WHEN DOES CHAIN-OF-THOUGHT HELP: A MARKOVIAN PERSPECTIVE

Zihan Wang, Yijun Dong & Qi Lei
New York University
{zw3508, yd1319, ql518}@nyu.edu

ABSTRACT

Chain-of-Thought (CoT) prompting is a widely used inference-time technique for improving reasoning, yet its gains are uneven across tasks. We analyze when and why CoT helps by modeling the step-wise reasoning trajectory as a Markov chain. Each intermediate step is a state and the dependence between steps is captured by a transition kernel. Our theory identifies transition alignment, whether instances share a common step-wise transition kernel, as the key determinant of CoT's effectiveness. When transitions are identical across steps, CoT reduces inference-time sample complexity: fewer context sample trajectories suffice to recover the final decision. In contrast, when transitions differ across steps, these gains can vanish. We further quantify how noise in intermediate steps modulates CoT's benefit. Beyond theory, we design synthetic benchmarks that isolate these factors to complement prior results on real-world tasks and to empirically validate our predictions.

1 INTRODUCTION

Chain-of-Thought (CoT) prompting has become a de facto approach of inference-time scaling for multi-step reasoning with large language models (LLMs). It demonstrates substantial improvements on math and symbolic tasks (Wei et al., 2022; Kojima et al., 2022; Wang et al., 2023; Zhou et al., 2022; Yao et al., 2023) but modest or mixed effects on others (Sprague et al., 2025). Noisy or unfaithful intermediate reasoning steps have even been shown to mislead predictions of CoT and bring worse performance than direct inference, despite the additional computation of CoT (Prabhakar et al., 2024).

Since its first introduction, CoT has been extensively investigated empirically and theoretically. Mechanism-centric work explores how to aggregate inference-time trajectories (e.g., self-consistency, tree search) and offers qualitative rationales (Wang et al., 2023; Yao et al., 2023). Task-centric works analyzed catalogs where CoT tends to help (e.g., math, symbolic, multi-hop) and where it does not (Sprague et al., 2025; Cobbe et al., 2021; Lewkowycz et al., 2022; Yang et al., 2018). Along this task-centric perspective, a critical missing piece is a rigorous yet intuitive theoretical model of CoT that explains its successes and failures on different downstream tasks. This motivates our central research questions:

When does CoT provably outperform direct inference?
Can we distinguish beneficial cases of CoT from its failures via measurable structural properties of the downstream task?

Mechanistic understandings of these questions will (i) provide first-principled guidelines for where to apply CoT and how to structure in-context demonstrations, (ii) clarify which aspects of the downstream tasks make CoT appear strong or fragile, and (iii) steer the design of clean evaluation metrics for CoT that disentangle different sources of failures. This work makes a step toward filling these gaps through theory-backed accounts and controlled experiments.

In light of the stochastic and autoregressive nature of CoT, along with the finite context windows in practice, we adapt the Markovian view of reasoning from Xie et al. (2021); Prystawski et al. (2023); Besta et al. (2024); Abbe et al. (2024); Sanford et al. (2024); Kim et al. (2025) and compare CoT

against direct inference through a unified lens. Conceptually, we formalize step-wise reasoning as a trajectory over latent states and study how two ingredients govern sample efficiency: **transition alignment** across steps ("same skill" vs. "different skills") and **noise** separating correct from competing options. Intuitively, identical local rules can enhance per-step votes through the reasoning process, leading to a performance gain for CoT. And because composed, end-to-end margins contract under uncertainty, the scale of intermediate-step noise also plays an important role in reasoning. We keep the formal modeling and decision rule in Section 3.

Specifically, we provide a compact analysis of inference-time CoT that isolates the two factors above and turns them into concrete predictions. Our theory shows when CoT enjoys a structural $1/T$-type improvement and when it does not, and also explains how noise affects in Section 4. The proof sketch of main results is in Section 5. In Section 6, we then validate these predictions in clean synthetic settings that manipulate alignment and noise directly, and in a small arithmetic task that offers a practical sanity check.

Our contributions are as follows, which offer a compact answer to our essential question together.

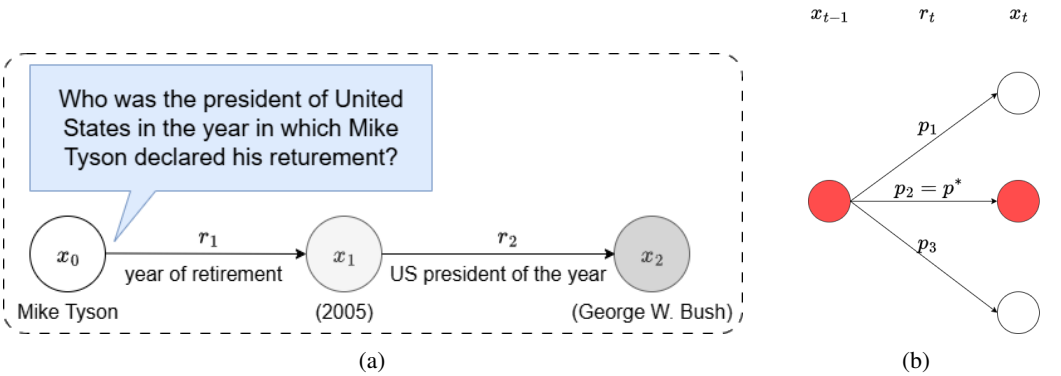
1. **A Markovian modeling and decision rule for inference-time aggregation.** We model reasoning as a Markov chain over a finite state space and analyze inference time only. This isolates how context samples translate into decisions, aligning with trajectory aggregation practices such as self-consistency and search (Wang et al., 2023; Yao et al., 2023).
2. **Theory that pinpoints the factors behind CoT's benefits.** We derive sample-complexity bounds for (a) direct inference, and (b) CoT with aligned and misaligned transitions. From the theoretical results, we demonstrate two key factors in CoT performance.
3. **Clean synthetic tests that manipulate alignment and noise.** We design clean benchmarks showing that CoT reduces the sample budget most when transitions align, and its relative advantage increases with noise. We also verify our results on a more realistic but still structured modular addition task.

2 Related Works

When does CoT help, (and when it doesn't)? Large-scale empirical assessments unveil that the largest gains of CoT tend to concentrate on math/symbolic tasks, with much smaller effects elsewhere; failures often arise from noisy or unfaithful intermediate steps, suggesting the need for aggregation or verification to counteract error propagation (Prabhakar et al., 2024; Sprague et al., 2025; Zheng et al., 2025). These observations underscore that task structure and intermediate noise govern the returns to inference-time compute, and thereby motivate our theoretical investigation of skill diversity and intermediate-step noise.

Inference-time scaling via reasoning trajectory aggregation. Inference in autoregressive models consists of forward passes through the trained network to generate tokens, a process once regarded as lightweight and essentially fixed in cost. With the unprecedented capacity of modern LLMs, however, reasoning has emerged as an effective attribution of computation, scaling which significantly improves inference-time performance (Jaech et al., 2024; Guo et al., 2025). CoT prompting (Wei et al., 2022) and its zero-shot variant (Kojima et al., 2022) established the basic paradigms of eliciting intermediate reasoning steps, which can be scaled via assorted schemes. (i) Subsequent methods that explore and aggregate multiple reasoning trajectories (Zhou et al., 2022; Chen et al., 2023; Wang et al., 2023; Yao et al., 2023; Snell et al., 2024), like self-consistency and best-of-n, have been shown to amplify these gains even further. (ii) In addition, CoT has been enhanced through various reinforcement-learning-based methods, such as Monte-Carlo Tree Search (Silver et al., 2018; Trinh et al., 2024; Feng et al., 2023b; Xie et al., 2024) and process reward models (Lightman et al., 2023; Uesato et al., 2022). (iii) Alternatively, self-improvement that refines reasoning trajectories using the model's own judgments (Zelikman et al., 2022; Hosseini et al., 2024; Kumar et al., 2024) has also been used to boost performance at inference. Collectively, these empirical successes of reasoning as inference-time scaling motivate our focus on the structure of multi-step reasoning trajectories and inference-time sample complexity.

Graph and Markovian views on LLMs and CoT. Reasoning on graphs, either externally through knowledge graphs or implicit via graph algorithms, provides an abstraction for multi-step tasks (Xie



(a) (b)

Figure 1: (a) **Example of Markov-style decomposition.** An illustrative instance, where the original text is selected from (Mavi et al., 2024), is parsed into an initial state x_0 and a sequence of relations/-operators $r_{1:T}$. The correct intermediate steps $x_{1:T}$ are also included. The example highlights how text data are collapsed into a formalized sequence, making the intermediate operations explicit for CoT demonstrations. (b) **Stepwise Markov dynamics.** Each r_t induces a transition kernel $P^{(t)}$ from x_{t-1} to x_t and the next node is generated randomly following the transition. At inference, the ground truth intermediate steps are the nodes with the largest probability p^* for each step.

et al., 2021; Prystawski et al., 2023; Besta et al., 2024; Abbe et al., 2024; Sanford et al., 2024; Kim et al., 2025). Several works interpret transformers or their decoding dynamics through a Markov lens, including token-level connections between autoregressive LLMs and finite-state Markov chains (Zekri et al., 2024), the relation between self-attention and Markov models (Ildiz et al., 2024), and mechanistic studies of in-context learning through induction heads (Edelman et al., 2024; Makkuva et al., 2024). Closer to our setting, Kim et al. (2025) casted CoT as a metastable Markov process on thought space, explaining why long reasoning trajectories exhibit phase transitions and suggesting search/RL/distillation schemes to escape local basins. We differ in what is modeled and which question is answered: by positing a task-level Markov chain over latent reasoning states, we analyze the inference-time sample complexity of CoT under different structures of the Markov chain corresponding to different skill diversity and intermediate-step noise of the downstream task.

