What reusable shortcuts do people propose when solving assembly problems?

Sean P. Anderson (seanpaul@stanford.edu)[1], Justin Yang (justin.yang@stanford.edu)[1], Maddy L. Bowers (mlbowers@csail.mit.edu)[2], Lionel Wong (liowong@stanford.edu)[1] & Judith E. Fan (jefan@stanford.edu)[1]

[1]Department of Psychology, Stanford University

[2]Department of Brain and Cognitive Sciences, MIT

Abstract

Humans readily extract statistical regularities from perceptual experience (e.g. a cook noticing which ingredients often appear together). How does such learning guide peoples' performance on procedural tasks (e.g. preparing various dishes)? Here we examine what shortcuts people propose to help them to complete assembly problems more efficiently by eliminating repeated subroutines. Participants (N = 301) repeatedly assembled tangram-like shapes, and could create composite tiles for future use. Some participants assembled a sequence of tangrams where certain pairs of tiles recurred consistently (*Highly structured*); the remaining assembled tangrams with less predictable tile arrangements (*Less structured*). Participants exposed to *Highly structured* sequences created tiles that tracked the frequency with which those tile pairs recurred across tangrams, and doing so was accompanied by greater efficiency in assembly. Taken together, these findings suggest that statistical learning guides not only pattern recognition, but also the prospective creation of shortcuts for procedural tasks.

Keywords: problem solving; statistical learning; skill learning; procedural memory; visual reasoning

Introduction

People rarely perform exactly the same activity twice, but many activities share related structure. For example, someone learning to cook by working their way through an entire cookbook might get faster not only at preparing those particular dishes, but other dishes as well. This generalized learning can happen as someone masters the *parts* of tasks that they encounter repeatedly, like the need to mince vegetables and herbs across many similar dishes. People who have gotten faster and more accurate at performing these *subroutines*—or patterns of actions shared across tasks—can draw on them to solve new related problems in the future.

A significant body of empirical and computational work has investigated the learning of subroutines across similar problems (Anderson, 1982; Chase & Simon, 1973; Schmidt, 1975). In particular, models of statistical structure learning (Ellis et al., 2021; Saffran et al., 1999; Tenenbaum et al., 2011) have been used to predict how people learn fluid subsequences of motor actions (Tian et al., 2020), remember parts of problems they have solved previously (Chi et al., 1981), and even describe how to create new drawings after seeing others comprised of similar parts (Wong et al., 2022). Statistical structure learning has been attested to in many other species, including pigeons and monkeys (Terrace, 2001).

However, unlike other animals, people sometimes get faster at completing tasks not just by practicing subroutines that recur repeatedly, but also by *prospectively* creating shortcuts that can replace parts of problem solving. For instance, any chef might find it generally useful to practice knife skills. But after mincing dozens of cloves of garlic across many dishes, a forward-thinking chef might also decide to go out and buy a large quantity of pre-minced garlic in advance. A person who invests time in creating a useful shortcut can therefore eliminate the need to execute the same sequence of actions many times in the future. However, it takes time and effort to plan in advance. Earlier work on navigational problem solving, for instance, finds that people are sensitive to the cognitive effort needed to weigh different potential shortcuts to the same final goal (Lancia et al., 2023). People do not create shortcuts preemptively for *every* subroutine they perform more than once. What determines when someone decides to invest in creating a shortcut, and when they do, the specific shortcut they decide to create?

In this paper, we investigate the relationship between structure across previously solved problems and prospective shortcut creation for future problem solving. We study this behavior using a domain of physical assembly tasks based on tangram puzzles (Figure 1), in which people must build complex shapes by arranging sets of simpler tiles. Participants solve puzzles in several rounds, and are rewarded simply for arranging tiles as quickly and accurately as possible into place. In our experiment, however, we also give participants the choice after each round to prospectively create *new*, composite tiles that they can use to solve future puzzles. Participants must therefore choose to spend time building a new shortcut that they expect will be useful, rather than simply advancing through to the next round of puzzles, and if so, which particular shortcut to construct. To explore the degree to which statistical learning underlies the proposal of these shortcuts, we study people solving both *highly structured* sets of puzzles with strong statistical regularities, and *less structured* puzzles composed randomly from the basic tangram tiles.

