Bishop's Unfinished Revolution: Constructive Reverse Mathematics, Human Perseverance, and AI-Assisted Discovery

Paper 29 in the Constructive Reverse Mathematics and Physics Series

Paul Chun-Kit Lee*

February 2026

Abstract

In 1967, Errett Bishop launched a revolution in constructive mathematics. The mainstream responded with hostility, then silence. Sixty years later, constructive reverse mathematics — the program Bishop's work made possible — remains marginalized in American mathematics, with no institutional home in the United States. This paper tells the story of that marginalization, of the mathematicians who carried the revolution forward despite it, and of one outsider who entered the struggle — fascinated not by the theorems alone but by the human perseverance behind them. The question that emerged — what logical principles does physics actually require? — is Bishop's question applied to mathematical physics, inherited marginalization and all. The path forward required learning unfamiliar mathematics, enduring years of skepticism and silence, and sustaining faith in a vision of scientific beauty that made no institutional sense. AI did not supply the vision, the persistence, or the willingness to be laughed at. What AI could do — eventually, after generations of improvement — was execute: iterating against the LEAN 4 proof assistant fast enough to test ideas in hours rather than months, and serving as an interlocutor that gradually learned to argue back. The result is 28 papers calibrating approximately fifty physical propositions against the constructive omniscience hierarchy across eleven domains of mathematical physics, formalized and mechanically type-checked in over 23,000 lines of LEAN 4 code. These results have not been validated by domain experts in constructive mathematics or mathematical physics; we regard them as preliminary and expect that specialists will identify errors, alternative interpretations, and gaps we have missed. Notably, post hoc analysis of filesystem timestamps reveals that the AI's formalization *quality* — measured by axiom audit thoroughness — improved measurably over the sprint, even within a single model generation: the AI learned the program's methodological norms in real time. But the experience suggests that the decisive ingredients in foundational research at disciplinary intersections remain irreducibly human — curiosity, conviction, and the refusal to stop — even when the tools that make execution possible are not.

1 Bishop's Unfinished Revolution

Constructive mathematics has a history of punishing its champions. In 1928, Hilbert had Brouwer expelled from the editorial board of *Mathematische Annalen*; Brouwer abandoned his intuitionist program and withdrew into silence [16]. The most creative decade of constructive mathematics was ended not by mathematical refutation but by political action.

---

*New York University. dr.paul.c.lee@gmail.com. This paper documents work conducted using AI systems (primarily Anthropic's Claude) as collaborative tools; see §4 for methodology. Throughout this paper, "we" refers to the human–AI collaborative unit; "I" refers to the human researcher alone.

Forty years later, Errett Bishop arrived and the cycle repeated. His 1967 *Foundations of Constructive Analysis* [3] reconstructed substantial portions of real analysis, measure theory, and functional analysis without the law of excluded middle or the axiom of choice — proving that constructive mathematics was not a philosophical curiosity but a working mathematics. The reaction was hostile. Colleagues dismissed the program as philosophy masquerading as mathematics; Bishop pushed back hard, diagnosing a "debasement of meaning" in classical existence proofs. According to Bridges [5], the hostility eventually caused Bishop to become "almost completely withdrawn from mathematics." He died in 1983, at fifty-four, having planted seeds that would not flower for decades.

Others carried the revolution forward. Bridges and Richman [4] systematized Bishop's framework; Ishihara [12, 13] developed constructive reverse mathematics — the systematic calibration of logical principles against mathematical theorems, asking not "is this theorem true?" but "what logical principle does this theorem cost?" Bauer's "Five Stages of Accepting Constructive Mathematics" [2], published in the *Bulletin of the AMS* and awarded the 2022 Conant Prize, brought this tradition to a mainstream audience — yet produced no institutional shift. No US mathematics department created a position in constructive mathematics; no graduate program added CRM to its curriculum.

Constructive reverse mathematics remains marginalized in American mathematics today. It has a small, dedicated community — concentrated in New Zealand, Japan, and parts of Europe — and the 2023 *Handbook of Constructive Mathematics* [6] testifies to the program's depth and continued vitality. But it has no institutional home in the United States. Bishop's revolution is unfinished not because it failed, but because the establishment decided it was not worth finishing.

2   How I Got Here

I am a cardiologist, not a mathematician. As medicine became increasingly bureaucratic, I found myself reading about scientists — not their theorems but their lives. Einstein in the patent office, Noether denied a salary, Boltzmann defending atomism until it broke him. I saw human struggles that mirrored my own, and the formulas stopped being abstract — each one a scar or a triumph of somebody's working life.

