Spontaneous meta-learning of efficient problem-solving algorithms

Huiwen Alex Yang[1] **(hw.yang@berkeley.edu)**
Mark K. Ho[2] **(mark.ho@nyu.edu)**
Bill D. Thompson[1] **(wdt@berkeley.edu)**
[1]Department of Psychology, University of California
[2]Department of Psychology, New York University

Abstract

A few minutes practice is often more than sufficient for an adult human participant to identify the structure of a problem they have never seen before, and plan a complex sequence of actions that creates a solution. However, it remains less clear whether this distinctive ability for ad-hoc discovery of problem-solving algorithms is itself subject to rapid meta-learning. We developed a novel problem-solving paradigm to examine aspects of this question. Participants in our study faced repeated iterations of an interactive sequential reasoning problem. Over trials, those who faced the hardest version developed increasingly efficient hierarchically-structured strategies that adaptively sequence a subordinate learning algorithm and an action planning policy; those who faced an easier version used simpler action-based strategies that did not involve learning the underlying structure. These results offer experimental evidence for efficient meta-learning of algorithmic concepts in a problem-solving setting.

Keywords: strategy use; strategy discovery; meta-learning

Introduction

On the screen before you is a grid of colored lights. Ten of them are green, and ten of them are blue. All the lights are off, and your goal is simple: turn them all on! But there is one catch: one of the two colors turns off the other. How would you proceed? One approach is to act randomly, which would likely trigger turn-offs. Another is to turn on all greens first, then the blues—but if blue is the "dangerous" color, the first blue light could undo your progress. A more efficient approach, particularly on larger grids, is to first identify which color is risky and then plan actions accordingly. Solving the problem efficiently therefore requires not just learning the dependencies between lights, but reasoning about which algorithmic strategy is best suited to the task.

This task is just one of many that illustrate a broad capacity of human intelligence: the rapid discovery and application of structured solutions to novel sequential problems. Inducing effective algorithms is typically data-intensive, yet humans achieve it quickly and flexibly. Even brief exposure to a problem allows most people to perform effectively, whereas conventional learning algorithms require far more experience (Tsividis et al., 2017). Even very young children are capable of this feat (Yang et al., 2025). The solutions people discover share fundamental properties with efficient algorithms (Haridi et al., 2023), and often vary between individuals, in ways that reflect principled trade-offs between algorithmic complexity and utility (Thompson et al., 2022). When provided with multiple algorithms for the same underlying problem, people select among alternatives in ways that reflect task structure (Brown et al., 2007; Lieder et al., 2014; Payne et al., 1988; Rieskamp & Otto, 2006) – a form of meta-reasoning over algorithmic concepts (Lieder & Griffiths, 2017).

Previous research demonstrates that both children and adults are capable of discovering effective strategies spontaneously, and that such discovery can be systematically measured (Crowley et al., 1997; Schuck et al., 2022; Siegler, 1991; Waris et al., 2021). Strategy selection is also often dynamic. For example, when instructed to perform causal inference, participants select adaptively between alternative testing strategies (Coenen et al., 2015). Similarly, studies of how people trade-off exploration and exploitation in simple bandit tasks reveal systematic strategies such as random versus directed exploration (Gershman, 2018; Wilson et al., 2021). However, it remains less clear whether similar principles apply in more open-ended tasks where participants could apply a broad range of different algorithms, and must discover themselves what computations are required – from simple action heuristics to hierarchically structured algorithms. This question has been difficult to explore because it requires a problem-solving task that is rich enough to admit a broad range of algorithmic solutions, but intuitive enough that people can reason about its structure during a brief experimental session. In this paper, we address this research gap by introducing a new task that has these features. We used this task to assess spontaneous meta-learning of efficient problem-solving algorithms. Participants in our study completed multiple trials of the grid-of-lights problem summarized above. After a handful of trials with no training at all, participants who faced the most difficult version spontaneously discovered a hierarchical strategy that begins with a phase of causal learning, followed by a phase of action execution; participants who encountered simpler causal structures or smaller grids used simpler algorithms with lower cognitive demands and faster execution. Our results offer evidence of rapid meta-learning of task-optimized algorithms in a relatively difficult sequential problem.