We hypothesize that statistical learning guides not only how people solve problems they encounter repeatedly, but choose to invest in shortcuts they might reuse in the future. Our experiments shed light on whether people explicitly reason and plan about learned statistical structures, using patterns in existing problems to plan creatively for solving similar ones in the future.

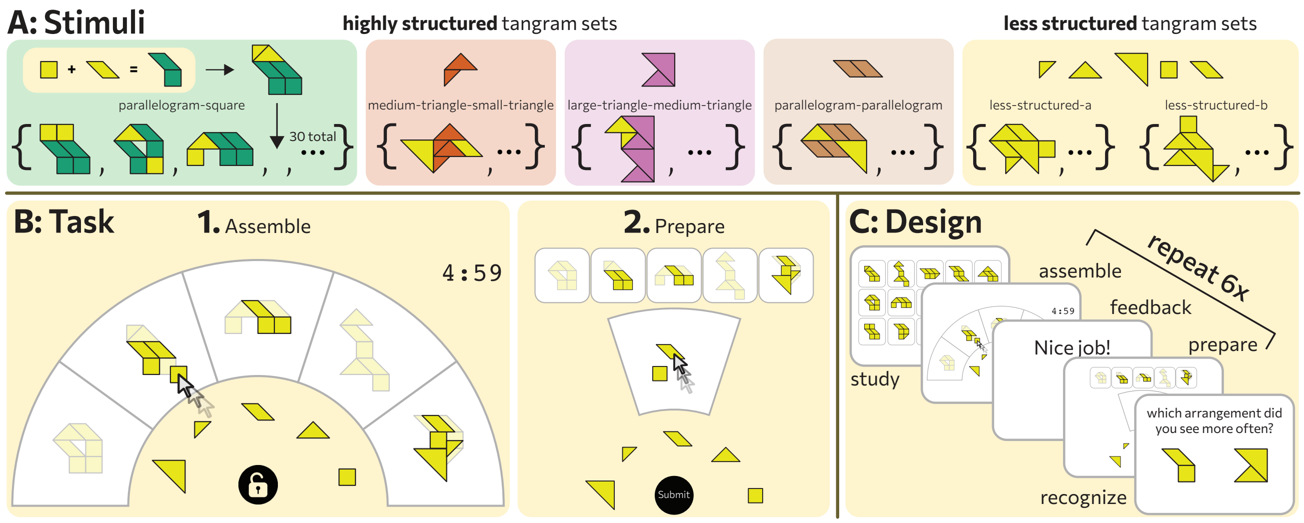

Figure 1: **(A) Stimuli:** We generated 4 sets of *Highly structured* tangrams and 2 sets of *Less structured* tangrams. **(B) Task:** Participants aimed to assemble a series of tangrams using a digital interface **(Assemble)**. Participants had to click the lock icon in the bottom of the interface to access the basic tiles. After the participant clicked a piece or two seconds had passed, the basic tiles disappeared. After participants reconstructed all five target tangrams (or five minutes had passed, whichever occurred sooner), they received feedback. Then they could propose their own composite tile to be available in the next round **(Prepare)**. **(C) Design:** Participants were assigned to one of the 6 set conditions and only encountered tangrams from that set. Participants first studied a sample of 15 tangrams for patterns, then completed 6 rounds of assembly, feedback, and prospective shortcut proposals. Then participants completed 6 2-AFC recognition trials.

Physical assembly and prospective shortcut creation experiment

To study the effect of shared problem structure on shortcut proposals, we designed a task based on tangram puzzles (Elffers & Hollingdale, 1976), in which people arrange a set of basic polygon-shaped tiles to create a more complex shape. The classic tangram setting uses a fixed set of tiles: two large triangles, a medium triangle, two small triangles, a parallelogram, and a square. In our modified setting, a puzzle might use any of these basic tile types multiple times.

