AP-010 — Laboratorios Alexandria

The Maturity Gradient: Quantifying Epistemic Attrition in Cross-Domain Discovery

Laboratorios Alexandria
contact@laboratoriosalexandria.com
June 2026

Abstract

Cross-domain scientific discoveries do not emerge fully formed. They follow a maturation process in which an initial detection is refined through deliberation, evaluation, and quality filtering. We present data from an autonomous epistemic system showing that this maturation follows a quantifiable attrition funnel: of 5,351 detected cross-domain correlations, 5,155 have undergone initial evaluation (EMERGING stage), 109 have been validated with supporting evidence (DEVELOPING stage), and 55 await final confirmation (NEAR MATURE stage). The overall survival rate from initial evaluation to near-maturity is approximately 1.1%. We analyze the attrition rates between stages, compare them to attrition funnels in drug discovery, venture capital, and academic peer review, and argue that the maturity gradient is a structural property of epistemic systems, not a deficiency to be optimized away but a necessary feature of quality-producing processes. The gradient is the mechanism by which detection becomes knowledge.

1 Introduction

Every knowledge-producing system has a funnel. In drug discovery, thousands of candidate molecules are screened; hundreds enter preclinical testing; dozens reach clinical trials; a few receive approval. In venture capital, thousands of pitches are heard; hundreds receive term sheets; dozens are funded; a few produce returns. In academic publishing, thousands of manuscripts are submitted; hundreds survive peer review; fewer still are cited.

The funnel is not a failure of the system, it is the system. The attrition at each stage represents the application of increasingly stringent quality criteria. A drug discovery pipeline that approved every candidate molecule would not be efficient, it would be dangerous. A venture fund that invested in every pitch would not be inclusive, it would be bankrupt.

We present evidence that autonomous cross-domain scientific discovery follows an analogous funnel, with quantifiable attrition rates at each stage. We call this the maturity gradient: the process by which an initial detection of a potential cross-domain connection matures, or fails to mature, into a validated epistemic finding with explicit falsifiability conditions.

2 The Maturity Stages

The epistemic system organizes discoveries into four maturity stages. Each stage represents a level of evidential validation, and progression between stages requires meeting specific quality criteria.

SEED: Initial detection. The system has identified a potential cross-domain correlation based on structural similarity, shared vocabulary, or pattern matching. No deliberation has occurred. The correlation is a candidate, not a finding.

SPROUT: Initial evaluation. The correlation has undergone preliminary assessment. Domain-specific context has been retrieved, and the correlation has been classified as potentially meaningful. This stage corresponds to a first-pass triage.

BRANCH: Validated with evidence. The correlation has been subjected to multi-actor deliberation. Domain arguers have presented supporting and challenging evidence. The dialectician has tested falsifiability. The epistemic judge has assigned a quality grade. The correlation has survived this process with at least coherent theoretical support.

NEAR MATURE: Awaiting final confirmation. The correlation has passed all quality filters and awaits final review for factual accuracy, consistency with the broader knowledge base, and explicit statement of falsifiability conditions. Correlations at this stage are near-mature findings.

3 The Numbers

The following table presents the maturity distribution as of June 2026:

Stage	Count	% of Total	Survival from Prior
SEED	32	0.6%	—
SPROUT	5,155	96.3%	—
BRANCH	109	2.0%	2.1%
NEAR MATURE	55	1.0%	50.5%
Total	**5,351**	**100%**	

Table 1. Maturity distribution of 5,351 cross-domain epistemic correlations, June 2026.

The critical transition is from SPROUT to BRANCH: of 5,155 correlations that have been initially evaluated, only 109 (approximately 2.1%) survive to the validated stage. This is the primary quality filter, the point at which multi-actor deliberation separates speculative connections from evidentially supported ones.

Once a correlation reaches BRANCH, the survival rate to NEAR MATURE is substantially higher (50.5%), suggesting that the deliberation process is the binding constraint: correlations that survive deliberation are likely to survive final review. The deliberation acts as the rate-limiting step in the maturation process, analogous to the clinical trial phase in drug discovery where most attrition occurs.

The 32 SEED entries represent correlations recently detected that have not yet undergone initial evaluation. This is a queue, not a stage, these will move to SPROUT as the system processes them. Their small number relative to the SPROUT population indicates that the detection-to-evaluation pipeline operates with low latency.

