Transdimensional Painter — Sample Chapter 2

The Long Hunch

By Kenneth A. Mendoza

2026-05-17

The Long Hunch

It is Tuesday morning on the Oregon coast. The fog has not lifted; it does this most mornings here, sitting at the front edge of the spruce line and then deciding, around ten, to either burn off or commit. I have not yet decided what kind of working day this is going to be, and the fog seems to be in the same state.

There are four windows open on the screen in front of me. One of them is a Rust process I started about an hour ago; it is running a multi-precision Newton refinement against thirteen candidate three-body orbital initial conditions and the verdict is hours away. One of them is the Erd■s Atlas workbench — a tool I built across about eighteen months that displays the relational structure of Paul Erd■s's published list of unsolved problems as a navigable graph. One of them is the John Baez tuning-book draft I have been reading since February. One is a half-finished paragraph I will not finish today.

The hunch is in the background of all four. It has been in the background of every working day for the last several years. The hunch is not technical; it is structural. It is the suspicion that the same small idea — that nature prefers short descriptions, and that the preference shows up as detectable structure in data of every kind — explains both why a coupled-oscillator resonance settles into a small-integer ratio and why an Erd■s problem from 1947 turns out to be the same problem as an Erd■s problem from 1968 in a different mathematical vocabulary. I have been working on this hunch in different forms since before I had a name for it. The Erd■s Atlas is the most recent material expression of it; the orbital-discovery work running in the other window is the working-out of a specific consequence.

This chapter is about how the hunch turned into a research practice. It is partly autobiographical, because the practice is shaped by the path that produced it. The path is not the one most working scientists take. I want to lay it out honestly here, because the rest of the book is going to ask the reader to follow me through some technical terrain, and the reader has a right to know who is doing the leading.

I. The microbiology years

I started in cancer virology. The specific lab was at UCLA, working on the molecular biology of papillomaviruses — the family of small DNA viruses that, in certain serotypes, cause cervical cancer. I was a research associate, not a graduate student. I cut my teeth on PCR optimization, plasmid construction, expression assays. None of this is the kind of work that the rest of this book is about. But the *habits* are.

A papillomavirus genome is about 8,000 base pairs. That is small. The genome encodes maybe eight proteins. The genome does an enormous amount of work for that little raw material — it manages its own replication, it manipulates the host cell's machinery, it occasionally integrates into the host genome and produces oncogenic transformation. The thing the genome does that I have come back to mentally a hundred times since then is the way it *re-uses* itself. The same stretch of DNA encodes one protein on one reading frame, a different protein on a different reading frame, a regulatory element on a third pass. The compression ratio between what the virus does and what its sequence is, is breathtaking. The virus is, in an information-theoretic sense, an exquisitely-compressed program.

The other thing I remember from those years is how a working molecular biologist actually reasons. You do not, in general, derive predictions from a master theory. You build up a mental library of patterns — "this kind of binding site usually means that kind of regulation," "this sequence motif usually means that protein family" — and you reason by pattern-match against the library. The library is the working knowledge. Whether the master theory exists or not, the library is what gets you through Tuesday.

I did not know it at the time, but I was being trained to do exactly what I would later do with mathematics: build a library of structural patterns and reason by pattern-match. Different vocabulary. Same shape of practice.

I moved from UCLA to a drug-discovery company called Arbor Vita Corporation, working on PDZ-domain protein-protein interactions. PDZ domains are short protein modules that mediate the way one protein in a cell binds to another protein's C-terminal tail. The library of PDZ-domain interactions across the human proteome is enormous; the rules that govern which PDZ binds which C-terminus are subtle. I spent several years building exactly that kind of pattern library at the molecular level — running peptide-binding assays, fitting binding affinities, comparing them across organisms.

What I learned at Arbor Vita that I did not learn at UCLA is what *industrial* pattern-finding looks like. UCLA was patient; academic biology lets you spend two years on a single regulatory element. Arbor Vita was not patient. The company had a pipeline, the pipeline had quarterly milestones, and the milestones required that the pattern library be *operational* — that you could query it for a useful answer in a reasonable amount of time, not just admire its structure. That distinction — between a pattern library you can admire and a pattern library you can query — turned out to be central to everything I have done since.