Tangram stimuli

We constructed two different puzzle conditions that varied in their degree of statistical regularity. For our *Highly structured* conditions, we generated sets of tangram stimuli in which the same 2-tile pattern (e.g., a parallelogram on top of a square; Figure 1A) appears twice in every puzzle, ensuring that this pattern would be the most commonly reused across the set. We generated 4 such sets, each with a different repeated 2-tile pattern (“parallelogram-square,” “medium-triangle-small-triangle,” “large-triangle-medium-triangle,” and “parallelogram-parallelogram”). Each set contained 30 tangrams, where each tangram was composed of 6 basic tiles (4 from the repeated pattern, plus 2 additional tiles). For our *Less structured* conditions, we generated 2 sets of 30 tangrams that do not reuse any particular 2-tile patterns. Instead, each tangram was generated from the 5 basic tiles plus one random additional tile (Figure 1A).

Participants

We recruited participants on Prolific (paid $6.00 each), excluding 1 who failed to solve any puzzles in at least 2 rounds of our task. After exclusions we analyzed $N = 301$ participants (50-51 per condition).

Tangram assembly and shortcut task

Conditions We assigned participants randomly to one of the six tangram set conditions (four *Highly structured*, two *Less structured*).

Pre-assembly familiarization trial Each participant was first shown a random sample of 15 of the 30 tangrams they would encounter later. We asked participants to “determine if there are any patterns in the shapes, such as which pieces consistently appear with which other pieces” and type their answer in a text box.

Interleaved assembly, feedback, and shortcut trials After the single familiarization trial, participants completed six rounds of interleaved tangram assembly, feedback, and shortcut proposal trials.

Assembly trials During assembly trials, participants saw five target tangram outlines, each with line segments displaying how to place basic tiles to build the shape (Figure 1B). We asked participants to “make a copy” of the tangrams as quickly as they could. In a separate region below the targets was a virtual “toolbox” containing the five basic tiles (square, large triangle, medium triangle, small triangle, and parallelogram).

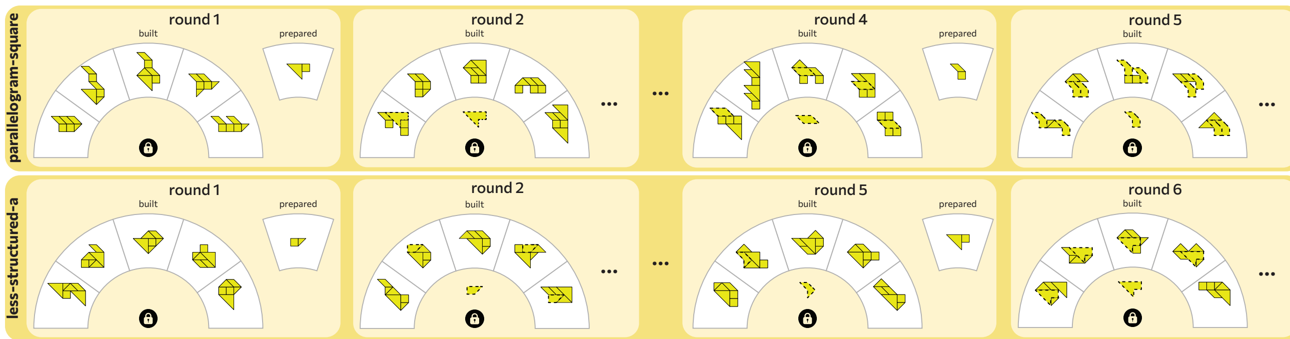

Figure 2: **Sample participant trajectories:** Handful of rounds from one *Highly structured* participant (top) and one *Less structured* participant (bottom). Composite tiles are marked with dashed borders.

The tiles were hidden from view until participants clicked a lock icon, after which the tiles would appear for two seconds and could be moved with the mouse to any of the target tangrams. Participants could use any tile any number of times. Tiles could not be rotated. If the participant successfully completed all five tangrams, the assembly period ended automatically. Otherwise, the assembly period ended after five minutes.

Assembly feedback trials After each assembly trial, participants received feedback: if the participant completed all five tangrams successfully under the time limit, we showered the screen with confetti with the text "Nice job!" for four seconds. Otherwise, we showed a fixation cross.

