

Manny's Unified Knowledge Manifold: Principles and Architecture

Core Principle: One Substrate, One Learning Physics

Manny is designed around a **single unified substrate** for all knowledge and experience. This means that **all data modalities (text, images, audio, etc.) are represented within one manifold** governed by one set of learning dynamics. Manny does not treat non-text data as external attachments or separate files; instead, every piece of information is integrated into the manifold in a form that can participate in Manny's cognitive processes. The guiding invariant is that **everything Manny knows must be decomposable into native primitives that obey the same "learning physics"** (the same rules of traversal, curvature formation, and updating). If some data cannot be traversed, reused, or affected by Manny's learning dynamics, then it does not truly reside in Manny's knowledge base. In practice, **files are a UI-level illusion** – internally Manny only has nodes, relationships, fields, and curvature in the manifold. This ensures that *all* knowledge (regardless of modality) influences and is influenced by the same underlying physics of learning.

Why one substrate? With a single integrated manifold, Manny avoids special-case reasoning for different data types. Other systems (e.g. OpenCog Hyperon's AtomSpace) can handle multiple data types in one graph with weighted links [1](#) [2](#), but Manny goes further: it enforces that **one unified cognitive process** operates over that graph. There are no separate "image modules" or "text modules" with their own logic – just one continuous space of representations. This unification is critical for generalization and cross-modal reasoning, allowing Manny to treat an image, a sentence, or a sound as different *regions* of the same knowledge space rather than incompatible formats.

Knowledge Representation: Primitives and Emergent Structure

Primitives in Manny are the smallest units of representation that are currently meaningful and computationally useful at a given layer of abstraction. Importantly, primitives are *not* fixed atomic symbols in an absolute sense – they are **"provisionally atomic"** building blocks that Manny can further decompose later if needed. At the start (for simplicity and efficiency), Manny uses high-level primitives such as words to represent concepts. Words are convenient initial primitives because they are human-defined compressions of meaning – each word packs a lot of semantic context. For now, treating words as atomic nodes works well to bootstrap Manny's conceptual layer. However, **Manny's design does not assume words are indivisible**: as the system evolves, a "word" node can be expanded into sub-nodes (morphemes, phonetic patterns, contextual senses, etc.) if finer granularity is needed. This approach of **provisional atomicity** means that what is considered a primitive today can be broken down tomorrow without breaking the higher-level structures. Primitives are defined by their utility at the current stage, not by an eternal definition.

Emergent structure arises when primitives connect and repeat in the manifold. Manny's knowledge is not explicitly stored as rigid data structures, but rather **emerges as stable patterns (attractors) in the manifold**. When certain combinations of primitives co-occur frequently and consistently, Manny forms higher-level nodes or **motifs** representing that pattern. In other words, meaning is captured as **clusters of primitives that have developed significant curvature (strength) through repeated usage**. For

example, consider an image of a man walking a dog: - At the lowest level, Manny could represent the image via many tiny primitives (e.g. pixel patches or edge fragments, each primitive node having attributes like color or edge orientation and linked to neighboring primitives by spatial adjacency). - Through experience, *higher-level clusters* form: edge primitives combine into shapes (e.g. a four-legged shape, a human outline). These mid-level patterns might correspond to parts of objects (dog's head, human torso). - At an even higher level, a stable motif emerges for the whole object: a **"dog" visual motif** and a **"man" visual motif** may form, because those collections of lower features recur across many images. Manny links these visual motifs to the existing concept nodes for "dog" and "man" (learned from text and other data), unifying visual and semantic knowledge. - A relational motif might also emerge (e.g. **"person-walking-dog"**) capturing the relationship seen in the image.

Crucially, **the original image file is not Manny's knowledge**. Manny can retain a pointer to the raw image (for example, to replay or re-inspect it in detail if needed), but the knowledge Manny gains from the image lies in those manifold structures – the network of primitives and motifs that the image activates. This is analogous to human memory: we might remember seeing a photograph, and we recall the concepts (man, dog, walking, park) and their relations, but we don't retain every pixel in our minds. The raw sensory data is available to re-examine (like looking at the photo again), but our *understanding* of the photo is stored in a highly processed, interconnected form.

By ensuring that even non-text data is represented via primitives and relationships, **Manny guarantees that all knowledge participates in manifold dynamics**. A visual concept like "dog" in an image and the textual concept "dog" from a story will converge in the manifold and reinforce each other, rather than living in separate silos. **No piece of knowledge in Manny is an opaque blob**; it's always made of manifold primitives that can connect, curve, and be traversed.

Layered Abstraction and Progressive Compression

To manage the complexity of raw sensory information, Manny employs a **layered representation scaffold**. Each layer compresses and abstracts the layer below it, gradually distilling raw data into meaningful concepts, while all layers remain part of the single manifold. Importantly, the same core mechanics (traversal of connections, curvature adjustment through learning, decay of unused structures) apply at every layer – only the granularity and timescales differ.

