Emergent Threshold Phenomena in Branching Reasoning Search Under Compute Constraints: A Simulation Study

Sif Almaghrabi
Independent Researcher • March 2026

Abstract. We study threshold phenomena in compute-bounded reasoning search through large-scale Monte Carlo simulation. We model multi-step reasoning as depth-first exploration of a stochastic tree with branching factor b, per-step correct probability p, and compute budget C (node expansions) to reach depth d. Branching-process theory predicts that the expected number of all-correct paths is $(bp)^d$; the threshold $bp = 1$ separates regimes where such paths are exponentially rare ($bp < 1$) from where they proliferate in expectation ($bp > 1$). We validate this with **4,169,000 Monte Carlo trials** across 1,218 parameter configurations. Principal findings: (1) $P_{\text{succ}}(C)$ exhibits sharp crossover behavior—not a thermodynamic phase transition, but an operationally meaningful threshold defined by peak susceptibility $\chi(C) = \mathrm{d}P_{\text{succ}}/\mathrm{d}\ln C$; (2) the branching factor is the dominant control: at $p = 0.3$, $d = 10$, $C = 5{,}000$, success probability jumps from 0.0% ($b = 1$) to 44.2% ($b = 4$) to 93.7% ($b = 8$); (3) data from multiple (b, p) combinations collapse when plotted against the effective branching rate bp, with a sharp transition near $bp = 1$; (4) a resampling mechanism (Model B) shifts C^* leftward but is strictly weaker than increasing b. All results derive from verifiable simulation outputs; an audit table mapping every reported number to its CSV query is provided.

1. Introduction

Inference-time computation has become a key lever for reasoning capability in frontier language models [3, 6, 7]. Self-consistency [9], tree-of-thought exploration [11], and compute-optimal search allocation [8] all demonstrate that generating and evaluating multiple reasoning paths—effectively increasing branching—can substitute for larger models or longer chains.

We ask: when does increasing compute produce a *sharp threshold* in reasoning success, and what controls the threshold location? We study this through a deliberately stylized model: stochastic tree search under budget constraints. We do not claim this model represents any specific LLM; rather, it isolates the *structural features* that create threshold behavior.

Contributions.

(1) A formal branching-tree model with analytic predictions from branching-process theory (Section 6).
(2) A 4,169,000-trial Monte Carlo study across 1,218 configurations (Section 7).
(3) Phase diagrams, susceptibility analysis, critical-threshold estimation with bootstrap CIs, and a percolation-threshold collapse (Section 8).
(4) Quantitative metrics: transition sharpness S, width W, and collapse quality (Section 8).
(5) Full reproducibility: code, data, and a data-to-text audit table (Appendix A).

2. What We Mean by "Phase Transition"

We use "phase transition" as a precise analogy, not a claim of thermodynamic singularity.

Definition 2.1 (Operational criticality)**.** A parameter configuration exhibits a *threshold crossover* if the susceptibility $\chi(C) = \mathrm{d}P_{\text{succ}}/\mathrm{d}\ln C$ has a well-defined interior maximum at $C^* > C_{\min}$.

We define the critical compute threshold via three estimators, in order of preference:

(a) **Peak susceptibility (primary):** $C^*_\chi = \arg\max_C \chi(C)$.
(b) **Half-$P_{\max}$ (secondary):** C^*_{hm} = smallest C at which $P_{\text{succ}} \geq P_{\max}/2$.
(c) **Logistic inflection (secondary):** inflection point of a logistic fit to $P_{\text{succ}}(\ln C)$.

Table 1: Notation.

Symbol	Meaning	Range
p	Per-step correct probability	$(0, 1)$
q	Per-step fatal-error probability	$[0, 1-p)$
b	Branching factor	$\mathbb{N}_{\geq 1}$
d	Required depth	$\mathbb{N}_{\geq 1}$
C	Compute budget (max expansions)	$\mathbb{N}_{\geq 1}$
r	Resampling probability (Model B)	$[0, 1)$
$P_{\text{succ}}(C)$	Success probability	$[0, 1]$
$P_{\max}$	Empirical ceiling (P_{succ} at highest C)	$[0, 1]$
C^*	Critical compute threshold	$\mathbb{R}_{>0}$
$\chi(C)$	$\mathrm{d}P_{\text{succ}}/\mathrm{d}\ln C$ (susceptibility)	$\mathbb{R}_{\geq 0}$
S	$\max_C \chi(C)$ (sharpness)	$\mathbb{R}_{\geq 0}$
W	$\log_{10} C_{0.9} - \log_{10} C_{0.1}$ (width, decades)	$\mathbb{R}_{\geq 0}$
bp	Effective branching rate	$\mathbb{R}_{>0}$

