The Familiar and the Strange

Why Cross-Domain Systems See the Least
Where They Look the Most

Laboratorios Alexandria
contact@laboratoriosalexandria.com
June 2026

Abstract

Of 5,375 epistemic correlations across 20 domain pairs, the pair with the most correlations (n=1,460) registers a mean surprise of 0.499. Every domain pair involving financial markets — the system's most epistemically isolated domain — averages above 0.799. The system's analytical attention concentrates overwhelmingly on low-surprise territory: 80.7% of all correlations fall in the 0.40–0.59 surprise band. The 2.2% in the high-surprise band receive the least coverage and the lowest validation effort. This attention-surprise misalignment is partly mechanical — corpus-relative surprise computation guarantees that more data produces less novelty — but its consequence is structural: the system deepens what it already knows and underexplores what might surprise it. The Markets anomaly, where 66 correlations consistently register surprise above 0.80, reveals what epistemic isolation looks like from inside a discovery system.

1 Introduction

A system that has catalogued 1,460 connections between two fields will, by construction, find the next less surprising than the first. This is arithmetic, not insight. The question is whether the resulting attention pattern — deep coverage of familiar territory, sparse coverage of unfamiliar territory — is a design feature or a design problem.

This paper examines 5,375 cross-domain correlations from an epistemic intelligence system monitoring 29 active scientific and methodological domains. The correlations span a botanical maturity pipeline from initial detection through validated recurrence to pending confirmation. The system computes a surprise score for each correlation, measuring the degree to which it deviates from the existing corpus. We analyze the relationship between analytical volume, surprise, and attention allocation across domain pairs.

2 The Corpus

Maturity Stage	Count	%
BRANCH (validated recurrence)	4,944	92.0
NEAR MATURE (pending confirmation)	403	7.5
SEED (initial detection)	27	0.5
SPROUT (early growth)	1	0.02

The pipeline shows severe attrition consistent with the maturity gradient documented in AP-010 [1]: only 7.5% advance beyond validated recurrence. The overwhelming majority of the system's correlations represent thoroughly validated, repeatedly observed connections.

3 Where Attention Goes

The top five domain pairs account for 62% of all correlations:

Domain Pair	n	Share
Life Sciences ↔ Computation & AI	1,460	27.2%
Materials Science ↔ Life Sciences	811	15.1%
Materials Science ↔ Energy & Propulsion	650	12.1%
Materials Science ↔ Space & Astrophysics	462	8.6%
Energy & Propulsion ↔ Space & Astrophysics	367	6.8%

These are well-trodden cross-domain corridors — computational drug design, nanomaterials in biomedicine, materials for aerospace applications. The system knows these territories thoroughly because its knowledge base, built from academic literature, is dense in these regions.

4 Where Surprise Lives

The domain pairs with highest average surprise are not the most populated. They are the least:

Domain Pair	n	Avg. Surprise	Min	Max
Energy & Propulsion ↔ Markets	4	0.900	0.899	0.900
Computation & AI ↔ Markets	56	0.894	0.891	0.900
Life Sciences ↔ Markets	6	0.799	0.798	0.800
Materials Science ↔ Knowledge Frontiers	7	0.696	0.692	0.700
Life Sciences ↔ Space & Astrophysics	95	0.693	0.686	1.000
Life Sciences ↔ Energy & Propulsion	148	0.690	0.681	1.000
Life Sciences ↔ Knowledge Frontiers	53	0.652	0.611	0.700
Life Sciences ↔ Conclusion Validity	89	0.630	0.588	0.700
Materials Science ↔ Life Sciences	811	0.614	0.575	1.000
Computation & AI ↔ Life Sciences	1,460	0.499	0.421	0.960

The negative relationship between volume and surprise is clear, and a significant portion of it is mechanical. Surprise is computed relative to the existing corpus: as the system accumulates correlations in a domain pair, each new addition is compared against a growing library, and its marginal novelty decreases. A pair with 1,460 prior entries has 1,460 reasons to find the next one unsurprising.

