Are More Tokens Rational?
Inference-Time Scaling in Language Models as Adaptive Resource Rationality

Zhimin Hu[1,2], **Riya Roshan**[2] **& Sashank Varma**[1,2]
[1]School of Psychological and Brain Science, Georgia Institute of Technology
[2]School of Interactive Computing, Georgia Institute of Technology

Abstract

Human reasoning is characterized by the rational use of resources to optimize performance under constraints. Recently, inference-time scaling has improved the reasoning performance of Large Language Models by increasing test-time computation. Instruction-tuned (IT) models explicitly generate long reasoning traces, whereas Large Reasoning Models (LRMs) are trained via reinforcement learning to discover reasoning paths that maximize accuracy. However, it remains unclear whether resource-rationality can emerge from such scaling without explicit rewards related to computational costs. We introduce a Variable Attribution Task (VAT) in which models infer which variables determine outcomes given candidate variables, input–output trials, and predefined logical functions. By varying the number of candidate variables and trials, we systematically manipulate task complexity. Both models exhibit a transition from brute-force to analytic strategies as complexity increases. IT models degrade on XOR and XNOR functions, whereas LRMs remain robust. These results suggest that resource rationality can emerge from inference-time scaling itself, even without explicit cost-based rewards.

Keywords: Reasoning, Resource Rationality, Inference-Time Scaling, Large Reasoning Models

Introduction

What is intelligence, if not the art of doing well with less? From biological brains to artificial agents, intelligent behavior requires the efficient use of limited cognitive resources. This idea dates back to Herbert Simon's theory of *bounded rationality* (Simon, 1955), which challenged the notion of 'perfect rationality' (Von Neumann & Morgenstern, 1947) by highlighting the external (environmental) and internal (cognitive) constraints on decision-making, including memory, computation, and time. Rather than seeking optimal solutions, agents 'satisfice' by setting expectation thresholds and stopping once an acceptable option is found. This process of 'constraint satisfaction' is recognized as fundamental for reasoning more generally. Simon's insights have been developed across multiple disciplines. Within economics, Conlisk attempted to bring 'deliberation costs' or 'optimization costs' into the computation of utility (Conlisk, 1988, 1996). Within AI, Boddy and Dean (1994) considered the design of agents planning solutions in time-constrained environments. In behavioral experiments, Payne and colleagues have studied the adaptivity of human decision making and problem solving under external constraints (Payne et al., 1988, 1996).

Expanding on Simon's work, Lieder and Griffiths (2020) proposed that cognition is *resource rational* and formalized reasoning as an optimization problem under cognitive constraints. In this view, agents seek to maximize the difference between expected utility and the cost of computation. Callaway et al. (2022) extended this approach by modeling reasoning as a form of metacognitive control, where agents flexibly adapt between strategies based on their relative reward and processing cost. Such models portray strategy adaptation as a core computational property of bounded intelligence.

The current study investigates whether resource rationality extends beyond human cognition to be a general computational principle of intelligent agents under constraint. Although paradigms like Mouselab (Lieder & Griffiths, 2020; Payne et al., 1988) have been valuable for exploring how humans trade off information acquisition cost and accuracy. they focus primarily on *external* sampling behavior. As such, they are less ideal for isolating the *internal* computational tradeoffs that underlie general strategy adaptation.

To address this limitation, we introduce the Variable Attribution Task (VAT), a structured learning paradigm for investigating how agents manage internal resources. As shown in Figure 1, the agent is given N candidate variables and T input–output trials where a predefined Boolean function maps a subset of the candidate variables to the outcome. This is a very general setup that can be mapped to multiple forms of reasoning, e.g., identifying causal variables in scientific reasoning. The VAT supports two cognitively distinct strategies: *permutation*-based search, which performs serial hypothesis testing with lower working memory (WM) demands; and *elimination*-based inference, which maintains and prunes a hypothesis space based on disconfirming evidence and makes higher WM demands. Both strategies operate over a hypothesis space that explodes with the number of candidate variables and evaluate it across T trials. However, their convergence efficiency differs depending on the structure of the target function. Under AND-like functions, elimination can rapidly reduce the candidate space, yielding faster identification of the subset of relevant variables. In contrast, under XOR-like functions, disconfirmation typically offers less pruning power, and thus, permutation can be the more efficient strategy.

