency (# of uses), flexibility (# of distinct categories those uses fall into – which we can assess by checking how far apart the answers' nodes are in the manifold's domain graph), and originality (perhaps using a reference: if Manny's knowledge includes common uses, those are low originality, whereas if it comes up with something not in its knowledge base until created via combination, that's high). We compare Manny's performance to human results on the same prompt (if available) and to an LLM's results. We anticipate that with sufficient knowledge learned, Manny will list a comparably rich set of ideas, but importantly, we can examine how it came up with them: e.g., did it systematically traverse each neighbor of "brick" in the manifold (material->paperweight, building->house, weapon->throwing, etc.), indicating a breadth-first search of associations? Or did it do deeper, multi-hop explorations (brick->shape->rectangle->domino effect -> use bricks as dominoes in an art installation)? The pattern will tell us how constrained or free its divergent thinking is. If Manny truly supports lateral expansion, we'd see multi-hop distant uses, not just immediate obvious ones. This task would also reveal if Manny falls into functional fixedness (only seeing brick as building material) or if it breaks out (which creative thinkers do). We could manipulate valence in this task: instruct Manny that "novel ideas are highly rewarded" and see if that yields more unusual uses than the default.

4. Edge-of-Chaos Parameter Sweep: We design an experiment to vary Manny's key parameters systematically and observe creativity metrics. For example, choose a set of creative challenges (like those above or even coding puzzles, riddles) and run Manny under various settings: η (learning rate) low vs high, τ (consolidation) low vs high, valence bias for novelty low vs high. In each condition, measure outcomes: success rate on tasks, and creativity metrics of solutions. Plotting these will test our earlier prediction of an optimal middle ground. We expect to see something like: as η increases from 0 upwards, initially Manny's flexibility and novelty increase (it's learning more from each experience), but beyond a point, performance might degrade or become erratic (network oscillations, knowledge not settling – Manny might contradict itself or forget earlier learning due to instability). Similarly for τ: as we decrease τ (meaning less consolidation), we expect a rise in idea variety generated, but if τ is too low (no pruning at all), the manifold might become noisy and filled with spurious edges, possibly reducing the quality of reasoning (the system might start making connections that lead nowhere productive). We can identify the "edge of chaos" by looking at when Manny's solution originality is high yet solution validity remains acceptable. That combination could be plotted as a creativity score. If the theory holds, the graph of creativity score vs. parameter will have a peak (an inverted U shape). If instead it's monotonic (e.g., more novelty valence always just decreases performance without any sweet spot), that would suggest perhaps our model of how these parameters affect creativity is wrong or too simplistic.
5. Baseline Comparisons: For each type of task, we include baseline models for context. For analogies, perhaps a fine-tuned GPT-4 on analogies, or a simple embedding-based analogical solver; for divergent thinking, GPT-4 as mentioned; for concept blending, maybe ConceptNet-based heuristics. The goal is to see if Manny's unique features actually give it an edge. One possible outcome: Manny may not match an LLM in sheer knowledge or fluency (since LLM has seen far more data), but Manny might generate more explainable and conceptually grounded creative answers. For example, GPT-4 might list 20 uses for a brick (many of which it likely saw in training data or variations thereof), whereas Manny, with much less data but an adaptive graph, might come up with 10 uses – some overlapping common ones, but maybe one or two truly odd ones that GPT-4 didn't list, because Manny connected "brick" to a part of its knowledge a language model wouldn't normally associate (due to lack of textual co-occurrence). If those odd ones are also useful (coherent), that's a big win: it means Manny discovered a niche idea through its structure (for instance, Manny might have learned about "heat retention" in a domain and connect brick to that to suggest using a brick as a thermal battery – a creative use that requires multi-hop reasoning).
6. Self-Reflection and Iterative Creativity: Another experiment could tap into Manny's bicameral loop more explicitly. We can have Manny attempt a task, then use its executive subsystem to "critique" or analyze its own result (perhaps via the /why command or by checking path valences), and then have the experiencer try again based on that feedback. This mimicry of human creative process (ideate, evaluate, refine) can be tested by comparing one-shot outputs vs. iterative outputs. We predict that allowing Manny a self-reflection iteration will improve the creativity metrics – e.g., it might initially give a very wild answer, then its executive notes "this part is incoherent", and in a second try Manny finds a compromise that's still novel but more coherent. There's evidence from human creativity that alternating between generative and evaluative modes yields the best results (A, psychology of creativity). Manny's architecture was explicitly designed to enable that (with the executive being able to regulate the experiencer). We'd like to see that materialize in practice.