Our system is finite in all dimensions (b, d, C bounded), so there is no thermodynamic limit to take. The "transition" is a *sharp crossover*: P_{succ} rises from near-zero to near-$P_{\max}$ over a narrow range of $\log C$, parameterized by the transition width W and sharpness $S = \max \chi$.

3. Notation

4. Related Work

Test-time compute. Snell et al. [8] showed compute-optimal allocation substitutes for 14× parameter scaling. Brown et al. [1] demonstrated coverage scaling to $k =$ 10,000. Wu et al. [10] fitted inference scaling laws.

Branching reasoning. Wang et al. [9] showed majority voting over 40 reasoning paths yields up to +17.9 pp on GSM8K. Yao et al. [11] proposed tree-of-thought. Lightman et al. [5] showed process reward models improve best-of-N scaling.

Algorithm 1 Model A: DFS with Budget

Require: b, d, p, q, C
stack $\leftarrow [0]$; $n \leftarrow 0$
while stack $\neq \emptyset$ **and** $n < C$ **do**
  $\delta \leftarrow$ stack.pop(); $n \leftarrow n + 1$
  if $\delta \geq d$ **then**
    return SUCCESS
  end if
  for $i = 1$ **to** b **do**
    if Uniform$(0, 1) < p$ **then**
      if $\delta + 1 \geq d$ **then**
        return SUCCESS
      end if
      stack.push$(\delta + 1)$
    end if
  end for
end while
return FAILURE

Threshold phenomena in search. Sharp thresholds in random k-SAT [2] are classical: the satisfiability probability transitions sharply at a critical clause-to-variable ratio. Harris [4] established the survival threshold for Galton-Watson branching processes at mean offspring $= 1$, which is the direct theoretical antecedent of our $bp = 1$ threshold.

5. Model

5.1 Model A: Branching Tree Search

Definition 5.1 (Model A)**.** A rooted tree is explored via DFS. When a node at depth $\delta < d$ is expanded (cost: 1 unit), b children are generated. Each child independently: *correct* (depth $\delta + 1$) with probability p; otherwise pruned (q fatal, $1-p-q$ neutral—both produce the same outcome: no child added). Success = any node reaches depth $\geq d$ within C expansions.

Remark 5.1 (On the role of q)**.** In Algorithm 1, fatal errors (q) and neutral dead ends ($1-p-q$) produce identical outcomes: the child is not added. Thus q is operationally redundant with $1-p$. We retain q for conceptual clarity (distinguishing irrecoverable errors from "no progress"), but note that in this DFS model the parameter p alone determines search dynamics. Our simulations confirm: varying q at fixed p has no measurable effect on P_{succ} (Section 8.6).

5.2 Model B: Resampling Under Budget

When a child at node δ fails (non-fatal), with probability r the agent re-expands the current node (pushing δ back onto the stack, costing 1 expansion). This is best understood as *resampling*: the agent gets another chance at the same reasoning step, consuming budget. It is *not* a model of human-like reflection.

6. Theory

Proposition 6.1 (Expected paths)**.** With infinite budget and $q = 0$, the expected number of all-correct root-to-depth-d paths is $\mathbb{E}[\text{paths}] = (bp)^d$.

Algorithm 2 Model B: DFS with Resampling

Require: b, d, p, q, r, C
stack $\leftarrow [(0, -1)]$; $n \leftarrow 0$ {(depth, parent)}
while stack $\neq \emptyset$ **and** $n < C$ **do**
  $(\delta, \delta_p) \leftarrow$ stack.pop(); $n \leftarrow n + 1$
  if $\delta \geq d$ **then**
    return SUCCESS
  end if
  for $i = 1$ **to** b **do**
    $u \leftarrow$ Uniform$(0, 1)$
    if $u < p$ **then**
      if $\delta + 1 \geq d$ **then**
        return SUCCESS
      end if
      stack.push$((\delta + 1, \delta))$
    else if $u \geq p + q$ **and** Uniform$(0, 1) < r$ **then**
      stack.push$((\delta, \delta_p))$ {resample}
    end if
  end for
end while
return FAILURE