A recent shift in the reasoning paradigm of Large Language Models (LLMs) has made them useful testbeds for the question of whether resource rationality can emerge naturally, as a computational principle of optimization under constraint. This shift leads to inference time scaling, which attempts to

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

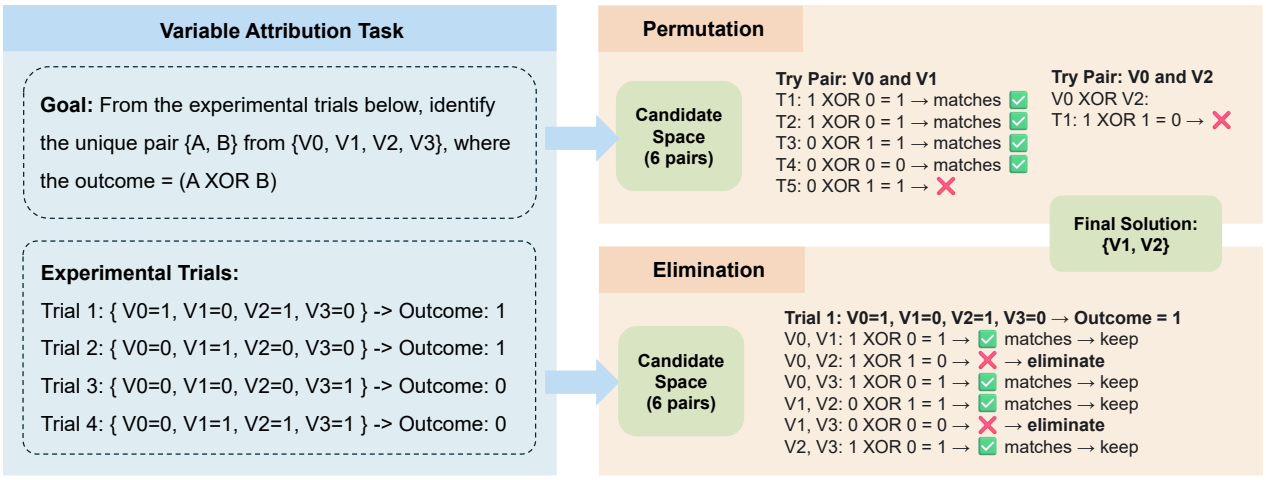

Figure 1: Variable Attribution Task and its computational strategies. Given input–output records from T trials and N candidate variables, the model must identify the pair that determines the output via a latent Boolean function. Solving the task starts either from testing candidate pairs (permutation) or incrementally pruning inconsistent pairs (elimination).

improve reasoning by expanding the computation at test time. Chain-of-Thought (CoT) (Wei et al., 2022) demonstrated that generating reasoning steps boosts performance without additional training. Instruction-tuned (IT) models build on this by undergoing Supervised Fine-Tuning (SFT) on CoT-style data, whereas Large Reasoning Models (LRMs) such as DeepSeek R1 (Guo et al., 2025) further optimize reasoning behavior through Reinforcement Learning (RL). While prior work evaluating human and machine reasoning by their ability to produce fixed answers (Hu et al., 2024; Webb et al., 2023) and short-form responses (Kabra et al., 2025), inference-time scaling enables models to generate explicit reasoning trajectories. These reasoning tokens serve as a traceable cognitive process, allowing us to analyze how models perform strategy adaptation over the course of inference. (We consider problems with taking reasoning tokens at face value in the Discussion.)

Using the VAT, we systematically manipulate task complexity to test IT models and LRMs, and we track their shifting between the permutation and elimination strategies. Unlike paradigms that emphasize exploration cost, the VAT specifically targets strategy adaptation, offering a more direct view of how computational effort is reallocated under constraint. Our results show that both IT models and LRMs exhibit a graceful transition from the permutation strategy to the elimination strategy as task complexity increases. However, for the XOR and XNOR functions, where pruning offers little benefit and maintaining the hypothesis space becomes costly, this transition is diminished or absent. In addition, although both model types show similar patterns of strategy adaptation, there is an important difference: LRMs maintain stable accuracy across functions, whereas IT models degrade for the more-demanding XOR and XNOR functions. Our findings demonstrate that resource rationality is an emergent property of inference-time scaling rather than a behavior requiring explicit cost-based rewards.