II. The graduate-school detour

After Arbor Vita I went to Cornell — not for biology, but for philosophy of science. I had decided that the questions I cared about were not really biology questions. They were epistemology questions. *How do we know what we know? What is the structure of a working scientific theory? What does it mean for a description to be a good description?* The biology was a substrate for these questions; the questions themselves belonged to a different field.

The advisor I worked with longest had a quiet, almost off-hand way of phrasing his fundamental point. He used to say, *you have only so many coat hangers; use them wisely; make them count.* The coat hangers, in his metaphor, were the abstractions a working scientist gets to keep in their head — the load-bearing concepts that organize the rest of the work. Most theories that are popular at the moment are made of too many coat hangers. They sag under their own weight. A good theory uses fewer coat hangers to hold up the same number of facts. That is the whole content of Occam's razor when you strip the formality off it. It is also, although he did not phrase it this way, the Minimum Description Length principle stated in plain English.

I spent five years at Cornell and most of those five years thinking about that one sentence. I did not finish a doctorate. I read enough of the Bayesian-statistics and information-theory literature to know what MDL formally was — the principle, due to Jorma Rissanen and developed through the 1980s and 1990s, that the best model of a dataset is the model that minimizes the total description length of the data plus the model. I read enough Kolmogorov-complexity theory to know why this was deep — the formal connection between MDL and the universal prior under Solomonoff induction is one of the most beautiful pieces of mid-twentieth-century mathematics. I read enough Chaitin-Schnorr-Levin material to know that the actual numbers are uncomputable in general, and that what survives computationally is a working approximation that biologists, physicists, and statisticians have all converged on independently under different names. MDL. Bayesian information criterion. Akaike information criterion. Hidden in the metric a working data scientist uses to choose between two model fits, the same idea: prefer the shorter description.

What I did not know how to do, at Cornell, was *operationalize* the idea on a specific scientific question I cared about. I could write essays about MDL. I could not yet build something with it. That gap was the next several years of my life.

III. The independent-researcher decision

I left Cornell without finishing and went home to the Oregon coast. There is a story I could tell about why, involving health, family circumstance, and the post-2008 academic job market; I will not tell it here, because the story does not actually matter for the work. What matters is the decision I made about how to keep working.

The decision was: do not wait for an institution to license the work. Build the work directly, in public, with whatever tools and computational substrate I can afford. If the work is good, the institutional certification will catch up. If the institutional certification never catches up, the work is still good — or, more honestly, the work is what it is, and the certification was never going to change that.

This sounds romantic. In practice it meant Oregon Coast AI LLC, a small research entity I set up to handle the formal side of things; a series of solitary working days; and several years of expecting nobody to read anything I wrote. I do not recommend the choice to other people who have the option of an institutional path. I am also not in a position to wish I had made it differently. The work that came out of those years was not work I could have done inside an institutional framework that expected quarterly deliverables or a particular field's standard methodology. I needed the room to do cross-domain pattern-matching at my own tempo, and the only way to get that room was to make it myself.

Two things kept me from drifting during the long stretch where nothing I was working on was visible to anyone else. The first was the daily discipline of writing things down. I kept a working notebook — eventually a series of working notebooks, eventually a Cloudflare-hosted infrastructure that backed up the working notebooks across machines — and I wrote in it every day, whether or not I had results. The second was the public-artifact rule. Anything I claimed had to be tied to an artifact someone else could reproduce. A figure had to be backed by a script that produced it from raw data. A statement about a number had to be backed by a hash on the file that contained the number. The artifact rule was, in retrospect, what kept the practice from drifting into mysticism. Without an artifact, a hunch is just a feeling. With an artifact, a hunch is at least a falsifiable proposition.