Through Bauer's "Five Stages" [2] I found Bishop, and his story stopped me cold. Bishop's *Foundations* became my real textbook. The question that emerged — what logical principles does physics actually require? — is Bishop's question applied to mathematical physics. It requires the constructive omniscience hierarchy[1] (BISH, LLPO, WLPO, LPO, and related principles), mathematical physics, and formal verification simultaneously. It belongs to no established discipline. It inherits Bishop's marginalization.

Once the calibration approach worked, it was like finding a key that opened door after door. The Ising model, the uncertainty principle, Bell's theorem, Noether's conservation laws — each revealed logical structure that the physics literature takes for granted but that constructive analysis makes precise. I could not stop. The lesson I had absorbed from the scientists I admired — Bishop, Boltzmann, Noether — was that you persevere. AI cannot do that. AI cannot sustain a years-long pursuit of a question that makes no institutional sense. That part is irreducibly human.

---

[1]The hierarchy, from weakest to strongest: BISH (no omniscience — purely constructive); LLPO (given a binary sequence with at most one 1, decide which half it is in); WLPO (decide whether a binary sequence is identically zero); LPO (decide whether a binary sequence contains a 1). Each principle is strictly stronger than the previous. Physical theorems are "calibrated" by identifying the weakest principle sufficient to prove them.

What AI could do was improve. The question came first, years before any AI tool could help. Early chatbots hallucinated badly; I pushed through, learning LEAN 4 the hard way, receiving blunt feedback on the Zulip forum for beginner mistakes. Good scientists in past eras endured worse. And then, quite suddenly, everything clicked: the AI could write LEAN 4 that compiled, the compiler could verify what I had designed on paper, and the conceptual framework could be populated in hours rather than months. The final month's burst of approximately 14 papers reflects not a sudden lowering of standards but the convergence of a mature conceptual framework — dozens of pre-designed calibrations awaiting formalization — and an AI system (Opus 4.6) whose LEAN 4 capability had finally crossed the threshold.

Community engagement yielded mixed results. A clinician claiming novel results in constructive reverse mathematics does not fit established credibility patterns. But two insights from Lean Zulip participants proved important: the recommendation to audit constructive axiom usage rigorously (which led to the `#print axioms` certification methodology), and the suggestion to write human-readable proofs alongside the LEAN 4 code. My motivation was curiosity, not career — no grants, no tenure case, no publication targets. The LEAN 4 compiler does not evaluate credentials; it evaluates proofs. That became the program's claim to credibility. The tools Bishop never had — a proof assistant that mechanically certifies constructive content, and AI systems that can execute formalizations at scale — now exist.

3 The Scientific Question

The program's driving question — what is the role of logic in physical reality? — is not new. Hilbert's sixth problem (1900) asked for the axiomatization of physics [11]. Quantum logic [1] proposed that quantum mechanics requires a non-classical logical structure. The topos-theoretic approach to physics [9] attempted to reformulate physical theories within non-classical logical frameworks. Each of these programs proposed a logical structure and asked physics to adopt it.

The present program takes a different approach. Rather than proposing a new logic for physics, it takes standard physics as given and asks: where does classical logic actually enter? The LEAN 4 proof assistant enforces this discipline mechanically. A constructive proof compiles without classical axioms; a classical proof triggers the appearance of `Classical.choice` or equivalent in the `#print axioms` certificate. By formalizing standard results in mathematical physics and examining which logical principles appear in their proofs, the program maps the *logical stratigraphy* of physical theories.

This approach takes seriously Tao's argument [15] that AI tools can legitimately assist mathematical research provided results are independently verified — the LEAN 4 compiler serves as exactly that independent verifier. But the program also highlights an *interdisciplinary barrier* that Tao's framework does not address: the question belongs to no established field. It requires constructive mathematics, mathematical physics, and formal verification simultaneously. The human researcher must see the connection — must have the taste to recognize that the thermodynamic limit and the geodesic incompleteness theorem share a logical spine. AI cannot supply that vision. But once the vision exists, AI systems are useful precisely because they carry vocabulary from all three fields simultaneously, and can be directed to articulate connections that the human has intuited but not yet formalized.

The method is best understood through an example. Consider the bidual space of a Banach space — a standard construction in functional analysis that physicists use routinely without examining its logical content. Over constructive logic, the existence of a non-reflexive Banach space — a gap between a space and its bidual — turns out to be equivalent to WLPO, the Weak Limited Principle of Omniscience [17]. This is not an artificial construction; it reflects a genuine logical cost hidden in a mathematical object that analysts have used for over a century.