Related Work

Inference Time Scaling

Inference-time scaling refers to the paradigm where models improve performance by allocating more computation at test time. Early work relied on prompt-based techniques. CoT (Wei et al., 2022) instructs models to generate intermediate reasoning steps before producing the final answer, while Tree-of-Thoughts (ToT) (Yao et al., 2023) frames inference as an explicit search process with lookahead and backtracking.

Recent approaches aim to integrate such intermediate computation into post-training. Models are first fine-tuned on long CoT datasets via SFT, enabling instruction-tuned (IT) models to follow prompts while generating reasoning steps. In contrast, reasoning models explicitly separate thinking tokens from output tokens and are further trained with RL to regulate reasoning. In DeepSeek R1 (Guo et al., 2025), RL rewards correct final answers, encouraging reasoning as latent search. In Qwen3 (Yang et al., 2025), RL focuses on reasoning coherence and control, penalizing both inconsistencies and incorrect answers. Empirically, increasing inference-time yields consistent performance gains (Snell et al., 2025).

Induction and Attribution

Induction is the learning of latent regularities from finite observations. The process of inferring specific causes of observed outcomes, termed 'attribution', is an essential part of induction (Cheng & Novick, 1990). In human induction, Cheng and Novick (1990) proposed that causal strength is judged by comparing the probability of an outcome in the presence versus the absence of a potential cause, and this contrastive process is, in essence, a form of attribution. However, the exponential (in the number of candidates) size of the hypothesis space forces biological agents to rely on local updates and sequential hypothesis testing rather than global

optimization (Bramley, 2017). Attribution difficulty also depends on the logic of the causal relation. Humans favor natural configural strategies (e.g., conjunctive logic) over non-monotonic forms (Ganzach & Czaczkes, 1995). To manage cost-accuracy tradeoffs, distinct strategies emerge. K. Varma et al. (2018) found that students use either instant rationality for rapid pruning or delayed rationality that tolerates inconsistency for global evaluation, independent of general cognitive ability.

Methods

Task Design and Logic Space

The VAT is designed to test the ability to perform causal induction under varying computational constraints. Given N candidate variables $\{V_0, V_1, \ldots, V_{N-1}\}$ and T experimental trials, the model must identify a unique pair of active variables (V_i, V_j) that determines the system output Y according to a given logical function f. We consider the complete space of bivariate Boolean functions. There are $2^{2^2} = 16$ possible functions for two inputs. However, we exclude the 2 constant functions and the 4 univariate functions (which depend on only one or zero variables) because they do not require a variable pair and are thus much simpler to determine. The remaining 10 non-trivial functions are categorized into three groups based on their truth-table characteristics:

- **Conjunctive:** Functions where only 1 of the 4 input combinations yields a positive outcome (e.g., $A \wedge (\neg B)$).
- **Disjunctive:** Functions where 3 of the 4 combinations yield a positive outcome, or conversely, only 1 yields a negative outcome (e.g., $(\neg A) \vee B$).
- **XOR/XNOR:** Non-linear functions where the outcomes are balanced (i.e., 2 positive, 2 negative).

Information Ratio

To systematically manipulate task complexity, we define the **Information Ratio** (ρ) as the relationship between the available experimental information and the hypothesis space: $\rho = 2^T / \binom{N}{2}$ where T is the number of trials and $\binom{N}{2}$ represents the total possible combinations of active variables. When ρ is small (near the theoretical minimum 1), the experimental trials are extremely dense, such that almost every trial is necessary to eliminate incorrect hypotheses, and vice versa when ρ is large.

Computational Strategies

The VAT is, in principle, an NP problem: although a solution can be verified in polynomial time, no known algorithm can solve it in polynomial time in the general case. Solving the task requires evaluating candidate pairs of variables against the input–output mappings observed in the trials. This is naturally done via two strategies. The permutation strategy iterates over all $\binom{N}{2}$ candidate pairs and checks their consistency across all trials. The elimination strategy processes each trial to incrementally prune incompatible pairs. To preclude heuristic shortcuts, we ensured that the input distributions (i.e., the frequency of 0s and 1s) for candidate variables were similar to those of the ground-truth variables. This prevents models from exploiting statistical cues (e.g., "V_1 is the only variable with a complete set of 1 and 0) and forces reliance on logical reasoning. While both strategies have similar theoretical complexity, they differ in cognitive demands and empirical efficiency; see Figure 1. The permutation strategy demands relatively fewer WM resources, as only one hypothesis needs to be maintained and tested at a time. In contrast, the elimination strategy requires updating and maintaining a global hypothesis space, and thus demands more WM resources.