IV. The 2023 conversation

In early 2023 I had a conversation that changed the shape of my working practice. It was with Aravind Srinivas, who at the time was building the company that would become Perplexity. The conversation was about ontology-aware AI — what it would look like to have a large-language-model system that knew the *structure* of a domain, not just the surface vocabulary. We talked for about two hours and most of what we talked about was orthogonal to the project Aravind was building. What I took away from the conversation was something different. It was the realization that the hunch I had been carrying around for ten years — that there is a domain-independent structural prior that explains why some patterns repeat across unrelated scientific fields — had just been handed a working substrate.

The substrate was the LLM. A large language model trained on enough text contains, in some form, the union of every cross-domain pattern anyone has ever published. The model itself does not know this in any rigorous sense; what it has is a high-dimensional similarity space in which structurally analogous patterns from different domains end up nearby. If you can query that similarity space — *what does this thing remind you of, across every field?* — you have an instrument for the kind of cross-domain pattern-matching I had been doing by hand. The instrument does not give you the answer. It gives you the candidates. The candidates still have to be checked, hashed, and bound to artifacts. But the candidate-generation step, which had previously cost me weeks of reading, suddenly cost me minutes.

In retrospect this conversation is when the practice flipped from "solitary cross-domain reading" to something more like an industrial pipeline. The hunch was still mine. The library was still mine. But the candidate-generation step had a new tool, and the tool was the kind of tool an independent researcher in 2023 could actually use.

V. Why Erd■s

By late 2023 I had decided that the most productive corpus to work on was the Erd■s problem list.

The reasoning was straightforward. The Erd■s problems are a set of about a thousand unsolved or recently-solved questions, accumulated by Paul Erd■s over roughly sixty years of writing problems on the backs of napkins and posting them in journals. The problems span number theory, combinatorics, geometry, analysis, and graph theory. Each problem is small enough to be stated in a few sentences. Each problem has, in most cases, an extensive history of partial results and near-solutions. The corpus is large enough to support pattern-matching, small enough to be navigable in finite time, and structured enough that the relationships between problems can be made explicit.

What I started building, in late 2023 and through 2024, was a navigable representation of those relationships. The Erd■s Atlas. The first version was a flat catalog — one row per problem, with the problem statement, the standard references, and my own notes. The second version added morphisms: explicit links between problems where one problem's solution would reduce or constrain another. The morphisms turned the flat catalog into a graph. The graph had structure; the structure was visible. Some problems clustered together because they were all really the same problem in different vocabularies. Other problems sat alone, on their own islands, with no obvious morphisms in or out.

The MDL scoring layer came next. For each candidate morphism, the Atlas computes a description-length score: roughly, how much information does the morphism save versus stating both problems as separate units. High-scoring morphisms are ones where two problems are nearly the same; low-scoring morphisms are ones where the link is more decorative than structural. The scoring is, in effect, a learned prior over which conjectured relationships are worth attention. It is the operationalization, on a specific corpus, of the MDL hunch I had been carrying around since Cornell.

By mid-2024 the Atlas had about a thousand problems, several thousand morphisms, and a working query interface. It is hosted on a private Cloudflare deployment because the morphism graph contains some unpublished structural conjectures I am not yet ready to put on the open web. The public-facing version, when it is ready, will be at a stable URL with the structural-conjecture layer redacted to defensible-only relationships.

The Atlas is a workbench. It does not solve problems. What it does is decide which problem to work on next, and what tools to bring to it. That second function is the one that matters. An independent researcher's scarcest resource is attention. The Atlas is, in effect, a structured attention-allocation tool.

VI. Erd■s problem 114

The first specific Erd■s problem I worked on with the Atlas guiding me was problem 114, the Erd■s-Herzog-Piranian lemniscate arc-length conjecture.

The conjecture, in plain English, is about the length of a curve called a lemniscate. A lemniscate is the set of points in the complex plane where the absolute value of a monic polynomial equals one. For a polynomial of degree n, the lemniscate is some union of curves whose total arc length can be measured. The question Erd■s, Herzog, and Piranian asked in 1958 was: as n grows, what is the maximum possible arc length of an n-degree lemniscate, taken over all monic polynomials of that degree?