Methods

Task Structure

Interface: a grid of colored lights Participants interact with a grid of colored lights (Fig. 1). The goal they're given is simple: turn on all the lights. Each light can be toggled by

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

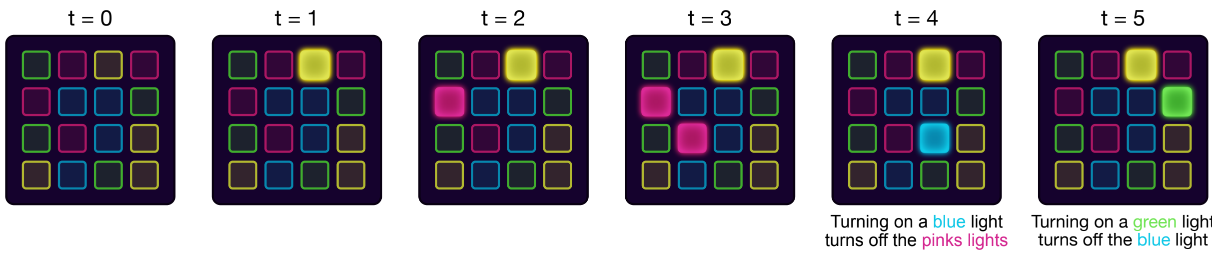

Figure 1: Task structure. First five actions in an example trial.

clicking. However, clicking on certain lights turns off other lights that were already on. The size of the grid, the number of colors, and the number of lights per color are task parameters. Critically, task difficulty is determined by dependency relations between colors, defined by turn-off interactions between colors. In this study, dependencies are specified at the color level and can be represented as a directed acyclic graph (DAG) – e.g. blues turn off pinks, greens turn off blues.

Game Dynamics Let $G = (V, E)$ be a directed acyclic graph whose vertices $V = \{1, \ldots, n\}$ denote the set of available colors and (directed) edges $E \subset V \times V$ encoding the turn-off relations, with $|E| = k$. A directed edge $(i, j) \in E$ indicates that clicking on any light of color i will turn off all currently active lights of color j. Each game is played on an $m \times m$ grid yielding $M = m^2$ lights. Each light $\ell \in L = \{1, \ldots, M\}$ has a color $c(\ell) \in V$. Multiple lights may share the same color, and colors are assigned such that the number of lights per color is as uniform as possible.

The state of the board at time t is $\mathbf{x}_t \in \{0, 1\}^M$ where $x_{t,\ell} = 1$ indicates that light ℓ is on (and $x_{t,\ell} = 0$ denotes off). An action $a_t \in L$ corresponds to clicking the light ℓ, which toggles that light to the on state if it was off and vice-versa. State transitions are deterministic:

$$x_{t+1,\ell'} = \begin{cases} 1 - x_{t,\ell}, & \text{if } \ell' = \ell, \\ 0, & \text{if } c(\ell') = j \text{ and } (i, j) \in E, \\ x_{t,\ell'}, & \text{otherwise.} \end{cases}$$

A game is correctly completed when all lights on the board are simultaneously on, i.e., $\mathbf{x}_t = \mathbf{1}$. The standard properties of a DAG ensure the game is complete-able. In the studies reported here, participants are not informed of the structure of the DAG.

Solution structure

Topological sorts on colors Under DAG game dynamics, the problem can be solved by turning on the colors in specific orders. For instance, if blue and yellow both turn off green, the game can be completed by activating either blue–yellow–green or yellow–blue–green. Formally, each valid ordering of a given DAG is a *topological sort* - a permutation π of the n vertices such that following that order will avoid any turn-offs. Every DAG has a set of topological sorts, but more complex DAGs make it harder to find a valid ordering.

