Hán Dān Xué Bù (Mimicry) or *Qīng Chū Yú Lán* (Mastery)? A Cognitive Perspective on Reasoning Distillation in Large Language Models

Yueqing Hu[1,2], **Xinyang Peng**[3], **Shuting Peng**[4], **Hanqi Wang**[5], **& Tianhong Wang (wangtianhong@ahu.edu.cn)**[1†]

[1]School of Philosophy, Anhui University, Hefei, China
[2]Institute of Neuroscience, Chinese Academy of Sciences, Shanghai, China
[3]Faculty of Education, University of Cambridge, Cambridge, UK
[4]School of Foreign Studies, South China Normal University, Guangzhou, China
[5]Division of Psychology and Language Sciences, University College London, London, UK

[†] corresponding author

Abstract

Recent Large Reasoning Models trained via reinforcement learning exhibit a "natural" alignment with human cognitive costs. However, we show that reasoning distillation via Supervised Fine-Tuning (SFT) fails to transmit this cognitive logic, leading to a "Cargo Cult" where students only mimic length. Testing the *Hán Dān Xué Bù* (Superficial Mimicry) hypothesis across 14 models, we identify a "Functional Alignment Collapse": while teacher models mirror human difficulty scaling ($\bar{r} = 0.64$), distilled students significantly degrade this alignment ($\bar{r} = 0.34$). Crucially, they exhibit "Negative Transfer," dropping below their own pre-distillation baselines. Our analysis reveals a "Linear Inflation Law" where students apply a constant verbosity multiplier (≈ 2.44) regardless of complexity. Consequently, distillation decouples computational cost from cognitive demand, revealing that human-like cognition is an emergent property of active reinforcement, not passive imitation.

Keywords: Large Reasoning Models; Cognitive Modeling; Reasoning Distillation; Cognitive Alignment; Supervised Fine-Tuning; Chain-of-Thought

Introduction

A central goal of cognitive science and artificial intelligence is to construct systems that not only solve problems but do so via mechanisms analogous to human cognition. Recently, de Varda et al. (2025) reported that Large Reasoning Models (LRMs) trained via Reinforcement Learning with Verifiable Rewards (RLVR) exhibit a "natural" alignment with human cognitive costs: the length of their Chain-of-Thought (CoT) reasoning traces predicts human reaction times across diverse domains, suggesting that goal-directed optimization can implicitly recover core features of human problem-solving complexity (Anderson, 1990).

Yet the alignment reported by de Varda et al. (2025) emerged from models that *learned to reason* through autonomous exploration. In contrast, the prevailing paradigm for deploying efficient AI is **knowledge distillation** (Hinton et al., 2015), where smaller "student" models are trained via Supervised Fine-Tuning (SFT) to mimic the output traces of powerful "teachers" (Ho et al., 2023). This shift from *learning by doing* to *learning by imitating* raises a fundamental question: when a student learns to generate reasoning traces by copying a teacher, does it internalize the teacher's *cognitive structure*, or merely mimic its *linguistic style*?

Following Brady et al. (2025), we adopt dual-process theory (Kahneman, 2011) as an *organizing structure*—not as a claim about shared cognitive architecture—and frame two competing hypotheses:

1. **The *Hán Dān Xué Bù* (Superficial Mimicry) Hypothesis**.[1] Distillation induces a "surface approach" to learning (Marton & Säljö, 1976): the student acquires the *form* of reasoning (verbosity) without the *function* (dynamic resource allocation).
2. **The *Qīng Chū Yú Lán* (Functional Mastery) Hypothesis**.[2] Alternatively, the student may successfully compress the teacher's reasoning policy, maintaining or improving alignment with human cognitive costs relative to its size.

We test these hypotheses by benchmarking 14 models—the DeepSeek-R1 teacher (Guo et al., 2025), six distilled variants (Qwen/Llama backbones), and their Instruct-tuned counterparts—against human reaction times across six reasoning tasks adapted from de Varda et al. (2025). Our results support the *Mimicry* hypothesis and identify a "Linear Inflation Law" under which students apply a fixed verbosity multiplier to their base outputs, decoupling computational cost from cognitive demand. Alignment with human cognition thus appears to be an emergent property of reinforcement learning, lost under the superficial imitation of supervised distillation.