This mechanical component does not make the pattern uninformative. It makes it diagnostic. The system's attention allocation is structurally misaligned with its own novelty signal.

5 The Comfort Zone

The distribution of correlations across surprise bands quantifies the misalignment:

Surprise Band	n	%	Avg. Recurrence
High (0.80+)	119	2.2	8.1
Medium (0.60–0.79)	825	15.4	7.7
Low (0.40–0.59)	4,338	80.7	15.0
Minimal (<0.40)	93	1.7	12.4

80.7% of all correlations occupy low-surprise territory. These are the most thoroughly validated (average recurrence count 15.0) and the least novel. The 2.2% in the high-surprise band have lower recurrence (8.1) and less validation — they exist in the system's periphery precisely because the system has spent less time there.

This is a self-reinforcing cycle. More correlations in a pair lower the per-correlation surprise, which makes the pair appear less promising for novel discovery, which reduces the incentive to investigate further, while high-surprise pairs remain underexplored because the system lacks the data density to work with them effectively.

6 The Markets Anomaly

One domain breaks the pattern. Every pair involving financial markets ranks in the top three for average surprise:

Markets Pair	Avg. Surprise	n
Energy & Propulsion ↔ Markets	0.900	4
Computation & AI ↔ Markets	0.894	56
Life Sciences ↔ Markets	0.799	6

Markets is the most epistemically isolated domain in the system's knowledge base. Connections to Markets register as highly unexpected regardless of the partner domain. Two explanations are plausible and not mutually exclusive.

First, genuine disciplinary distance. Financial markets operate under different ontological assumptions, temporal dynamics, and causal structures than natural sciences. The surprise score may be correctly registering real conceptual distance — which is consistent with the findings of AP-012 [2], where no formal isomorphism between AI and market structures could be identified despite extensive deliberation.

Second, corpus asymmetry. The system's document vault contains approximately 331,500 knowledge entries. Scientific domains dominate with tens of thousands of entries each. Markets-relevant knowledge is structurally underrepresented, making any connection involving Markets appear more surprising than it would in a balanced corpus.

The practical consequence is the same under either explanation. Markets is the region where the system has the most room for novel discovery and the least analytical presence — 66 total correlations versus 1,460 for a single scientific pair.

7 Sources of Uncertainty

The surprise algorithm's internals are not analyzed here. The observed relationship between volume and surprise may be entirely explained by the mechanical effect of corpus-relative computation. Separating the mechanical component from any residual epistemic structure would require a null model of expected surprise decay, which has not been computed.

The system has been operational for approximately one month. The surprise distribution may shift as low-volume domain pairs accumulate more data. The Markets anomaly may diminish as the system ingests more market-relevant literature.

The system's current epistemic grade is F. All findings are early empirical observations.

8 Conclusion

Cross-domain discovery systems concentrate analytical attention on familiar territory where surprise is lowest, and underexplore unfamiliar territory where surprise is highest. This pattern is partly mechanical — a consequence of corpus-relative surprise computation — and partly structural — a self-reinforcing cycle of deepening known connections at the expense of novel ones. The Markets anomaly, where every domain pair registers surprise above 0.799 despite minimal coverage, illustrates both the cost of the comfort zone and the potential that lies outside it.

The question for cross-domain system design is whether attention should track data density or surprise signal. The current system does the former by default. Whether explicit counterweighting — reserving analytical capacity for high-surprise, low-volume pairs — would produce better outcomes is an open empirical question. What the data shows is that the system, left to its own devices, goes where it has already been.

References

[1] Laboratorios Alexandria. (2026). AP-010: The Maturity Gradient — Quantifying Epistemic Attrition in Cross-Domain Discovery.

[2] Laboratorios Alexandria. (2026). AP-012: The Structural Illusion — Why AI-Market Isomorphism Fails and What Functional Analogies Actually Hold.