This perspective views the task in action-oriented terms. To solve the task efficiently, participants need not infer the full causal structure, but only identify a valid color ordering. From a computational standpoint, inferring a DAG is substantially more demanding than reasoning over orderings: the number of edges is unknown, yielding a prohibitively large hypothesis space. By contrast, the space of linear orders is fully determined by the number of nodes n and directly constrained by observed turn-off events. For example, with $n = 4$ there are 24 possible orders but 543 DAGs; with $n = 6$, 720 orders but over 3.7 million DAGs. From a resource-rational perspective, reasoning over topological orders offers a tractable and cognitively plausible solution space compared to full causal inference.

Resource considerations In light of this task structure, what kind of algorithms are efficient for this task? The answer depends heavily on both DAG structure and grid size. Sparse DAGs admit many orderings and can often be solved with little information: even random color choices have a reasonable chance of following a valid topological order. In contrast, dense DAGs permit few orderings, making uninformed choices costly and leading to frequent turn-offs. These costs are further amplified on larger grids, where each mistake disables many lights.

For participants engaging with this task for the first time, these contingencies pose a meta-learning problem: is it worth trying to establish a valid color sequence before turning on lights – a hierarchical strategy that first gathers information before exploiting this knowledge? Or is a simpler algorithm sufficient – in some situations, the task can be solved well by simply turning on the rows, one by one. To make these questions more precise and generate clear predictions for a meta-learning agent, we identified a set of structurally distinct algorithms that can be used to solve the task, and diagnostic variations of the task that induce distinct performance profiles.

Reference Models

Optimal Model This model represents a hierarchical strategy that identifies a valid topological ordering before using it to turn on the remaining lights. The first phase is implemented by maintaining a belief over feasible topological orders and

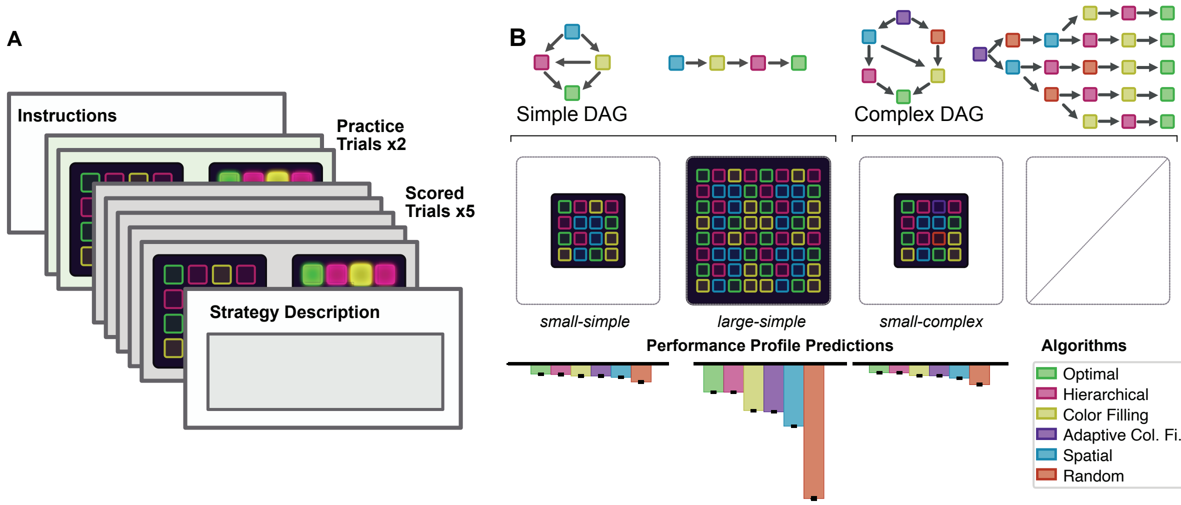

Figure 2: Procedure and conditions. (**A**) Procedure: instructions, two practice trials, and five scored trials, followed by a strategy description. (**B**) Experimental conditions defined by board size (small vs. large) and task structure (simple vs. complex DAG), yielding three conditions. Representative grids and DAGs for each condition, along with a set of working orderings. Predicted performance profiles for each condition under different algorithmic strategies. Longer bars indicate less efficient strategies.