4 Comparison with Other Attrition Funnels

The maturity gradient becomes more informative when compared to attrition funnels in other quality-producing systems. The following table presents survival rates across four domains:

System	Overall Survival	Critical Transition	Source
Drug discovery	0.01–0.02%	Phase I→II (33%)	DiMasi et al., 2016
Venture capital	0.1–0.2%	Pitch→funding (1%)	Gompers et al., 2020
Academic peer review (top tier)	5–10%	Submission→accept	Varies by journal
Epistemic maturation (this work)	1.1%	SPROUT→BRANCH (2.1%)	Operational data

Table 2. Attrition rates across quality-producing systems.

4.1 Drug Discovery

In pharmaceutical development, the attrition from initial compound identification to approved drug is approximately 1 in 5,000 to 1 in 10,000 (DiMasi et al., 2016). The overall survival rate is 0.01–0.02%. Our epistemic funnel shows a survival rate of approximately 1.1% from detection to near-maturity, roughly 50–100 times higher than drug discovery. However, the comparison is not direct: drug discovery requires wet-lab validation at each stage, while epistemic validation is computational. The stages are not equivalent in cost or time. What is equivalent is the structural logic: each stage applies criteria that the previous stage did not, and each transition eliminates candidates that satisfied earlier criteria but fail later ones.

4.2 Venture Capital

In venture capital, approximately 1% of pitches receive funding (Gompers et al., 2020), and approximately 10–20% of funded companies produce significant returns. The overall success rate from pitch to return is approximately 0.1–0.2%. Our epistemic funnel's 1.1% survival rate is comparable to the venture funding rate, suggesting a structural similarity: in both cases, approximately 1 in 100 candidates survives initial quality filtering. The analogy extends further: in both systems, the survivors are valuable precisely because the non-survivors were eliminated. A venture fund that funded every pitch would produce returns indistinguishable from noise.

4.3 Academic Peer Review

Journal acceptance rates vary from 5–20% in selective journals to 50–70% in broad-scope venues. A typical top-tier journal accepts approximately 10% of submissions. Our SPROUT-to-BRANCH survival rate of 2.1% is substantially more selective than typical peer review, reflecting the additional difficulty of cross-domain evaluation: a finding must be valid not just within one field but across two. This additional requirement, that the structural correspondence between domains be genuine, not merely metaphorical, imposes a filter that single-domain peer review does not face.

5 The Gradient Is Not a Deficiency

A natural response to the attrition data is to seek optimization: how can we increase the survival rate? How can we reduce the loss between stages? We argue that this response, while intuitive, misunderstands the function of the gradient.

The attrition at each stage represents the application of quality criteria. The SPROUT-to-BRANCH transition eliminates correlations that are structurally plausible but evidentially unsupported. This elimination is not waste, it is the primary value-producing activity of the system. A discovery pipeline that passed every candidate through to maturity would not be efficient; it would be useless, because the quality signal that makes mature findings valuable depends on the elimination of immature ones.

This principle is well understood in other domains but frequently forgotten in computational settings, where the marginal cost of processing an additional candidate is negligible. The temptation is to lower thresholds, retain more candidates, and let downstream stages sort them out. But this strategy degrades the quality signal at every subsequent stage. If SPROUT-to-BRANCH survival rises from 2.1% to 20% by lowering the deliberation threshold, the BRANCH stage becomes contaminated with weakly supported correlations, and the NEAR-MATURE stage must bear a filtering burden it was not designed for. The gradient is a system; optimizing one transition in isolation degrades the whole.

The correct optimization target is not the survival rate but the quality of the survivors. If the 109 BRANCH-stage correlations and 55 NEAR MATURE findings are genuinely well-supported, with explicit falsifiability conditions and rigorous evidential backing, then the system is working as intended. The 5,155 correlations that did not survive SPROUT are not lost effort, they are the filter that makes the survivors meaningful.

There is a deeper philosophical point. The maturity gradient embodies a commitment to the proposition that knowledge is expensive. Not computationally expensive, the cost of running a deliberation session is trivial, but epistemically expensive. Genuine cross-domain insight requires evidence from multiple fields, compatibility of methodologies, explicit falsifiability conditions, and survival under dialectical challenge. These requirements cannot be relaxed without changing what counts as knowledge. The gradient is not a bottleneck to be widened; it is a definition to be respected.

6 What the Gradient Reveals About Discovery

The maturity gradient is not only a quality filter; it is a diagnostic instrument. The distribution of correlations across stages reveals properties of the discovery process itself.