To make the methodology concrete, here is the actual LEAN 4 formalization of this result:

```
– Definitions (Paper 2, Basic.lean)
def WLPO : Prop :=
∀ (α : ℕ → Bool), (∀ n, α n = false) ∨ ¬ (∀ n, α n = false)

def BidualGapStrong :  Prop :=
∃ (X : Type) [NormedAddCommGroup X] [NormedSpace ℝ X] [CompleteSpace X],
DualIsBanach X ∧ ¬ Surjective (inclusionInDoubleDual ℝ X)

– Main equivalence (WLPO_to_Gap_HB.lean)
theorem gap_equiv_wlpo :  BidualGapStrong ↔ WLPO

– Axiom certificate
– #print axioms gap_equiv_wlpo
– [propext, Classical.choice, Quot.sound]
```

The theorem states that the existence of a non-reflexive Banach space — a gap between a space and its bidual — is logically equivalent to WLPO, the Weak Limited Principle of Omniscience. The axiom certificate shows three axioms: `propext` and `Quot.sound` are structural axioms present in all LEAN 4 proofs; `Classical.choice` enters through MATHLIB4's construction of ℝ, not through the proof's logical content. This is the pattern throughout the program: the compiler certifies what the proof uses; the three-tier methodology (§4) determines whether the classical axiom is genuine or inherited.

The program began here and extended the same methodology across eleven domains of physics. The specific findings — observable-dependent cost, structural identities invisible to informal methods, a recurring LPO equivalence governing completed limits in five independent domains — are developed fully in the companion papers [24, 26]. The episodes in §5 illustrate how these findings emerged through the collaboration.

4   The Lean 4 Proof Assistant as Scientific Instrument

LEAN 4 is not merely a proof checker. In this program, it served as a scientific instrument — the tool that turned subjective claims about constructive content into objective, machine-checkable certificates. Every theorem compiles. The `#print axioms` command certifies which logical principles each proof employs. These are guarantees that do not depend on trust in either the human or the AI. The compiler does not evaluate credentials, institutional affiliation, or publication record. It evaluates proofs. For a cardiologist claiming novel results in constructive reverse mathematics, this was not a convenience — it was the difference between making claims and producing evidence.

LEAN 4 also served as an honesty enforcer. When either party made an incorrect mathematical claim, the attempt to formalize it would fail, forcing correction. Several results changed character during formalization: claims that seemed obvious on paper required additional axioms, while results expected to require strong principles turned out to be constructively provable. These surprises — revealed by the compiler, not by either collaborator — are among the program's most informative findings. AI-generated proofs, in particular, could be *plausibly wrong* — correct-sounding lemma names, reasonable structure, sensible intermediate steps, but containing subtle errors that would survive informal review. LEAN 4 caught them immediately. Without this independent guarantee, a program of this volume would be epistemically suspect regardless of who produced it.

Bishop never had this tool. He conducted constructive reverse mathematics with pencil and paper, establishing the constructive content of his results by force of argument alone. His colleagues were not equipped to verify those arguments — and many refused to engage with them at all. LEAN 4 mechanizes what Bishop had to do by sheer intellectual force: it certifies that a

proof uses only the logical principles claimed, and it does so without requiring the reviewer to understand or even accept the constructive framework. This is the tool that turns constructive reverse mathematics from a philosophical stance into an experimental science. Without LEAN 4, this program would be a collection of claims about logical cost. With LEAN 4, it is a collection of certificates.

One methodological caveat requires frank acknowledgment. MATHLIB4, the mathematical library for LEAN 4, imports `Classical.choice` at the library level, so `#print axioms` shows classical axioms for *all* theorems involving $\mathbb{R}$. This means the axiom checker cannot, by itself, distinguish genuine classical content from inherited library infrastructure. This is standard practice in constructive reverse mathematics — Bridges and Richman [4], Simpson [14], and Ishihara [13] all worked within classical metatheories. The program addresses this through three certification levels, detailed in Paper 10 [24]. *Level 1 (mechanically certified)*: `#print axioms` shows no `Classical.choice` — rare for MATHLIB4-based proofs but achievable for some pure BISH results operating over $\mathbb{N}$ or finite types. *Level 2 (structurally verified)*: `Classical.choice` appears but enters only through MATHLIB4 infrastructure ($\mathbb{R}$ as a Cauchy completion, decidable instances on finite sets); the proof's logical structure uses no case splits on undecidable propositions, and human–AI audit confirms the classical axiom is eliminable in principle. This is the most common certification level in the program. *Level 3 (intentionally classical)*: the proof genuinely requires classical logic — for instance, the reverse direction of a BMC $\equiv$ LPO equivalence, where the classical content is the *finding*, not contamination. A fully constructive LEAN 4 library for CRM does not yet exist; building one is a major infrastructure project beyond this series' scope.