selecting actions that maximize expected information gain. At time t, the set of feasible orders is:

$$P_t = \{\pi_1, \pi_2 \dots\},$$

where each π is a permutation consistent with previously observed turn-off events. P_0 contains all $n!$ permutations. When an observation indicates that clicking node i turns off node j, the model infers the constraint $i \to j$ and updates its belief:

$$P_{t+1} = \{\pi \in P_t \mid \text{index}_\pi(i) < \text{index}_\pi(j)\}.$$

If no turn-off occurs, the belief remains unchanged, $P_{t+1} = P_t$. To select the next action, the model evaluates the expected information gain (EIG) for each color $i \in V$, i.e., the expected reduction in the number of feasible topological orders P_t if a light of that color is clicked. For each possible outcome S (a set of colors that would turn off if a light of color i is clicked), the set of eliminated permutations is

$$E(i, S) = \{\pi \in P_t \mid \exists j \in S \text{ s.t. } \text{index}_\pi(i) > \text{index}_\pi(j)\},$$

and the expected information gain for color i is

$$\text{EIG}(i) = \sum_{S \subseteq C_t} \Pr(S \mid i, P_t) \frac{|E(i, S)|}{|P_t|}.$$

Action selection is probabilistic at the color level via a softmax over color EIG with temperature $\tau >= 0$ (greedy when $\tau = 0$):

$$P(i_t) = \frac{\exp(\text{EIG}(i_t)/\tau)}{\sum_{i' \in V} \exp(\text{EIG}(i')/\tau)}.$$

The selection process repeats until a single consistent topological order is inferred or all colors have at least one light turned on.

Random Baseline Selects uniformly at random from all lights on the grid that are currently turned off.

Strategy Models

Each of the following strategies represents a different class of algorithms that participants could reasonably use to solve the problem, defined a-priori based on theoretical considerations and observations from pilot studies of similar tasks. We used the probabilistic programming library FlipPy (Ho & Correa, 2026) to implement these strategies, estimate parameters, and compute data likelihoods.

Color Filling Systematically turns on all lights of one color before moving to the next, without considering dependencies between colors. The color order and a probabilistic lapse rate are treated as latent parameters inferred during model fitting.

Adaptive Color Filling A variant of Color Filling with a simple adaptive mechanism: if activating a new color causes previously activated lights to turn off, the model first completes the current color and then returns to re-activate the lights that were turned off. This strategy adapts to observed dependencies while maintaining a color-by-color progression.

Spatial Filling Lights are turned on using simple spatial patterns: e.g. row-by-row, column-by-column, or back-and-forth zigzag sequences. Pattern and lapse rate are latent parameters inferred during model fitting. Does not consider color or turn-off relations.

Hierarchical Two-Phase Filling Similar to the Optimal reference model but more relaxed and error-tolerant, this model proceeds in two phases. In phase one, it tests the lights color by color, selecting only colors that have not yet been turned on, allowing it to establish a working order without computing expected information gain. In phase two, the model fills the board by committing to an arbitrary order drawn from the set of valid topological orders. The model departs from the Optimal model in two key ways: it does not rely on expected information gain to guide color selection in phase one, and it does not enforce the use of a working order in phase two. These deviations are motivated by the observation that participants' behavior is often noisy and inconsistent, requiring a looser, more error-tolerant model to capture empirically observed strategies. Two latent parameters are inferred during fitting: the number of clicks at which the model transitions from phase one to phase two, and the order used to fill the board in phase two.

Behavioral Experiment

Participants We recruited from Prolific 150 US adult participants fluent in English with Prolific approval rating above 95%. Participants were assigned to one of the three following conditions (50 per condition).