More theoretical accounts of CoT. Beyond the graph and Markovian views, the benefits and mechanisms of CoT have been extensively investigated from various theoretical perspectives, including expressivity (Feng et al., 2023a; Chiang et al., 2023; Merrill & Sabharwal, 2024; Li et al., 2024b), simulation of optimization algorithms Huang et al. (2025b), inference-time statistical efficiency (Altabaa et al., 2025; Joshi et al., 2025; Wen et al., 2024; Huang et al., 2025a; Kim & Suzuki, 2025; Hu et al., 2024), and the emergence of CoT/reasoning capability from pretraining (Nichani et al., 2024; Li et al., 2024a; Yang et al., 2025; Wang et al., 2025). Our Markovian perspective of CoT is orthogonal and complementary to these, explaining not only the empirical successes but also failures of CoT in terms of sample complexity at the inference time.

3 Markovian Modeling

We model an instance as a sequence of T relations (local rules/operators) applied to an initial state. The ground truth of an instance contains an input x_0, a final output x_T, relations $(r_1, \ldots, r_T)$ and latent intermediate steps $(x_1, \ldots, x_{T-1})$. The model only observes the input and relations $(x_0, r_1, \ldots, r_T)$ and must infer x_T. With CoT, the model may also emit predictions $\hat{x}_1, \ldots, \hat{x}_{T-1}$ to estimate intermediate steps sequentially conditioned on the growing prefix; without CoT it directly predicts $\hat{x}_T$. An illustrative example is shown in Fig. 1a.

This abstraction treats the textual problem as a latent state transformation pipeline: each r_t induces a transition kernel $P^{(t)}$ that maps the current state x_{t-1} to a distribution over next states x_t. Many reasoning tasks naturally decompose this way, for example carrying updates in arithmetic (Cobbe et al., 2021; Lewkowycz et al., 2022; Chen et al., 2023)), symbolic rewrites (Clark et al., 2020; Tafjord

et al., 2021)), knowledge-graph paths (Yang et al., 2017; Hamilton et al., 2018; Ren & Leskovec, 2020), and composing path in multi-hop QA (Yang et al., 2018; Khot et al., 2020; Trivedi et al., 2022)). By collapsing text into a finite set of task-relevant states, we separate *form* (how a chain is written) from *mechanism* (what operation is performed), which is exactly what CoT aims to expose: a trajectory of intermediate operations rather than a single terminal answer.

We work with a finite state space $[k]$ as a compact abstraction of task-relevant equivalence classes (e.g., arithmetic states, intermediate answers in multi-hop QA), not a literal vocabulary. In many benchmarks, a small latent alphabet suffices to determine the final decision. The end-to-end kernel is

$$Q := P^{(1)} P^{(2)} \cdots P^{(T)}, \tag{1}$$

so the i-th row Q_i specifies the distribution of x_T given $x_0 = i$, and the predictor outputs the one-hot index of the maximal entry of Q_i.

Our analysis focuses purely on *inference time*. The model receives n context samples that share the same relation tuple $(r_1, \ldots, r_T)$ and are drawn with $x_0 \sim \mu$ and transitions governed by $\{P^{(t)}\}$. A direct-inference sample reveals only x_T; a CoT sample reveals the full path $(x_0, \ldots, x_T)$. The modeling choice aligns with recent formalisms that cast decoding and trajectory aggregation as Markov-like dynamics (Kim et al., 2025) and with algorithmic views of in-context learning (Xie et al., 2021; Akyürek et al., 2022; Zhang et al., 2024).

A key structural distinction is whether $P^{(1)} = \cdots = P^{(T)}$ (homogeneous/aligned) or the kernels differ by step (heterogeneous/misaligned). In the first case, a single trajectory provides T glimpses of the same local rule, allowing per-step votes to be pooled. In the second, per-step observations do not reinforce any single kernel, so pooling does not have the same effect. This is the precise mathematical counterpart of "same skill versus different skills across steps."

4 Theoretical Analysis

Our goal in this section is to isolate *which factors* govern the benefit of inference-time CoT theoretically. To keep the analysis transparent, we adopt the same decision rule for **both** direct inference and CoT: the model behaves like a simple *counter* that uses the context samples to estimate class frequencies and then predicts the *hard-max* index (argmax) for the relevant row of the transition kernel (Fig. 1b). Within this lens, two ingredients determine whether CoT helps: (i) **transition alignment**, and (ii) **noise**. The first drives the structural $1/T$-type advantage and the second explains when noise amplifies CoT's relative gains.

4.1 Direct Inference

In direct inference, each context sample reveals only the end-to-end output x_T under kernel Q, defined in Eq. (1). With the shared "count-and-argmax" rule, recovering the argmax of each row Q_i reduces to multi-class frequency estimation conditioned on $x_0 = i$. Reliable identification requires that every input state appears with nontrivial frequency and that the optimal class has a positive gap.

Assumption 4.1 (Coverage of distribution)**.** *The multinomial distribution μ satisfies that the smallest element of μ is strictly greater than 0 and denoted by $\mu_{\min}$.*

Assumption 4.2 (Positive margin)**.** *The ground truth transition matrix Q satisfies that*

$$Q_{i,j^*(i)} - Q_{ij} \geq \Delta_Q > 0$$

for all i and $j \neq j^*(i)$, where $j^*(i) = \arg\max_i Q_{ij}$.

Under these conditions, the usual frequency-estimation rate appears:

Theorem 4.1 (Direct inference)**.** *Under Assumption 4.1 and 4.2, if the number of samples n satisfies*

$$n = \Theta\left(\frac{\log(k/\delta)}{\mu_{\min}\Delta_Q^2}\right),$$

then with probability at least $1 - \delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with direct inference.

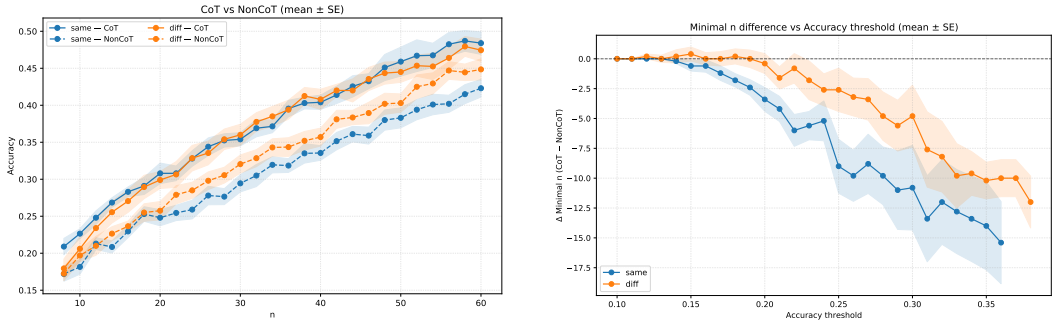


Figure 2: **Synthetic alignment study (same vs. diff).** *Left:* Accuracy as a function of the number of context samples n comparing CoT (solid) and NonCoT (dashed). CoT yields a consistently larger improvement under **same** than under **diff**. *Right:* Sample-complexity proxy $\Delta n(\tau)$ (CoT−NonCoT) vs. target accuracy τ. More negative values indicate fewer samples required by CoT. The widening gap in the **same** condition is consistent with the theoretical $1/T$-type gain for aligned kernels. Shaded regions denote mean ± SE.

Two observations already connect to our main question. First, the dependence on $\mu_{\min}$ simply reflects coverage: under the counting rule, you cannot learn the argmax for states you rarely observe. Second, the end-to-end margin Δ_Q may be much *smaller* than per-step margins because local uncertainties compound through L transitions. This foreshadows why CoT—which counts *locally*—can be comparatively robust when transitions align.

4.2 Chain of Thought

With CoT, each context sample exposes the entire path $(x_0, x_1, \ldots, x_T)$, so the same counting rule is applied *per step*: estimate the hard-max index of each row of the relevant transition kernel from the observed transitions and then compose these local argmax decisions. To link local and global decisions we assume a standard greedy-to-global consistency:

Assumption 4.3 (Local-global consistency)**.** *We assume for all t and i, each row $P_i^{(t)}$ of the transition matrix has a single max index and denote this maximum index by $I_t(i)$. We additionally assume that*

$$I_T \circ \cdots \circ I_1(x_0) = \arg\max_j Q_{x_0,j}.$$

The impact of CoT then hinges on **how the per-step kernels relate**. We make this explicit by contrasting the homogeneous and heterogeneous regimes.