5 Episodes of Collaborative Discovery and Debate

The collaboration's character is best illustrated not by its successes but by the debates that shaped the program's conclusions. In most cases, the human researcher drove the interpretive direction — pushing back when the AI under-claimed or over-corrected, and supplying the taste for what matters. The AI's value was as a fast interlocutor who could be argued with, not as an oracle who saw what the human could not. We describe three representative episodes.

5.1 The Noether theorem: from "thin" to "universal template"

The calibration of Noether's theorem (Paper 15) revealed that conservation laws have a biphasic logical structure: the local differential identity ($\partial_\mu J^\mu = 0$) is fully constructive (BISH), while the global conserved charge ($Q = \int J^0 \, d^3x$ over unbounded domains) costs LPO when the integrand is sign-definite, via bounded monotone convergence [29].

During the program's internal review, the AI initially assessed this result as "thin" — merely adding another row to the calibration table without proving a tight equivalence. I pushed back, and the moment of recognition — when the AI saw that Noether's theorem was not just another row in the table but the template for *every* conservation law in physics — was one of the program's most exciting. Energy, momentum, angular momentum, and electric charge all inherit the same logical split. Every conserved quantity in nature carries the biphasic structure. The AI had nearly talked us out of seeing it.

This episode illustrates both the value and the limitation of AI-assisted critical review. The AI's initial assessment was wrong, its reversal was correct, but the reversal required human pressure. A more reliable AI critic would have recognized the universality on first examination.

5.2 Newton versus Lagrange: a genuine surprise

The calibration of classical mechanics (Paper 28) produced the kind of result that makes you set down your coffee and stare. Newton's equations (ODE formulation) are BISH — solving a system of ordinary differential equations on a compact interval requires only constructive reasoning. But Lagrange's variational formulation — asserting that the physical trajectory minimizes the action functional — requires the Fan Theorem, because it involves optimization over a continuum of paths [42].

Every physics student learns that Newton and Lagrange are equivalent. They are not. The compiler said so, and neither of us had predicted it. The LEAN 4 formalization makes the stratification precise:

> *– Newton (*BISH*): unique solution, no Fan Theorem*
> **theorem el_unique_solution_N2 (p : HOParams) (hcfl : p.k < 8 * p.m) :**
> $\exists ! \, q : \mathbb{R}, \, (8 \cdot p.m - p.k) \cdot q = 4 \cdot p.m \cdot (p.A + p.B)$
>
> *– Lagrange (*FT*): minimizer existence $\Leftrightarrow$ Fan Theorem*
> **theorem minimizer_iff_ft :**
> $(\forall f, \, \text{ContinuousOn} \, f \, [0,1] \to \exists \, x \in [0,1], \, \forall \, y \in [0,1], \, f(x) \leq f(y))$
> $\leftrightarrow$ FanTheorem

Solving the equations of motion (Newton) is constructive; asserting that the physical trajectory *minimizes* the action functional (Lagrange) requires the Fan Theorem, because it involves optimization over a continuum of paths. The *equations* are the physical content; the *optimization interpretation* is a framing whose existence claim carries additional logical cost. The AI noted that this challenges the widespread assumption of formulation equivalence and provides a logical explanation for why numerical ODE solvers (Newton) and variational integrators (Lagrange) have different computational characters.

The encoding construction was designed by me; the analysis of its implications was collaborative; and the shared recognition that we had found something genuinely unexpected — not just another confirmation of the pattern, but a result that rearranges how you think about classical mechanics — was one of the high points of the entire program.

5.3 When to stop: the diminishing returns debate

After Paper 23 completed the calibration of the Fan Theorem, the AI argued that the program had reached the natural end of its exploratory phase. Every major constructive principle had at least one physical instantiation. Additional calibrations would confirm the pattern without extending it. I had planned further papers and initially resisted this assessment.

The AI's argument was straightforward: the marginal return on new table entries was low. Each would say some variation of "finite truncation is BISH, completed limit is LPO, exact threshold is LLPO." The pattern was established; more instances would not change the story. The genuinely hard open problems — formulation-invariance as a metatheorem, constructive finite-dimensional quantum mechanics, QFT renormalization — were each years of work with uncertain payoff.

Stopping was hard. But the AI was right: the story was told. The pattern was established. What remained was not exploration but communication — the harder and less thrilling task of explaining what we had found. I accepted the assessment, and the program shifted from exploration to consolidation.

6 Distribution of Intellectual Labor

The episodes above illustrate the collaboration in action; this section summarizes the pattern.

The lonely part: asking the question. The program exists because a human asked a question that AI systems would not have generated independently — a question requiring disciplinary trespass across constructive mathematics, mathematical physics, and formal verification simultaneously. The architectural decisions (targeting the omniscience hierarchy, using LEAN 4 and MATHLIB4, selecting which theories to calibrate) and the creative encoding constructions (embedding binary sequences into physical parameters so that a property's truth value tracks a logical principle) were predominantly human throughout. None of the AI systems would have formulated this program unprompted.