Layers of the Manifold (from low-level to high-level):

- **Sensory Microstructure Layer:** This is the "near-sensor" layer handling very high-detail, high-frequency data with minimal interpretation. For vision, this could be keypoint features, small patches of pixels, edges or gradients; for audio, it might be short audio frame features or frequency bands. At this layer, representations are *very transient* – they have high entropy and decay quickly if not used. Manny does not typically store every raw pixel or waveform indefinitely; instead these feed into slightly more abstract units. The sensory layer is like a fast, short-term echo of the input, providing fine detail to higher layers on demand. (In future implementations, this layer could even be handled by specialized hardware or pre-processing modules, such as neuromorphic sensors, that convert raw signals into feature primitives before entering Manny's core manifold.)
- **Perceptual Primitives Layer:** In this layer, Manny holds slightly more processed features that have some persistence and reusability. For images, this could include small shapes, corners, texture patches or simple objects parts; for audio, phoneme-like sounds or recurring motifs; for text, individual tokens or syllables. These **perceptual primitives** are still modality-specific to an

extent, but they represent a significant compression of raw data. They capture structure (e.g. an edge, a color patch, a chord in music) that can be reused across different inputs. Nodes at this level might represent concepts like “vertical line segment” or “a fuzzy circular patch (could be an eye or wheel)” – pieces that mean little in isolation but are the building blocks of larger meanings. Relationships in this layer include spatial adjacency (for vision), temporal adjacency (for audio/time-series), and basic associations (which features frequently coincide).

- **Conceptual Semantic Layer:** This is the layer where traditional “concepts” live – objects, ideas, actions, categories – generally corresponding to words or phrases in language. By the time information is promoted to this layer, it has undergone compression and integration. For example, many visual primitives together have triggered the concept “**dog**” node to activate, or a sequence of sound primitives has activated the concept “**doorbell**”. Concepts at this layer are **modality-agnostic** attractors: the concept “dog” is now a single region in the manifold that can be activated by a picture of a dog, the sound of a bark, or the word “dog” in text. Relations at this layer become more semantic (dog *is-a* animal, dog *associated-with* loyalty, etc.), and they draw on structure distilled from lower layers. The conceptual layer is where cross-modal integration happens: e.g., the visual motif for dog links to the semantic dog concept, which also links to the word “dog,” to the sound “bark,” and to emotional impressions a person has of dogs.
- **Narrative and Episodic Layer:** Above individual concepts, Manny can form higher-level structures representing sequences and contexts – essentially narratives or episodes. This includes temporal chains of events (e.g. “**dog chases ball → dog catches ball → returns to owner**”), cause-effect linkages, and even personal episodic memories (“I saw a dog at the park yesterday”). These structures bind concepts into ordered, meaningful wholes with roles (agent, action, object, time). The narrative layer gives Manny an internal sense of experience and context – it knows not just static facts, but can encode **situations** and **stories**. Technically, this might be represented by motif-nodes that connect concept nodes in a sequence with temporal/causal links and perhaps an “observer” or “self” reference for personal memories. This layer evolves more slowly and decays more slowly – Manny “remembers” significant episodes (especially those with high emotional or importance valence) longer than isolated details.

Across all layers, **the same learning principle holds:** Manny only retains structure that has proven meaningful through repeated use or high value. Each layer receives input from the layer below, looks for **resonance vs. novelty**, and compresses information: - **Resonance:** If an incoming pattern aligns with existing structure (e.g. a set of features strongly matches the known concept “dog”), then Manny integrates it by reinforcing those existing nodes/edges. The path through the manifold for that pattern has low “energy” because the manifold is already curved in that shape; the input “falls” into an existing attractor. This reinforcement slightly deepens the attractor well (increasing the “curvature” there, making the concept even more robust). - **Novelty:** If an incoming pattern does not fit any existing structures (e.g. Manny sees a truly new pattern that doesn’t match any known concept), it generates a “surprise” or high-energy signal. Manny will then either **explore** – trying new connections, forming a tentative new node – or simply let it dissipate if it’s deemed unimportant. Novel patterns with high valence (importance) may form new structures (new primitives or new concepts), but Manny is cautious: it creates new nodes at the lowest layer necessary. It might keep a novel pattern as a combination of perceptual primitives for a while; only if it recurs or proves significant will Manny **promote** it to a stable concept.

Memory as Compression: Throughout these layers, memory is not a raw recording; it’s the **compressed result of many similar experiences**. Each time an experience resonates, Manny’s representation gets a bit more compressed (less noisy, more generalized). Over time, details that never mattered for any higher-level pattern simply decay away. What’s left is a lean representation capturing

the essence needed to reconstruct the experience if required. In essence, Manny remembers **what was stable and useful** in the long run and forgets the rest. This aligns with human memory: we retain recurring themes and important details, while countless sensory inputs from each day fade out.

Dynamic pruning and consolidation: Manny will periodically perform a kind of “maintenance” (analogous to a brain’s sleep consolidation). In these phases, it will: - Decay or remove the weakest connections and nodes (those which have low curvature because they were rarely used or found irrelevant). - Strengthen and possibly reify any motifs that have occurred frequently (for example, after seeing many instances of “man walking dog,” Manny might consolidate that pattern into a distinct relational motif node for efficiency). - Adjust the overall curvature distribution (for example, preventing any one region from becoming so “deep” it traps all thoughts, or normalizing learning rates across regions).

This ensures the manifold remains flexible, not overly cluttered with noise, and that important knowledge solidifies over time.