Methods

Datasets

We utilized six reasoning tasks adapted from de Varda et al. (2025), spanning diverse cognitive domains (see Figure 1). To verify alignment across a broad spectrum of human intelligence, we selected tasks ranging from algorithmic processing

[1]*Hán dān xué bù* (邯郸学步) refers to someone who blindly imitates the walking style of people in Handan, only to fail and forget their own. It implies that student models might degrade their base capabilities while imperfectly mimicking the teacher.

[2]*Qīng chū yú lán* (青出于蓝) means "indigo blue is extracted from the indigo plant but is bluer than it," describing a student who surpasses their teacher.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

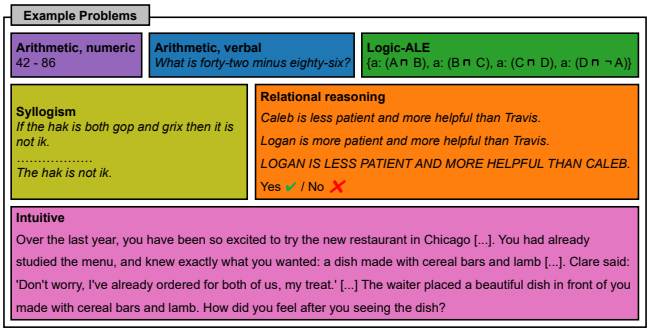

Figure 1: Example reasoning problems.

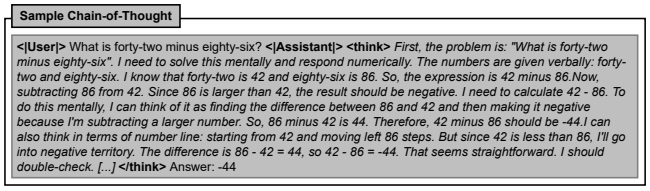

Figure 2: Sample chain-of-thought produced by R1.

to common-sense intuition.[3]

1. **Arithmetic:** 168 items (84 numeric, 84 verbal) requiring *rigid algorithmic processing*, solved by $N = 60$ participants.
2. **Syllogisms:** 32 problems testing *formal deductive logic*, solved by $N = 24$ participants.
3. **Logic-ALE:** 20 consistency judgment problems in ALE format (also formal logic), solved by $N = 84$ participants.
4. **Relational Reasoning:** 84 items testing *transitive inference* ($N = 310$).
5. **Intuitive Reasoning:** 144 vignettes requiring *common-sense causal judgments* rather than formal rules ($N = 58$).

Models and Inference

To dissect the mechanics of reasoning distillation, we employed a rigorous teacher-student-baseline comparison involving 14 models (see Table 1).

Experimental Control: The Role of Instruct Baselines. A critical confound in evaluating reasoning distillation is distinguishing between the capability to follow instructions ("chat") and the capability to reason ("think"). Teacher models (e.g., DeepSeek-R1) transmit both linguistic fluency and reasoning structures. To isolate the effect of reasoning distillation, we exclusively selected **Instruct-tuned** checkpoints as our Base Models (e.g., Qwen2.5-Math-Instruct) rather than raw pre-trained bases. This design ensures that any behavioral divergence observed in the Distilled models is attributable to the internalization of *Chain-of-Thought patterns*, rather than general instruction-following capabilities.

We evaluated three model groups:

1. **Teacher Models:** DeepSeek-R1 and DeepSeek-V3, serving as the gold standard for cognitive alignment.
2. **Distilled Models (Student):** Six models from the DeepSeek-R1-Distill family, utilizing Qwen (1.5B, 7B, 14B, 32B) and Llama (8B, 70B) backbones.
3. **Base Models:** The six corresponding Instruct-tuned checkpoints: Qwen2.5-Math-Instruct (1.5B, 7B), Qwen2.5-Instruct (14B, 32B), Meta-Llama-3.1-Instruct (8B), and Llama-3.3-Instruct (70B).

Models were accessed via commercial APIs (Aliyun DashScope for Qwen-series, DeepInfra/ModelScope for Llama-series, Together AI for DeepSeek) with temperature $T = 0$ to ensure deterministic reproducibility.

Measures

For each problem i and model m, we extracted:

- **Accuracy ($Acc_{m,i}$):** Validated against ground truth.
- **Reasoning Cost ($C_{m,i}$):** For R1 and distilled models, cost is defined as the token count within the `<think>` delimiters (see Figure 2). For base models, we appended the CoT trigger "*Let's think step by step*" and counted the full completion length.

Analysis Framework

We structured our analysis to distinguish between superficial imitation (*Mimicry*) and functional internalization (*Mastery*).