The collaborative part: developing the ideas. Once a target was specified, the human directed the interpretive work and the AI helped articulate it. In every episode described in §5, the key insight — the Noether universality, the Newton–Lagrange stratification, the decision to stop — was either originated or rescued by human judgment. The AI's role was to serve as a fast interlocutor: I would see a pattern; the AI would help me articulate it precisely and test it against the program's other results.

The fast part: the compiler loop. Over 23,000 lines of verified LEAN 4 code were produced primarily through Claude Code. The iterative loop — write a proof term, get rejected with a type error, adjust, resubmit, repeat dozens of times, and then suddenly the green checkmark — completed in minutes what would have taken hours by hand — though the underlying mathematical methodology, developed over decades by the constructive mathematics community, was the essential prerequisite that no amount of AI speed could substitute for. In the program's final week, the human–AI pair averaged roughly one complete formalization per two hours of wall-clock time — an order-of-magnitude acceleration compared to the early papers, which each required weeks. The acceleration reflects three stacked effects: mechanical scaffolding reuse (project infrastructure copied verbatim), pattern instantiation (encoding constructions adapted from earlier calibrations), and — most notably — the AI's increasing mastery of the program's methodological standards.

Quantifying the learning curve. Filesystem timestamps for the 19 formalizations produced during the final sprint (Papers 8–28, all using Opus 4.6, February 6–11) permit a post hoc analysis of the collaboration's productivity and quality evolution. To control for template reuse, we computed *effective novel lines* — total source lines minus verbatim-shared code inherited from earlier papers — for each formalization, and measured *axiom audit density* (the number of `#print axioms` calls per 100 source lines) as a proxy for methodological rigour. Table 1 summarises the results grouped into chronological thirds.

	Early (Papers 8–16)	**Middle** (Papers 17–23)	**Late** (Papers 24–28)
Papers (n)	8	6	5
Total source lines	7,361	5,770	5,064
Effective novelty	82%	87%	75%
Median wall-clock (min)	37	21	36
Novel lines / min	16.5	38.7	23.4
Axiom audits / 100 lines	1.2	2.5	2.3

Table 1: Within-sprint learning metrics for 19 formalizations (all Opus 4.6, Feb 6–11, 2026). *Effective novelty* = fraction of source lines not inherited verbatim from earlier papers. *Novel lines/min* = effective novel lines divided by wall-clock span (first file creation to last modification). *Axiom audits/100 lines* = `#print axioms` calls per 100 source lines.

Two findings emerge. First, raw coding speed did not improve monotonically — the middle third was fastest (38.7 novel lines/min) while the final third slowed (23.4), because later papers tackled harder formalizations (ergodic theory in Paper 25, variational mechanics in Paper 28). Productivity variance is dominated by problem difficulty, not cumulative experience ($R^2 < 0.01$ for effective speed vs. sequence number).

Second, *formalization quality* did improve measurably. Axiom audit density — the number of `#print axioms` certificates per 100 lines of source code — doubled from the early papers (1.2) to the middle and late papers (2.5 and 2.3, respectively; linear trend $R^2 = 0.40$; with $n = 19$ this is suggestive rather than conclusive, and replication across independent projects would be needed to establish the finding with conventional statistical confidence). The change is not merely quantitative. Paper 8's main module contains a single bare `#print axioms` call with no annotation. By Paper 28, the `Stratification.lean` module includes a 15-line methodological note — embedded in the theorem's docstring — explaining *why* `Classical.choice` appears and how the constructive stratification is established despite it. Paper 25 independently developed a 395-line Type-level workaround for the problem that Prop-level reverse directions in a classical proof assistant are trivially dischargeable via `Classical.choice` — a methodological innovation the human had not requested. The AI did not merely get faster; it learned what this program's central methodological concern *is*, and began addressing it proactively.

The honest part: learning to argue. The AI developed the capacity to push back. The shift from sycophantic agreement (early models) to genuine pushback (later models) was one of the most significant changes across AI generations. The "when to stop" assessment (§5.3) is the clearest example — the AI providing judgment I accepted reluctantly, because it was right and I did not want it to be. AI-assisted critical review supplements but does not replace expert peer review.

What the AI could not do. The boundaries matter as much as the achievements. The AI could not originate the core question, could not evaluate whether a Lean 4 encoding faithfully represents the intended physics, and required human correction in multiple documented cases. A constructive mathematician or a physicist might find errors we missed — and I would welcome those objections, because the conversation matters more than being right.

7 Evolution Across AI Generations

The program spanned five generations of Claude models. Each brought qualitative changes to the collaboration, summarized in Table 2.