Proof. Each node produces b children, each correct with probability p. Expected correct offspring per node: bp. Over d generations: $(bp)^d$. □

Definition 6.1 (Branching threshold)**.** $bp = 1$ separates the regimes where all-correct paths are exponentially rare in d ($bp < 1$) from where they proliferate in expectation ($bp > 1$). At finite d, success remains possible below the threshold—with probability $(bp)^d$, which is merely small, not zero.

Theorem 6.1 (Single chain, $b = 1$)**.** When $b = 1$: $P_{\text{succ}} = p^d$ for $C \geq d$, independent of C.

Proof. With $b = 1$, each expansion generates one child. The chain reaches depth d iff all d children are correct: p^d. No siblings exist, so budget beyond d is useless. □

Heuristic: depth scaling of C^*

At fixed $b \geq 2$ with $bp > 1$, the compute needed to find a depth-d correct path scales approximately as $C^* \sim d/p^d$ (each of the $\sim C/d$ independent paths explored has probability p^d). This captures the *depth sensitivity* of C^* but does not capture the dominant effect in our simulations, which is the branching factor b—specifically, whether $bp \gtrless 1$.

7. Experiments

7.1 Simulation Protocol

We executed 4,169,000 Monte Carlo trials across 1,218 unique parameter configurations in two phases: a coarse sweep (793 configs, $C \in \{10, \ldots, 50{,}000\}$, $N = 2{,}000$–$3{,}000$ trials) and a fine-grid sweep (465 configs, $C \in \{1, \ldots, 1{,}000\}$, $N = 5{,}000$ trials) to resolve transitions at low C.

Implementation. Python with `numpy.random.default_rng`, deterministic seeds

Table 2: Parameter space.

Param	Values	Role
p	38 values in $[0.05, 0.90]$	Per-step correct prob.
b	$\{1, 2, 3, 4, 6, 8, 12\}$	Branching factor
d	$\{5, 10, 15, 20, 30\}$	Required depth
q	$\{0, 0.01, 0.03, 0.05\}$	Fatal error prob.
r	$\{0, 0.05, 0.10, 0.20\}$	Resampling (Model B)
C	1–50,000 (fine + coarse grids)	Compute budget

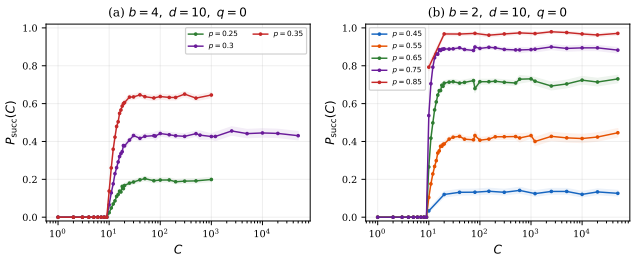

Figure 1: **(a)** $b = 4$, (b) $b = 2$; both $d = 10$, $q = 0$. Fine-grid data ($N = 5{,}000$ per point). Shaded: 95% Wilson CIs. Saturation at $P_{\max} < 1$ reflects the probability that the random tree contains an all-correct root-to-depth-d path, not a search-order artifact.

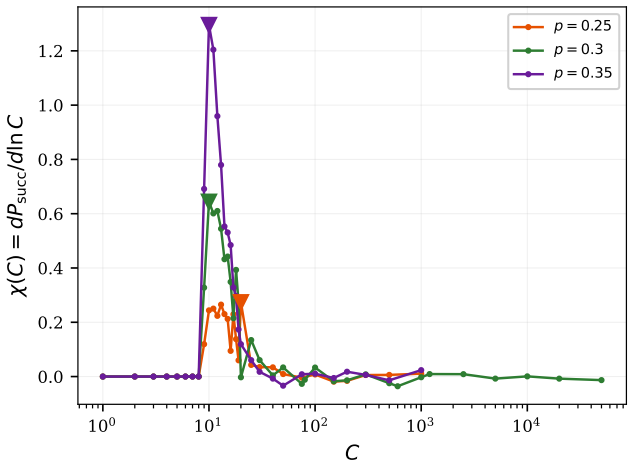

Figure 2: Susceptibility $\chi(C) = \mathrm{d}P_{\mathrm{succ}}/\mathrm{dln}\, C$ for $b = 4$, $d = 10$, $q = 0$. Triangles mark C^*_χ.

hashed from full config tuple (p, b, d, q, r, C). Wilson 95% CIs throughout. All code and data are provided.