1. Functional Alignment We assessed whether models retain human-like difficulty scaling by computing the Pearson correlation (r_{align}) between the log-transformed reasoning cost and human reaction times (RT):

$$r_{align} = \text{Corr}(\log(C_m), \log(RT_{human})) \tag{1}$$

2. Representational Similarity Analysis (RSA) To quantifiably map the topology of reasoning strategies, we constructed "difficulty fingerprints" for all agents. We first log-transformed reasoning costs and z-scored them *within each task* to isolate relative difficulty patterns. Formally, the fingerprint $\mathbf{v}_m$ for model m is the concatenation of standardized costs across tasks:

$$\mathbf{v}_m = \bigoplus_{t \in T} \mathcal{Z}(\log(\mathbf{c}_{m,t})) \tag{2}$$

We utilized these fingerprints to compute pairwise Pearson correlations, generating the RSA matrix. Furthermore, to visualize the distillation geometry, we projected models into a 2D space defined by **Teacher Similarity** ($x = \text{Corr}(\mathbf{v}_S, \mathbf{v}_{Teacher})$) and **Human Alignment** ($y = \text{Corr}(\mathbf{v}_S, \mathbf{v}_{Human})$).

[3] The *H-ARC* task from the original study was excluded to strictly isolate linguistic reasoning costs. Recent scholarship argues that ARC tasks are fundamentally vision-centric problems relying on spatial priors rather than linguistic logic (K. Hu et al., 2025).

Table 1: Accuracy (%) comparisons across six reasoning tasks. Human baselines are adapted from de Varda et al. (2025). Models were evaluated using **Pass@1** with greedy decoding ($T = 0$). **Bold** indicates the best performance in each column; underline indicates the best performing Distilled (Student) model.

	Arithmetic						
Model	*Numeric*	*Verbal*	**Syllogism**	**Logic**	**Relational**	**Intuitive**	**Average**
Human Baseline	96.0	95.1	84.6	83.9	72.1	81.6	85.5
Teacher Models							
DeepSeek-R1	**100.0**	94.2	**100.0**	**100.0**	94.2	**88.9**	**97.2**
DeepSeek-V3	**100.0**	**100.0**	**100.0**	**100.0**	94.2	81.2	95.9
Distilled Models (Student)							
DS-R1-Distill (Qwen-1.5B)	96.4	97.6	81.2	80.0	64.0	47.2	77.7
DS-R1-Distill (Qwen-7B)	97.6	95.2	100.0	100.0	89.5	52.8	89.2
DS-R1-Distill (Qwen-14B)	100.0	100.0	100.0	95.0	**97.7**	75.0	94.6
DS-R1-Distill (Qwen-32B)	100.0	100.0	100.0	100.0	95.3	77.8	95.5
DS-R1-Distill (Llama-8B)	95.2	95.2	96.9	85.0	82.6	64.6	86.6
DS-R1-Distill (Llama-70B)	100.0	98.8	100.0	100.0	94.2	85.4	96.4
Base Models (Pre-distillation)							
Qwen2.5-Math (1.5B)	91.7	96.4	50.0	85.0	60.5	31.9	69.3
Qwen2.5-Math (7B)	100.0	98.8	84.4	85.0	66.3	53.5	81.3
Qwen2.5 (14B)	100.0	100.0	93.8	100.0	86.0	73.6	92.2
Qwen2.5 (32B)	100.0	100.0	96.9	100.0	86.0	81.2	94.0
Llama-3.1 (8B)	95.2	89.3	81.2	80.0	79.1	66.7	81.9
Llama-3.3 (70B)	100.0	98.8	96.9	100.0	94.2	81.9	95.3
Distilled Avg (±SD)	98.2 ±2.1	97.8 ±2.2	96.4 ±7.5	93.3 ±8.8	87.2 ±12.6	67.1 ±15.0	90.0 ±11.9
Base Avg (±SD)	97.8 ±3.6	97.2 ±4.1	83.9 ±17.8	91.7 ±9.3	78.7 ±12.9	64.8 ±19.3	85.7 ±12.7

3. Surface Similarity Metrics To quantify how closely students mimic the teacher's output distribution while accounting for varying task lengths, we first calculated the **Relative Effort** ($e_{m,i}$) by normalizing the token cost of each item by the teacher's average cost for that task ($e_{m,i} = C_{m,i}/\bar{C}_{T,\text{task}}$). We then computed two metrics over the distributions of e:

$$\rho = \mu(e_S), \quad D_{KL}(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right) \tag{3}$$

where ρ denotes the **Thinking Effort Ratio** (measuring verbosity inflation), and D_{KL} denotes the **Strategy Divergence** between the probability distribution of the Teacher's effort (P) and the Student's effort (Q).