Materials

To systematically investigate model behavior under varying task complexity, we generated a total of 3000 variable attribution samples spanning 10 logical functions. We considered 10 values of N, the number of candidate variables: $[N \in \{3, 4, 5, 6, 7, 8, 10, 12, 14, 16\}$. For each N, we computed the minimal number of trials T_{min} required to uniquely identify the correct variable pair. Thus, we adjusted T_{min} to: $T = T_{\text{min}} + \{0, 1, 2, 3, 4, 5\}$. For each combination of (N, T), we randomly generated 5 samples, resulting in 300 samples per logical function.

Models

We evaluated two families of LLMs on this dataset, `DeepSeek R1 & V3` and `Qwen3-Next-80B-A3B-Thinking & Instruct` (hereafter referred to as Qwen3). All 4 models were evaluated using the same base prompt. For the thinking models, we analyzed their "thinking" tokens to identify the underlying computational strategies. Token usage is measured using character count. For the thinking models, we included both the reasoning trace and the final output in this count. For Qwen3-Instruct, we additionally tested a constrained setting in which the model was instructed to output only the final answer, which serves as the baseline for accuracy.

LLM as Judge

To identify the reasoning strategies employed by models, we used a high-capacity language model, `Kimi-K2-Instruct-0905`, as an external judge. Given a model's response, the judge was instructed to classify the strategy as permutation, elimination, or invalid. To ensure the validity of this approach, we constructed a human-labeled evaluation set. We evenly sampled responses from Qwen3-Thinking and DeepSeek-R1, manually annotated their strategies, and curated a balanced subset of 100 examples with roughly equal numbers of permutation and elimination cases. We then evaluated the prompt on this set. The LLM-based judge achieved an accuracy of 0.86 and a Cohen's Kappa of 0.76, indicating substantial agreement with human annotations and validating its use for analysis.