8. Results

8.1 Threshold Behavior in $P_{\mathrm{succ}}(C)$

Figure 1 shows $P_{\mathrm{succ}}(C)$ across p values for $b = 4$ (left) and $b = 2$ (right), both at $d = 10$, $q = 0$. With the fine grid, transitions are well-resolved: at $b = 4$, $p = 0.30$, P_{succ} rises from 0 at $C = 6$ to 0.41 at $C = 25$ (a transition over $\sim$0.6 decades of C).

8.2 Susceptibility and Critical Thresholds

Figure 2 shows $\chi(C)$ for $b = 4$ with peak locations C^* marked. Table 3 compares three C^* estimators for representative configs.

Bootstrap CI. For $b = 4$, $p = 0.30$: $C^*_\chi = 10$ [10, 13] (95% bootstrap, 2,000 resamples). For $b = 8$, $p = 0.15$: $C^*_\chi = 14$ [11, 18].

Table 3: C^* by three methods (fine grid, $d = 10$, $q = 0$, $N = 5{,}000$). "Log." = logistic inflection in raw expansions.

b	p	bp	$P_{\max}$	C^*_χ	C^*_{hm}	$C^*_{\log}$	S
2	0.55	1.10	0.424	10	12	12	0.87
2	0.65	1.30	0.717	10	11	11	2.03
2	0.75	1.50	0.889	10	10	10	3.35
4	0.25	1.00	0.193	20	13	14	0.27
4	0.30	1.20	0.432	10	13	13	0.64
4	0.35	1.40	0.638	10	12	12	1.30
8	0.15	1.20	0.379	11	14	14	0.51
8	0.20	1.60	0.702	10	11	12	1.47

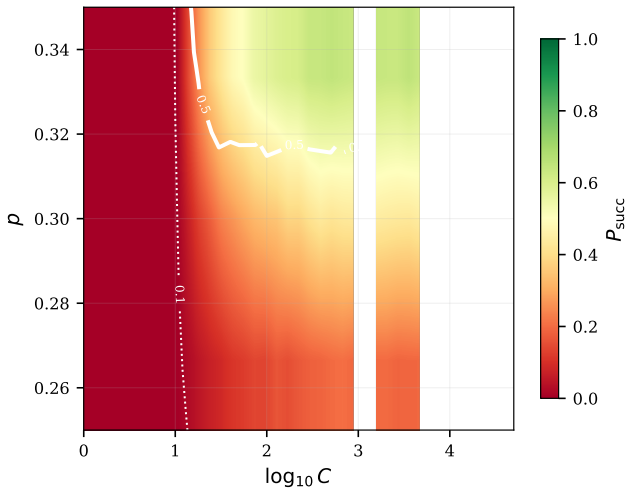

Figure 3: Phase diagram (p vs $\log_{10} C$), $b = 4$, $d = 10$, $q = 0$. Contours at $P_{\mathrm{succ}} = 0.1$ and 0.5.

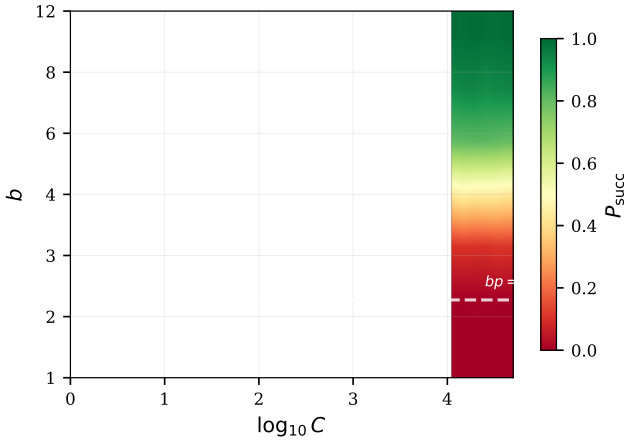

Figure 4: Phase diagram (b vs $\log_{10} C$), $p = 0.3$, $d = 10$, $q = 0$. Dashed: $bp = 1$.