Each of these experiments will produce data we can analyze. Particularly, we'll gather quantitative evidence like correlation between Manny's internal measures (curvature changes, motif usage count, valence patterns) and the creativity outcomes. For instance, does the number of motifs reused correlate with solution flexibility? Does high average path valence correlate with judges' ratings of answer quality (it should, if valence is doing its job as a confidence metric)? By tying the observed creative success back to Manny's mechanics, we not only validate creativity in this system but also contribute insights to the broader question of how creativity can be quantified and recognized in machines.

Novelty and Feasibility Assessment

Is a geometric manifold architecture uniquely capable of supporting lateral creative processes? After this deep analysis, we can venture an assessment: Manny Manifolds offers several unique advantages for modeling creativity, as well as some challenges and open questions regarding its novelty and feasibility.

On the positive side, Manny's design embodies many principles that creativity research has identified as important:

- Structured yet flexible knowledge: Manny's manifold is not rigid: it can grow (new nodes/edges), re-weight connections, and even reshape via consolidation. This addresses the knowledge inflexibility seen in many AI models. Unlike an LLM frozen in its training, Manny can truly incorporate a new idea and its entire topology shifts accordingly. This continual learning is crucial for creativity – creative agents learn from each attempt, often radically changing strategy, something static models struggle with. Manny does this by default (every interaction curves the space). The feasibility of continual learning without catastrophic forgetting is a challenge, but Manny's use of consolidation and bounded curvature suggests it can manage stability–plasticity balance (supported by testing so far in its development, per project documentation).
- Explainability and introspection: Creativity in AI is often criticized because we don't see the "thinking process." Manny inherently logs its reasoning path. This means any creative leap can be examined post-hoc: we can see which concepts were connected and even identify a missing link ("I jumped from A to C via a new edge because no existing path sufficed"). This is scientifically valuable: if Manny generates a creative solution, we can trace its genesis – something rarely possible in subsymbolic networks. It also means Manny's creativity is falsifiable in a Popperian sense: if it makes a strange connection, we can inspect whether that was due to a bug, a single erroneously high valence, or truly a clever multi-step analogy. This transparency could make Manny a unique tool for cognitive science, to simulate how ideas form and spread in a network, and to compare with human data (like spreading activation in semantic priming experiments).

- Integration of symbolic and subsymbolic: Manny uses an LLM as a lens (subsymbolic pattern matcher) but keeps the main knowledge base symbolic (graph). This hybrid may offer the best of both worlds: the LLM provides a rich associative soup to draw candidate links (like a creative spark), while the graph provides a disciplined structure to test and reinforce those links. This is a feasible approach given current AI – LLMs are readily available and Manny uses them only as helpers, so it's cost-effective and safe (not ceding full control to the LLM avoids the pitfalls of uninterpretable leaps or toxic outputs). The novelty here is using the LLM not as the generator of final answers, but as a generator of new edges and perspectives (a "suggestive muse" rather than an oracle). If it works, this pattern could be applied broadly: any AI system could use a language model to brainstorm internally while keeping a structured world model to verify and implement ideas – somewhat akin to how humans use intuitive thought vs. analytical thought.

However, there are open questions and potential limitations:

- Scaling and Complexity: Manny is currently conceptual – the real implementation might face scalability issues as the knowledge manifold grows. Creativity often comes from vast associative richness. Will Manny's manifold become too large or slow to traverse if it learns hundreds of thousands of concepts? There's a question of whether geometry-based heuristics (like using curvature to guide searches) scale well. If each creative query requires exploring many paths, computational cost could explode. LLMs have the advantage that their "associations" are implicit and massively parallel in the network weights, allowing them to produce an answer in one forward pass which already blends a lot of knowledge. Manny might have to explicitly traverse combinations, which is exponential in worst case. We need to find out if motif caching, heuristic search, and perhaps parallel thread exploration can make Manny efficient enough. Feasibility will depend on clever engineering (perhaps pruning the search space by valence so it doesn't try obviously low-valence paths). The feasibility is promising in moderate domains (the developers have metrics showing near-linear scaling in active region size), but for human-level creativity (which draws on very large knowledge), further optimizations or hierarchies might be needed.
- Evaluating Creativity Objectively: Even if Manny behaves in ways we consider creative, there's an epistemological question: how do we know it's not just following its programming? Critics might say: "Manny isn't really creative, it's just rearranging what it learned." Of course, one could say the same for humans – we recombine knowledge too. To address this, we proposed metrics and comparisons. If Manny consistently outperforms baseline systems on tasks intended to require creativity (and does so with novel solutions, not just faster), that is evidence that its architecture adds something. One falsifiable prediction: Manny will solve certain analogies or insight problems that a pure neural model like GPT-4 fails, despite GPT-4 having more knowledge. If that happens, it strongly indicates Manny's representational bias (geometry + plasticity) confers a real creative advantage, not just a different style. If it doesn't happen – if GPT-4 still solves everything better – then Manny might not yet have a "niche" where it's uniquely powerful, although it could still be useful for interpretability reasons.
- Novelty vs. Randomness: A concern is whether Manny's creative outputs will genuinely be meaningful novel combinations or if there's a risk of producing

incoherent nonsense when pushed for creativity (a common failure mode in naive combinatorial creativity systems). The design tries to mitigate this via valence (which should act like a coherence filter) and the executive oversight. But we should scrutinize results: we might find that Manny sometimes forms edges that are artifacts (e.g., if two unrelated concepts happen to share a neighbor, Manny might create a motif that actually doesn't make sense, a kind of false analogy). An example: Manny learns "apple is round" and "earth is round" and thus links apple↔earth in a motif, then suggests an apple can be orbited like a planet – a playful but scientifically silly idea. Is that creative or just a misunderstanding? Human creativity can be zany, but we usually require a certain logic to the nonsense (surrealist art still has internal consistency). Manny's outputs will need human evaluation to judge this. One could imagine adding constraints to Manny (like consistency checks via an ontology) to prevent completely absurd blends, but doing so might also limit the freedom that yields creativity. This tension will be an ongoing theme in development: how to keep Manny's creativity "within bounds" of usefulness without squashing the very exploration that produces original ideas.

- Uniqueness of the Geometry Approach: Could a simpler approach achieve the same thing? For instance, one might ask: what if we just took a normal knowledge graph and ran a Monte Carlo random walk to generate ideas – would that not yield lateral thinking too? Possibly, but the curvature and valence in Manny imbue the graph with a form of memory and preference that a static graph lacks. Manny remembers paths taken (via curvature) and learns which are promising (via valence). A random walk on a static graph will keep rediscovering the same obvious connections unless artificially forced to jump. Manny, by contrast, gradually warps the graph so that newly promising connections become shorter (easier to find) and unhelpful ones effectively lengthen (harder to stumble into). This self-modifying aspect is indeed unique and testable. Over time, Manny's manifold should become an increasingly tailored "creativity space" for its domain – like an assistant that has learned how to make unusual links that turned out useful in the past. Other systems would have to be explicitly reprogrammed or given new data to do that. Manny does it on the fly. This could be its signature strength if realized: adaptive creativity – improving its lateral thinking with experience, not just its task performance. It's an open question if this will really materialize; it might require a lot of experience for patterns to emerge. But if, say, after solving 50 problems Manny starts solving the 51st in a more inventive way than it solved the 1st, qualitatively, that suggests a creativity learning curve is present.

Novelty of Manny's Approach: In summary, Manny Manifolds' approach to creativity – treating "thought as movement through a curved space" – is a novel paradigm in AI. It leverages emergent geometry in a way no mainstream model does. If successful, it could illuminate how understanding and creativity are two sides of the same coin: as the documentation eloquently put, "motion through the manifold equals learning, and learned curvature equals understanding" . We would add: creative insight equals finding a new low-energy path in that manifold that wasn't there before. Manny is built to literally create those paths. The feasibility of reaching human-level creativity is uncertain – human knowledge and intuition are immense – but Manny can start in narrower domains and scale up. Each success (even recreating a known creative solution independently) will bolster the case that geometric cognitive modeling is a fruitful route for AI. And beyond AI, it serves as a computational hypothesis for cognitive science: maybe our own brains form ideas by a similar mechanism of a self-curving conceptual space (the "brain on the edge of chaos" idea combined with conceptual spaces theory). Manny could thus be a testbed for theories of insight (e.g., the sudden restructuring of a problem – in Manny terms, a sudden curvature change connecting two regions that solves the problem).

