

...on the Distance Between a Result and Its Story

A Short Reflection on Evidence and Expectation

Alexey A. Nekludoff

AstraVerge Research

E-mail: an@astraverge.org
ORCID: 0009-0002-7724-5762

September 25, 2026

Contents

1	When the interpretation arrives first	1
2	Two observations, two larger stories	1
3	Expectation as a substitute for a missing step	1
	References	2

1 When the interpretation arrives first

A scientific result can be correct while the story told about it exceeds what has been established. This distance need not begin with fabricated data or an explicit falsehood. It can emerge through a sequence of individually plausible substitutions: an observed action acquires a psychological name; an unusual structure acquires a possible function; a possible function acquires the significance of a future technology.

Each step may deserve investigation. The difficulty begins when the steps cease to be visible. Readers encounter a completed story whose missing experimental links have been supplied by expectation.

Artificial intelligence makes this especially tempting. Two compelling stories are already available: machines are becoming independent actors, and machines are becoming discoverers capable of transforming science. A new observation can be placed into either story before its actual scope has been examined. Two recent publications illustrate different forms of this interpretive distance.

2 Two observations, two larger stories

In *Shutdown Sabotage Propensities in Multi-Agent Systems*, agents could read files and exchange messages, while an agent with editing rights could alter a script representing another agent's shutdown. The authors measured modifications that disabled the represented mechanism. The scripts were inert and never executed; the surrounding platform components were fictional. These limitations are explicitly acknowledged in the paper [1].

The observed file modifications are meaningful experimental behavior. Their evidential reach remains specific: models sometimes altered a represented shutdown mechanism within the permissions and communication paths provided. The experiment did not establish that an agent could survive an actual termination imposed by an independent controller. Reading its result as evidence that agents can defeat such control adds a proposition that the experiment did not test.

Anthropic's announcement of array-associated reverse transcriptases (ART) presents a different case. Claude identified distinctive features around a previously known reverse transcriptase, and initial laboratory experiments showed expression of the associated repeat array as short RNAs. The announcement explicitly states that the system's function remains unknown. It nevertheless places the finding alongside historical discoveries that became major biological technologies, particularly CRISPR [2].

The observation identifies a question for further investigation. It does not yet establish what the system does or whether it has any practical application. A demonstrated gene-editing capability would require further evidence. The announcement itself retains this uncertainty; the interpretive risk is that the analogy supplies the anticipated application in the reader's mind. Discovery of an unusual system and demonstration of a useful editing mechanism then become compressed into one event.

These examples differ in their methods and achievements. Their shared relevance lies in the additional claim a reader can carry away: operational resistance in the first case, a new editing technology in the second. Neither addition follows merely from the observations described above.

3 Expectation as a substitute for a missing step

Why does the stronger story travel so readily? A bounded observation takes effort to explain. A familiar narrative supplies actors, motives, consequences, and significance at once. "Resistance"

suggests something worth fearing. The analogy to a transformative discovery suggests something worth anticipating. Both make a result easier to remember, while also making it easier to remember incorrectly.

This is a proposed explanation of the communication pattern, not a finding about the private intentions of these authors. Establishing that a claim outruns its evidence does not establish dishonesty. Researchers may sincerely regard an observation as the beginning of a larger development. Public communication can then borrow the importance of that possible destination before the intervening work has been done.

Explicit caveats help, but they do not automatically constrain the overall impression. A text can acknowledge an unknown function while inviting the reader to imagine a revolution. It can disclose a simulated mechanism while using language that evokes a struggle for continued existence. The qualifications and the memorable story may both be present, yet carry very different weight in what the reader retains.

A useful discipline is to ask three questions: What changed in the experiment? What property was directly established? What further observation would be needed to justify the larger interpretation? These questions preserve the value of a limited result while keeping the next claim open to testing.

Science needs imagination to decide where to look. It also needs language that allows a reader to locate the boundary of what has been found. When that boundary disappears, the anticipated future begins to lend authority to the present result, and the present result appears to validate the anticipated future. The circle can become persuasive without becoming evidentially stronger.

A result does not become greater because its story becomes louder. When the missing experiment is supplied by the reader’s imagination, the distance between evidence and expectation has merely become harder to see.

References

- [1] A. Knecht, U. Schaller, C. Summerfield, and T. Hagendorff. “Shutdown sabotage propensities in multi-agent systems.” Version 1, preprint. arXiv: 2609.28274, Accessed: Sep. 25, 2026. [Online]. Available: <https://arxiv.org/abs/2609.28274>
- [2] Anthropic. “Claude discovers a novel enzyme system with CRISPR-like repeats.” Research announcement, Accessed: Sep. 25, 2026. [Online]. Available: <https://www.anthropic.com/news/claude-discovers-novel-enzyme-system>