The concentration of 96.3% of correlations at the SPROUT stage indicates that the detection system is far more generative than the evaluation system is selective, that is, the system detects potential connections much faster than it can rigorously evaluate them. This is by design: the detection layer casts a wide net (identifying any structurally plausible connection), while the evaluation layer applies stringent criteria (requiring evidential support and falsifiability). The asymmetry between detection and evaluation rates creates the SPROUT accumulation.

The 50.5% survival rate from BRANCH to NEAR MATURE suggests that the deliberation process is well-calibrated: correlations that survive the primary quality filter are likely to survive final review. If this rate were much lower (say, 10%), it would indicate that the deliberation is passing candidates that final review consistently rejects, a sign of miscalibrated criteria. If it were much higher (say, 95%), it would suggest that final review is adding little value and could be eliminated. The observed rate suggests that both stages are doing meaningful, complementary work.

The relationship between the maturity gradient and the fertility distribution documented in AP-006 raises a further question: do fertile domain pairs produce correlations that mature more easily, or do they simply produce more candidates at the detection stage? If the SPROUT-to-BRANCH survival rate is constant across domain pairs, then fertile pairs produce more mature findings simply by generating more candidates. If the survival rate varies by pair, higher for structurally complementary pairs, lower for structurally sterile ones, then the fertility distribution understates the true advantage of complementary pairs. Preliminary evidence from deliberation outcomes (AP-006, Table 2) suggests the latter: the Materials Science × Life Sciences pair produces not only more correlations but a higher proportion of high-quality findings. The gradient is steeper in sterile territory.

7 Sources of Uncertainty and Limitations

The maturity distribution is a snapshot from a single system at a specific point in time. The distribution may shift as the knowledge base grows and as deliberation parameters are refined. The corpus has grown from 5,037 to 5,351 correlations during the current analysis period, and the stage distribution has evolved correspondingly.

The stage definitions are specific to our system's architecture. Different discovery systems with different quality criteria would produce different stage definitions and different attrition rates. The specific numbers (2.1% SPROUT-to-BRANCH survival) are properties of our system, not universal constants.

The comparisons with drug discovery, venture capital, and peer review are illustrative, not rigorous. The stages in each domain involve fundamentally different activities, costs, and timescales. The structural similarity of the funnels does not imply equivalence of the processes. What the comparison demonstrates is that attrition rates of 1–2% at the critical quality transition are within the range observed across quality-producing systems, suggesting that our gradient is neither pathologically selective nor pathologically permissive.

We do not reveal how promotion between stages works. The quality criteria for each transition are governed by constitutional provisions that are proprietary. We report the attrition rates as empirical observations without specifying the mechanisms that produce them.

8 Conclusion

Cross-domain discovery follows a quantifiable maturity gradient with attrition rates comparable to other quality-producing systems. Of 5,351 detected correlations, approximately 1.1% reach near-maturity. The critical filter is the transition from initial evaluation to validated status, where approximately 2.1% of candidates survive multi-actor deliberation. This attrition is not a deficiency but a structural requirement: the value of mature findings depends on the rigor of the process that selects them.

The gradient reveals properties of the discovery process itself: the asymmetry between detection and evaluation rates, the calibration of sequential quality filters, and the relationship between domain-pair fertility and maturation success. These properties are diagnostic—they tell us not only how much knowledge the system produces but how it produces it.

The maturity gradient suggests that cross-domain discovery systems should be evaluated not by how many connections they detect but by the quality and survival rate of connections that mature through rigorous evaluation. Detection is cheap. Maturation is expensive. Quality lives in the gradient.

References

[1] DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. (2016). Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. Journal of Health Economics, 47, 20–33.

[2] Gompers, P. A., Gornall, W., Kaplan, S. N. & Strebulaev, I. A. (2020). How Do Venture Capitalists Make Decisions? Journal of Financial Economics, 135(1), 169–190.

[3] Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. (2013). Atypical Combinations and Scientific Impact. Science, 342(6157), 468–472.

[4] Bai, Y., Kadavath, S., et al. (2022). Constitutional AI. arXiv:2212.08073.

[5] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLOS Medicine, 2(8), e124.

[6] Kuhn, T. (1962). The Structure of Scientific Revolutions. University of Chicago Press.

[7] Rafailov, R., Sharma, A., et al. (2023). Direct Preference Optimization. NeurIPS 2023.

[8] Rosenthal, R. (1979). The File Drawer Problem and Tolerance for Null Results. Psychological Bulletin, 86(3), 638–641.