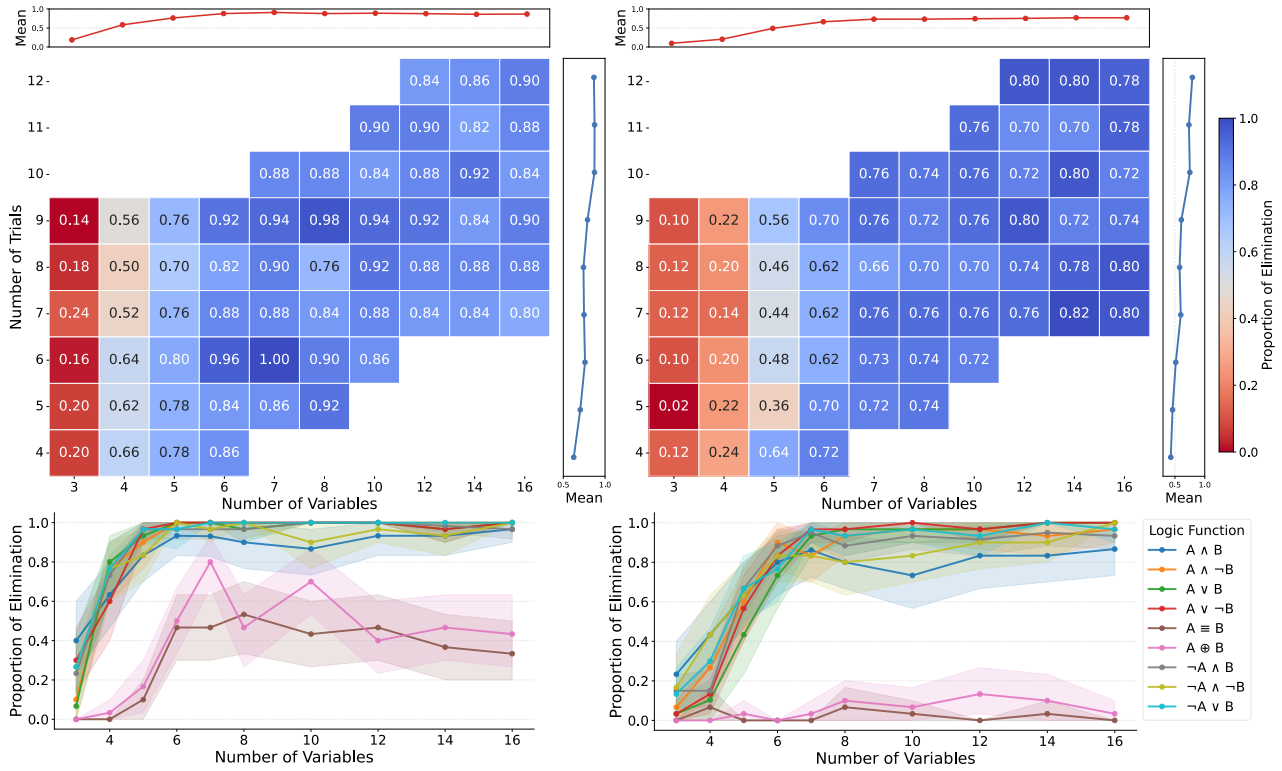

Figure 2: Strategy distribution across task complexity. (Top) Proportion of permutation strategies as N (Red line) and T (Blue line) increase. (Bottom) Proportion of elimination for specific logical functions. (Left) Deepseek-R1. (Right) Qwen3-thinking.

Results

Emergent Strategy Selection

We first investigated whether LLMs adaptively switch between reasoning strategies. As shown in Figure 2, we observed a clear transition in strategy selection for both the DeepSeek and Qwen families. When the number of candidate variables N is small (e.g., $N < 4$), models predominantly employ a permutation strategy, iterating through possible pairs to check consistency. However, as N increases, the hypothesis space $\binom{N}{2}$ grows quadratically. The models adapt and transition toward an elimination strategy, which processes trials one-at-a-time to prune the search space. This trend is observed across both Large Reasoning Models (LRMs) and their instruction-tuned (IT) counterparts (DeepSeek V3 and Qwen-Instruct).

Notably, we observed a divergence in strategy selection for the XOR and XNOR functions. While DeepSeek models show a moderate shift toward elimination for these functions, the proportion remains lower than for conjunctive/disjunctive tasks. In contrast, the Qwen family almost entirely eschews elimination for XOR-like logical functions, sticking with permutation even as N increases. From a resource rational perspective, this divergence is highly instructive. In conjunctive logic (e.g., AND), a single positive trial can eliminate a vast region of the hypothesis space, making the "WM cost" of elimination highly efficient. By contrast, the XOR/XNOR functions are non-monotonic and "information-dense"; they do not allow for rapid pruning of the variable space following single trials. If the computational cost of maintaining a very large and only slowly shrinking hypothesis space exceeds the cost of a linear permutation search, remaining in the permutation regime is the resource-optimal choice. The fact that the models modulate their strategy based on the prunability of the logic—rather than just the raw size of N—suggests they are sensitive to the difference between expected utility and cost.

Model	**AIC**	**Pseudo R^2**
Elimination ~ $\log\binom{N}{2} \times T$	2727.29	0.160
Elimination ~ $\log\binom{N}{2} + T$	2751.10	0.152
Elimination ~ $\log\binom{N}{2}$	2765.76	0.147
Elimination ~ ρ	3138.23	0.032
Elimination ~ T	3178.55	0.019

Table 1: Model comparison for predicting the frequency of applying the elimination strategy. $\log\binom{N}{2}$ denotes hypothesis space size, T denotes the number of trials, and ρ denotes the information ratio.

Sensitivity to Task Complexity

We next examine which components of task complexity drive this shift in strategy selection. To prevent multicollinearity, we centered the predictors before performing a logistic regression