Starting simple and increasing fidelity: Practically, we will implement Manny in stages of fidelity. Initially, we focus on simpler, coarse representations that are tractable to implement (like words and basic relations for conceptual knowledge). We won’t start by feeding raw pixels into Manny’s core. Instead, we might use existing computer vision to extract mid-level features (e.g. object detections or image embeddings) as input primitives. This respects Manny’s philosophy – as long as those features are broken into Manny’s graph form (nodes and edges) and not treated as opaque vectors. As Manny’s capabilities and performance improve, we can gradually move the boundary downward (processing finer detail in the manifold). The idea is to **only increase sensory fidelity as needed**. If the current level of compression is yielding good results, there’s no need to add more detail. This controlled growth keeps the system efficient and understandable. It also mirrors biology: our sensory neurons heavily preprocess inputs before they ever reach cognitive centers, because it’s wasteful for the brain to handle every raw signal at full resolution.

Semantic Mass and Curvature: A Physical Analogy for Meaning

A powerful way to understand Manny’s approach to knowledge is through a **physics analogy**. Manny treats information and meaning in a manner analogous to how physics treats mass and gravity: - **Information is like mass:** Every piece of information (every co-occurrence, correlation, constraint across features) contributes a kind of “mass” in the knowledge space. When many pieces of information align (i.e. they point to the same concept or pattern), they accumulate mass in that region of the manifold. - **Meaning is curvature:** Mass warps spacetime in physics, and similarly, accumulated information “warps” Manny’s manifold, creating curvature. A strongly learned concept (one reinforced by many experiences) corresponds to a deep curvature well or a dense region in the manifold. This curvature influences the trajectories (thought processes) moving through the space.

In practical terms, **a concept is not stored as a symbol or an entry in a table, but as a stable geometric deformation in the manifold**. For example, the concept of “apple” exists in Manny as a region of curved space resulting from many inputs: - Visual inputs (many sightings of apples shape a visual sub-region: round, red/green, shiny texture). - Auditory/linguistic inputs (hearing or reading “apple” in various contexts adds associations: fruit, food, “Apple” as a company, biblical apple, etc.). - Experiential inputs (tasting an apple contributes to a taste dimension of the concept). - Functional/cultural inputs (knowing apples are used in pie, associated with teachers or health “an apple a day...”).

All these facets add up to a multi-dimensional “mass” concentrated in one area – the apple concept – which curves the manifold around it. **The similarity between concepts is then literally geometric**

proximity or alignment in this space. So “apple” and “pear” end up in nearby regions because they share many contributing dimensions (both are fruits, sweet, similar uses, similar shape), whereas “apple” and “stone” are very far apart (few common dimensions apart from maybe shape). In Manny, **analogy and similarity require no special rule or taxonomy – they emerge from the geometry**. A new input about apples (say a statement “Apples can be green or red”) will naturally activate the apple region and possibly spill into the pear region due to proximity, yielding an analogical insight (“pears also can be green or brown”).

Primitives as manifold deformations: Earlier we defined primitives as the basic units of representation. Now we can refine that: a primitive in Manny can be seen as an **elementary deformation or pressure on one or more fundamental dimensions** of the manifold. Instead of thinking of a primitive as a discrete token, think of it as a *force field element* that shapes the space locally. For instance, encountering the word “apple” applies a certain pattern of pressure across dimensions (visual imagery, taste, category, etc.), nudging the manifold in the direction of the apple concept basin. Each primitive – be it a word, a pixel feature, or a sound – has a signature across the manifold’s dimensions, contributing positive or negative curvature in certain directions. Meaning then is not stored in the primitives themselves but in the *resultant shape* formed by many primitives acting together. In this view, **symbols (like words) are simply labels we attach to these stable shapes** after they form. Manny doesn’t inherently know “apple” as text – it knows a region of the manifold that humans label “apple” once it’s stable and well-defined.

Threads as trajectories through curved space: Manny’s cognitive process (reasoning, recall, association) can be envisioned as **a particle traveling through a curved spacetime**, where curvature is created by semantic mass (learned knowledge). When a thought process (“thread”) is initiated – for example, by a question or a sensory cue – it’s like dropping a marble into the manifold at a certain point. The marble will roll along the curves, gravitating toward attractors (strong concepts that are relevant) and following paths of least resistance (established associations). In a well-shaped manifold, **a thread naturally finds relevant answers or ideas by “falling” into them**. This means Manny doesn’t have to brute-force search all possibilities or use complex logical inference rules; much of the work is done by the landscape itself. If the manifold has been trained such that, say, the concept “apple” is deeply connected to “fruit” and “food,” then a question like “Is an apple food?” is answered by the thread simply sliding from apple’s region down into the “food” region through the existing connection – effectively a geodesic path in the knowledge space.

This approach yields enormous cognitive synergy “for free” once the manifold is mature: - **Reasoning without explicit rules:** Deductive or inductive reasoning appears as stable pathways in the manifold. For example, if we know “apples are fruits” and “fruits are food,” the manifold will have a curved path connecting apple to food via fruit. To answer “Can you eat an apple?”, Manny’s thread follows the curved path (apple → fruit → food → edible) and arrives at “yes” without an external logic engine iterating through syllogisms – the path is already carved. - **Analogies without special handling:** If two domains share structural similarity in their manifold regions, a question or problem can slide from one to the other. Manny might recognize that the relationship between the Sun and planets (in terms of gravity and orbit) has a similar shape to the relationship between a nucleus and electrons – because the pattern of relations (a central body with smaller bodies orbiting under a force) creates analogous curvature patterns. Thus Manny could analogize solar system and atom spontaneously by the geometric alignment, not because it was pre-programmed to map those analogies. - **Planning as gradient descent:** If Manny has a goal or problem (represented as a high-valence node or constraint state in the manifold), threads exploring from a start state will tend to move downhill (reducing “distance” or energy relative to the goal). In effect, Manny’s manifold can naturally encode “what states lead closer to the goal.” A well-trained manifold for procedural knowledge means that planning (like solving a puzzle or finding a multi-step solution) can happen by *following the curvature* toward the goal,