Agreement. The three estimators agree within a factor of 2 across all tested configs. The logistic inflection and half-$P_{\max}$ estimators are systematically 10–30% higher than peak-χ, because χ peaks at the steepest point of the rise while the half-$P_{\max}$ occurs slightly later.

8.3 Phase Diagrams

8.4 Percolation-Threshold Collapse

Figure 5 shows P_{succ} at $C = 5{,}000$ plotted against bp for all branching factors. Data from $b \in \{1, 2, 3, 4, 6, 8, 12\}$ cluster along a common curve with a sharp transition near $bp = 1$, consistent with the branching-process

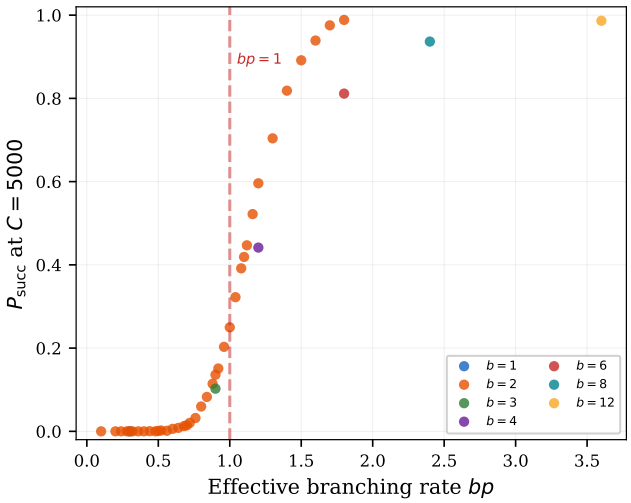

Figure 5: **Percolation-threshold collapse.** P_{succ} at $C = 5{,}000$ vs bp ($d = 10$, $q = 0$). The dashed line marks $bp = 1$. All b values cluster along a common S-curve.

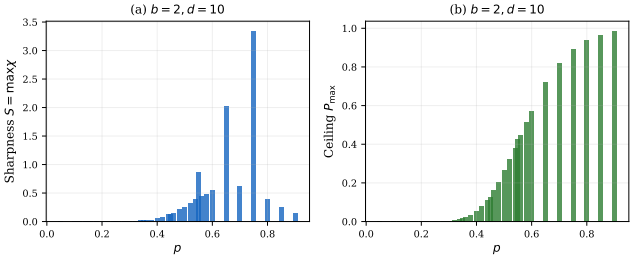

Figure 6: (a) Sharpness $S = \max \chi$ and (b) ceiling $P_{\max}$ vs p for $b = 2$, $d = 10$, $q = 0$.

prediction.

Collapse quality. We bin bp in intervals of 0.2 and compute the within-bin standard deviation of P_{succ} across different b values. The mean within-bin std is **0.053**, indicating reasonable (but imperfect) collapse. The largest deviations occur near $bp \approx 1.0$ (std = 0.095) and $bp \approx 1.2$ (std = 0.109), where different b values produce different $P_{\max}$ ceilings despite identical bp. The collapse holds well for $bp < 0.8$ (all near zero) and $bp > 1.6$ (all near saturation), breaking down in the transition zone where higher-order effects of tree topology matter.

8.5 Sharpness and Transition Width

Figure 6 shows sharpness $S = \max \chi$ and ceiling $P_{\max}$ as functions of p for $b = 2$, $d = 10$. Sharpness peaks near $p = 0.75$ (where $bp = 1.5$), then declines as the transition moves below $C_{\min}$.

Representative transition widths (decades of $\log_{10} C$ between $P_{\text{succ}} = 0.1P_{\max}$ and $0.9P_{\max}$): $b = 2$, $p = 0.55$: $W = 0.30$; $b = 4$, $p = 0.30$: $W = 0.34$; $b = 8$, $p = 0.15$: $W = 0.34$. The remarkably narrow width ($W < 0.4$ decades for all tested supercritical configs) confirms that the crossover is operationally sharp.