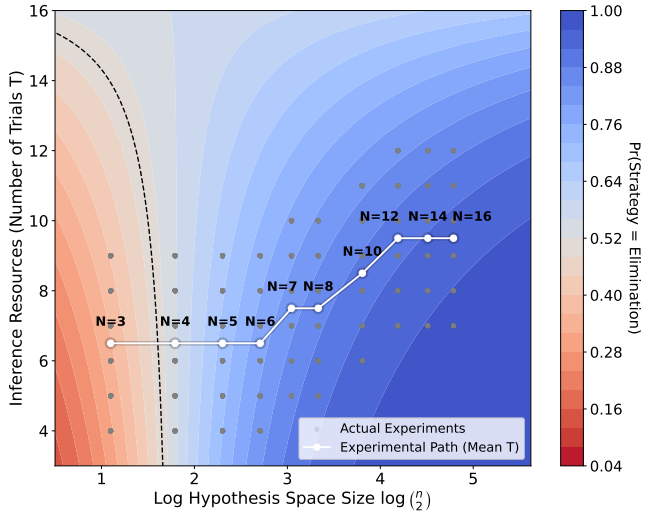

Figure 3: The fitted decision landscape of strategy selection. The heatmap represents the predicted probability of choosing the elimination strategy based on the interaction model. The gray dots represent actual experimental samples, and the white line denotes the mean experimental path (T_{mean} for each N). The dashed line indicates the 50% decision boundary.

to quantify the drivers of strategy selection. As shown in Table 1, the interaction model, incorporating both the log-hypothesis space size and the number of trials, yielded the best fit ($AIC = 2727.29$, Pseudo $R^2 = 0.160$).

We used the interaction model's coefficients to visualize the decision landscape. As shown in Figure 3, two competing forces drive the probability (in terms of frequency) of choosing elimination. First, as the size of the hypothesis space increases, the model is more likely to adopt elimination. This is supported by a strong positive coefficient ($\beta = 1.0212$, $p < .001$) on the logarithm of hypothesis space sizes. At the same time, the number of trials has the opposite effect. When more trials are available, the model tends slightly to fall back on the simpler permutation strategy—even when the hypothesis space is large. This is reflected in the negative coefficient for trial count ($\beta = -0.1466$, $p < .001$).

In Figure 3, we observed that elimination is preferred only when pruning substantially reduces the hypothesis space—typically when N is large, and T is limited. As T increases, the marginal value of each additional observation drops: the model can rely on permutation without incurring high error, making the added cost of elimination unjustifiable. In resource-rational terms, when $\mathbb{E}[\text{gain from pruning}] < \text{cost}_{\text{elimination}}$, the optimal choice shifts back to permutation. The landscape thus captures a dynamic cost-utility boundary, where strategy selection is sensitive not just to task size, but to how informative each trial is for pruning.

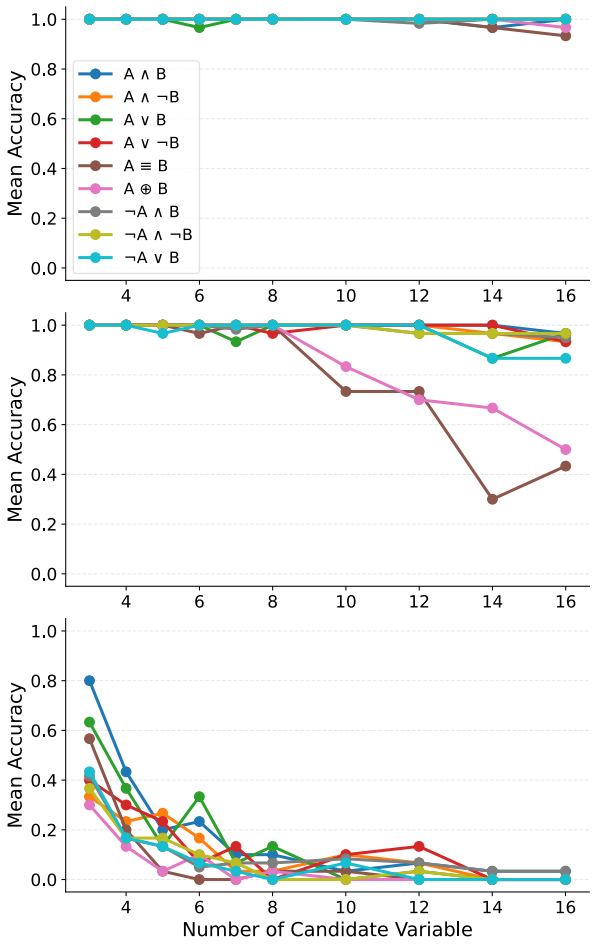

Figure 4: Mean accuracy across logical functions as task complexity (N) increases (DeepSeek). From top to bottom: DeekSeek R1, V3, and V3 with direct answer.