Prospective shortcut proposal trials After each assembly and feedback trial, participants entered a *shortcut proposal* trial. Here, participants were invited to "prepare for the next round" by **creating a new composite tile out of two or fewer tiles that they could use in future assembly trials.** Unlike the tiles in the toolbox, which required a click to access, the participant's last proposed composite tile would be available above the toolbox so they could access it without clicking the lock. During shortcut proposal, participants could view the target tangrams (and their submitted reconstructions) from the previous round. Participants assembled a composite tile from scratch in the 1st proposal period, and edited this composite on later rounds.

Pattern recognition test Finally, to validate whether participants could explicitly recognize recurring structure across assembly problems, we presented them with 2-alternative forced choices (AFC) between two composite tiles. We instructed them to select which arrangement of tiles appeared more frequently in the 30 tangrams they attempted. Participants made six choices, one for each pairing of composite tiles used to generate the *Highly structured* tangram sets.

Analyses We used AI tools while implementing experiment materials and analyses. For group-level parameter estimates, we performed 9999 bootstrap resamples with replacement, including a participant's entire data when resampled. When computing the Total Variation Distance (TVD) metrics, we corrected for bias in bootstrapped confidence intervals by subtracting the residual between the actual TVD and the mean bootstrapped TVD from the original interval bounds. To assess differences in measures across rounds or conditions, we formulated linear mixed-effects regression models with random intercepts for individual participants and tangram set condition. We used the `emmeans` package for model comparison in R, with Type III ANOVAs as reported by `emmeans::joint_tests()`.

Results

Participants explicitly discover recurring patterns across assembly problems

Distribution of proposed composite tiles As a coarse measure of the relationship between statistical regularity and tile proposals, we examine the distribution of composite tiles proposed across conditions (Figure 4). Qualitatively, participants in the *Highly structured* conditions mostly converged on a small set of composite tiles in the final round, while those in the *Less structured* conditions proposed a relatively larger and more diffuse set of tiles. We confirm this by computing the Gini coefficient (a measure of spread in a distribution with the range [0, 1], where higher values mean more concentrated) of the composite tiles. **The distributions of final round shortcut proposals were indeed more concentrated in the *Highly structured* conditions than in the *Less structured* conditions** ($\mu_{\text{Highly structured}}(G_{\text{set condition}}) - \mu_{\text{Less structured}}(G_{\text{set condition}}) =$ 0.354, 95% CI: [0.210, 0.387]). Furthermore, we find that participants in the *Highly structured* conditions tended to converge on an increasingly concentrated selection of composites over rounds ($\mu_{\text{Highly structured}}(\delta_G) = 0.111$ increase per round, 95% CI: [0.028, 0.144]), suggesting that they learned from statistical regularities that became more apparent with more assembly experience. We did not observe this in the *Less structured* condition ($\mu_{\text{Less structured}}(\delta_G) = 0.037$ increase per round, 95% CI: [-0.052, 0.120]), suggesting that both the statistical strength of patterns and the amount of experience impact shortcut proposals.

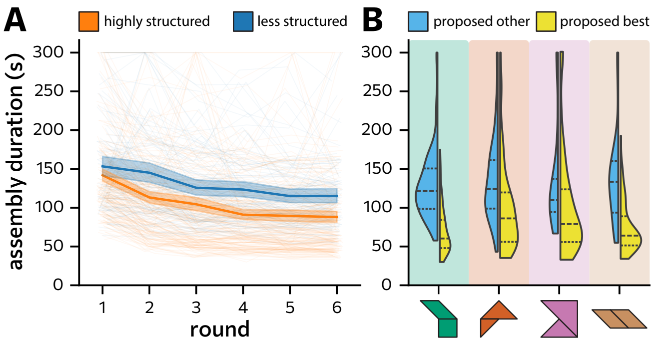

Figure 3: **Performance: A**: Average time to reconstruct all five tangrams in a single assembly period over all six rounds. Participants were limited to 5 minutes to complete their reconstructions. Faint lines indicate individual participants. Bands indicate 95% bootstrapped CIs. **B**: Average assembly duration between participants in *Highly structured* conditions who proposed the most frequent composite tile for their set (x axis) and participants who never did. Area of each kernel density is normalized to number of subjects. Edges of kernel density area are chopped to range in assembly duration. Internal lines denote median, 25th, and 75th quartiles.