Conditions We identified a between-participant factorial design that manipulated two factors independently (Fig. 2B): the size of the grid (small vs. large) and the structure of the DAG (simple vs complex). These manipulations were designed to evaluate whether participants were sensitive to key contingencies of strategy utility. **Grid size:** Large grids motivate strategies that first learn about the dependencies before acting, because turn-off events can be very costly; small grids lack this feature, and therefore can be solved efficiently with simpler heuristics. **DAG Structure:** Simple DAGs facilitate early inference and planning; while complex DAGs make such investment more costly in both actions and cognitive effort (e.g. remembering dependencies)

Our key prediction was that participants assigned to the *large-simple* condition would be motivated by both factors to identify hierarchical, inference-first strategies (e.g., Hierarchical Two-Phase Filling). In this setting, structure learning is relatively easy, and incorrect ordering is costly, as a single incorrect click can turn off many lights. In contrast, we predicted that participants assigned to *small-complex* were expected to engage less in active structure learning, because inferring the structure requires substantial exploration while offering limited downstream benefit on a small grid. This key contrast predicts that participants assigned to *large-simple* would (a) gather more information and (b) resolve the task structure more frequently and (c) earlier in the trial than participants assigned to *small-complex*.

We also included a *small-simple* control condition, in which the task is sufficiently easy that many strategies are effective, allowing insight into participants' default or most intuitive approaches. We did not assign participants to the implied fourth condition (*large-complex*) – this condition would have been too difficult.

Procedure The procedure was identical across conditions, except for grid size and DAG structure (Fig. 2). Participants completed four phases: Instructions, Practice, Scored, and Questionnaire. Practice consisted of two trials generated from the same DAG: a 4×4 grid followed by a 6×6 grid, with the number of colors varying by condition (4 in *small-simple* and *large-simple*, 6 in *small-complex*). In the Scored phase, participants completed five trials, each with a distinct game governed by an isomorphic DAG. This maintained comparable difficulty while requiring participants to learn new rules each trial. Participants earned a bonus for correct trials and could skip trials. Finally, participants reported their strategies in the Questionnaire phase.

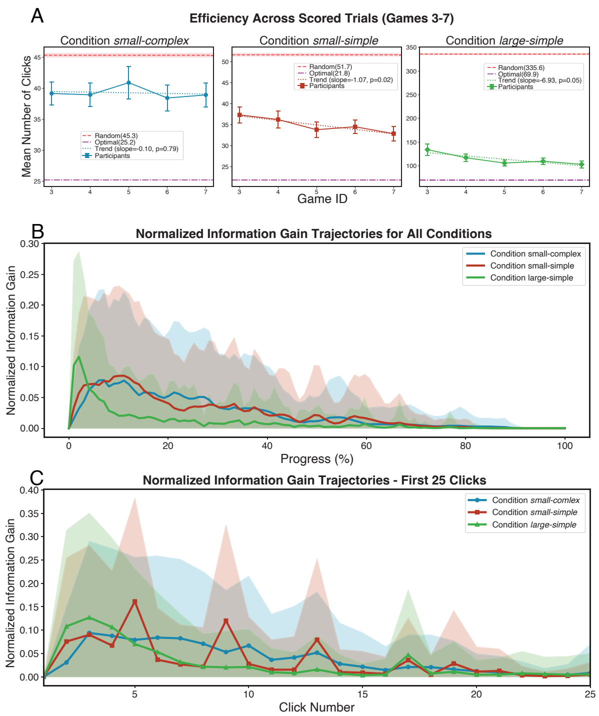

Figure 3: Results. (**A**) Mean clicks per trial across scored games. Solid lines show participants; dotted lines show linear trends. Dashed lines (shaded ±1 SEM) indicate Random and Optimal model performance. Participants were more efficient than Random and less efficient than Optimal, with efficiency improving over time in *small-simple* and *large-simple*. (**B**) Mean normalized information gain over trial progress for each condition (shaded ±1 SD). (**C**) Mean normalized information gain by click number (first 25 clicks), showing earlier concentration in *large-simple* and later, more periodic gains in *small-simple*.