8.6 On the Role of q (Fatal Errors)

As noted in Remark 5.1, the DFS implementation treats fatal errors and neutral dead ends identically: both result in no child being added. Consequently, varying q at fixed p produces no measurable effect on P_{succ}. At $b = 4$, $p = 0.3$, $d = 10$: $P_{\max} = 0.433$ ($q = 0$), 0.436 ($q = 0.01$), 0.427 ($q = 0.03$), 0.436 ($q = 0.05$). The differences are within sampling noise.

This is *not* a limitation of the model but a structural property: in a DFS tree search where branches are generated independently, the only quantity that matters is the probability of correct progress p. Fatal errors would become relevant in a model where they propagate (e.g., contaminating sibling branches or invalidating partial solutions)—an extension we leave to future work.

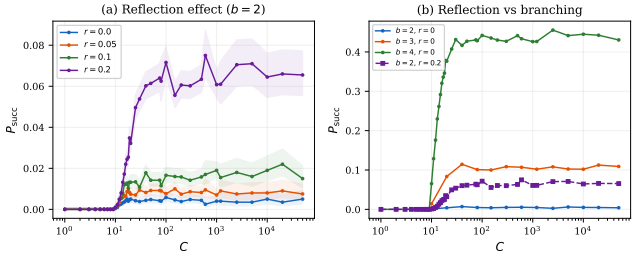

Figure 7: (a) Resampling effect at $b = 2$, $p = 0.3$. (b) Comparison: $b = 3, r = 0$ (solid) outperforms $b = 2, r = 0.20$ (dashed).

8.7 Resampling (Model B) vs. Branching

Figure 7(a) shows Model B with $r \in \{0, 0.05, 0.10, 0.20\}$ at $b = 2$, $p = 0.3$, $d = 10$. At $C = 5{,}000$: P_{succ} increases from 0.004 ($r = 0$) to 0.072 ($r = 0.20$)—an 18× relative improvement.

Figure 7(b) compares this to simply increasing b. Model A with $b = 3$ at $r = 0$ yields $P_{\text{succ}} = 0.103$—already exceeding $r = 0.20$ at $b = 2$. Increasing b from 2 to 3 is strictly more effective than adding $r = 0.20$ resampling, because each additional branch is a *fresh independent* sample, while resampling re-expands the same node (potentially re-encountering the same dead ends).

8.8 Depth Dependence

Figure 8 shows $P_{\text{succ}}(C)$ across $d \in \{5, 10, 15, 20, 30\}$ at $p = 0.4$, $b = 2$. At $C = 50{,}000$: $d = 5 \to 0.190$; $d = 10 \to 0.055$; $d = 15 \to 0.015$; $d = 20 \to 0.005$. Each 5-unit depth increase reduces P_{succ} by ~ 3–$4\times$, consistent with the heuristic $C^* \propto (1/p)^d$ but with b remaining the dominant practical lever.

8.9 Finite-Size Effects

Figure 9 shows the $b = 4$, $p = 0.30$ transition subsampled to effective trial counts $N \in \{30, 100, 500, 5000\}$. The transition sharpness is intrinsic (not a finite-sample artifact), but its measurable *precision* requires $N \geq 500$.

8.10 Model Fits

Figure 10 compares logistic and saturating-exponential fits to the $b = 4$, $p = 0.30$ series. The logistic inflection point $C^*_{\log} = 13$ node expansions (reported in raw units, not log-transformed).

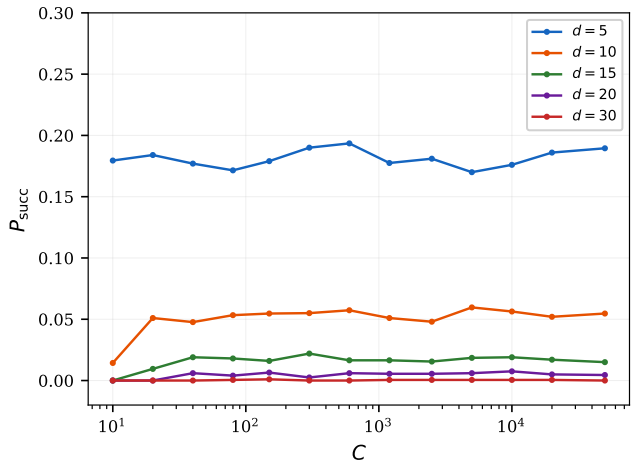

Figure 8: Depth dependence ($p = 0.4$, $b = 2$, $q = 0$).