The conjectured sharp bound is $2\pi n$. The lemniscate's arc length grows linearly with the degree of the polynomial, with a coefficient exactly equal to two pi. Pommerenke proved in 1961 that the arc length is at most $2\pi e n$ — the same form but with an extra factor of $e \approx 2.718$ floating around in the constant. The gap between the conjectured 2π and the proven $2\pi e$ has remained open since.

The reason the lemniscate is the right object for this problem is older than the problem itself. A lemniscate is the geometric-mean version of a conic. An ellipse is the set of points whose *arithmetic mean* of distances from two fixed foci is constant. A lemniscate is the set of points whose *geometric mean* of distances from n fixed foci is constant — equivalently, the level set where the absolute value of a monic polynomial equals a constant. Replace arithmetic averaging with geometric averaging and the ellipse becomes a lemniscate. This is not a curiosity. The natural logarithm of the geometric mean is the arithmetic mean of the logarithms, which means the lemniscate is the *equal-information* level set of the polynomial — the curve along which the entropic distance to the polynomial's roots is constant. Hilbert's lemniscate theorem, from 1897, says that lemniscates are universal approximators of plane curves: any compact set in the complex plane with connected complement can be approximated, to any desired tolerance, by a lemniscate of high enough degree. Walsh extended this in the 1920s and 1930s into the polynomial-approximation theory that organizes most of modern complex analysis. Polynomials approximate holomorphic functions; lemniscates approximate the level curves of those holomorphic functions; and the description length of the polynomial controls the description length of its lemniscate. The EHP problem, read in this light, is a question about the minimum description length needed to encode a closed curve of given arc length. The conjectured $2\pi n$ bound, if it holds, would say that description length grows exactly linearly with arc length — the cleanest possible MDL identity in this corner of the complex plane.

This is also why the lemniscate sits at the structural midpoint of Kepler's *Harmonices Mundi* and modern information geometry. Kepler's orbits are ellipses; ellipses are arithmetic-mean conics. Musical intervals are multiplicative ratios — `3:2`, `4:3`, `2:1` — not additive intervals; multiplicative averaging is geometric averaging. The lemniscate is the geometric-mean cousin of the Keplerian ellipse, in the same way the musical interval is the multiplicative cousin of the additive distance. The duality runs the whole way down. This is the connection that made me first suspect, several years before I had a name for the suspicion, that the same small idea would explain both orbital resonance and harmonic resonance — and that the lemniscate flash I would not yet have seen on Mathstodon, the one that opens Chapter 1, was an instance of a much older mathematical pattern.

I am not going to claim to have closed the gap. The work I have done on EHP is partial: I have an extension of the Pommerenke argument to a slightly more general class of polynomials, an analysis of the structure of the extremal lemniscates for small n, and a computational program that surfaces specific polynomials whose lemniscates approach the bound on small n. The combined work was deposited on Zenodo as a preprint with the concept DOI 10.5281/zenodo.19184467 — a versioned identifier that updates as the work is extended and that institutional readers can use to track the most recent version.

The interesting thing about EHP is not the specific bound. It is what working on EHP revealed about the broader Erd■s corpus. The Pommerenke argument involves a balance between the polynomial's degree, the complex geometry of its level sets, and a specific information-theoretic quantity (the logarithmic capacity of the level set) that is *also* the relevant quantity in several other Erd■s problems in different areas. The capacity-like quantity is one of those domain-independent structural priors that the MDL hunch predicts should exist. EHP gave me a concrete instance of one. The instance turned out to be transferable to several other problems in the Atlas, which is how a single problem's work ends up touching ten others.

VII. The DeepMind handshake

In late 2025 the next layer of infrastructure became available, and the layer was Google DeepMind's `formal-conjectures` repository.

`formal-conjectures` is a public Lean 4 repository, maintained by DeepMind's mathematics group, that contains formal statements (in the Lean 4 theorem-proving language) of approximately four hundred open problems in mathematics. The Erd■s problems are the largest single block of the repository. Each problem is encoded as a Lean statement that says, in the formal language, exactly what the problem claims; the statements compile against the current Mathlib stable release; and the repository accepts pull requests from outside contributors who want to add new problems, sharpen existing statements, or — eventually — attach formal proofs.