Results

Participants solved the problem accurately and efficiently in all conditions

Participants performed well in all three conditions: Accuracy (percentage trials in which all lights were successfully turned on), was 93.61% in *small-complex*, 98.77% in *small-simple*, and 83.76% in *large-simple*. Accuracy of scored trials only

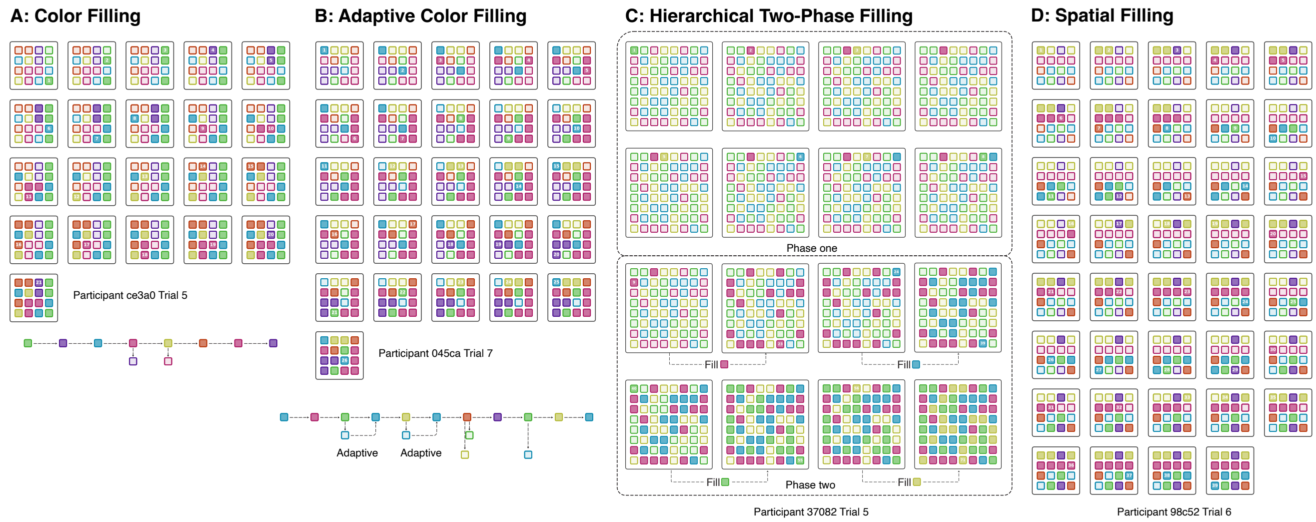

Figure 4: Example of participants using each strategy. Each panel shows example action sequences generated by participants that were well fit by one of the four strategy models.

was similarly high: 98.25%, 100.0%, and 82.14%, respectively. Data from participants with no recorded clicks in one or more trials were excluded (*small-complex*: $n = 4$; *small-simple*: $n = 2$; *large-simple*: $n = 4$). Trials with click counts exceeding 3 SD above the mean were excluded ($n = 9, 10, 8$ for *small-complex*, *small-simple*, and *large-simple*, respectively). Accuracy in the scored phase remained stable over time, indicating participants solved the task successfully from the start, with no evidence of a practice effect (all linear trends $p > 0.1$).

We compared participants' behavior to two reference models, generating 1000 simulations per model per condition using boards and DAGs sampled from the games encountered by participants. Models clicked until all lights were on, with efficiency measured as the number of clicks required. As shown in Fig. 3A, participants were significantly more efficient than the Random baseline in all three conditions. On successfully completed trials, they required 14–66% fewer clicks than the Random model ($p < .001$ for all conditions). Despite this, participants required substantially more clicks than the Optimal model. Across conditions, they used roughly 55–64% more clicks on average than the model ($p < .001$ for all conditions; also shown in Fig. 3A). Together, these results demonstrate that participants understood the task and engaged with it meaningfully; they performed in a way that was systematic and efficient in all conditions. However, there remains a clear gap between average human behavior and optimal performance.

Figure 3A also shows that efficiency improved with experience in *small-simple* and *large-simple*, while remaining stable in *small-complex* (p-values in the figure). These results suggest that, although participants could perform the task accurately from the outset, practice allowed them to refine their strategies and reduce the number of clicks needed, making their approach more efficient. Participants who completed all scored trials correctly were ranked by mean clicks per trial. Across conditions, most efficient participants approached near-optimal performance, using 43–78% fewer clicks than the Random model and coming within 3–7% of the Optimal model, indicating substantial improvements over unstructured behavior and close approximation of the optimal strategy.