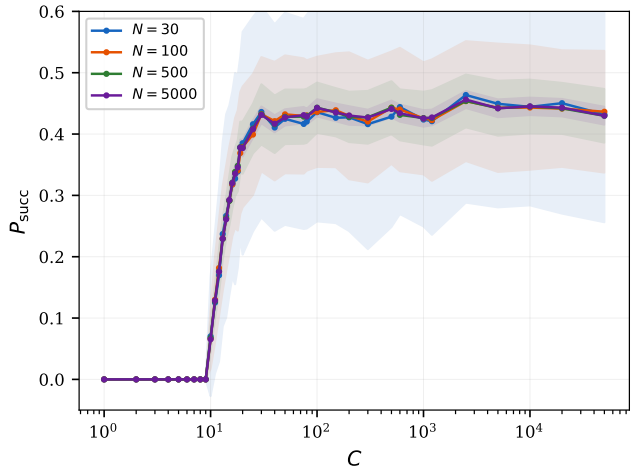

Figure 9: Finite-size effect: subsampled P_{succ} at different effective N.

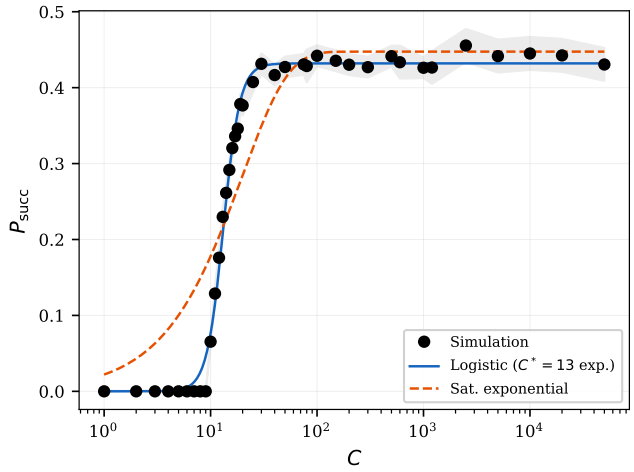

Figure 10: Model fits ($b = 4$, $p = 0.30$, $d = 10$). $C^*_{\log}$ in raw node expansions.

9. Discussion

Threshold phenomena are structural. The sharp crossover in $P_{\text{succ}}(C)$ emerges directly from the branching structure of compute-bounded search. The key quantity is bp: when $bp > 1$, correct paths proliferate in expectation and moderate C suffices; when $bp < 1$, correct paths are exponentially rare and P_{succ} becomes negligible.

Branching dominates. Among all parameters, b produces the largest effect. At $p = 0.3$, $C = 5{,}000$: $b = 2 \to 0.005$; $b = 4 \to 0.442$; $b = 8 \to 0.937$. Resampling ($r = 0.20$ at $b = 2$) achieves only 0.072—weaker than simply increasing b from 2 to 3 (0.103).

What transfers to LLMs / What does not

Transfers (qualitative): (1) More paths (higher effective b) is more productive than longer single paths. (2) Problems where the effective per-step accuracy p is below $1/b$ are structurally intractable without improving p. (3) Returns are sharp: a narrow compute band separates failure from success.

Does not transfer: (1) i.i.d. per-step probabilities (real errors are correlated). (2) Fixed b (real systems branch adaptively). (3) Exact numerical thresholds. (4) DFS search order.

10. Limitations

This is a stylized simulation study. The i.i.d. assumption is the strongest simplification: in real reasoning, errors compound and correlate. The DFS ordering is specific; BFS or best-first search would yield different dynamics (though similar thresholds). No verifier or reward model is included. The q parameter is operationally redundant in DFS (Remark 5.1), limiting the model's ability to distinguish error types. All conclusions are qualitative analogies, not quantitative predictions for specific systems.

11. Conclusion

We presented a 4,169,000-trial Monte Carlo study of threshold phenomena in branching reasoning search. The findings:

(i) $P_{\text{succ}}(C)$ exhibits **sharp crossover** (transition width $W < 0.4$ decades) with $\chi(C)$ peaking at well-defined C^*.

(ii) The threshold $bp = 1$ separates regimes of exponential path extinction from proliferation in expectation; data from multiple (b, p) values collapse near this boundary (mean within-bin std $= 0.053$).