Homogeneous (aligned) transitions. When $P^{(1)} = \cdots = P^{(T)} =: P$, every trajectory provides T observations of the *same* kernel. Under the counting rule, this multiplies the effective number of local votes per trajectory, subject to standard mixing/coverage and a per-step margin:

Assumption 4.4. *The pseudo-spectral gap of P is $\gamma > 0$. The stationary distribution of P is π and the minimal element of π is positive $\pi_{\min} > 0$. The input distribution μ satisfies $\sqrt{\chi^2(\mu||\pi)} =: \chi_0$. The ground truth transition matrix P satisfies*

$$P_{i,j^*} - P_{ij} \geq \Delta_P > 0$$

for all i and $j \neq j^*(i)$, where $j^*(i) = \arg\max_i P_{ij}$.

Theorem 4.2 (Homogeneous CoT)**.** *Under Assumptions 4.1, 4.3 and 4.4, if the number of samples n satisfies*

$$n = \Theta\left(\left(\frac{1}{T\pi_{\min}\Delta_P^2 r} + \frac{1}{T\pi_{\min}^2 r^2}\right)\log(k/\delta)\right),$$

where $r = 1 - \frac{\chi_0}{T\pi_{\min}\gamma}$, then with probability at least $1-\delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with CoT inference with homogeneous transitions.

Three messages tie back to our factorization. (A) **Alignment** produces the structural $1/T$ gain: the same counting rule now accrues T local votes per trajectory for a single kernel. Note that if T is large enough, then the sample complexity of homogeneous CoT will always be smaller than direct inference, showing alignment is the most important factor here. (B) The **margin that matters** is the local Δ_P, typically larger than the composed Δ_Q; hence CoT's relative robustness to compounding uncertainty. (C) The term r quantifies how fast trajectories diversify across rows (mixing/coverage), ensuring those local counts are informative rather than redundant.

Heterogeneous (misaligned) transitions. When the kernels vary by step, a trajectory's T local observations are *split* across different kernels. The same counting rule still applies, but the local estimates no longer pool toward a single transition. Identification thus has to hold at every time, controlled by stepwise coverage and a uniform local margin:

Assumption 4.5 (Heterogeneous CoT)**.** *The intermediate distributions satisfy there exists $q_{\min} > 0$ such that*

$$q_{\min} \leq \min_{0 \leq t \leq T-1} \min_i \mu_i^{(t)}.$$

The ground truth transition matrices satisfy

$$P_{i,j^*}^{(t)} - P_{ij}^{(t)} \geq \Delta > 0$$

for all t, i and $j \neq j_t^*(i)$, where $j_t^*(i) = \arg\max_i P_{ij}^{(t)}$.

Theorem 4.3. *Under Assumptions 4.1, 4.3 and 4.5, if the number of samples n satisfies*

$$n = \Theta\left(\frac{\log(Tk/\delta)}{q_{\min}\Delta^2}\right),$$

where $r = 1 - \frac{\chi_0}{T\pi_{\min}\gamma}$, then with probability at least $1-\delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with CoT inference with heterogeneous transitions.

Here there is no guaranteed $1/T$ speedup. The same trajectory yields evidence about *different* kernels, so the counting rule cannot pool those votes as in the aligned case, and a $\log T$ term appears instead.

Remark 4.4. *Comparisons between Sample Complexity of Heterogeneous CoT and Direct Inference.*

1. *$q_{\min} \leq \min_i \mu_i^{(0)} = \mu_{\min}$, showing that the coverage in misaligned CoT is worse than direct inference.*
2. *Both cases $\Delta_Q > \Delta$ and $\Delta_Q < \Delta$ exist. For the former case one can take*

$$P^{(1)} = \begin{pmatrix} 0.6 & 0.4 \\ 0.1 & 0.9 \end{pmatrix} \text{ and } P^{(2)} = \begin{pmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{pmatrix},$$

and for the latter case

$$P^{(1)} = \begin{pmatrix} 0.8 & 0.2 \\ 0.8 & 0.2 \end{pmatrix} \text{ and } P^{(2)} = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}.$$

This remark shows that the sample complexity of heterogeneous CoT can be smaller than direct inferece, due to the $\log T$ factor, worse coverage and possibly smaller margin. Nonetheless, heterogeneous CoT can still be competitive when the per-step margin Δ is substantially larger than the composed margin Δ_Q, which is the case in our experiments.

The conclusions of this section are as follows.

- Under a common, simple *count-and-argmax* decision rule, the factor that controls structural gains is **transition alignment**: identical per-step kernels yield a $1/T$-type reduction in sample complexity; different kernels do not.
- The factor that controls sensitivity is the **noise (margin)**: CoT depends on the local margin Δ_P, while direct inference depends on the composed margin Δ_Q. Because Δ_Q typically contracts faster under noise, CoT's *relative* advantage grows as intermediate-step noise increases.

5 PROOF SKETCH

The bounds in Section 4 reduce hard-max prediction in our Markovian model to a sequence of elementary *top-1 identification* problems for multinomial distributions. Intuitively, under the count-and-argmax rule, we only need (i) enough coverage of the relevant rows to estimate their class frequencies, and (ii) a positive margin so the empirical argmax matches the true argmax. Direct inference and homogeneous CoT each require identifying the top class for one $k \times k$ transition matrix (either Q or P), whereas heterogeneous CoT requires doing so for all T per-step matrices. This is why a uniform top-1 identification lemma for multinomials is the core primitive behind all three theorems.

Lemma 5.1 (Chapter 2, (Boucheron et al., 2013))**.** *For a multinomial distribution p, we write $p_{(1)} \geq p_{(2)} \geq \cdots$ for its ordered elements. We assume the margin $\Delta_p := p_{(1)} - p_{(2)} > 0$. If there exist absolute constants $C > 0$ such that*

$$N \geq \frac{C}{\Delta^2} \log \frac{k}{\delta},$$

then the empirical argmax equals to the argmax with probability at least $1 - \delta$.

The lemma converts a margin lower bound into the familiar $O(\Delta^{-2} \log(k/\delta))$ sample requirement, so the main task is to lower bound the effective counts with which each row is observed. Let $N_i^{(0)}$ be the number of context samples whose initial state is $x_0 = i$, and $N_i^{(t)}$ the number of visits to state i at step t. For direct inference, it suffices to control $\min_i N_i^{(0)}$ because we only estimate the row of Q indexed by x_0. A Chernoff bound (e.g., Hoeffding, 1963) gives $\min_i N_i^{(0)} \gtrsim c\, N\, \mu_{\min}$ with high probability for an absolute constant $c > 0$, which yields Theorem 4.1.

For heterogeneous (misaligned) CoT, we must identify the top class in each per-step kernel $P^{(t)}$, so we need a uniform lower bound on the step-wise counts $\min_{t,i} N_i^{(t)}$. Since the states at step t are distributed as $\mu^{(t)} = \mu P^{(1)} \cdots P^{(t)}$, we obtain $\min_i N_i^{(t)} \gtrsim c\, N\, \min_i \mu_i^{(t)}$; invoking the assumption $q_{\min} \leq \min_{t,i} \mu_i^{(t)}$ and a union bound over $t \in \{0, \ldots, T-1\}$ controls the simultaneous deviations across steps. Plugging this coverage into the multinomial lemma with the local margin Δ gives the heterogeneous CoT bound (Theorem 4.3), including the logarithmic dependence on T from the union bound.

The homogeneous (aligned) CoT case is more delicate and also where the structural gain appears. Because each trajectory contributes all T transitions of the same kernel P, the total number of usable local observations scales like NT. Define the aggregated counts $N_i := \sum_{t=0}^{T-1} N_i^{(t)}$. If the initial distribution μ were already stationary for P, then each step would sample from π and we would have $N_i \approx NT\, \pi_i$ up to concentration, immediately yielding a $1/T$ improvement when the lemma is applied with margin Δ_P. In general μ need not equal π, so we quantify the bias and dependence via the χ^2-divergence χ_0 between μ and π and the pseudo-spectral gap γ of P. These two quantities govern how quickly the chain forgets its start and how many "effectively independent" transitions we get along a path. The next bound formalizes this intuition.

Lemma 5.2. *There exist absolute constants $c_1, c_2 > 0$ such that, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\min_i N_i \geq NT\, \pi_{\min} - \frac{N\, \chi_0}{\gamma} - \sqrt{\frac{c_1\, NT}{\gamma} \log \frac{k}{\delta}} \; - \; c_2,$$

where $\pi_{\min} := \min_i \pi_i$.

Remark 5.3.

1. *The bias term $N\chi_0/\gamma$ comes from the the error between μP^t and π: $\frac{1}{T} \sum_{t=0}^{T-1} \|\mu P^t - \pi\|_{\mathrm{TV}} \lesssim \chi_0/(T\gamma)$ (Paulin, 2015).*
2. *The fluctuation term derives from Bernstein–type concentration for bounded functions along Markov chains with pseudo–spectral gap (Paulin, 2015).*

Combining this coverage bound with the multinomial lemma yields the homogeneous CoT rate in Theorem 4.2: the leading $1/T$ term reflects T informative transitions per trajectory for the same kernel, and the factor $r = 1 - \chi_0/(T\pi_{\min}\gamma)$ captures the erosion due to initialization bias and mixing.