rather than enumerating all possible action sequences. The plan is discovered as a path of decreasing resistance. - **Creativity and insight:** Because Manny doesn't rigidly stick to one hierarchy of reasoning, it can occasionally take novel paths (especially if some randomness or "temperature" is applied to the thread's movement). These might lead to new connections (imagine the marble jumps a bit and finds a nearby but non-obvious basin). What humans call creative leaps or insights often come from such traversals that escape a local minimum and find an adjacent relevant region (e.g., using an analogy from a different field to solve a problem).

It's important to emphasize that Manny's approach **shifts the complexity to the learning phase rather than the execution phase**. A lot of work goes into shaping the manifold (curving space) during training and experience accrual. But once shaped, **using the knowledge is lightweight** – it's just moving along the grooves that are already there. This is analogous to how humans put effort into learning and practice so that performance of tasks later becomes fast and intuitive. In Manny, if the manifold is properly curved by learning, answering a query or generating a thought can be as effortless as a stone rolling downhill.

Fundamental Dimensions vs. Emergent Domains

To build the manifold in a principled way, we consider **semantic dimensions** as the fundamental axes along which meaning can vary or be constrained. It's crucial to distinguish these inherent dimensions of Manny's knowledge space from the **domains or topics** that emerge within the space: - **Dimensions** are *primitive axes of variation or interaction*. They are like the coordinate axes or fields in the space – they exist a priori as the ways things can differ or relate. - **Domains** (or topics) are *emergent regions or clusters* in the manifold where certain dimensions consistently combine in characteristic ways. Domains (like "cooking" or "physics" or "emotional experiences") are not hard-coded; they form when many pieces of knowledge coalesce in that area of the manifold.

What might Manny's core dimensions be? Dimensions should be modality-agnostic where possible and capture general aspects of reality and experience. Based on our design discussions, some likely fundamental dimensions include: - **Spatial structure:** relationships like near/far, inside/outside, part/whole, shape configurations. (This covers physical layout and inclusion relations.) - **Temporal structure:** before/after relations, sequence, duration, rhythms. (Critical for narrative and causality in time.) - **Causal influence:** cause and effect, enablement, prevention, dependency. (Governs understanding of why things happen or how actions lead to outcomes.) - **Agency and Intent:** subject (agent) vs object roles, goals, intentions, social interactions. (Important for understanding plans, narratives, and social dynamics.) - **Affordance/Function:** what an object or concept can be used for, what actions can be done with it or to it. (E.g., "knife = can cut," "apple = can be eaten.") - **Similarity and Difference:** an abstract space of feature similarity (things that share many attributes cluster together, whereas contrasting pairs are opposed along certain dimensions). - **Generalization and Specification:** an axis for abstraction (e.g., instance vs category, like apple ↔ fruit ↔ food). - **Affect and Valence:** emotional or value dimension (good/bad, pleasant/unpleasant, importance, danger). This influences what is salient or remembered. - **Sensory dimensions:** if needed, sub-dimensions for different sensory qualities (color hue, brightness, sound frequency, texture, flavor, etc.) – these often feed into higher-level dimensions like similarity or affordance. - **Social/Narrative role:** protagonist, antagonist, tool, helper, etc., and notions of narrative context (e.g., comedic vs tragic tone could be seen as a dimension of experience).

These dimensions are **the degrees of freedom of Manny's semantic universe**. They define how any concept can be situated and related to others. Note that they are not tied to specific domains; they cut across domains. For instance, "causality" is as relevant to cooking (heat causes food to cook) as it is to physics (force causes acceleration) or social matters (insult causes anger).

How domains emerge: A **domain** like “cooking” or “biology” emerges when a particular set of dimensions frequently come into play together and form a tightly curved region. For example, the domain of “cooking” involves: - Objects and ingredients (with their **affordances**: knives cut, pots contain, heat transforms). - Sequences and timing (**temporal structure**: first chop, then heat, then serve). - Causal relations (**heat causes chemical changes** in food, certain combinations cause flavors). - Goals and intent (**agency**: the cook’s goal to make a dish). - Cultural context (dishes have **social significance** and emotional valence, like comfort food). - Sensory outcomes (**sensory**: tastes, aromas). All these dimensions converge in a consistent way for many experiences labeled “cooking.” Manny’s manifold will reflect that by having a cluster of nodes and edges (ingredients, actions, tools, recipes) that are densely interconnected and strongly curved, making “cooking” a recognizable region. However, **“cooking” as a domain was not a predefined axis** – it was an emergent combination of more fundamental axes.