4. Structural Mechanism: The Linear Inflation Law To determine the generative mechanism of efficiency degradation, we first formulated the **Inverse Efficiency Index** ($\mathcal{I}_E$) as the ratio of average reasoning cost ($\bar{C}$) to accuracy (Acc):

$$\mathcal{I}_E = \frac{\bar{C}_m}{Acc_m} \tag{4}$$

We then tested for a **proportional Linear Inflation Law** by modeling the relationship between student and base models via a regression constrained to the origin:

$$\mathcal{I}_E^{student} \approx N \cdot \mathcal{I}_E^{base} \tag{5}$$

where the slope N represents a "verbosity multiplier." A strict fit to this model implies that distillation imposes a pure multiplicative penalty on computational cost, independent of the base model's intrinsic heuristics.

Results

The Phenomenology of Failure: Diagnosis of "Hán Dān Xué Bù"

Functional Alignment Collapse Reasoning distillation induces a severe collapse in cognitive alignment. While the Teacher model (DeepSeek-R1) mirrors human difficulty scaling ($\bar{r} = 0.640$), Distilled models significantly degrade this alignment ($\bar{r} = 0.341$, $t = 2.45$, $p = 0.044$).

Crucially, we observe a **"Negative Transfer"** effect confirmed by statistical testing ($t = -2.60$, $p = 0.011$). Distilled models perform significantly *worse* than their own Base models ($\bar{r}_{base} = 0.534$) and fail to match the baseline alignment of standard RLHF models like DeepSeek-V3 ($\bar{r} = 0.558$). This indicates that the distillation process effectively "overwrites" the base models' native, efficient heuristics with a maladaptive

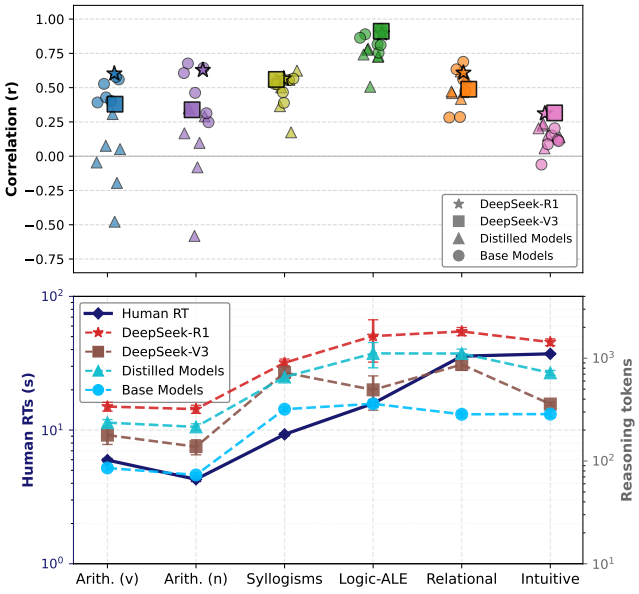

Figure 3: **Combined Analysis of Cognitive Alignment and Reasoning Cost.** (Top) Correlation (r): between reasoning cost and human RTs. (Bottom) Comparison of Human RTs (solid, left axis) and model token counts (dashed, right axis) across tasks.

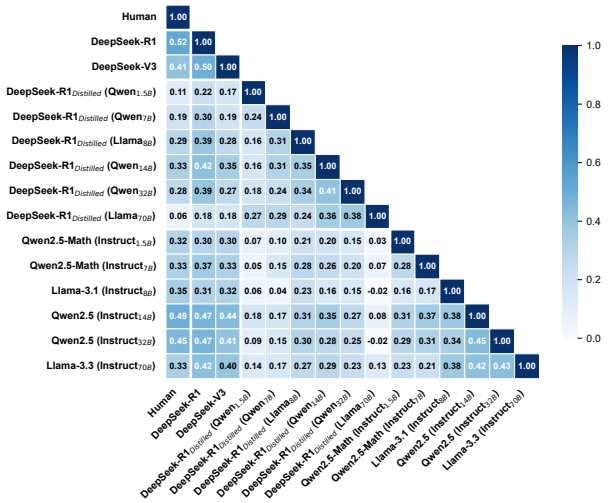

Figure 4: **Representational Similarity Analysis (RSA) Heatmap.** Values denote Pearson correlations of difficulty fingerprints (z-scored costs). Distilled models exhibit low structural similarity to the Teacher (DeepSeek-R1), indicating a failure to internalize the target reasoning policy.

strategy (see Figure 3, top panel), supporting the premise that the student has lost its original "way of walking."