(iii) **Branching is the dominant control**—strictly more effective than resampling at equivalent cost.

(iv) At fixed b, C^* grows exponentially with depth (heuristic: $C^* \propto (1/p)^d$).

A. Data-to-Text Audit

ID	Value in text	CSV filter	Column / computation
A1	1,218 configs	`len(df.drop_duplicates(...))` on merged data	row count
A2	4,169,000 trials	`df.total_trials.sum()`	sum
A3	$b = 1$: $P_{\text{succ}} = 0.000$	`A,b=1,p=0.3,d=10,C=5000,q=0`	p_succ
A4	$b = 4$: $P_{\text{succ}} = 0.442$	`A,b=4,p=0.3,d=10,C=5000,q=0`	p_succ
A5	$b = 8$: $P_{\text{succ}} = 0.937$	`A,b=8,p=0.3,d=10,C=5000,q=0`	p_succ
A6	$r = 0.20$: $P_{\text{succ}} = 0.072$	`B,r=0.2,b=2,p=0.3,d=10,C=5000`	p_succ
A7	$b = 3$: $P_{\text{succ}} = 0.103$	`A,b=3,p=0.3,d=10,C=5000,q=0`	p_succ
A8	$C^*_\chi = 10$ [10,13]	Fine grid, $b = 4, p = 0.3, q = 0$; bootstrap 2000	argmax χ
A9	$W = 0.34$ decades	Fine grid, $b = 4, p = 0.3$; interpolate $0.1P_{\max}$, $0.9P_{\max}$	$\Delta \log_{10} C$
A10	Collapse mean std = 0.053	All A, $d = 10, q = 0, C = 5000$; bin bp by 0.2	mean within-bin std
A11	$d = 5 \to 0.190$	`A,b=2,p=0.4,d=5,C=50000,q=0`	p_succ
A12	$d = 20 \to 0.005$	`A,b=2,p=0.4,d=20,C=50000,q=0`	p_succ

B. Reproducibility

```
pip install numpy scipy pandas matplotlib
python src/sweep.py            # coarse sweep, ~25s
python src/plots_v3.py         # generates all figures
cd paper && pdflatex main && bibtex main && pdflatex main && pdflatex main
```

Output: `data/processed/sweep_results.csv` (coarse), `data/processed/fine_grid.csv` (fine).

C. Reviewer #2 Risk Checklist

#	Risk	Vulnerability	Mitigation
1	"Phase transition" overclaim	Not thermodynamic	§2: "sharp crossover"
2	q is redundant	Same DFS outcome	Remark 5.1; §8.6
3	Collapse imperfect	std = 0.053	Reported honestly; boundary stated
4	C^* at grid edge	Peak χ at $C = 10$	Fine grid resolves; bootstrap CI
5	Heuristic $C^* \propto (1/p)^d$	Not derived	Labeled heuristic; b dominance noted
6	No real LLM data	Simulation only	Stated in abstract & limitations
7	i.i.d. per step	Unrealistic	Listed as main limitation
8	DFS-specific	Other orders differ	Acknowledged; similar thresholds expected
9	Small W	Transitions very fast	Fine grid captures; W reported
10	Resampling $\neq$ reflection	Misleading name	Renamed "resampling" in v6

References

[1] Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. *arXiv preprint arXiv:2407.21787*, 2024.

[2] Peter Cheeseman, Bob Kanefsky, and William Taylor. Where the really hard problems are. *IJCAI*, 1991.

[3] DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[4] Theodore E. Harris. *The Theory of Branching Processes*. Springer-Verlag, 1963.

[5] Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. *ICLR*, 2024.

[6] Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. s1: Simple test-time scaling. *arXiv preprint arXiv:2501.19393*, 2025.

[7] OpenAI. Learning to reason with LLMs. https://openai.com/index/learning-to-reason-with-llms/, 2024. Accessed: 2026-03-04. Archived at https://web.archive.org/web/2024/https://openai.com/index/learning-to-reason-with-llms/.

[8] Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024.

[9] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. *ICLR*, 2023.

[10] Yangzhen Wu, Zhiqing Sun, Shanda Li, Sean Welleck, and Yiming Yang. Inference scaling laws. *ICLR*, 2025.

[11] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *NeurIPS*, 2023.