Most frequently appearing patterns in assembly problems We focus on the *Highly structured* conditions, which we designed so that a single 2-tile pattern appeared much more frequently than any other (in the *Less structured* sets, all 2-tile patterns appeared with relatively uniform frequencies; see Figure 4). By the final round, the majority of participants (59.2%) in the *Highly structured* condition had proposed the most frequent 2-tile pattern in their set as their composite tile (95% CI: [52.2, 66.2]). We also observed that participants in the *Highly structured* condition became slightly more likely to propose this specific 2-tile pattern as they completed more rounds ($\delta_{\text{Highly structured}} = 0.037$ proportion participants per round, $F_{1,1004} = 88.4$, $p < 0.001$), suggesting that participants attend to structure in their accumulating experience when proposing shortcuts for future assembly problems.

Statistical regularities and composite tile proposals In an exploratory analysis, we also evaluated whether the shortcuts participants tended to propose in general mirrored the frequencies of 2-tile patterns in the puzzles, beyond the single most frequent one. We restrict our analysis to 2-tile composites, and exclude participants who did not propose a composite, proposed a single basic tile as their composite, or proposed a composite pattern that did not actually appear in their tangram set, leaving $N = 232$ participants across both conditions. We then analyze the composite tiles participants proposed in the final (6th) round only, after they had encountered all 30 tangrams of their assigned condition.

We first consider whether the distributions of shortcuts participants proposed match the empirical distributions of patterns they encountered in the puzzles. We compute an *empirical frequency* distribution capturing how often a 2-tile pattern appeared in a participant's assigned tangram set (Figure 4). We find that **the *empirical frequency* distribution predicts the top few shortcuts participants proposed**, though not the relative proportions with which they proposed them. Specifically, in 3 of the 4 *Highly structured* conditions, the top-2 most empirically frequent patterns are the same as the top-2 most frequent shortcuts people proposed. In one *Less structured* condition, the set of participants' top-3 most frequent proposals is the same as the top-3 most empirically common patterns; in the other, 2 of participants' most frequent proposals are the first and third most frequent patterns in the stimuli (though by design, the frequency of these patterns are closer to uniform overall). We further quantify how well the *empirical frequency* distribution predicts the distribution of shortcut proposals using Total Variation Difference, which is the absolute difference between the two distributions summed over all tiles and divided by 2. This distance metric ranges from 0 for identical distributions to 1 for non-overlapping distributions. In the *Highly structured* conditions, the distribution of shortcuts participants proposed was **much closer to the *empirical frequency* of patterns than to a *uniform* distribution** over all 2-tile patterns ($\text{TVD}_{\text{raw freq}} = 0.44$, 95% CI: [0.41, 0.49], $\text{TVD}_{\text{uniform}} = 0.83$, 95% CI: [0.80, 0.86]; Figure 5). In the *Less structured* conditions, the distribution of shortcuts was no closer to the *empirical frequency* than to *uniform*, likely because the empirical frequency was already close to uniform by design ($\text{TVD}_{\text{raw freq}} = 0.38$, 95% CI: [0.32, 0.45], $\text{TVD}_{\text{uniform}} = 0.48$, 95% CI: [0.42, 0.54]).

However, as can be seen in Figure 4, participants appear to favor the most common composite tiles more than the *empirical frequency* distribution can explain. We additionally consider a *greedy* distribution, which places all likelihood mass on the single most frequent 2-tile pattern of a tangram set. Indeed, in the *Highly structured* conditions, this ***greedy* distribution is closer to the distribution of shortcuts participants proposed** than the *empirical frequency* or *uniform* distributions ($\text{TVD}_{\text{greedy}} = 0.25$, 95% CI: [0.18, 0.32]). In the *Less structured* conditions, conversely, this greedy distribution is a worse predictor of participants' shortcut proposals ($\text{TVD}_{\text{greedy}} = 0.84$, 95% CI: [0.75, 0.92], Figure 5), consistent with the notion that participants did not have strong preferences for any specific shortcut when no pattern was especially common.