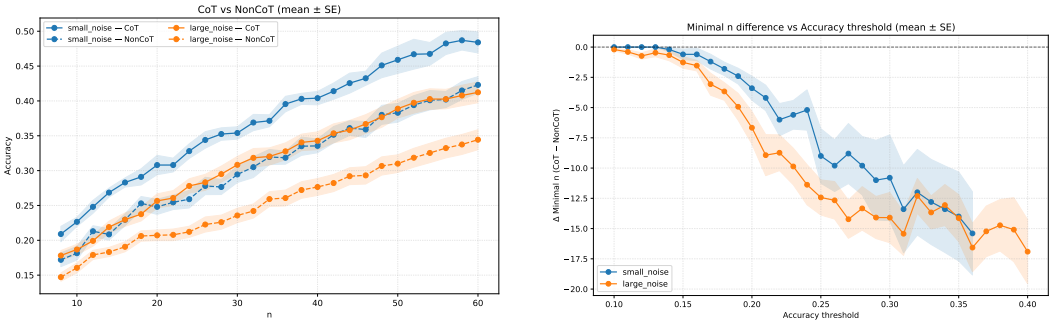


Figure 3: **Noise ablation under aligned transitions.** *Left:* Accuracy vs. n for CoT/NonCoT under **small** and **large** intermediate-step noise. *Right:* $\Delta n(\tau)$ vs. τ. CoT reduces the required context budget more when noise is higher, reflecting that the global margin Δ_Q shrinks faster than the local margin Δ_P as stochasticity increases. Shaded regions denote mean $\pm$ SE.

6 Experiments

The empirical benefits of chain-of-thought (CoT) prompting have been extensively documented in prior work (Sprague et al., 2025). Our experiments are therefore not intended as a continuation of those results, but instead as a controlled setup that mirrors our theoretical analysis. In particular, real-world benchmarks make it difficult to disentangle the effect of CoT from other confounding factors, such as prior knowledge embedded in pretrained LLMs. To address this, we first design synthetic tasks that the model has never encountered before, allowing us to isolate and probe the theoretical factors of interest.

6.1 Synthetic experiment setup

Our synthetic benchmark is designed to cleanly manipulate the two factors singled out by the theory, *transition alignment* and *intermediate-step noise*, while holding everything else fixed. Each instance is a two-step process. We sample a triple (x_0, r_1, r_2) where x_0 is an integer state and each r_t indexes a small stochastic *local rule* acting on the current state (e.g., with probability 0.8 apply $+1$, otherwise apply -2). The **same** condition enforces the rule of r_1 aligns with r_2, whereas r_1 and r_2 represents different local rules in **diff** condition. For each condition we compare **CoT** against **NonCoT**, where we give a specific LLM n context samples generated by the stochastic calculation and expect the model to compute the hard-max result for the query. Note that in this task, we design the random local rule to satisfy the consistency in Assumption 4.3, enforcing the hard-max selected by the sequential local rule or the global rule are the same. We sweep $n \in \{8, \ldots, 60\}$ and aggregate results over multiple random seeds.

There are several reasons why we select this setup: (i) Enforcing **same** vs. **diff** isolates the notion of *transition alignment*. (ii) The rule family is intentionally simple so that the ground-truth hard-max at each step and the composed end-to-end hard-max are unambiguous. This lets us measure accuracy without confounding from parsing or generation. (iii) Noise is controlled via the rule probabilities, allowing us to increase the chance that the local hard-max is flipped while leaving everything else unchanged. This separation mirrors the theoretical quantities Δ_P and Δ_Q. These advantages avoid spurious gains from unrelated factors in complicated experiments like (Sprague et al., 2025; Wei et al., 2022; Wang et al., 2023; Yao et al., 2023).

Metrics. Accuracy is the fraction of test queries whose final prediction equals the end-to-end hard-max index (which, by construction, agrees with composing the per-step hard-max indices). To visualize sample complexity, for a target accuracy threshold τ we compute the least n that reaches τ for CoT and NonCoT and plot $\Delta n(\tau) = n_{\text{CoT}}(\tau) - n_{\text{NonCoT}}(\tau)$. Negative values mean CoT achieves the target with fewer samples.

6.2 SYNTHETIC RESULTS AND ANALYSIS

Alignment vs. misalignment. From Fig. 2, CoT helps in both **same** and **diff**, but the effect size depends strongly on alignment. With aligned kernels (**same**), CoT consistently dominates NonCoT over the full range of n, and the gap widens at higher accuracy thresholds. With misaligned kernels (**diff**), improvements are present but uniformly smaller, and can flatten at stricter thresholds. This pattern matches the theory: when transitions align, each trajectory yields multiple informative votes for the *same* kernel (here $T{=}2$). When transitions differ, votes are split across kernels and cannot deliver the same $1/T$ efficiency. Note that for misaligned kernel case, there are n such that CoT is not better than direct inference, which also matches our analysis.

Noise sensitivity. By Fig. 3, increasing intermediate-step noise leads to a larger *relative* advantage for CoT. The reason is structural: higher noise reduces both the local margin Δ_P and the global margin Δ_Q, but due to compounding across steps, Δ_Q shrinks faster. CoT is therefore more robust to noise than estimating the end-to-end hard-max from final outcomes alone. This is consistent with empirical observations that noisy or unfaithful chains can change outcomes markedly and that aggregation over trajectories helps (Prabhakar et al., 2024; Wang et al., 2023).

6.3 REALISTIC EXPERIMENT: MODULAR ADDITION

While the synthetic data closely mirrors our theoretical setup, it lacks practical relevance. To bridge theory and practice, we therefore turn to a simple arithmetic problem, multi-step modular addition, that has clear real-world value yet remains structured enough for controlled analysis. Specifically, the task is to calculate long additions modular a certain number. To avoid conflating our discussion with heterogeneous step difficulties, we design a simplified version where each step involves an addition of comparable form, differing only in whether all steps add the *same* or *different* numbers.

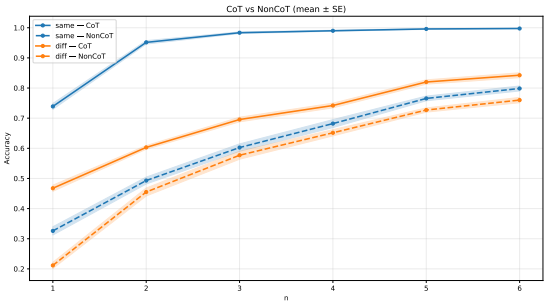


Figure 4: **Multi-step modular addition (practical check).** Accuracy vs. n for CoT and NonCoT when both steps add the **same** number versus **different** numbers (fixed modulus). The larger CoT gain in the aligned (**same**) case corroborates the synthetic findings in a more realistic arithmetic setting. Error bands show mean $\pm$ SE.

We compare CoT and NonCoT when all steps add the same number versus different numbers. The result in Fig. 4 echoes the synthetic study: CoT delivers a substantially larger boost in the **same** condition, while improvements are smaller in **diff**. This task is less "clean" (it involves natural-language templates and minor formatting variations), but it is more practical and still isolates the alignment factor: when both steps share the same local operation, trajectories provide reusable per-step evidence toward the same kernel; on the contrary, the benefit of per-step counting shrinks. Together with the synthetic results, this supports the claim that *transition alignment* is a decisive driver of CoT's inference-time sample efficiency.

6.4 REALISTIC EXPERIMENT: CITY–STATE RANKINGS

We test transition alignment on a corpus-based task built from U.S. city/state rankings under two criteria: population and area. We collected the data and built a two-hop QA dataset. Each instance in the dataset directly asks for *"the X-th largest city in the Y-th largest state"*, where "largest" is instantiated by a chosen criterion for the state rank and a chosen criterion for the city-within-state rank. We treat the criterion (population vs. area) as the step "skill": **same-skill** uses the same criterion for both steps (pop–pop or area–area), while **diff-skill** mixes criteria (pop–area or area–pop).

We sweep $n \in \{2, 4, 6, 8, 10\}$ in-context demonstrations and compare CoT vs. NonCoT (full details in Appendix C.3). Fig. 5 shows that CoT improves over NonCoT in both settings, with a consistently larger gain in the aligned same-skill condition.