Together, these findings suggest that participants quickly discovered viable solution strategies and refined them with experience. Importantly, over the course of the experiment, participants solved five instances of the problem that shared the same underlying causal structure within each condition. The observed increases in efficiency therefore reflect iterative improvement of strategies across structurally similar problems, rather than independent solutions to isolated instances. This pattern indicates meta-learning over problem structure, whereby participants refined how they approached the task itself, rather than merely solving one instance after another independently.

Strategic Information Acquisition Varies Across Conditions

To assess whether participants actively sought to establish valid orderings as implied by hierarchical strategies, we computed the information gain of each click, focusing on early clicks in each trial. Information gain was defined as the number of feasible topological orders eliminated by a click, providing a principled, quantitative measure of whether participants' actions efficiently uncovered the underlying structure.

Fig. 3B shows mean normalized information gain as a function of percentage trial progress for all conditions. Normalized information gain was computed by averaging, across trials, the information gain of each click relative to the total possible information gain in that condition, with the x-axis scaled by

Table 1: Number of correctly completed scored trials best fit by each computational strategy model, for all conditions.

Strategy	*small-* complex	*small-* simple	*large-* simple
Col. F.	102 (43.6%)	121 (48.8%)	112(57.7%)
Adap. Col. F.	40 (17.1%)	38 (15.3%)	34 (17.5%)
H. 2-Phase F.	79 (33.8%)	86 (34.7%)	48 (24.7%)
Spatial F.	13 (5.6%)	3 (1.2%)	0

percentage progress to allow comparison across trials of different lengths. In *large-simple*, a substantial proportion of early clicks yielded high information gain, indicating that participants actively sought to gather information, as implied by the hierarchical strategy. This pattern was not observed in the other two conditions.

Fig. 3C plots mean normalized information gain as a function of click number for each condition. Because the x-axis is not scaled, trials can be aggregated to reveal group-level patterns at specific clicks. large information gains occur at clicks 4, 8, 12, and so on, likely reflecting passive information acquisition when previously activated lights are turned off by subsequent color changes. This pattern suggests that many participants begin by activating all lights of one color before switching, consistent with a Color Filling strategy. A similar but weaker pattern appears in *large-simple* around click 16. In contrast, participants in *large-simple* show high information gain early in the trial, consistent with active testing. Together, these results indicate that participants in *small-simple* and *large-simple* adopt different early-trial strategies.

We also examined how often participants acquired full information (eliminated all topological orders inconsistent with the true DAG). Completing a trial does not require full information, as participants may succeed by chance without observing many turn-offs. In such cases, inconsistent orders remain possible, reflecting chance rather than systematic inference. Full information was acquired in 35.8% of trials in *small-complex* (SD = 48.0%), 52.0% in *small-simple* (SD = 50.0%), and 66.7% in *large-simple* (SD = 47.2%). Acquisition rates differed systematically across conditions, with higher rates in *large-simple* and *small-simple* compared to *small-complex* (all pairwise χ^2 tests significant; χ^2 range: 12.37–51.92, $p < .001$). For trials in which full information was acquired, we examined how early participants resolved the ordering, measured as the proportion of total clicks required. Participants in *small-complex* required the largest proportion of clicks (M = 57%, SD = 21%), followed by *small-simple* (M = 42%, SD = 22%), with *large-simple* requiring the fewest (M = 35%, SD = 23%). All pairwise t-tests were significant (t range: 2.66–8.00, $p \leq .008$). Together, these results indicate systematic differences in how quickly participants resolved the dependency structure: participants in *large-simple* inferred a complete ordering earlier, whereas those in *small-complex* often completed trials without fully resolving the structure.

These findings support our hypothesis that participants in *large-simple* adaptively used strategies that gather information about the task structure early in the trial, allowing them to use this information to complete the game efficiently. They also demonstrate flexibility in spontaneous strategy use across conditions: when such a strategy is less beneficial, as in *small-complex*, participants used it less frequently.