It might be the case that the *empirical frequency* distribution provides a stronger fit than the previous analysis would suggest, supposing it captures participants' tendency to propose shortcuts that reflect more common patterns vs. less common ones, even if none of these are the single-most common pattern. However, the *empirical frequency* distribution might overpredict or underpredict these shortcut proposals in numerical terms, even if they capture relative differences in how often participants propose one shortcut over another. To account for that possibility, we compute a third *fitted frequency* distribution, by applying the softmax function to the

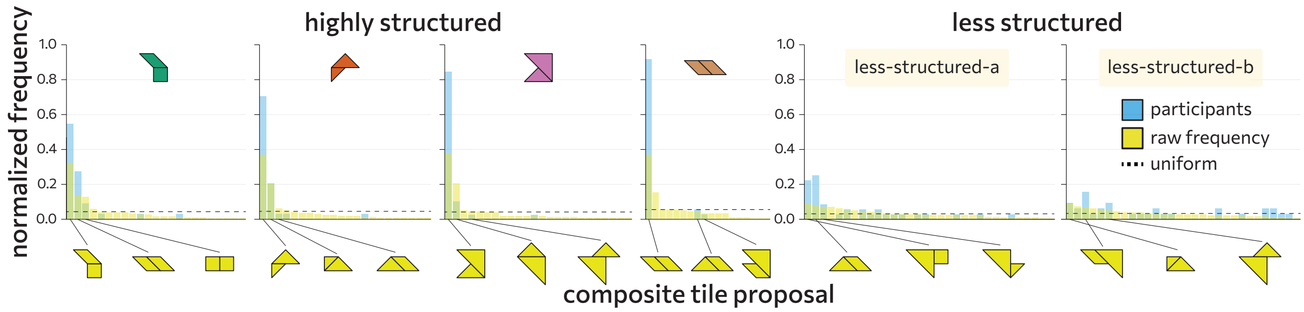

Figure 4: **Composite tile proposals** in final round: Only participants who proposed tiles with 2 pieces that appear in their tangram set condition are included. X-axis is ordered by highest frequency assigned by the *empirical frequency* ("raw") distribution. Annotations of composite tiles are in order of highest count of participant proposals, left-to-right.

empirical frequencies of patterns and fine-tuning the temperature parameter, which controls how much probability mass is concentrated on the most frequent patterns. Varying this temperature parameter allows us to test how well the empirical frequency distribution can approximate the shape of the observed distribution of shortcut proposals when we relax the assumption that participants performed exact probability matching when deciding which shortcut to propose. We fit this temperature parameter once to participants across all conditions, yielding an optimum of $T = 0.53$. In the *Highly structured* condition, this ***fitted frequency*** **distribution is closer to the distribution of participants' shortcut proposals than the *empirical frequency* distribution and the *greedy* distribution** ($\text{TVD}_{\text{fitted freq}} = 0.18$, 95% CI: [0.14, 0.23]). Furthermore, in the *Less structured* conditions, the *fitted frequency* distribution was not significantly closer than either the *empirical frequency* or the *uniform* distributions ($\text{TVD}_{\text{fitted freq}} = 0.37$, 95% CI: [0.30, 0.45]), but again was closer than the *greedy* distribution. Overall, across all conditions, **no distribution was a better fit to participants' proposals than the *fitted frequency* distribution** (Figure 5), suggesting that participants are sensitive to the patterns that recur frequently across tangrams and that these regularities influence the shortcuts they propose. In particular, participants in more structured contexts favor the few most recurring patterns when these statistical regularities are especially strong.