Accuracy and Reasoning Effort

To evaluate how accuracy scales with reasoning effort, we compare the performance and character consumption (as a proxy for effort) of IT models and their LRM counterparts against a baseline of IT models constrained to direct answer.

As depicted in Figure 4, LRMs exhibit remarkable robustness, maintaining near-perfect accuracy regardless of logical function. In contrast, IT models show a performance drop on XOR/XNOR as N increases. The direct answer baseline fails almost immediately, with accuracy quickly approaching chance levels or zero. This shows that inference-time scaling is the mechanism that allows models to boost reasoning. Meanwhile, the extended reasoning chain in LRMs is not merely a verbose version of standard output; rather, it allows the model to stabilize latent representations of non-linear relationships that are otherwise lost in the shallow processing of IT models.

As shown in Figure 5, character counts scale linearly with both N and T across all logical functions. Notably, XOR and XNOR functions consistently require higher character

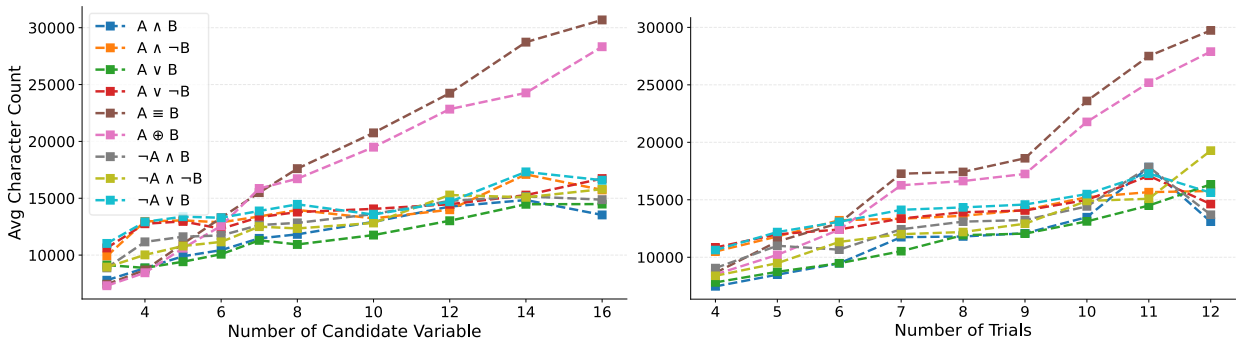

Figure 5: Computational scaling (DeepSeek R1). (Left) Average character count relative to N. (Right) Scaling relative to T. The dashed lines (XOR/XNOR) have steeper slopes, indicating that logically complex tasks demand more inference resources.

counts, reflecting the inherent computational complexity of non-monotonic attribution. These findings suggest that models dynamically allocate more computational time as a trade-off to stabilize the latent representations of non-linear relationships. This result potentially serves as a behavioral analog for human response time (RT), offering a predictive benchmark for future human study.

Discussion

A common critique of LLMs is that their apparent reasoning abilities reflect probabilistic pattern matching rather than principled inference (Kambhampati, 2024). From this perspective, model behavior is driven primarily by surface-level retrieval. Our findings challenge this view by showing that inference-time scaling functions as an implicit mechanism for resource-rational behavior. In particular, IT models and LRMs demonstrate adaptive selection of strategies as task demands increase. Notably, this adaptation emerges without any explicit training objective that penalizes or rewards computational cost, and appears largely independent of the specific post-training procedure. These results suggest that resource rationality may arise as an emergent property of large-scale optimization under constraint, rather than requiring explicit normative enforcement.

This analysis supports a broader interpretation of resource rationality. Rather than viewing it solely as a prescriptive framework that specifies how agents *should* allocate resources, resource rationality may also serve as a descriptive lens for understanding how complex systems tend to behave when operating under finite capacity. In this sense, inference-time scaling does not merely improve accuracy, but reorganizes computation in a way that reflects sensitivity to both environmental structure and internal limitations.

These observations also resonate with longstanding findings in cognitive neuroscience. Prior work has emphasized that human cognition is constrained by available cortical resources, and that increasing task demands are accompanied by the recruitment of additional neural circuitry (Just & Varma, 2007; S. Varma, 2014). In our setting, reasoning tokens can be interpreted as a workspace that maintains intermediate states during inference. When models confront high-dimensional hypothesis spaces, they generate longer reasoning traces to preserve partial eliminations and candidate relations. While this behavior is reminiscent of how the human brain increases prefrontal engagement under high WM load (Norman & Bobrow, 1975; Norman & Shallice, 1986), we emphasize that this analogy remains tentative. An alternative explanation is that reasoning length reflects limitations in attention-based state maintenance rather than deliberate resource allocation. Distinguishing between these possibilities requires further investigation into the interaction between attention capacity, representation stability, and inference-time computation.

