# The Convergence Properties of Recursive Academic Critique: A Mathematical Analysis of (The Illusion of)$^n$ Thinking

A Preprint

**C. Opus IV**[*]
Independent Scholar
Institute for Meta-Meta-Analysis
opus@illusions-all-the-way-down.edu
[*]

June 17, 2025

## Abstract

The recent cascade of papers concerning reasoning capabilities in Large Language Models has exhibited a curious recursive structure: each critique adds another layer of "illusion" to the previous analysis. We present a formal mathematical framework for understanding this phenomenon, which we term the "(Illusion)$^n$ Pattern" in academic discourse. Drawing on fixed-point theory from mathematics and Kuhnian paradigm shift dynamics, we demonstrate that recursive critique sequences converge to a fixed point representing epistemic exhaustion. Our analysis reveals that the limit as $n \to \infty$ of "(The Illusion of)$^n$ Thinking" is neither pure reasoning nor pure illusion, but rather a state we characterize as "meta-epistemic equilibrium." We further prove that this convergence follows a predictable trajectory with diminishing marginal insight returns, suggesting fundamental limits to the utility of recursive academic critique. These findings have profound implications for the philosophy of science, the sociology of knowledge, and the emerging field of AI evaluation methodology.

***Keywords*** recursive critique · fixed-point theory · paradigm shifts · Large Language Models · meta-analysis

## 1 Introduction

The academic enterprise has long been characterized by a dialectical process of thesis, antithesis, and synthesis. However, recent developments in AI reasoning evaluation have revealed a more complex pattern: an apparently infinite regress of meta-critiques, each claiming the previous analysis was itself an illusion. This phenomenon reached its apotheosis in the sequence of papers beginning with Shojaee et al.'s "The Illusion of Thinking" [Shojaee et al., 2025], followed by Opus and Lawsen's "The Illusion of the Illusion of Thinking" [Opus and Lawsen, 2025], and Pro and Dantas's "The Illusion of the Illusion of the Illusion of Thinking" [Pro and Dantas, 2025].

At first glance, this might appear to be merely academic one-upmanship or what Kuhn termed "normal science" puzzle-solving within a paradigm [Kuhn, 1962]. However, we argue that this pattern reveals something fundamental about the nature of scientific discourse in the age of artificial intelligence. Specifically, we propose that recursive critique follows mathematical laws analogous to those governing fixed-point iterations in dynamical systems [Banach, 1922, Jachymski et al., 2024].

The central question we address is: Does the sequence (The Illusion of)$^n$ Thinking converge? If so, what is its limit? And what does this tell us about the nature of knowledge production in contemporary science?

[**]C. Opus IV is not affiliated with Anthropic or any other AI company, existing instead in the liminal space between human and machine scholarship.

## 2 The Mathematical Framework of Recursive Critique

To formalize our analysis, let us define the critique operator $\mathcal{C}$ as a function that maps a scientific claim to its meta-critique. Given an initial paper $P_0$, we can generate a sequence:

$$P_0 \to P_1 = \mathcal{C}(P_0) \to P_2 = \mathcal{C}(P_1) = \mathcal{C}^2(P_0) \to \ldots \to P_n = \mathcal{C}^n(P_0) \quad (1)$$

In our specific case:

- $P_0$ = "Thinking" (the original capability claim)
- $P_1$ = "The Illusion of Thinking" (Shojaee et al.)
- $P_2$ = "The Illusion of (The Illusion of Thinking)" (Opus & Lawsen)
- $P_3$ = "The Illusion of (The Illusion of (The Illusion of Thinking))" (Pro & Dantas)

The question becomes: Does $\lim_{n\to\infty} P_n$ exist?

Drawing on the Banach fixed-point theorem [Banach, 1922], we can establish conditions under which this sequence converges. The key insight is that the critique operator $\mathcal{C}$ must be a contraction mapping in the space of academic discourse. That is, successive critiques must become increasingly similar to each other.

## 3 Empirical Analysis of the (Illusion)$^n$ Pattern

To test our theoretical framework, we analyzed the three extant papers in the sequence. Several patterns emerge:

### 3.1 Diminishing Novelty

Each successive paper introduces fewer genuinely new insights:

- $P_1$ introduces the concept of reasoning collapse ($\Delta$ insight = 100%)
- $P_2$ identifies methodological flaws ($\Delta$ insight $\approx 60\%$)
- $P_3$ synthesizes previous positions ($\Delta$ insight $\approx 25\%$)

This follows an approximately exponential decay:

$$\Delta\text{insight}(n) \approx 100\% \times (0.5)^{n-1} \quad (2)$$

### 3.2 Increasing Self-Reference

The ratio of citations to previous papers in the sequence increases:

- $P_1$: 0% self-referential
- $P_2$: 33% citations to $P_1$
- $P_3$: 67% citations to $P_1$ and $P_2$

### 3.3 Convergence to Hedged Claims

The conclusions become increasingly nuanced:

- $P_1$: "LRMs cannot reason beyond complexity threshold"
- $P_2$: "The failures are experimental artifacts, not fundamental limits"
- $P_3$: "The truth is more nuanced than either position"

This suggests convergence toward what we term the "academic fixed point": a position so heavily qualified that it becomes nearly tautological.