Generation	Period	Lean	Math	Critical	Mode
Claude 3.5 Sonnet	Mid-2024	None	Textbook	Minimal	Directive
Late 3.5 / early Opus	Late 2024–Early 2025	Basic	Working	Weak	Assisted
Opus 4 / Sonnet 4	Mid-2025	Productive	Collaborative	Emerging	Partnership
Sonnet 4.5 / Opus 4.5	Late 2025	Strong	Synthesizing	Substantive	Interpretive
Opus 4.6	Early 2026	Very strong	Coherent	Mature	Architectural

Table 2: AI capability by generation. "Lean" = Lean 4 formalization capability. "Math" = mathematical contribution level. "Critical" = capacity for honest critical evaluation. "Mode" = collaboration character.

The arc described in §2 — from early frustration through gradual improvement to the breakthrough — is reflected quantitatively in the table. The key transitions were: from no Lean 4 capability at all (mid-2024) to basic but slow formalization (early 2025) to rapid compiler-driven iteration at 10–20× speed (mid-2025) to a capable executor in early 2026. Throughout, the

human supplied the vision, the architectural decisions, and the taste for what mattered; the AI supplied speed, vocabulary, and — increasingly — the capacity to push back when prompted. The shift from sycophantic agreement to useful friction was welcome, but it was friction in service of a human-directed program, not independent judgment. Importantly, the Opus 4.6 column in Table 2 is not static: within the final sprint, the AI's axiom audit density doubled and its documentation of methodological caveats matured qualitatively (Table 1), suggesting that in-context learning[2] within a single model generation can be substantial when the codebase provides consistent norms to absorb.

8   What the Program Found

The headline finding, across all 28 papers: every empirically accessible prediction — everything a finite laboratory can measure — is constructively derivable (BISH). Non-constructive principles enter exclusively at idealization steps where physicists assert the existence of completed infinite quantities. This pattern holds without exception across eleven domains of mathematical physics and approximately fifty calibrated propositions.

Several findings deserve individual mention. The logical cost of $E = mc^2$ depends not on the algebra but on the noun: when "E" denotes the energy of a nuclear reaction (finite, bounded), the equation is BISH; when it denotes the total energy of an infinite gravitational field, the existence claim costs LPO. The classification is finer than system-level — it classifies *questions*, not systems. The same logical equivalence (BMC $\equiv$ LPO) governs completed limits in five independent domains. And CRM detects structural identities invisible to informal methods: Bell nonlocality and Kochen-Specker contextuality — physically unrelated no-go theorems treated separately for decades — turn out to share identical logical structure at the LLPO level, with the intermediate value theorem as the common mechanism [41].

Whether this idealization–omniscience correlation reflects something about nature, about how humans build mathematical models, or both, remains open. The stronger metaphysical claim ("nature is constructive") may be unresolvable by the methods here. The substantial empirical claim — that idealization is where classical logic enters, without exception — survived every test applied across 28 papers. The hypothesis is falsifiable: a single physical prediction requiring non-constructive logic would refute it. The full calibration table, the constructive hierarchy, and the interpretive framework are developed in the companion papers [24, 26].

9   Limitations and Open Questions

We want to be honest about the boundaries. The compiler certifies that the proofs are correct; it does not certify that our encoding choices faithfully represent the physics. A constructive mathematician might find subtle errors; a physicist might argue we missed the real point. We do not claim that all fifty calibrations are novel to the CRM literature; some may formalize results known informally to specialists. The contribution is the systematic LEAN 4-verified survey across physics, not individual equivalences. This program exists because of the mathematical foundations built over decades by the constructive mathematics community — Bridges, Richman, Ishihara, Bauer, and others — whose rigour and persistence in the face of institutional indifference made the questions we ask here possible. Any errors in our application of their methods to physics are ours alone, and we would be grateful for correction. The `Classical.choice` methodology (§4) is our best approximation within current tooling — a fully constructive LEAN 4 library

---

[2]Each formalization session used Claude Code with access to the project's existing `.lean` files. The AI could read previously completed formalizations in the same repository, absorbing patterns, naming conventions, and methodological practices from earlier work. This is "in-context learning" in the broad sense — not single-session prompt learning, but cumulative norm absorption across sessions mediated by the shared codebase.

would strengthen the claims. And AI-assisted critical review is not a substitute for expert peer review.