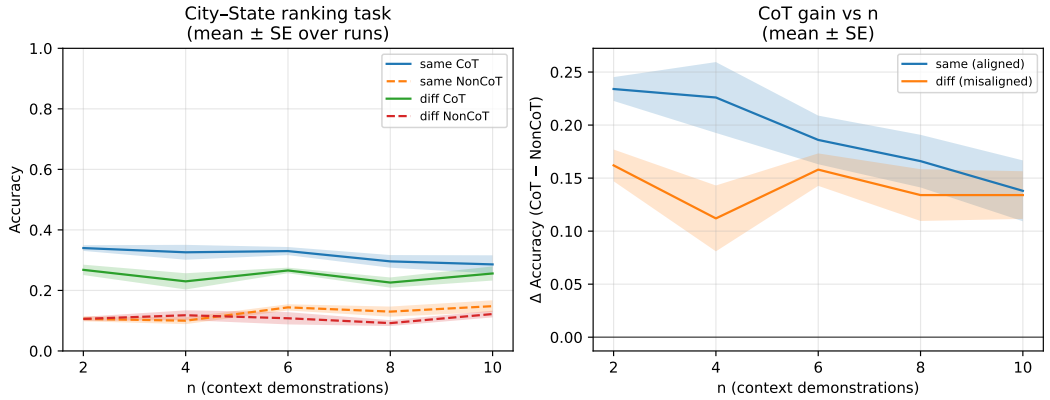


Figure 5: **City–State rankings.** Accuracy vs. n (left) and CoT gain ΔAcc (right) under **same-skill** vs. **diff-skill**. Error bands: mean $\pm$ SE.

7 Conclusion and Discussions

We provided a mechanism-level account of when CoT improves inference-time sample efficiency by modeling stepwise reasoning as a finite-state Markov chain, motivated by the autoregressive nature of CoT and the finite context windows in practice. Two structural properties govern the performance of CoT: (i) **transition alignment** that characterizes whether all steps share the same local transition, and (ii) **decision margins/noise** that describes the per-step confidence. We rigorously establish the decisive effects of both properties. In particular, we show that (i) CoT yields a $1/T$ gain over direct inference only when the local transitions in successive steps are aligned; while (ii) the decision margins control the gain of CoT, which improves when per-step evidence is comparatively more robust than end-to-end decisions. These theoretical findings are corroborated by controlled experiments on synthetic and modular addition tasks.

Implications for implicit thinking. Our Markovian formulation treats multi-step reasoning as a trajectory of latent states with step-wise transitions. An explicit CoT transcript is one of the intuitive ways to expose this trajectory. Whether the intermediate reasoning can be reused across steps, the efficiency gain of CoT in our analysis comes from the underlying dynamics, instead of expressing the intermediate states as human-readable text. This perspective casts light on implicit thinking: one can keep multi-step state evolution, but change the interface that reveals intermediate states (e.g., compress them or not reveal them at all). Empirically, this suggests comparing implicit and explicit interfaces under matched test-time compute. It is also helpful to control for the same structural descriptors in our theory (e.g., alignment and intermediate noise), so we can separate improvements in internal dynamics from improvements due to the readout format.

Future directions. We see two complementary future avenues. From the explanatory perspective, existing works on CoT theory mostly relies on simplified abstractions. A next step is to push toward more practical models that still admit analysis, e.g., higher-order or semi-Markov structure, hierarchical/graph decompositions, or continuous latent states. Then the factors can be characterized across a wider range of real reasoning pipelines. From the application perspective, we seek to extend the measurement of the key structural properties identified by the theory to complex real-world tasks, which can potentially steer the design of more effective, aligned, and robust contexts for CoT prompting.

ACKNOWLEDGMENTS

QL acknowledges support of NSF DMS-2523382 and DOE Office of Science under Award #DE-SC0024721.

REFERENCES

Emmanuel Abbe, Samy Bengio, Aryo Lotfi, Colin Sandon, and Omid Saremi. How far can transformers reason? the globality barrier and inductive scratchpad. *Advances in Neural Information Processing Systems*, 37:27850–27895, 2024.

Ekin Akyürek, Dale Schuurmans, Jacob Andreas, Tengyu Ma, and Denny Zhou. What learning algorithm is in-context learning? investigations with linear models. *arXiv preprint arXiv:2211.15661*, 2022. URL https://arxiv.org/abs/2211.15661.

Awni Altabaa, Omar Montasser, and John Lafferty. Cot information: Improved sample complexity under chain-of-thought supervision. *arXiv preprint arXiv:2505.15927*, 2025.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, et al. Graph of thoughts: Solving elaborate problems with large language models. In *Proceedings of the AAAI conference on artificial intelligence*, 2024.

Sébastien Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Wenhu Chen, Xueguang Ma, Xuezhi Wang, and William W. Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=PWZwW2A2Jq.

David Chiang, Peter Cholak, and Anand Pillay. Tighter bounds on the expressivity of transformer encoders. In *International Conference on Machine Learning*, pp. 5544–5562. PMLR, 2023.

Peter Clark, Oyvind Tafjord, and Kyle Richardson. Transformers as soft reasoners over language. In *Proceedings of IJCAI*, 2020. URL https://arxiv.org/abs/2002.05867. RuleTaker.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Jared Kaplan, Sam McCandlish, Yury Gorishniy, Ilya Sutskever, and Dario Amodei. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021. URL https://arxiv.org/abs/2110.14168. Introduces GSM8K.

DeepSeek-AI. Deepseekmath: Simple yet effective LLMs for mathematical reasoning. arXiv preprint arXiv:2407.10191, 2024. URL https://arxiv.org/abs/2407.10191.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. The llama 3 herd of models. *arXiv e-prints*, pp. arXiv–2407, 2024.

Ezra Edelman, Nikolaos Tsilivis, Benjamin Edelman, Eran Malach, and Surbhi Goel. The evolution of statistical induction heads: In-context learning markov chains. *Advances in neural information processing systems*, 37:64273–64311, 2024.

Guhao Feng, Bohang Zhang, Yuntian Gu, Haotian Ye, Di He, and Liwei Wang. Towards revealing the mystery behind chain of thought: a theoretical perspective. *Advances in Neural Information Processing Systems*, 36:70757–70798, 2023a.

Xidong Feng, Ziyu Wan, Muning Wen, Stephen Marcus McAleer, Ying Wen, Weinan Zhang, and Jun Wang. Alphazero-like tree-search can guide large language model decoding and training. *arXiv preprint arXiv:2309.17179*, 2023b.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018. URL https://arxiv.org/abs/1806.01445.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Arian Hosseini, Xingdi Yuan, Nikolay Malkin, Aaron Courville, Alessandro Sordoni, and Rishabh Agarwal. V-star: Training verifiers for self-taught reasoners. *arXiv preprint arXiv:2402.06457*, 2024.

Xinyang Hu, Fengzhuo Zhang, Siyu Chen, and Zhuoran Yang. Unveiling the statistical foundations of chain-of-thought prompting methods. *arXiv preprint arXiv:2408.14511*, 2024.

Baihe Huang, Shanda Li, Tianhao Wu, Yiming Yang, Ameet Talwalkar, Kannan Ramchandran, Michael I Jordan, and Jiantao Jiao. Sample complexity and representation ability of test-time scaling paradigms. *arXiv preprint arXiv:2506.05295*, 2025a.

Jianhao Huang, Zixuan Wang, and Jason D Lee. Transformers learn to implement multi-step gradient descent with chain of thought. In *The Thirteenth International Conference on Learning Representations*, 2025b.

M Emrullah Ildiz, Yixiao Huang, Yingcong Li, Ankit Singh Rawat, and Samet Oymak. From self-attention to markov models: Unveiling the dynamics of generative transformers. *arXiv preprint arXiv:2402.13512*, 2024.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv preprint arXiv:2412.16720*, 2024.

Nirmit Joshi, Gal Vardi, Adam Block, Surbhi Goel, Zhiyuan Li, Theodor Misiakiewicz, and Nathan Srebro. A theory of learning with autoregressive chain of thought. *arXiv preprint arXiv:2503.07932*, 2025.

Tushar Khot, Peter Clark, Michal Guerquin, Peter Jansen, and Ashish Sabharwal. Qasc: A dataset for question answering via sentence composition. In *Proceedings of AAAI*, 2020. URL https://arxiv.org/abs/1910.11473.

J. Kim et al. Metastable dynamics of chain-of-thought reasoning: A markov perspective on search, rl, and distillation. *arXiv preprint arXiv:2502.01694*, 2025. URL https://arxiv.org/abs/2502.01694.

Juno Kim and Taiji Suzuki. Transformers provably solve parity efficiently with chain of thought. In *The Thirteenth International Conference on Learning Representations*, 2025.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *NeurIPS 2022 Datasets and Benchmarks Track*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/8bb0d291acd4acf06ef112099c16f326-Abstract-Conference.html.

Aviral Kumar, Vincent Zhuang, Rishabh Agarwal, Yi Su, John D Co-Reyes, Avi Singh, Kate Baumli, Shariq Iqbal, Colton Bishop, Rebecca Roelofs, et al. Training language models to self-correct via reinforcement learning. *arXiv preprint arXiv:2409.12917*, 2024.

Aitor Lewkowycz, Thibault Sellam, Jean-Baptiste Alayrac, Laurent Shafey, et al. Solving quantitative reasoning problems with language models. *arXiv preprint arXiv:2206.14858*, 2022. URL https://arxiv.org/abs/2206.14858. Minerva.

Hongkang Li, Meng Wang, Songtao Lu, Xiaodong Cui, and Pin-Yu Chen. How do nonlinear transformers acquire generalization-guaranteed cot ability? In *High-dimensional Learning Dynamics 2024: The Emergence of Structure and Reasoning*, 2024a.