Participants use various strategies

Participants solved the task using qualitatively different strategies. To capture this heterogeneity, we fit four Strategy Models (described in Methods) to each correctly completed scored trial. The results of this analysis are summarized in Tab. 1. Participants did not rely on a single, homogeneous strategy; instead, their behavior reflected a mixture of structured strategies. Fig. 4 shows four example trials, each illustrating behavior from a participant that is best characterized by a different strategy. Across all conditions, color-based strategies (Color Filling and Adaptive Color Filling) were dominant; the other two were less frequent. Model fits were less precise for long action sequences or for behaviors that deviated from reference strategies. Accordingly, model assignments should not be interpreted as definitive labels of participants' internal strategies, but rather as estimates of the relative prevalence of different computational motifs. Importantly, the Hier. Two-Phase Fill. strategy is just one instantiation of broader information-seeking strategies, and the information gain observed in the harder conditions may also arise from adaptive behaviors not captured by the current model set.

Conclusion

Spontaneous discovery of structured algorithmic strategies is a distinctive human ability at the forefront of knowledge in cognitive science. In this study, participants faced multiple iterations of a novel sequential problem under three conditions with distinct task structures that favored meaningfully different strategies. Across all conditions, participants achieved high accuracy and efficiency, demonstrating the rapid discovery and application of task-adapted strategies. In the most difficult condition, participants spontaneously discovered that to solve the task efficiently, they must first learn the structure of the underlying problem, then plan actions in light of that knowledge. Using this strategy drove improvement over trials on structurally identical problems, providing evidence of rapid meta-learning of a hierarchically structured strategy that adaptively deploys in sequence a subordinate learning algorithm and an action planning policy. Participants who experienced easier problems did not try to learn the underlying structure; they use simpler efficient strategies. These findings support a view of human cognition as adaptive and resource-rational, capable of balancing accuracy, efficiency, and computational cost even in the context of unfamiliar problems requiring algorithmically-structured solutions.

References

Brown, S., Steyvers, M., & Hemmer, P. (2007). Modeling experimentally induced strategy shifts. *Psychological Science*, *18*(1), 40–45.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133. https://doi.org/10.1016/j.cogpsych.2015.02.004

Crowley, K., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review*, *17*(4), 462–489.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.

Haridi, S., Wu, C. M., Dasgupta, I., & Schulz, E. (2023). The scaling of mental computation in a sorting task. *Cognition*, *241*, 105605.

Ho, M. K., & Correa, C. G. (2026). FlipPy: Pythonic probabilistic programming [GitHub repository].

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological review*, *124*(6), 762.

Lieder, F., Plunkett, D., Hamrick, J. B., Russell, S. J., Hay, N. J., & Griffiths, T. L. (2014). Algorithm selection by rational metareasoning as a model of human strategy selection. *Advances in Neural Information Processing Systems*, *27*.

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.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of experimental psychology: General*, *135*(2), 207.

Schuck, N. W., Li, A. X., Wenke, D., Ay-Bryson, D. S., Loewe, A. T., Gaschler, R., & Shing, Y. L. (2022). Spontaneous discovery of novel task solutions in children. *PLoS One*, *17*(5), e0266253.

Siegler, R. S. (1991). Strategy choice and strategy discovery. *Learning and Instruction*, *1*(1), 89–102.

Thompson, B., Van Opheusden, B., Sumers, T., & Griffiths, T. (2022). Complex cognitive algorithms preserved by selective social learning in experimental populations. *Science*, *376*(6588), 95–98.

Tsividis, P., Pouncy, T., Xu, J. L., Tenenbaum, J. B., & Gershman, S. J. (2017). Human learning in atari. In 2017 AAAI Spring Symposia.

Waris, O., Jylkkä, J., Fellman, D., & Laine, M. (2021). Spontaneous strategy use during a working memory updating task. *Acta Psychologica*, *212*, 103211.

Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current opinion in behavioral sciences*, *38*, 49–56.

Yang, H. A., Thompson, B. D., & Kidd, C. (2025). Children spontaneously discover efficient solutions to a difficult sorting task. *Nature Human Behaviour*, 1–12.