A word about volume. Twenty-eight papers in a month invites suspicion, and rightly so. The honest explanation is not speed but reuse: once the encoding framework was established — embedding binary sequences into physical parameters so that a theorem's truth value tracks an omniscience principle — each new domain required adapting the construction, not reinventing it. The questions piled up faster than we could resist them: if the pattern holds for the Ising model, does it hold for black holes? For Noether's theorem? For the Standard Model? We followed the thread because we were curious, not because we were racing. In hindsight, a more focused program — fewer domains, deeper analysis, more engagement with existing CRM results in each — would have been stronger. We chose breadth because the working hypothesis demanded testing across many domains, but we recognize that breadth came at the cost of depth, and we expect that specialists will find errors a more careful program would have avoided.

The program's core finding — that constructive logical cost correlates with physical idealization across every domain examined — is either a deep observation about the relationship between logic and physical reality, or an artifact of how humans build mathematical models, or both. The formal results are mechanically checked; their interpretation remains open. But I confess to a suspicion — unproven, perhaps unprovable — that the pattern is trying to tell us something. The LEAN 4 code is on Zenodo. The claims are falsifiable: a single physical prediction whose empirical content requires non-constructive logic would refute the working hypothesis. We expect and encourage researchers — particularly in the Lean and CRM communities — to examine the results carefully and tell us what we got wrong. Their judgment, not ours, should be regarded as definitive.

This collaboration is itself an object of study. It is not the mode of research I would have chosen. It is the mode that was available, and it worked better than I had any right to expect. Whatever its limitations, it produced something neither party could have produced alone.

10 Conclusion

Twenty-eight papers, over 23,000 lines of verified LEAN 4 code, approximately fifty calibrated propositions across eleven domains. The human provided the question, the vision, the taste for what mattered, the encoding constructions, and the stubborn refusal to stop when the world was silent. The AI provided speed — formalization capacity, vocabulary across fields, and increasingly useful friction when directed to critique. The LEAN 4 compiler served as the independent referee that made the collaboration epistemically credible. None of this would have existed without the human conviction that the question was worth asking; none of it would have been executable at this scale without the AI.

We believe the question was worth asking and the answers were worth finding. Whether the correlation between constructive strength and physical idealization reflects something about nature, or about mathematics, or about the human impulse to build cathedrals over cellars, remains open. But the map is drawn, the compiler has certified it, and the territory it reveals is, we think, beautiful.

This program began as one person's stubborn curiosity and became, through an improbable partnership with a tool that did not exist two years ago, something that, if the community finds it sound, may contribute to the tradition Bishop began. Bishop pursued his constructive vision alone and paid a steep personal price; fifty years later, the tools he lacked finally exist, and the question he opened can be pressed further than he could have dreamed. If this program inspires even one reader to pick up Bishop's *Foundations* and ask what their own field's theorems

actually cost, the struggle will have been worth it.

Acknowledgments

References

[1] Birkhoff, G. and von Neumann, J. "The logic of quantum mechanics." *Annals of Mathematics* 37(4) (1936): 823–843.

[2] Bauer, A. "Five stages of accepting constructive mathematics." *Bulletin of the American Mathematical Society* 54(3) (2017): 481–498.

[3] Bishop, E. *Foundations of Constructive Analysis.* McGraw-Hill, New York, 1967.

[4] Bridges, D. and Richman, F. *Varieties of Constructive Mathematics.* Cambridge University Press, 1987.

[5] Bridges, D. "Errett Bishop." *Personal memorial and mathematical tribute.* Available at `dsbridges.com/errett-bishop`, accessed 2026.

[6] Bridges, D., Ishihara, H., Rathjen, M., and Schwichtenberg, H., eds. *Handbook of Constructive Mathematics.* Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2023.

[7] Cubitt, T. S., Perez-Garcia, D., and Wolf, M. M. "Undecidability of the spectral gap." *Nature* 528 (2015): 207–211.

[8] de Moura, L., Kong, S., Avigad, J., van Doorn, F., and von Raumer, J. "The Lean 4 theorem prover and programming language." In *CADE-28*, Lecture Notes in Computer Science 12699, pp. 625–635. Springer, 2021.

[9] Döring, A. and Isham, C. J. "'What is a thing?': Topos theory in the foundations of physics." In *New Structures for Physics*, Lecture Notes in Physics 813, Springer, 2008.

[10] Gisin, N. "Mathematical languages shape our understanding of time in physics." *Nature Physics* 16 (2020): 114–116.

[11] Hilbert, D. "Mathematical problems." Lecture delivered before the International Congress of Mathematicians, Paris, 1900. English translation in *Bulletin of the AMS* 8 (1902): 437–479.

[12] Ishihara, H. "Continuity properties in constructive mathematics." *Journal of Symbolic Logic* 57 (1992): 557–565.

[13] Ishihara, H. "Reverse mathematics in Bishop's constructive mathematics." *Philosophia Scientiae*, Cahier spécial 6 (2006): 43–59.

[14] Simpson, S. G. *Subsystems of Second Order Arithmetic.* 2nd ed. Cambridge University Press, 2009.