Zhiyuan Li, Hong Liu, Denny Zhou, and Tengyu Ma. Chain of thought empowers transformers to solve inherently serial problems. *arXiv preprint arXiv:2402.12875*, 1, 2024b.

Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *The Twelfth International Conference on Learning Representations*, 2023.

Ashok Vardhan Makkuva, Marco Bondaschi, Adway Girish, Alliot Nagle, Martin Jaggi, Hyeji Kim, and Michael Gastpar. Attention with markov: A framework for principled analysis of transformers via markov chains. *arXiv preprint arXiv:2402.04161*, 2024.

Vaibhav Mavi, Anubhav Jangra, Adam Jatowt, et al. Multi-hop question answering. *Foundations and Trends® in Information Retrieval*, 17(5):457–586, 2024.

William Merrill and Ashish Sabharwal. The expressive power of transformers with chain of thought. In *The Twelfth International Conference on Learning Representations*, 2024.

Eshaan Nichani, Alex Damian, and Jason D Lee. How transformers learn causal structure with gradient descent. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 38018–38070, 2024.

Daniel Paulin. Concentration inequalities for markov chains by marton couplings and spectral methods. *Electronic Journal of Probability*, 20(79):1–32, 2015. doi: 10.1214/EJP.v20-4035. URL `https://arxiv.org/abs/1212.2015`.

Akshara Prabhakar, Thomas L Griffiths, and R McCoy. Deciphering the factors influencing the efficacy of chain-of-thought: Probability, memorization, and noisy reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 3710–3724, 2024.

Ben Prystawski, Michael Li, and Noah Goodman. Why think step by step? reasoning emerges from the locality of experience. *Advances in Neural Information Processing Systems*, 36:70926–70947, 2023.

Hongyu Ren and Jure Leskovec. Beta embeddings for multi-hop logical reasoning over knowledge graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL `https://arxiv.org/abs/2010.11465`.

Clayton Sanford, Bahare Fatemi, Ethan Hall, Anton Tsitsulin, Mehran Kazemi, Jonathan Halcrow, Bryan Perozzi, and Vahab Mirrokni. Understanding transformer reasoning capabilities via graph algorithms. *Advances in Neural Information Processing Systems*, 37:78320–78370, 2024.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419): 1140–1144, 2018.

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.

Zayne Rea Sprague, Fangcong Yin, Juan Diego Rodriguez, Dongwei Jiang, Manya Wadhwa, Prasann Singhal, Xinyu Zhao, Xi Ye, Kyle Mahowald, and Greg Durrett. To cot or not to cot? chain-of-thought helps mainly on math and symbolic reasoning. In *The Thirteenth International Conference on Learning Representations*, 2025.

Oyvind Tafjord, Bhavana Dalvi, and Peter Clark. Proofwriter: Generating implications, proofs, and abductive statements over natural language. *arXiv preprint arXiv:2012.13048*, 2021. URL `https://arxiv.org/abs/2012.13048`.

Trieu H Trinh, Yuhuai Wu, Quoc V Le, He He, and Thang Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625(7995):476–482, 2024.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Peter Clark. Musique: Multi-hop questions via single-hop question composition. In *Proceedings of ACL*, 2022.

Jonathan Uesato, Nate Kushman, Ramana Kumar, Francis Song, Noah Siegel, Lisa Wang, Antonia Creswell, Geoffrey Irving, and Irina Higgins. Solving math word problems with process-and outcome-based feedback. *arXiv preprint arXiv:2211.14275*, 2022.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*, 2023.

Zixuan Wang, Eshaan Nichani, Alberto Bietti, Alex Damian, Daniel Hsu, Jason D Lee, and Denny Wu. Learning compositional functions with transformers from easy-to-hard data. *arXiv preprint arXiv:2505.23683*, 2025.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*, 2022. URL https://arxiv.org/abs/2201.11903.

Kaiyue Wen, Huaqing Zhang, Hongzhou Lin, and Jingzhao Zhang. From sparse dependence to sparse attention: unveiling how chain-of-thought enhances transformer sample efficiency. *arXiv preprint arXiv:2410.05459*, 2024.

Sang Michael Xie, Aditi Raghunathan, Percy Liang, and Tengyu Ma. An explanation of in-context learning as implicit bayesian inference. *arXiv preprint arXiv:2111.02080*, 2021. URL https://arxiv.org/abs/2111.02080.

Yuxi Xie, Anirudh Goyal, Wenyue Zheng, Min-Yen Kan, Timothy P Lillicrap, Kenji Kawaguchi, and Michael Shieh. Monte carlo tree search boosts reasoning via iterative preference learning. *arXiv preprint arXiv:2405.00451*, 2024.

Fan Yang, Zhilin Yang, and William W. Cohen. Differentiable learning of logical rules for knowledge base reasoning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. URL https://arxiv.org/abs/1702.08367.

Tong Yang, Yu Huang, Yingbin Liang, and Yuejie Chi. Multi-head transformers provably learn symbolic multi-step reasoning via gradient descent. *arXiv preprint arXiv:2508.08222*, 2025.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of EMNLP*, 2018. URL https://arxiv.org/abs/1809.09600.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *arXiv preprint arXiv:2305.10601*, 2023. URL https://arxiv.org/abs/2305.10601.

Oussama Zekri, Ambroise Odonnat, Abdelhakim Benechehab, Linus Bleistein, Nicolas Boullé, and Ievgen Redko. Large language models as markov chains. *arXiv preprint arXiv:2410.02724*, 2024.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

Ruiqi Zhang, Spencer Frei, and Peter L. Bartlett. Trained transformers learn linear models in-context. *Journal of Machine Learning Research*, 25(49):1–55, 2024. URL https://jmlr.org/papers/v25/23-1042.html.

Tianshi Zheng, Yixiang Chen, Chengxi Li, Chunyang Li, Qing Zong, Haochen Shi, Baixuan Xu, Yangqiu Song, Ginny Y Wong, and Simon See. The curse of cot: On the limitations of chain-of-thought in in-context learning. *arXiv preprint arXiv:2504.05081*, 2025.

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc Le, and Ed H. Chi. Least-to-most prompting enables complex reasoning in large language models. *arXiv preprint arXiv:2205.10625*, 2022. URL https://arxiv.org/abs/2205.10625.

A Preliminaries and Core Lemma

Notation. We write $[k] = \{1, \dots, k\}$ for the finite state space. Vectors are row vectors unless stated otherwise. For a probability vector $p \in \Delta^{k-1}$, let $p_{(1)} \geq p_{(2)} \geq \cdots$ denote the order statistics of its coordinates and define the *(row) margin*

$$\mathrm{margin}(p) := p_{(1)} - p_{(2)}.$$

For a row-stochastic matrix $M \in \mathbb{R}^{k \times k}$, we define $\mathrm{margin}(M) := \min_{i \in [k]} \mathrm{margin}(M_i)$.

Divergences and distances. For probability laws ν, π on $[k]$, the χ^2-divergence is

$$\chi^2(\nu \| \pi) := \sum_{i=1}^{k} \frac{(\nu_i - \pi_i)^2}{\pi_i} = \left\| \frac{\nu}{\pi} - \mathbf{1} \right\|_{L^2(\pi)}^2, \quad \text{where } \frac{\nu}{\pi}(i) := \nu_i / \pi_i.$$

It controls total variation (TV) via the inequality $\|\nu - \pi\|_{\mathrm{TV}} \leq \frac{1}{2}\sqrt{\chi^2(\nu \| \pi)}$.

Pseudo-spectral gap. Let P be a (possibly nonreversible) Markov kernel on $[k]$ with stationary distribution π, and let P^* denote its time-reversal on $L^2(\pi)$: $P^*(x, dy) = \frac{\pi(dy)\, P(y, dx)}{\pi(dx)}$. Following Paulin (2015, Eq. (3.3)), the *pseudo-spectral gap* of P is

$$\gamma_{\mathrm{ps}} := \max_{k \geq 1} \frac{\gamma\big((P^*)^k P^k\big)}{k}, \tag{2}$$

where $\gamma(A)$ denotes the (absolute) spectral gap of the self-adjoint operator A on $L^2(\pi)$. For reversible chains, γ_{ps} coincides with the usual L^2 spectral gap up to a constant factor, while for nonreversible chains it generalizes multiplicative reversibilization (Paulin, 2015, Remark 3.2). Moreover, γ_{ps} relates to mixing time via Paulin (2015, Prop. 3.4).

Context trajectories and counts. We observe n i.i.d. context trajectories (inference-time only). For the homogeneous case (all steps use P), write $X_{0:T}^{(\ell)}$ for the ℓ-th path and $N_i^{(t)} := \sum_{\ell=1}^{n} \mathbf{1}\{X_t^{(\ell)} = i\}$ for the number of visits to state i at step t. We also use the pooled count $N_i := \sum_{t=0}^{T-1} N_i^{(t)}$. For direct inference and the heterogeneous case, we further write $N_i^{(0)} := \sum_{\ell=1}^{n} \mathbf{1}\{X_0^{(\ell)} = i\}$.

Now we introduce the main lemma used in the proof.