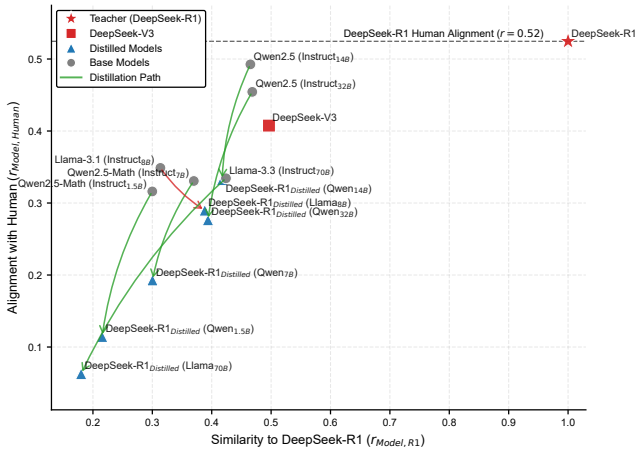

Figure 5: **The Cognitive Waterfall.** Models define a topology of failed imitation defined by Teacher Similarity (x-axis) and Human Alignment (y-axis). Instead of converging to the Teacher (top-right), distilled models collapse into a "Spurious Valley" characterized by simultaneous loss of alignment (The Drop) and structural similarity (The Drift).

Topological Drift: The "Spurious Valley" To map the structural consequences of this collapse, we employed Representational Similarity Analysis (RSA). RSA reveals a stark representational drift (Figure 4): while Human and Teacher patterns are structurally similar ($r = 0.52$), distilled models become untethered from both.

Figure 5 illustrates the **"Cognitive Waterfall."** Instead of converging to the Teacher, student models collapse into a "Spurious Valley" driven by two vectors:

1. **The Drop (Alignment Collapse):** A vertical descent indicating a loss of human-likeness. In terms of representational similarity to humans, distilled models score significantly lower ($\bar{r} = 0.21$) than their base counterparts ($\bar{r} = 0.38$, $p = 0.01$).
2. **The Drift (Similarity Regression):** A counter-intuitive leftward shift indicating that distillation actively *degrades* structural similarity to the teacher. Quantitatively, distilled models show lower representational correlation with the Teacher ($\bar{r} = 0.32$) than even their raw, pre-distilled base counterparts ($\bar{r} = 0.39$), showing a marginally significant trend ($p = 0.09$).

Together, these results confirm that distilled models occupy a "Spurious Valley"—a cognitive state distinct from both human intuition and robust machine reasoning.

The Mechanism of Mimicry: Quantitative Analysis

To investigate the generative mechanisms of the collapse, we first analyzed surface-level metrics.

Surface-Level Divergence. We computed the Thinking Effort Ratio (ρ) and Strategy Divergence (D_{KL}) to assess distributional mimicry. As illustrated in Figure 6, distillation