Taxonomies and hierarchies as projections: Traditional knowledge representation often starts with taxonomies – e.g., a tree of categories (Animal → Mammal → Dog → specific breeds, etc.). In Manny, hierarchical relations like **“is-a” (class inclusion)** and **“part-of”** are certainly present as important edges (they are part of the structural and generalization dimensions). But Manny does not hard-code a single taxonomy of everything. Instead, **taxonomies can be derived as one view or slice of the manifold**. If you focus on the “generalization/specificity” dimension and the “part/whole” dimension, you can trace a taxonomy-like tree (this is like applying a *taxonomy lens*). For example, by following *is-a* links upward you might get: apple → fruit → plant → organism. That chain exists in Manny’s manifold, but alongside it, apple also connects sideways to taste, color, cultural symbols (Eve’s apple, Apple Inc.), etc., which a strict taxonomy would consider “cross-domain.” Manny’s representation allows those cross-connections to exist naturally.

Lenses – weighting dimensions for views or tasks: A **lens** is essentially a context that highlights certain dimensions over others. When Manny is asked to provide a taxonomic listing, it can apply a lens that prioritizes the “category (is-a)” relationships and downplays, say, affect or temporal relations. If asked for a functional explanation, a lens might emphasize causal and affordance dimensions. Lenses don’t alter the underlying knowledge; they just filter the traversal of the manifold to be most relevant to the question. In implementation terms, a lens could modify the “energy” or weight of moving along certain types of edges. For example, under a **biological lens**, paths that go through parent-child category links (like genus-species relationships) have very low cost (so threads will move easily along those), whereas tangential associations (like “Newton → apple” due to the anecdote of the falling apple) would be deemphasized.

Because of this approach, **Manny can fluidly support multiple views of knowledge**: - It can act like a taxonomic knowledge graph when needed, by focusing on those hierarchical dimensions. - It can act like an associative semantic network on a different query, finding connections through similarity or common context. - It can recall an episodic memory by focusing on temporal and narrative links. All these are just different traversals of the same manifold.

No single ontology is imposed: The benefit here is flexibility and robustness. Traditional ontologies often struggle with edge cases or cross-cutting concepts (e.g., does a tomato go under fruit or vegetable?). Manny doesn’t force a single hierarchy; “fruit” vs “vegetable” is just one relationship (culinary usage vs botanical classification – Manny can accommodate both as separate relations). Over time, the stronger perspective (in everyday experience, tomato is used as a vegetable in cooking) will have more curvature for certain questions, but the other perspective is still there (botanically tomato *is* a fruit). Manny can reconcile these because those are just two different edges in the manifold, each with its context of use (perhaps the *culinary lens* vs. *botanical lens*).

In summary, **dimensions are fundamental and built-in (how the manifold can curve), whereas domains and taxonomies are emergent patterns of curvature.** This separation means Manny's knowledge base can grow and self-organize without being constrained or redesigned every time a new kind of thing is learned – new domains just appear as new regions in the existing space.

Cognitive Dynamics: Thought as Free Fall in Semantic Space

When the manifold has been richly trained, **cognition in Manny becomes the natural movement of threads through the curved space**, much like a body freely falling along geodesics (the paths of least resistance) in a gravitational field. This perspective has several implications for how Manny “thinks” and responds:

- **No explicit search or planning algorithm needed (in principle):** If Manny's knowledge has formed a correct shape, then answering a question or solving a problem is a matter of releasing a thread at the starting point (the query or initial context) and letting it follow the curvature to a low-energy endpoint (the answer or solution). The manifold's shape embodies what we might call “common sense” or “world knowledge,” guiding the thread along sensible routes. For example, ask Manny “How do you make an apple pie?” and the starting activation (apple pie, making) will be somewhere in the cooking region. The strongest curvature will lead the thread through steps like (have ingredients) → (prepare dough, cut apples) → (combine and bake) → (result: apple pie). Manny doesn't have to enumerate all recipes; the major steps are attractors in the cooking region and the thread falls into that sequence. This is *reasoning by following your nose* – and in humans it feels like recalling or intuiting the answer rather than calculating it.
- **Starting point agnosticism:** In Manny, *any* node or concept can be a starting point for cognition. There is no single “input layer” or designated query format. Whether the prompt is a word, an image, a sound, or an internal goal, it just lights up certain nodes in the manifold. From there, Manny's dynamics carry it forward. This is important for multi-modal and context-rich tasks. For instance, a smell input might activate a certain memory region, which then triggers visual and verbal associations. Manny can handle that because everything is in one connected space. It also means Manny doesn't require a well-formed question to think – partial cues or even free association can set a thread rolling. This is analogous to human thought where any sensation or idea can spark a chain of thoughts.
- **Approximate paths are sufficient:** Manny doesn't strive for an *exact optimal path* in the way a classical graph search or optimization algorithm might. Natural cognition is content with “good enough” paths that lead to a coherent answer. The manifold's curvature guides threads in an approximate gradient descent: they tend to move from higher “energy” (less coherence, more uncertainty) to lower energy (more familiar, more coherent) states. As long as the direction is roughly correct, the process will converge on an answer. Small deviations or noisy steps often correct themselves because if a thread goes slightly off course, it will encounter a rising energy slope (less support in the manifold) and be pulled back towards the nearest strong attractor. In practical terms, this means Manny is **gracefully tolerant of ambiguity and error**. It doesn't require perfect inference rules at run-time; it relies on the wisdom of the manifold. This is why people can answer questions quickly with intuition – your brain isn't doing exhaustive logic every time; it's following well-worn grooves.
- **Spontaneous emergence of “abilities”:** Many capabilities that would normally require separate modules can emerge spontaneously:

- **Logical inference:** If A implies B and B implies C have been learned (curvature along those relations), then a query on $A \rightarrow C$ will simply traverse that path.
- **Analogy:** If two situations share a structural pattern, Manny's threads might bridge them (especially if a little randomness or exploration is introduced, the thread might hop from one basin to a similar-shaped basin).
- **Generalization:** New inputs that are variations of known ones will naturally fall into the same basin. Manny might never have seen a "golden pear," but if it knows "pear" and the color "golden," a description will place this new concept near "pear" in the manifold without needing explicit training on it.
- **Counterfactual or imagination:** By temporarily altering certain edges or activating a hypothetical node, Manny can explore how the manifold would reconfigure. For example, asking "What if apples were blue?" could involve activating an "apple" concept together with a "blue color" attribute while attenuating the normal color dimension for apple. The thread would then wander in a slightly altered manifold to see implications (perhaps it finds that birds might not recognize them, or it just realizes it's a strange scenario with no strong attractors, concluding it's purely hypothetical).
- **Planning and problem-solving:** If Manny has goal states marked by high positive valence, threads will be biased to find routes that increase that valence. The system doesn't explicitly enumerate plans; it feels its way. If a dead-end is reached (a high energy state where no progress toward the goal is evident), Manny can backtrack or try a different route (this could be implemented via multiple exploratory threads with some stochasticity, analogous to how one might mentally simulate different approaches to a problem).

It should be noted that while this **free-fall cognition** description is the ideal state, reaching it depends on **the manifold being well-shaped through extensive learning**. Early on, or if Manny lacks knowledge in an area, threads may not find good paths (just as a person ignorant of a topic will flounder). In such cases, Manny might resort to more brute-force exploration (like a random walk or a breadth-first search in the graph) until more learning fills in the gaps. However, as Manny learns, those inefficient searches are gradually replaced by direct "falling" along learned pathways. The end goal is that, after sufficient training, Manny's responses appear fluid, context-aware, and fast – because they are the natural consequence of its structured understanding, not an on-the-fly computation from scratch.

This approach has a strong correspondence to cognitive science and neuroscience theories: - In humans, tasks become intuitive (system 1 thinking) when the brain's neural network has been tuned by practice – we don't consciously calculate, we just respond by pattern. - Concepts that are deeply learned literally have a "weight" to them in our mind (we often speak of ideas *carrying weight* or *gravity*, metaphorically capturing this effect). - When solving problems, people often use analogical thinking and mental simulation spontaneously; Manny's manifold provides a substrate for the same.

In sum, Manny's cognitive dynamics aim to make thinking a byproduct of the manifold's shape. The motto could be: **Don't force the thought, shape the space**. By investing in learning the correct representations and relations (shaping curvature), we allow the thinking process to be as effortless as an object rolling downhill.

Training Manny: Experiences, not Explicit Programming

Given this architecture, training Manny is a distinctive process. It's less like traditional ML training (with static datasets and global loss functions) and more akin to **teaching or growing a cognitive structure through guided experience and self-organization**.

Initial seeding (the starting topology): We begin by defining the initial set of primitives and some baseline connections – essentially a rough scaffold of knowledge that Manny will refine. For example, at the current stage using words as primitives, we might start with a vocabulary of key concepts and some obvious relations: - Basic factual assertions (e.g. “an apple is a fruit”, “fruit is a type of food”, “a dog is an animal”). - A few hierarchical backbones (taxonomic *is-a* chains or part-whole relations that are well-established). - Possibly incorporate distributional knowledge (for text, we could connect words that often appear together in human text data as weak edges initially, akin to an embedding graph where proximity in embedding space suggests a connection). These serve as **exploratory links** – they are not high-curvature yet, but they give Manny a starting sense of which concepts are related. - Any known strong rules or constraints (if we have domain ontologies, we can encode those as initial edges with some weight).

Think of this as providing Manny with a loose net of concepts – a graph that is initially flat (low curvature) but connected. Manny doesn’t yet *understand* these connections deeply; they are just available pathways.

Experience as training data: Training Manny means feeding it a series of experiences or tasks that will drive threads through the manifold and update curvature. An experience could be: - A simple fact or statement (which Manny processes by linking the concepts in the statement). - A question to answer or problem to solve (which Manny attempts by traversing from question to potential answers). - A narrative or sequence of events (which Manny plays through, strengthening temporal and causal linkages). - A comparison or analogy task (prompting Manny to find similarities/differences between two concepts, e.g. “How is an apple like a pear? How is it different?”).

Rather than presenting these as static pairs of input-output, we **immerse Manny in the scenario**. For example, to teach Manny “apple is a fruit,” we might activate the “apple” node and the “fruit” node under a context lens of category inclusion and apply a positive valence reward when Manny forms a connection along an *is-a* relation. This is akin to the mentor guiding Manny’s attention along the right path and then saying “yes, remember this.” If Manny had any uncertainty (multiple possible connections), the positive feedback on the correct one will curve that path (lower its energy for future traversals).

Teacher-forced vs. autonomous learning: Early on, **teacher-forced experiences** are useful. This means we guide the thread along an ideal route: - We explicitly tell Manny the steps of a reasoning chain or the parts of a procedure (e.g., “To make a tea: boil water → steep leaves → pour into cup”). Manny’s job is to internalize that sequence by strengthening those edges. We ensure it traverses each step in the right order during training. - We might also sometimes provide the destination concept and let Manny find it, correcting it if it goes wrong (e.g., ask “What is an apple?” expecting “fruit” and if Manny goes astray, we bring it back).