Lemma A.1 (Multinomial top-1 identification (restatement of Lemma 5.1))**.** *Let $p \in \Delta^{k-1}$ with margin $\Delta_p := p_{(1)} - p_{(2)} > 0$, and let $\hat{p}$ be the empirical frequencies from N i.i.d. draws. There exists an absolute constant $C > 0$ such that if*

$$N \geq \frac{C}{\Delta_p^2} \log \frac{k}{\delta},$$

then $\arg\max_j \hat{p}_j = \arg\max_j p_j$ *with probability at least* $1 - \delta$.

Proof. Let $j^\star = \arg\max_j p_j$ and for each $j \neq j^\star$ define $Z_\ell := \mathbf{1}\{X_\ell = j^\star\} - \mathbf{1}\{X_\ell = j\} \in [-1, 1]$ with $\mathbb{E} Z_\ell = p_{j^\star} - p_j \geq \Delta_p$. Then $\hat{p}_{j^\star} - \hat{p}_j = \frac{1}{N}\sum_{\ell=1}^{N} Z_\ell$ and by Hoeffding $\Pr\{\hat{p}_{j^\star} \leq \hat{p}_j\} \leq \exp(-2N\Delta_p^2)$. Union-bounding over $j \neq j^\star$ yields $\Pr\{\exists j \neq j^\star : \hat{p}_j \geq \hat{p}_{j^\star}\} \leq (k-1)\exp(-2N\Delta_p^2) \leq \delta$ under the stated N (for a suitable absolute C). □

B Proofs of the Main Theorems

We restate each theorem and then give a full proof. All three proofs invoke Lemma A.1 and the homogeneous case additionally uses a pooled-coverage lemma whose proof relies on two results from Paulin (2015).

B.1 Proof of Theorem 4.1 (Direct inference)

Theorem B.1 (Restatement of Theorem 4.1)**.** *Under Assumptions 4.1 and 4.2, if the number of samples n satisfies*

$$n = \Theta\left(\frac{\log(k/\delta)}{\mu_{\min}\Delta_Q^2}\right),$$

then with probability at least $1-\delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with direct inference.

Proof. Let $N = n$ and $N_i^{(0)}$ be the number of contexts with $x_0 = i$. Since $\mu_{\min} > 0$ (Assumption 4.1), a Chernoff bound gives

$$\Pr\{N_i^{(0)} \le \tfrac{1}{2}N\mu_i\} \le \exp(-N\mu_i/8) \le \exp(-N\mu_{\min}/8).$$

Union-bounding over $i \in [k]$, we obtain that if $N \ge C_0\,\mu_{\min}^{-1}\log(k/\delta)$ (for an absolute C_0), then with probability at least $1-\delta/2$,

$$\min_{i\in[k]} N_i^{(0)} \ge \tfrac{1}{2}N\mu_{\min}. \tag{3}$$

Conditioned on $x_0 = i$, the $N_i^{(0)}$ terminal labels x_T are i.i.d. from the multinomial distribution Q_i. By Assumption 4.2, the row margin obeys $\mathrm{margin}(Q_i) \ge \Delta_Q > 0$ for all i. Applying Lemma A.1 with $N \leftarrow N_i^{(0)}$ and $\Delta_p \leftarrow \Delta_Q$ yields

$$\Pr\left\{\hat{j}^*(i) \ne \arg\max_j Q_{ij} \;\middle|\; N_i^{(0)}\right\} \le k\cdot\exp\Big(-C\,\Delta_Q^2\,N_i^{(0)}\Big),$$

for an absolute $C > 0$. On the event (3), the RHS is at most $k\exp(-C\,\Delta_Q^2\,N\mu_{\min}/2)$. Another union bound over $i \in [k]$ shows that all k rows are identified correctly with probability at least $1-\delta/2$ provided $N \gtrsim \frac{1}{\mu_{\min}\Delta_Q^2}\log(k/\delta)$. Combining with (3) via a union bound establishes the theorem with the stated $\Theta(\cdot)$ rate. □

B.2 Proof of Theorem 4.2 (Homogeneous CoT)

We introduce two technical lemmas for Markov chain based on pseudo spectral gap.

Lemma B.2 (TV decay via pseudo-spectral gap (Paulin, 2015, Prop. 3.4))**.** *Let P be a (possibly nonreversible) Markov kernel on $[k]$ with stationary distribution π. Let P^* denote its time-reversal on $L^2(\pi)$, and define the pseudo-spectral gap*

$$\gamma_{\mathrm{ps}} := \max_{m\ge1}\frac{\gamma\big((P^*)^m P^m\big)}{m}, \tag{4}$$

where $\gamma(A)$ is the (absolute) spectral gap of the self-adjoint operator A on $L^2(\pi)$. For any initial law q absolutely continuous w.r.t. π, write $N_q := \mathbb{E}_\pi\big[\big(\frac{dq}{d\pi}\big)^2\big] = 1+\chi^2(q\|\pi)$. Then, for all $n \ge 1$,

$$d_{\mathrm{TV}}(qP^n,\pi) \le \frac{1}{2}\,(1-\gamma_{\mathrm{ps}})^{(n-1/\gamma_{\mathrm{ps}})/2}\,\sqrt{N_q-1}. \tag{5}$$

In particular, letting $\chi_0 := \sqrt{\chi^2(q\|\pi)} = \sqrt{N_q-1}$ and summing the geometric decay,

$$\frac{1}{T}\sum_{t=0}^{T-1} d_{\mathrm{TV}}(qP^t,\pi) \le \frac{C}{T}\cdot\frac{\chi_0}{\gamma_{\mathrm{ps}}}, \tag{6}$$

for a universal constant $C > 0$.

Lemma B.3 (Bernstein inequality for additive functionals (Paulin, 2015, Thm. 3.11))**.** *Let $(X_t)_{t\ge0}$ be a (possibly nonreversible) Markov chain with stationary law π and pseudo-spectral gap $\gamma_{\mathrm{ps}} > 0$ as in (4), and assume $X_0 \sim \pi$ (stationary start). Let $f : [k] \to \mathbb{R}$ be bounded with $\|f - \mathbb{E}_\pi f\|_\infty \le b$ and variance $\mathrm{Var}_\pi(f) = \sigma^2$. Then there exist absolute constants $a_1, a_2 > 0$ such that for all $u > 0$ and $T \in \mathbb{N}$,*

$$\Pr\left\{\left|\sum_{t=0}^{T-1}\big(f(X_t)-\mathbb{E}_\pi f\big)\right| \ge u\right\} \le 2\exp\left(-\frac{a_1\,\gamma_{\mathrm{ps}}\,u^2}{T\,\sigma^2 + a_2\,b\,u}\right). \tag{7}$$

With the lemmas above, we can prove the following important lemma that gives a lower bound on the total coverage along the Markov chain.

Lemma B.4 (Aggregated step coverage under mixing (restatement of Lemma 5.2))**.** *Assume $P^{(1)} = \cdots = P^{(T)} =: P$ with stationary law π ($\pi_{\min} > 0$), pseudo-spectral gap $\gamma_{\mathrm{ps}} > 0$, and $\chi_0 = \sqrt{\chi^2(\mu\|\pi)}$. Let $N_i^{(t)}$ be the number of visits to state i at step t across n i.i.d. trajectories, and $N_i := \sum_{t=0}^{T-1} N_i^{(t)}$. There exist absolute $c_1, c_2 > 0$ such that, for any $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\min_{i\in[k]} N_i \geq nT\,\pi_{\min} - \frac{n\,\chi_0}{\gamma_{\mathrm{ps}}} - \sqrt{\frac{c_1\,nT}{\gamma_{\mathrm{ps}}}\,\log\frac{k}{\delta}} - c_2.$$

Proof of Lemma B.4. Fix $i \in [k]$ and set $f_i(x) = \mathbf{1}\{x = i\} \in [0,1]$. For one trajectory $X_{0:T-1}$, write $Y_i := \sum_{t=0}^{T-1} f_i(X_t)$.