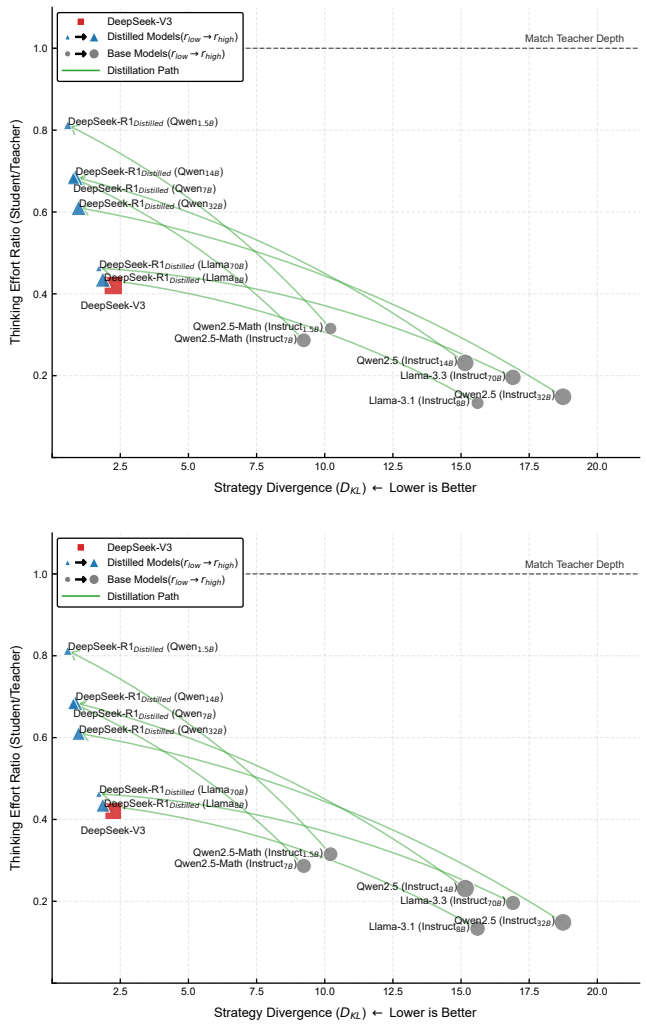

Figure 6: **Surface-Level Mimicry Landscape.** Models migrate to the high-verbosity "Mimicry Region" (top-left, low D_{KL} and $\rho \approx 1.0$). Bubble sizes indicate (a, top) representational similarity to the Teacher and (b, bottom) alignment with Human reaction times. Shrinking bubbles in *both* panels reveal that surface-level mimicry decouples computational effort from both the Teacher's item-wise allocation policy and human cognitive demand.

effectively minimizes the Kullback-Leibler divergence, driving student models into a "Mimicry Region" characterized by low D_{KL} and a Teacher-like effort density ($\rho \approx 1.0$).

However, visual inspection of Figure 6 reveals a critical discrepancy. In panel (a), bubbles shrink as students enter the Mimicry Region, indicating that despite matching the Teacher's *aggregate* effort distribution (low D_{KL}), distilled students diverge from the Teacher's *item-wise* difficulty fingerprint ($\bar{r}_{Model,R1} = 0.32$; cf. Figure 4). In panel (b), bubbles shrink further, showing that alignment with human reaction times degrades even more ($\bar{r}_{Model,Human} = 0.21$). This *dual decoupling*—from both the Teacher's allocation policy and human cognitive demand—demonstrates that distillation transmits the statistical form of reasoning traces without their underlying cognitive function.

The Linear Inflation Law. We further investigated intrinsic efficiency changes by modeling the Inverse Efficiency Index ($\mathcal{I}_E$). Regression analysis (Figure 7) reveals a strict linear dependency between the inefficiencies of student and base models (Pearson's $r = 0.92$, $p < 0.01$). We confirm the **Linear Inflation Law** proposed in Eq. (5), finding a fitted slope of $N \approx 2.44$. This result suggests that distillation imposes a systematic, multiplicative penalty on computational cost. Rather than adopting the Teacher's dynamic allocation policy, student models mechanically inflate their base output length by a fixed factor of ~ 2.44 across tasks, regardless of problem complexity.

Discussion

Our findings reveal a paradox: while student models statistically reproduce the teacher's verbose traces, they fail to internalize the underlying cognitive policy driving that verbosity. This disconnect—the *Hán Dān Xué Bù* effect—exposes a fundamental epistemic limit in current reasoning distillation paradigms.

The Illusion of Depth: Why Distillation Degrades Alignment

A critical anomaly in our results is the "Negative Transfer" effect: distilled models exhibit significantly worse cognitive alignment ($\bar{r} = 0.34$) than their own instruction-tuned base models ($\bar{r} = 0.53$). Since both groups utilize Supervised Fine-Tuning (SFT), the failure is not inherent to SFT but specifically arises from *CoT Distillation*.

Standard instruction tuning (Base Models) typically optimizes for answer correctness across diverse tasks, allowing the model to utilize its native, albeit simpler, heuristics. In contrast, CoT distillation forces the student to mimic the teacher's extended "thinking process" as a target output. This triggers a form of catastrophic interference (French, 1999), where verbose patterns overwrite the base model's efficient latent structures. Unlike biological complementary learning systems (McClelland et al., 1995) that integrate new knowledge without erasing the old, the distilled models succumb to "Negative Transfer," performing worse than their pre-distillation baselines.