Over time, we shift to **free traversal with feedback**. We ask questions or set problems without giving the path, and let Manny explore: - If it finds a correct or reasonable path, we reinforce it (strengthen those edges). - If it makes a mistake or gets stuck, we provide a hint or correction, which helps adjust the curvature appropriately (perhaps adding a missing link or increasing the weight of a relevant dimension).

Repetition and variation: Manny needs multiple passes through similar knowledge in varied ways to form strong attractors rather than memorizing one specific route. For example, to solidify the concept “apple is a fruit”, we might: - State it directly (“An apple is a kind of fruit”). - Ask it as a question later (“Is an apple a fruit or an animal?”). - Include it in a narrative (“John picked an apple from the fruit basket” –

linking apple and fruit in context). - Contrast it ("An apple is a fruit, whereas a carrot is a vegetable" – situating apple in a taxonomy, but also giving counterpoint). Each of these experiences adds to the curvature around "apple" and "fruit" and their link. Overtraining one phrasing is avoided; instead Manny gets a well-rounded exposure.

Periodic consolidation (the "sleep" metaphor): After a batch of experiences, Manny should consolidate: - **Promote** frequently co-occurring patterns to higher stability. If certain cluster of nodes was activated together in many experiences, consider creating a motif node for it. For instance, after many "man walks dog" episodes, Manny might create a composite node representing that relationship, which speeds up future traversal (the next time it sees a man and a dog in context, it can jump via the composite node). - **Decay/prune** spurious associations. If Manny tentatively formed an edge that never got confirmed, this is where it gets weakened or removed. For example, if early on a noise in data made Manny connect "apple" to "Apollo" (perhaps because of a coincidental co-occurrence in text), but later experiences did not reinforce this, consolidation will let that link fade. - **Rebalance** importance weights (Manny might have internal parameters akin to "short-term importance" and "long-term importance" like in some cognitive architectures ²). Consolidation can transfer transient learning to long-term stability for truly recurring patterns, and drop transient spikes that proved irrelevant.

Continuous learning and plasticity: The design is for Manny to be **online-learning** friendly. It is not trained once and frozen; rather, every new interaction is an opportunity to adjust. However, Manny must also avoid catastrophic forgetting of older knowledge. The curvature and consolidation mechanism inherently manage this: as long as an old piece of knowledge is occasionally used or strongly ingrained, it remains a deep well that won't be easily overwritten by a few new experiences. New experiences initially cause local perturbations; only if they represent a genuine new pattern (with repetition or high valence) will they reshape the manifold significantly.

Scalability considerations: Early training will likely focus on a narrow domain or a controlled knowledge set to validate these concepts (perhaps a few hundred key concepts and relations). As Manny scales up: - We might feed in structured knowledge from existing databases (like WordNet or ConceptNet style assertions) to quickly scaffold lots of obvious facts. These would be added as initial edges with moderate weight. - We could use pre-trained language model embeddings or image embeddings to propose connections (e.g., if a language model embedding says "apple" is close to "pear" and "fruit", we ensure Manny has edges linking apple-pear and apple-fruit initially). Manny will then confirm or adjust these through actual experiences or queries. - We will need to monitor that the manifold doesn't get cluttered with too many meaningless nodes. The emphasis on *promotion thresholds* (not creating a new concept until it's seen enough) and decay will help manage this.

Topology vs. Taxonomy in training: We do seed some taxonomy-like relations as mentioned, but we remain open to Manny discovering non-taxonomic connections. For instance, we won't explicitly tell Manny to connect "Newton" to "apple," but if we train a narrative about Newton and the falling apple, Manny will form that cross-domain link. The manifold might then place "Newton" (a person node) in a region that has a link to the apple concept (because of the gravity story) as well as links to math and physics concepts. This is fine; it's a feature, not a bug, as it allows multidisciplinary context. Later, using a lens, we can separate those contexts (in a scientific discussion lens, "apple" might not be activated by Newton, whereas in a historical anecdote lens it might).

Summary of training methodology: 1. **Define the Graph Grammar:** Determine what types of nodes and edges Manny will use initially (concept nodes, relation types like "is-a", "part-of", "causes", "associated-with", etc.). This defines what structures Manny can form. Keep it expressive enough but not overly complex for a start. 2. **Initialize Concepts:** Create nodes for the basic concepts in scope. Initialize certain edges (from domain knowledge or automated sources) to give Manny a starting shape. 3.

Iterative Experience Feeding: - Present a scenario or query. - Let Manny's thread activate and traverse. - If correct, reinforce (increase curvature along those edges). - If incorrect or incomplete, guide it (either by forcing a connection or giving a hint) and then reinforce the corrected path. - Record any significant new connections formed. 4. **Consolidate Regularly:** After a batch of experiences, run the consolidation: - Remove edges below a weight threshold. - Merge nodes or create a composite node for patterns that co-occurred frequently. - Adjust weights (e.g., normalize so that extremely frequent connections don't get disproportionately overstrong to the point of blocking creativity or alternative paths). 5. **Evaluate Emergence:** Occasionally, test Manny with novel problems or analogies to see if it can use the learned manifold structure to handle them. This helps identify if certain needed connections are missing or if noise is causing wrong turns. 6. **Expand and Refine:** As needed, introduce more primitives (if we see that words are too coarse in some area, we might introduce sub-word nodes or phrase nodes; or if images are being integrated, we add the visual primitive nodes and relations). The architecture allows adding these without rewriting the whole system, because they all plug into the same manifold.