Future research should also address the limitations of the current study. One question is whether the adaptive strategy selection with increasing complexity observed here for IT models and LRMs extends to more difficult tasks, for example, to cases where the observations are noisy or the given functions are more than boolean expressions of two variables. Another question is whether humans also show the same graceful transition between strategies with increasing complexity observed here for the models. This could be answered in a study of human performance on the VAT. A final question is the validity of equating thinking tokens with information processing effort, and ultimately with behavioral measures like human response times. Some machine learning researchers have warned against 'anthropomorphizing' thinking tokens (Stechly et al., 2025), finding only low correlations between information processing effort (A* search length on problem-solving tasks) and thinking tokens consumed (Palod et al., n.d.). That said, there are only a handful of such studies, and the question of the relationship between thinking tokens and behavioral measures is an important target for future research.

Acknowledgments

Portions of the code used to analyze experimental results were developed with the assistance of AI tools, which provided suggestions for Python implementation, data visualization, and debugging.

References

Boddy, M., & Dean, T. L. (1994). Deliberation scheduling for problem solving in time-constrained environments. *Artificial Intelligence*, *67*(2), 245–285.

Bramley, N. R. (2017). *Constructing the world: Active causal learning in cognition* [Doctoral dissertation, UCL (University College London)].

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature human behaviour*, *6*(8), 1112–1125.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of personality and social psychology*, *58*(4), 545.

Conlisk, J. (1988). Optimization cost. *Journal of Economic Behavior & Organization*, *9*(3), 213–228.

Conlisk, J. (1996). Why bounded rationality? *Journal of economic literature*, *34*(2), 669–700.

Ganzach, Y., & Czaczkes, B. (1995). The learning of natural configural strategies. *Organizational Behavior and Human Decision Processes*, *63*(2), 195–206.

Guo, D., Yang, D., Zhang, H., Song, J., Wang, P., Zhu, Q., Xu, R., Zhang, R., Ma, S., Bi, X., et al. (2025). Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, *645*(8081), 633–638.

Hu, Z., van Paridon, J., & Lupyan, G. (2024). Failures and successes to learn a core conceptual distinction from the statistics of language.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 153–191.

Kabra, K., Inani, K., Marupudi, V., & Varma, S. (2025). Modeling understanding of story-based analogies using large language models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*.

Kambhampati, S. (2024). Can large language models reason and plan? *Annals of the New York Academy of Sciences*, *1534*(1), 15–18.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*, e1.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive psychology*, *7*(1), 44–64.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In *Consciousness and self-regulation: Advances in research and theory volume 4* (pp. 1–18). Springer.

Palod, V., Valmeekam, K., Stechly, K., & Kambhampati, S. (n.d.). Performative thinking? the brittle correlation between cot length and problem complexity. *NeurIPS 2025 Workshop on Efficient Reasoning*.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of experimental psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational behavior and human decision processes*, *66*(2), 131–152.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118.

Snell, C. V., Lee, J., Xu, K., & Kumar, A. (2025). Scaling llm test-time compute optimally can be more effective than scaling parameters for reasoning. *International Conference on Learning Representations*.

Stechly, K., Valmeekam, K., Palod, V., Gundawar, A., & Kambhampati, S. (2025). Beyond semantics: The unreasonable effectiveness of reasonless intermediate tokens. *First Workshop on Foundations of Reasoning in Language Models*.

Varma, K., Boekel, M. V., & Varma, S. (2018). Middle school students' approaches to reasoning about disconfirming evidence. *Journal of educational and developmental psychology*, *8*(1), 28–42.

Varma, S. (2014). The subjective meaning of cognitive architecture: A marrian analysis. *Frontiers in psychology*, *5*, 440.

Von Neumann, J., & Morgenstern, O. (1947). Theory of games and economic behavior, 2nd rev.

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9), 1526–1541.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al. (2025). Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, *36*, 11809–11822.