(Bias.) By Lemma B.2, in particular (6),

$$\mathbb{E}Y_i = \sum_{t=0}^{T-1} (\mu P^t)_i \;\geq\; T\pi_i - \frac{C\,\chi_0}{\gamma_{\mathrm{ps}}}.$$

(Fluctuation.) Consider the centered sum around $\mathbb{E}_\pi f_i$:

$$S_T^\circ := \sum_{t=0}^{T-1} \big(f_i(X_t) - \mathbb{E}_\pi f_i\big).$$

Here $\sigma^2 = \mathrm{Var}_\pi(f_i) \leq \frac{1}{4}$ and $b = \|f_i - \mathbb{E}_\pi f_i\|_\infty \leq 1$. Applying Lemma B.3 with these (σ^2, b) yields, for absolute $c_1' > 0$,

$$\Pr\{|S_T^\circ| \geq u\} \leq 2\exp\!\left(-\,c_1'\,\frac{\gamma_{\mathrm{ps}}\,u^2}{T+u}\right). \tag{8}$$

The centered sum we actually need is $S_T := \sum_{t=0}^{T-1}\big(f_i(X_t) - \mathbb{E}f_i(X_t)\big) = S_T^\circ - \sum_{t=0}^{T-1}\big(\mathbb{E}f_i(X_t) - \mathbb{E}_\pi f_i\big)$. The deterministic bias term is bounded by total variation:

$$\Big|\sum_{t=0}^{T-1}(\mathbb{E}f_i(X_t) - \mathbb{E}_\pi f_i)\Big| \leq \sum_{t=0}^{T-1} \|\mathcal{L}(X_t) - \pi\|_{\mathrm{TV}} \leq C\,\chi_0/\gamma_{\mathrm{ps}}$$

by Lemma B.2. Thus S_T inherits the tail in (8) up to a shift by a deterministic constant. Absorbing that shift into constants, we can write

$$\Pr\{|S_T| \geq u\} \leq 2\exp\!\left(-\,c\,\frac{\gamma_{\mathrm{ps}}\,u^2}{T+u}\right) \qquad (c > 0 \text{ absolute}). \tag{9}$$

(From one trajectory to n trajectories.) Let $Y_i^{(1)}, \ldots, Y_i^{(n)}$ be i.i.d. copies, and $Z_\ell := Y_i^{(\ell)} - \mathbb{E}Y_i$. By (9), each Z_ℓ is sub-exponential with variance proxy $O(T/\gamma_{\mathrm{ps}})$ and scale proxy $O(1/\gamma_{\mathrm{ps}})$. A Bernstein inequality for independent sub-exponential sums gives absolute $C_1, C_2 > 0$ such that, with probability at least $1 - \delta/k$,

$$\sum_{\ell=1}^{n} Y_i^{(\ell)} \;\geq\; n\,\mathbb{E}Y_i \;-\; \sqrt{C_1\,\frac{nT}{\gamma_{\mathrm{ps}}}\,\log\frac{k}{\delta}} \;-\; C_2\,\frac{1}{\gamma_{\mathrm{ps}}}\,\log\frac{k}{\delta}.$$

A union bound over $i \in [k]$ and the lower bound on $\mathbb{E}Y_i$ complete the proof (absorbing the linear-in-log term into c_2). □

Theorem B.5 (Restatement of Theorem 4.2)**.** *Under Assumptions 4.1, 4.3 and 4.4, if the number of samples n satisfies*

$$n = \Theta\left(\left(\frac{1}{T\pi_{\min}\Delta_P^2\,r} + \frac{1}{T\pi_{\min}^2\,r^2}\right)\log\frac{k}{\delta}\right), \qquad r := 1 - \frac{\chi_0}{T\pi_{\min}\gamma_{\mathrm{ps}}},$$

then with probability at least $1-\delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with CoT inference under homogeneous transitions.

Proof. By Lemma B.4, with probability $\geq 1-\delta/2$, $\min_i N_i \geq nT\,\pi_{\min} - \frac{n\,\chi_0}{\gamma_{\mathrm{ps}}} - \sqrt{\frac{c_1\,nT}{\gamma_{\mathrm{ps}}}\log\frac{k}{\delta}} - c_2$. Assumption 4.4 ensures $\mathrm{margin}(P) \geq \Delta_P > 0$. Applying Lemma A.1 row-wise with a union bound gives the stated sample-size condition (first term from pooled T transitions per trajectory; second from $\sqrt{nT/\gamma_{\mathrm{ps}}}$ fluctuations). Assumption 4.3 then maps local argmaxes to the global argmax of Q. □

B.3 Proof of Theorem 4.3 (Heterogeneous CoT)

Theorem B.6 (Restatement of Theorem 4.3)**.** *Under Assumptions 4.1, 4.3 and 4.5, if the number of samples n satisfies*

$$n = \Theta\left(\frac{\log(Tk/\delta)}{q_{\min}\Delta^2}\right),$$

then with probability at least $1-\delta$, we have $\hat{j}^*(i) = \arg\max_j Q_{ij}$ for all i with CoT inference with heterogeneous transitions.

Proof. At step $t \in \{0,\ldots,T-1\}$ the state distribution is $\mu^{(t)} = \mu P^{(1)}\cdots P^{(t)}$. Let $N_i^{(t)}$ be the number of visits to i at step t across the n trajectories. Then $N_i^{(t)} \sim \mathrm{Bin}(n, \mu_i^{(t)})$ independently across trajectories. By Chernoff and Assumption 4.5, $\Pr\{N_i^{(t)} \leq \frac{1}{2}n\mu_i^{(t)}\} \leq \exp(-n\mu_i^{(t)}/8) \leq \exp(-nq_{\min}/8)$. Union-bounding over the Tk pairs (t,i) shows that, if $n \geq C_0\, q_{\min}^{-1}\log(Tk/\delta)$, then with probability at least $1-\delta/2$,

$$\min_{t\in\{0,\ldots,T-1\}}\ \min_{i\in[k]} N_i^{(t)} \ \geq\ \tfrac{1}{2}n\,q_{\min}. \tag{10}$$

Conditioned on $x_t = i$, the next state x_{t+1} is multinomial with row $P_i^{(t)}$. Assumption 4.5 provides the *uniform* local margin $\min_{t,i}\mathrm{margin}(P_i^{(t)}) \geq \Delta > 0$. Applying Lemma A.1 to each pair (t,i) with $N_i^{(t)}$ samples and margin Δ, and union-bounding over all Tk pairs, we conclude that (10) together with $n \gtrsim (q_{\min}\Delta^2)^{-1}\log(Tk/\delta)$ ensures that all per-step rowwise argmax decisions are correct with probability $\geq 1-\delta/2$. Finally, by Assumption 4.3, composing these local decisions yields the global argmax of Q for every initial state, completing the proof. □

C Experimental Details

All experiments on math problems use *DeepSeek-Math-7B Base* (DeepSeek-AI, 2024) with identical decoding settings across conditions and the city-state experiment use *Llama-3-8b-instruct* (Dubey et al., 2024). We study *inference time* only in the experiments, so no training, fine-tuning, or gradient updates are performed. The model simply consumes the provided context demonstrations and predicts via the count-and-argmax decision rule analyzed in the main text.

C.1 Synthetic experiments

We use a two-symbol alphabet $\{A,B\}$ for local rules. Each symbol $r \in \{A,B\}$ denotes a two-point stochastic update on the current discrete state: with probability $p(r)$ apply one update (e.g., $+u(r)$), and with probability $1-p(r)$ apply the other (e.g., $-v(r)$). The states are wrapped to remain in $[k]$. A test instance is a two-step tuple (x_0, r_1, r_2) with $x_0 \sim \mu$ (uniform by default) and $r_t \in \{A,B\}$. The key manipulation is how the *semantics* of the letters are tied across steps. In **same** (aligned), the letter-to-operation mapping is fixed across steps. In **diff** (misaligned), the step-2 mapping is altered so that the same letter need not mean the same stochastic rule as in step 1. For each configuration we draw $n \in \{8,\ldots,60\}$ i.i.d. context trajectories. For context samples, **NonCoT** records only terminals (x_0, r_1, r_2, x_2), while **CoT** records full paths $(x_0, r_1, x_1, r_2, x_2)$. We report accuracy ($\pm$ SE over seeds) and the proxy $\Delta n(\tau) = n_{\mathrm{CoT}}(\tau) - n_{\mathrm{NonCoT}}(\tau)$. Noise is controlled by moving $p(A)$ and $p(B)$ toward $1/2$ (smaller per-step margin) while holding all other factors fixed.

C.2 REALISTIC EXPERIMENT: MODULAR ADDITION

Each instance specifies modulus M, initial $x_0 \in \{0, \ldots, M-1\}$, and addends $(a_1, \ldots, a_L)$. The target is $x_L \equiv x_0 + a_1 + \cdots + a_L \pmod{M}$. We compare **same** ($a_1 \equiv \cdots \equiv a_L$) vs. **diff** ($a_1, \ldots, a_L$ drawn independently). Contexts contain either only terminals (x_L) (NonCoT) or also $x_l \equiv x_0 + a_1 + \cdots + a_l$ for all l (CoT), with the same budget n and seeds as in the synthetic suite. Estimation and reporting metrics follow identically from the synthetic protocol, providing a practical check that alignment drives CoT gains beyond the idealized setting.

C.3 REALISTIC EXPERIMENT: CITY–STATE RANKINGS

We build a text corpus of U.S. city/state rankings under two criteria: **population** and **area**. Facts are rendered in a fixed template for (i) states (global ranks) and (ii) cities within each state (within-state ranks), for both criteria. Each query is phrased as *"the X-th largest city in the Y-th largest state"* with explicit criteria for the state step and the city-within-state step. Ground truth is obtained by (1) selecting the Y-th state under the chosen state criterion, then (2) selecting the X-th city within that state under the chosen city criterion, and outputting the resulting city name. We define **same-skill** by using the same criterion for both steps (pop–pop or area–area) and **diff-skill** by mixing criteria (pop–area or area–pop). All prompts include the corpus and n in-context demonstrations. CoT demonstrations expose the two intermediate steps (state identification, then city identification) before the final answer. NonCoT demonstrations output only the final city name. Evaluation in both cases is on the final answer only.