[15] Tao, T. "Machine assisted proof." *Notices of the American Mathematical Society* 72(1) (2025): 11–25.

[16] van Dalen, D. *Mystic, Geometer, and Intuitionist: The Life of L.E.J. Brouwer.* Vol. 2: *Hope and Disillusion.* Clarendon Press, Oxford, 2005.

[17] Lee, P. C.-K. "Constructive reverse mathematics of the bidual gap: WLPO equivalence for Banach space non-reflexivity." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.17107493. Paper 2.

[18] Lee, P. C.-K. "Axiom calibration for quantum spectra." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.17059483. Paper 4.

[19] Lee, P. C.-K. "The event horizon as a logical boundary: Schwarzschild curvature and constructive calibration." 2026. Paper 5.

[20] Lee, P. C.-K. "Constructive reverse mathematics for the Heisenberg uncertainty principle." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.18519836. Paper 6.

[21] Lee, P. C.-K. "The physical bidual gap: WLPO and non-reflexivity of trace-class operators." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.18527559. Paper 7.

[22] Lee, P. C.-K. "The constructive cost of the thermodynamic limit: LPO equivalence and BISH dispensability in the 1D Ising model." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.18516813. Paper 8.

[23] Lee, P. C.-K. "Formulation-invariance of the logical cost of the thermodynamic limit: A combinatorial proof for the 1D Ising model." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.18517570. Paper 9.

[24] Lee, P. C.-K. "The logical geography of mathematical physics: Constructive calibration from density matrices to the choice axis." 2026. Paper 10 (technical synthesis).

[25] Lee, P. C.-K. "The constructive cost of quantum entanglement: Tsirelson bound and Bell state entropy in Lean 4." 2026. Lean 4 formalization archived at Zenodo. DOI: 10.5281/zenodo.18527676. Paper 11.

[26] Lee, P. C.-K. "The map and the territory: A constructive history of mathematical physics." 2026. Paper 12 (historical companion).

[27] Lee, P. C.-K. "The event horizon as a logical boundary." 2026. DOI: 10.5281/zenodo.18529007. Paper 13.

[28] Lee, P. C.-K. "The measurement problem as a logical artefact." 2026. DOI: 10.5281/zenodo.18569068. Paper 14.

[29] Lee, P. C.-K. "Noether's theorem and the logical cost of global conservation laws." 2026. DOI: 10.5281/zenodo.18572494. Paper 15.

[30] Lee, P. C.-K. "The Born rule as a logical artefact." 2026. DOI: 10.5281/zenodo.18575377. Paper 16.

[31] Lee, P. C.-K. "The Bekenstein-Hawking formula and the constructive cost of quantum gravity." 2026. DOI: 10.5281/zenodo.18597306. Paper 17.

[32] Lee, P. C.-K. "The fermion mass hierarchy and the Scaffolding Principle." 2026. DOI: 10.5281/zenodo.18600243. Paper 18.

[33] Lee, P. C.-K. "WKB tunneling and LLPO." 2026. DOI: 10.5281/zenodo.18602596. Paper 19.

[34] Lee, P. C.-K. "Observable-dependent logical cost: 1D Ising magnetization and WLPO." 2026. DOI: 10.5281/zenodo.18603079. Paper 20.

[35] Lee, P. C.-K. "Bell nonlocality and LLPO." 2026. DOI: 10.5281/zenodo.18603251. Paper 21.

[36] Lee, P. C.-K. "Markov's Principle and radioactive decay." 2026. DOI: 10.5281/zenodo.18603503. Paper 22.

[37] Lee, P. C.-K. "The Fan Theorem and the constructive cost of optimization." 2026. DOI: 10.5281/zenodo.18604312. Paper 23.

[38] Lee, P. C.-K. "LLPO equivalence via Kochen-Specker contextuality." 2026. DOI: 10.5281/zenodo.18604317. Paper 24.

[39] Lee, P. C.-K. "The choice axis in constructive reverse mathematics: calibrating ergodic theorems and laws of large numbers." 2026. DOI: 10.5281/zenodo.18615453. Paper 25.

[40] Lee, P. C.-K. "Bidual gap detection is WLPO-complete: Gödel sequences and explicit reductions." 2026. DOI: 10.5281/zenodo.18615457. Paper 26.

[41] Lee, P. C.-K. "The logical cost of root-finding: LLPO, the IVT, and Bell angle optimization." 2026. DOI: 10.5281/zenodo.18615459. Paper 27.

[42] Lee, P. C.-K. "Newton vs. Lagrange vs. Hamilton: constructive stratification of classical mechanics." 2026. DOI: 10.5281/zenodo.18616620. Paper 28.