Open Questions and Future Work: Some open questions remain that are worth stating explicitly (each of these is a falsifiable point for research):

- Will Manny spontaneously exhibit "aha" moments? (E.g., a rapid drop in surprise when it finally connects two strands of knowledge.) We can monitor its energy landscape; a true insight would be indicated by Manny quickly finding a much shorter path to goal than any tried before, corresponding to a sharp curvature update on some key edge (like a new bridge). If we never see such patterns distinct from gradual learning, then maybe Manny's process is more incremental than insightful.
- Can Manny handle multi-modal creativity? Right now, it's mostly textual/semantic. Human creativity often mixes visual, spatial, etc. Manny's architecture could in theory integrate other modalities as sub-manifolds connected via lenses (one could imagine a visual manifold and a verbal manifold linked by learned correspondences). Would that allow, say, creative visual analogies (seeing a solution to a mechanical problem because Manny's visual manifold recalls a similar shape)? This is future work, but feasible given Manny's modular design for domains.
- How to evaluate value of creativity in Manny's context? In practical terms, if Manny is used as a collaborative AI, we'd want to ensure its creative suggestions are not just novel but actually helpful or aligned with human needs (e.g., ethical, safe). Manny's explainability helps here, as a human can follow why it made a leap and intervene if it's going astray. Still, implementing a form of "sanity check" or ethical valence is an open task. A creative AI can also generate harmful or nonsensical ideas if not guided – we'll likely need to integrate constraints or moral circuits (perhaps via the valence system imposing negative valence on clearly bad outcomes) to keep creativity beneficial.

In conclusion, Manny Manifolds stands as a novel hypothesis: that an AI can be creative by design through a self-curving knowledge space, rather than as an accidental byproduct of massive data training. The evidence surveyed (A: neuroscience and cognitive studies, B: computational experiments, C: theoretical analogies) generally supports the components of this hypothesis – from the need for dual-process dynamics to the utility of adaptive network topology and the effectiveness of reusable structural patterns . The ultimate proof will come from implementing Manny and observing emergent creative behavior. If the predictions hold true, Manny could not only replicate known creative mechanisms but also shed light on new ones, offering a unique bridge between human creativity and machine cognition. If some predictions fail, that will equally inform us – perhaps indicating which aspects of human creativity require additional ingredients not yet in Manny (e.g., emotional depth or conscious deliberation). Either way, Manny Manifolds provides a fertile ground for experimenting with creativity as a computational phenomenon, with geometry as the canvas on which ideas form, connect, and evolve. The journey to test these ideas empirically is just beginning, and the coming experiments will tell us how close this geometric mind can come to thinking outside the box.

References: (Evidence grade in parentheses)

- Beaty, R.E. et al. (2015). “Creativity and the Default Network” – NeuroImage, 2015. (A)
- Constant, A., Friston, K., & Clark, A. (2024). “Cultivating creativity: predictive brains and the enlightened room problem.” Phil. Trans. Royal Soc. B, 379(1895):20220415. (A)
- Ruan, H. et al. (2014). “Dopamine-enabled Hebbian plasticity in prefrontal circuitry.” Front. Neural Circuits 8:38. (A)
- OpenCog Project (2009). “Economic Attention Networks: Associative Memory and Resource Allocation for General Intelligence.” (B)
- Laird, J., Newell, A., Rosenbloom, P. (1987). “Soar: An architecture for general intelligence (Chunking mechanism).” Artificial Intelligence. (A)
- Nickel, M. & Kiela, D. (2017). “Poincaré Embeddings for Learning Hierarchical Representations.” NeurIPS. (A)
- Choudhary, K. et al. (2025). “Local-Curvature-Aware Knowledge Graph Embedding via Ricci Flow (RicciKGE).” arXiv:2510.12345. (B)
- Mednick, S.A. (1962). “The associative basis of the creative process.” Psychological Review, 69(3), 220-232. (C)
- Lynn, C.W. et al. (2024). “Heavy-tailed neuronal connectivity arises from Hebbian self-organization.” bioRxiv preprint. (A)
- Cambridge Univ. (2009). “The human brain is on the edge of chaos” (news on Plos Comp Bio study by Kitzbichler et al.). (A)
- Manny Manifolds Project Docs (2025). System Design and Vision (G. Ribeiro). (C) (Internal design reference)

(Additional references included inline above as per citation marks.)