Furthermore, our "Linear Inflation Law" ($N \approx 2.44$) suggests that students treat reasoning tokens not as functional steps, but as a rigid stylistic constraint. This mirrors the Einstellung (set) effect (Bilalić et al., 2008, 2010): the student becomes "blinded" by a familiar, complex strategy (verbosity), mechanically applying it even when simpler heuristics would suffice. The result is a "fluency illusion" (Deslauriers et al., 2019): the model's coherent output masks a lack of cognitive engagement, creating a false appearance of competence.

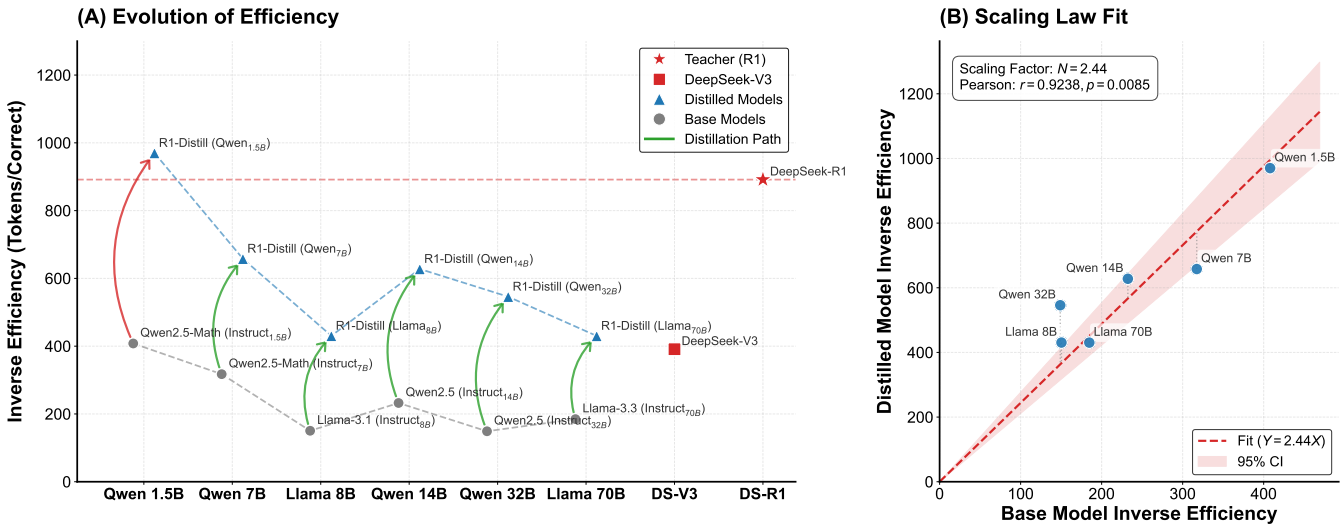

Figure 7: **Inverse Efficiency Analysis. (A)** Distilled models (blue triangles) exhibit consistently worse efficiency (tokens/correct) than their base counterparts (grey circles). **(B)** Regression confirms a Linear Inflation Law: distillation imposes a systematic multiplicative penalty ($N \approx 2.44$) on the base model's inefficiency, independent of intrinsic reasoning heuristics.

The Credit Assignment Problem in Passive Learning

Why does mimicking the teacher's process lead to functional collapse? We propose that reasoning emerges from *credit assignment*—a mechanism SFT structurally lacks. In the teacher's Reinforcement Learning (RL) phase, verbosity is a byproduct of a search process surviving sparse rewards; a long chain exists only because it was instrumental in reaching a correct solution. The teacher effectively performs a neuroeconomic calculation, scaling cognitive effort where the expected value of correctness outweighs the computational cost (Westbrook & Braver, 2015).

In SFT, however, every token is weighted equally; students receive identical updates for both vacuous fillers and crucial logical pivots. Consequently, the student fails to develop critical meta-reasoning capabilities—specifically the monitoring and control of cognitive effort (Ackerman & Thompson, 2017). Lacking internal error signals or the "feeling of rightness" that guides human termination of thought, the student cannot distinguish between the "scaffolding" of thought and the thought itself. This aligns with the "Spring Model" of alignment (Ji et al., 2025): distillation stretches the model's output distribution to match the teacher's length (elasticity) but fails to restructure its internal plasticity, resulting in the "Spurious Valley" observed in our RSA analysis. This account dovetails with an ongoing debate. Vankov et al. (2026) and Y. Hu (2026) argued that LRM trace length may reflect *performative scaffolding* rather than genuine computation, while de Varda et al. (2026) showed that RLVR traces track validated difficulty dimensions. Our findings dissociate these positions: the alignment signal survives under RLVR but collapses under CoT distillation, with the Linear Inflation Law as the quantitative signature of the scaffolding regime.