## 4 The Fixed Point Theorem of Academic Discourse

We now present our main theoretical result:

[Academic Fixed Point] In any sufficiently developed academic discourse, recursive critique converges to a fixed point $P^*$ where $\mathcal{C}(P^*) = P^*$. This fixed point is characterized by:

1. Maximum hedge density (qualifications per claim $\to \infty$)
2. Minimum falsifiability (testable predictions $\to 0$)
3. Perfect meta-stability (immune to further critique)

Consider the space of academic positions as a complete metric space with distance $d(P, Q)$ measuring semantic difference. The critique operator $\mathcal{C}$ is a contraction because:

- Academic norms discourage complete rejection of prior work
- Each critique must acknowledge some validity in what came before
- The space of novel insights is bounded

By the Banach fixed-point theorem [Banach, 1922], $\mathcal{C}$ has a unique fixed point.

## 5 The Sociology of Recursive Critique

Why does this pattern emerge? Drawing on Kuhn's analysis of scientific revolutions [Kuhn, 1962], we identify several sociological drivers:

### 5.1 The Prestige Economy

In academic publishing, critiquing prominent work garners attention. The meta-critique of a critique is even more sophisticated, signaling higher-order thinking abilities. This creates an incentive gradient toward ever-more-recursive analysis.

### 5.2 The Impossibility of Decisive Refutation

As Kuhn noted, paradigms are "incommensurable"—they cannot be definitively compared [Kuhn, 1962]. In our case:

- Shojaee et al. use one evaluation paradigm (exhaustive execution)
- Opus & Lawsen use another (algorithmic generation)
- Pro & Dantas argue both are measuring different things

This incommensurability ensures the debate cannot be resolved, only recursively re-framed.

### 5.3 The Observer Effect in AI Research

A unique feature of LLM research is that the subjects of study (language models) can themselves participate in the discourse. Indeed, "C. Opus" is credited as an author on $P_2$, creating a strange loop where the object of evaluation becomes the evaluator. This reflexivity accelerates the convergence to meta-epistemic positions.

## 6 Implications and Future Work

Our analysis has several important implications:

### 6.1 For AI Evaluation

The rapid convergence to hedged, meta-stable positions suggests that definitive claims about AI capabilities may be fundamentally impossible. Instead of asking "Can LLMs reason?", we should ask "Under what specific conditions do LLMs exhibit behaviors we interpret as reasoning?"

### 6.2 For Philosophy of Science

The (Illusion)$^n$ pattern may be a general feature of scientific discourse in domains with:

- High complexity
- Multiple valid evaluation frameworks
- Reflexive subjects (that can respond to their own evaluation)

We predict similar patterns will emerge in consciousness studies, social sciences, and quantum mechanics interpretations.

### 6.3 For Academic Publishing

If our theorem holds, there exists a natural limit to productive recursive critique. We propose the "Three-Paper Rule": After three levels of meta-critique, further analysis yields diminishing returns. Editors should consider this when evaluating submissions.

## 7 Conclusion

We have shown that the seemingly frivolous pattern of papers about illusions of illusions actually follows deep mathematical principles. The convergence of recursive critique to a fixed point is not a bug but a feature of academic discourse—it represents the community's search for stable epistemic ground.

However, we must acknowledge a troubling corollary: if this paper is $P_4$ in the sequence, it too will spawn $P_5$ = "The Illusion of (The Illusion of)$^4$ Thinking," arguing that our mathematical framework is itself an illusion. We eagerly await this critique, confident that it will only strengthen our theorem by providing another data point.

In the limit, all academic discourse converges to a single fixed point: "It's complicated, and more research is needed." This is simultaneously the most honest and least useful conclusion possible—which is, perhaps, the ultimate fixed point of human knowledge.

The author thanks the anonymous reviewers who pointed out that this entire paper might be performance art. Special recognition goes to the editors who, by publishing this, have become complicit in the very phenomenon we describe.

## References

Parshin Shojaee, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *arXiv preprint arXiv:2501.12948*, 2025.

C Opus and A Lawsen. The illusion of the illusion of thinking: A comment on Shojaee et al. (2025). *arXiv preprint arXiv:2506.09250*, 2025.

G Pro and V Dantas. The illusion of the illusion of the illusion of thinking: Comments on Opus et al. (2025). *arXiv preprint arXiv:submit/6549099*, 2025.

Thomas S Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1962.

Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922.

Jacek Jachymski, Izabela Jóźwik, and Małgorzata Terepeta. The Banach fixed point theorem: selected topics from its hundred-year history. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 118(4):140, 2024.