I submitted a pull request to `formal-conjectures` in March 2026. The pull request added a formal statement for Erd[redacted]s problem 505, Borsuk's conjecture on whether every bounded subset of n-dimensional Euclidean space can be partitioned into n+1 pieces of strictly smaller diameter. The statement was sharpened across two review rounds and merged as pull request #3746.

The pull request did not contain a proof. The merged contribution is a *statement* — a formalization that gives the conjecture a machine-checkable home in the Mathlib ecosystem. The actual proof, when it exists, will be a separate contribution. What the merge represented, for me, was the first time an independent researcher's work entered Google-DeepMind-maintained Lean infrastructure. I will not claim it is the first such contribution in some absolute sense — the repository is open, the contributor list is public, and the question of who counts as "independent" depends on definitions. What I can say is that the merge was the first contribution from an explicitly non-institutional, non-academic researcher that I know of, and that it changed what the Erd[redacted]s Atlas was for.

It changed what the Atlas was for because the merge made the Atlas's structural conjectures part of an audit trail that other people could follow. Before the merge, the Atlas's conjectured morphisms — the structural links between problems — were claims I was making in private documents. After the merge, they were proposed Lean statements that anyone with access to the public repository could read, evaluate, and either accept or push back on. The work moved from a private notebook to a public surface in a single step.

The reviewers of the pull request were not impressed by anything dramatic. They were impressed by the discipline. The DeepMind `formal-conjectures` maintainers — `mo271`, `Paul-Lez`, and the rest — review on multi-day cadences, push back on overclaim, demand explicit hypotheses in place of axioms, require sorry-with-external-certificate when a proof is gated on literature, and refuse to merge anything that does not pass the repository's compilation gate against current Mathlib. The discipline is the contribution. The ten-rule list of pull-request behavior I drafted after PR #3746 closed lives in my project notes and is the operating discipline I bring to every subsequent submission. The rules are not glamorous. They are the substrate on which the rest of the work rests.

VIII. Back to the desk

It is now almost one o'clock on the Oregon coast. The fog burned off about an hour ago and the spruce line is visible through the window for the first time today. The Rust process in the other window is still grinding through Newton iterations at multi-precision; the verdict on whether the orbital-family work I have been doing for the last three months survives a strict-periodicity test is hours away.

The Erd[redacted]s Atlas tab is open in front of me. The Atlas has flagged a new candidate morphism — between Erd[redacted]s problem 30 (a Sidon-set question Erd[redacted]s valued at one thousand dollars when he was alive) and a structurally adjacent problem from the lemniscate-capacity family I have been working on through EHP. The morphism scores high enough on the MDL prior that the Atlas has queued it for further work. I will look at it after I close this paragraph.

The Baez tuning-book tab is also still open. I have not made progress on the chapter about it that I owe myself. I will not make progress today.

The point of this chapter, the autobiographical chapter that lets the reader know who is narrating the rest of this book, is not the sequence of jobs or the choice to leave Cornell or the specific Erd■s problems I have worked on. The point is that there is a continuity between all of them, and the continuity is the hunch — that the structure nature settles into is the structure that minimizes its own description length, and that the structure shows up in unrelated domains because the structure is domain-independent. The hunch has been driving this practice for over a decade. I do not yet know whether the hunch is correct in the way I think it is. What I do know is that working from it produces work that holds up under audit. The lemniscate-arc-length result, the Atlas's morphism graph, the merged Borsuk formalization, the orbital-polyrhythm signature waiting on its multi-precision verdict in the next tab over — none of these are the same kind of thing. All of them came from the same hunch, applied with the same artifact discipline, at four different layers of the description-length stack.

The book you are reading is the inside view of what working from a long hunch is actually like, on a Tuesday morning when the fog has not yet decided what to do.

Sample chapter 2 ends. The book continues with Chapter 3 — Two Geometries, One Event, returning to the methodological setup that the lemniscate flash in Chapter 1 first hinted at.

For the full manuscript or for further sample chapters, please contact ken@kenmendoza.com.