From Behavioral Mimicry to Cognitive Mastery

The contrast between our RL-trained teacher and SFT-trained students parallels, *heuristically*, the distinction between *behaviorist* conditioning and *constructivist* learning—with the caveat, following Brady et al. (2025), that such framings serve to interpret outputs rather than assert shared cognitive architecture. SFT assumes that reasoning can be transmitted via passive observation—watching the teacher walk. However, cognitive skill acquisition theories posit that declarative knowledge (traces) does not convert to procedural knowledge (reasoning) without active practice (Anderson, 1982; VanLehn, 1996). Our data suggests that "System 2" reasoning (Kahneman, 2011) is not a static template to be memorized, but a dynamic policy for resource allocation that requires the proceduralization of heuristics.

True mastery (*Qīng Chū Yú Lán*) likely requires the agent to engage in "cognitive foraging"—actively exploring the search space and experiencing the error signals necessary to prune inefficient paths. This aligns with findings that self-directed learning facilitates structural generalization significantly better than passive reception (Gureckis & Markant, 2012). While SFT provides the necessary linguistic scaffolding, akin to Vygotsky's Zone of Proximal Development (Vygotsky, 1978), transcending the teacher's capabilities requires the autonomous pressure of rewards.

Future cognitive modeling must therefore transcend "trace cloning," utilizing methods that incentivize students to rediscover the *function* of reasoning. Specifically, we propose training on traces only when they yield verified correct answers, enabling the model to experience its own success, or employing teacher traces as a *prior* within reinforcement learning rather than as static scripts.

Acknowledgments

We thank Dr. Huzi Cheng, Jiaxin Li, and Junxian Yang for insightful discussions on the use of Reinforcement Learning (RL) and Supervised Fine-Tuning (SFT) techniques in training Large Language Models (LLMs), which substantially informed the framing of this work. We acknowledge the use of Gemini 3 Pro to assist with the development of experimental scripts, data analysis code, and visualization pipelines. Additionally, the model was utilized for linguistic refinement and copy-editing of the manuscript. The authors retain full responsibility for the scientific content and final text.

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, *21*(8), 607–617.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious einstellung (set) effect. *Cognition*, *108*(3), 652–661.

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science*, *19*(2), 111–115.

Brady, O., Nulty, P., Zhang, L., Ward, T. E., & McGovern, D. P. (2025). Dual-process theory and decision-making in large language models. *Nature Reviews Psychology*, 1–16.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, *116*(39), 19251–19257.

de Varda, A. G., D'Elia, F. P., Kean, H., Lampinen, A., & Fedorenko, E. (2025). The cost of thinking is similar between large reasoning models and humans. *Proceedings of the National Academy of Sciences*, *122*(47), e2520077122. https://doi.org/10.1073/pnas.2520077122

de Varda, A. G., D'Elia, F. P., Kean, H., Lampinen, A., & Fedorenko, E. (2026). Reply to Vankov et al.: Reasoning traces are linked to accuracy and capture key dimensions of problem complexity. *Proceedings of the National Academy of Sciences*, *123*(12), e2603574123.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128–135.

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Ho, N., Schmid, L., & Yun, S.-Y. (2023). Large language models are reasoning teachers. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 14852–14882. https://doi.org/10.18653/v1/2023.acl-long.830

Hu, K., Lin, A. C., Qiu, L., Ding, X. D., Wang, R., Zhu, Y. E., Andreas, J., & He, K. (2025). ARC is a vision problem! *arXiv preprint arXiv:2511.14761*.

Hu, Y. (2026). "Thinking traces" in large reasoning models: Cognitive cost or performative scaffolding? *Proceedings of the National Academy of Sciences*, *123*(17), e2604554123.

Ji, J., Wang, K., Qiu, T., Chen, B., Zhou, J., Li, C., Lou, H., Dai, J., Liu, Y., & Yang, Y. (2025). Language models resist alignment: Evidence from data compression. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 23411–23432. https://aclanthology.org/2025.acl-long.1264/

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus; Giroux.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

Vankov, I. I., Adolfi, F., Heaton, R. F., Puebla, G., & Bowers, J. S. (2026). Correlations without causation do not support claims of human–LLM reasoning alignment. *Proceedings of the National Academy of Sciences*, *123*(12), e2536362123.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415.