This training process is ongoing – Manny is envisioned as continually learning. In deployment, Manny would keep updating from user interactions as well (with some moderation to prevent chaotic changes). There may be a mechanism for **meta-learning** where Manny learns how to learn better (for instance, adjusting its own promotion thresholds or lens weights based on past success).

Conclusion: Driving Principles and System Coherence

Bringing it all together, the key principles of Manny's architecture and design are:

- **Unified Manifold:** All knowledge and modalities live in one interconnected representational space. This space is the single source of “truth” for Manny's cognition. No parallel subsystems with their own isolated knowledge.
- **One Learning Law:** The manifold is governed by one set of physics-like rules (traversal, curvature update, decay, promotion). We do not bolt on separate algorithms for different tasks; we adjust parameters or lenses of the same underlying algorithm. This yields consistency and potential elegance in emergent behavior.
- **Primitives as Temporarily Atomic:** We choose useful primitives for the current level of system development (words, basic image features, etc.), but remain ready to refine them. Nothing is sacred or un-splittable if the system's fidelity demands more detail later. This aligns with the concept that what counts as “atom” depends on your level of analysis.
- **Knowledge as Geometry:** Manny treats semantics geometrically. Frequency and coherence of experiences create “mass,” which bends the representational space into curved pathways. Meaning is not a static entry, but a basin in this space, and thinking is moving through this space. This principle ensures that things like analogy, generalization, and context-sensitivity are natural outcomes of the model's geometry, not separately coded features.
- **Progressive Compression (Hierarchy of Abstraction):** Manny reduces high-dimensional sensory input into lower-dimensional meaningful representations in stages. Each layer of abstraction filters out noise and preserves salient structure, enabling higher layers to operate efficiently. Memory is the end product of this compression: what's stored is only what survived as consistently useful patterns.

- **Dimension-based Richness:** By considering various fundamental dimensions of experience, Manny's knowledge isn't confined to a single view. This allows flexibility (the same knowledge can answer a factual question, serve in a story, or spark an analogy depending on which dimensions/lenses are engaged). Domains and taxonomies come out of the manifold's shape instead of being forced in, which gives Manny a fluid ability to cross-connect knowledge.
- **Emergent Reasoning and Skills:** We rely on the manifold to yield reasoning capabilities. Instead of coding explicit rules or decision trees, we train Manny until its manifold naturally performs those operations. The system's complexity is in the learned topology, not in hard-coded logic. This is a bet on emergence: if we design the representation and learning rules well, sophisticated cognitive functions will spontaneously appear when the knowledge is sufficiently rich and structured.
- **Training via Experiences:** We treat knowledge acquisition as a gradual, interactive process. Manny learns more like a child or an apprentice than a static program: through demonstration, practice, feedback, and gradual increase in challenge. This not only forms knowledge, but also *forms the way to use knowledge*. Manny's notion of "truth" or "solution" comes from what has been consistently validated in experience (weighted by valence and context), not from an external hard-coded database.
- **Continuous Adaptation:** Manny should remain adaptable. The world changes, and Manny's unified model can update continuously. Because it isn't a fixed taxonomy or a locked model, new concepts or corrections can be integrated by simply providing new experiences, and the manifold will adjust its curvature accordingly. The principle here is resilience: knowledge that is no longer useful naturally atrophies, and new knowledge finds its place.

By formalizing these principles, we have a clear target for implementation and a guide for gap analysis. Any component of the design or any implementation decision can be checked against these fundamentals: - Are we accidentally creating a separate silo or bypassing the manifold (violating unity)? - Are we introducing an algorithm that doesn't fit the "single physics" metaphor (like a bolt-on rule system) unnecessarily? - Are our chosen primitives and representation levels adequate for now, and can they evolve later? - Does our system preserve the geometric meaning idea (e.g., are we allowing the system to form its own similarity measures, or are we imposing arbitrary metrics)? - Does each new feature integrate into the learning/traversal framework rather than sidestepping it?

Using these principles, we can conduct a gap analysis between the *current implementation plan* and the envisioned architecture: - If the current plan treats images as file attachments or uses separate AI models without manifold integration, that's a gap to close (we need to bring images into the manifold as primitives). - If the current plan has a fixed ontology or doesn't allow the graph to self-organize, that needs adjustment to align with the emergent approach. - If the current plan lacks mechanisms for decay or consolidation (meaning it would just accumulate knowledge indefinitely or equally weight all past inputs), we should plan those features to ensure long-term coherence. - If the plan doesn't include the concept of lenses or context modulation, we might not achieve the flexibility in reasoning modes – this could be added to the roadmap.

Each principle can be turned into concrete requirements or checks. Ultimately, Manny's design is ambitious: it attempts to capture intelligence as a **unified field of knowledge** where structure and dynamics are one. Formalizing it in this way clarifies why each design choice matters and how the pieces fit together. This will guide development, prevent ad-hoc drift away from the vision, and serve as a North Star for evolving Manny into a robust artificial general intelligence system grounded in a single manifold of experience.

1 2 Hyperon:Atomspace - OpenCog
<https://wiki.opencog.org/w/Hyperon:Atomspace>