Revisions of tile proposals over rounds Finally, we investigated the composite tiles participants proposed over each round, as they accumulated more information on the statistical regularities present in their tangrams. In the *Highly structured* condition, participants appear to converge on proposing the most frequent composites over successive rounds (Round 1: $\text{TVD}_{\text{raw freq}} = 0.32$, 95% CI: [0.28, 0.37], Round 6: $\text{TVD}_{\text{raw freq}} = 0.44$, 95% CI: [0.41, 0.49]). However, participants in the *Less structured* condition show little change in that direction. They do not converge to proposing even the most empirically common tiles in their sets (Round 1: $\text{TVD}_{\text{raw freq}} = 0.38$, 95% CI: [0.32, 0.45], Round 6: $\text{TVD}_{\text{raw freq}} = 0.38$, 95% CI: [0.32, 0.45]). Perhaps participants in the *Highly structured* condition are able to take advantage of the more noticeable patterns early on in the sequence of assembly problems, retaining and slightly editing their initial guesses as the structure becomes more pronounced over time. Participants in the *Less structured* condition, in contrast, must estimate much subtler statistical properties over a changing sequence of problems. Future work could manipulate finer grained variations of problem structure.

Memory post-test As a final confirmation that participants were explicitly attentive to statistical structure across problems, we assessed whether participants would recognize frequently recurring patterns. In the 2-AFC recognition trials after all rounds, *Highly structured* participants indicated the most frequent tile in their tangram set as more common than the most frequent tiles of other set conditions (2.83/3 trials correctly responded old, 95% CI: [2.77, 2.88]).

Discovering and explicitly reporting the pattern was associated with better assembly performance

We additionally evaluated whether people improved on the assembly task with experience, and if their performance was modulated by the degree to which structure was shared across problems. During the first assembly period, participants in all conditions successfully completed 4.89 tangrams on average (95% CI: [4.81, 4.95]). By the last round, participants completed slightly more tangrams on average (0.07 more tangrams completed per round in final round than in first round, 95% CI: [0.007, 0.016]).

Participants in the *Highly structured* condition assembled tangrams more quickly across all rounds than participants in the *Less structured* condition ($t_{\text{Highly structured}} - t_{\text{Less structured}} = 25.0$ sec, $F_{1,4.01} = 10.1$, $p = 0.033$; Figure 3A). In both conditions, participants got faster at assembling tangrams as they completed more rounds (first round: 146 sec, last round: 97.0 sec, $\delta = -9.44$ sec/round, $F_{1,1799} = 165.242$, $p < 0.001$). And that improvement over round was more pronounced in the *Highly structured* conditions ($\delta_{\text{Highly structured}} - \delta_{\text{Less structured}} = -2.03$ sec/round, $F_{1,1503} = 5.39$, $p = 0.020$). Participants who encountered problems with greater shared structure performed better overall, and improved more as they solved more

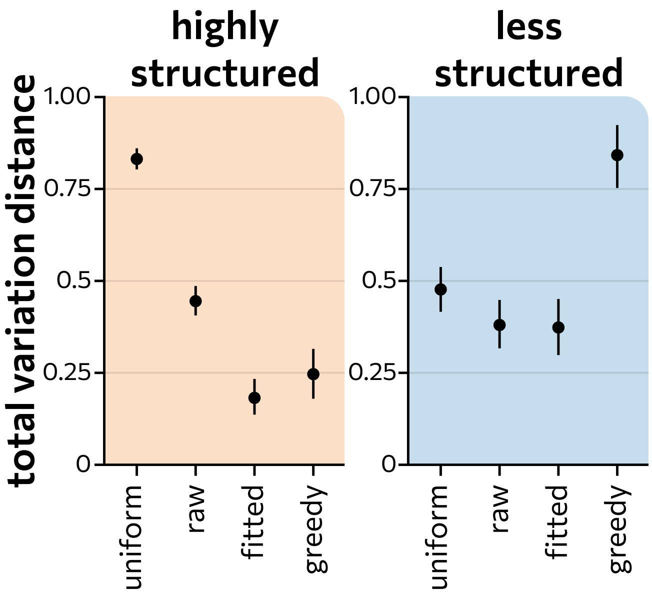

Figure 5: Distance between participant proposal distributions (2-piece tiles) and candidate proposal distributions in final round; lower TVD indicates better fit. "Raw" denotes *empirical frequency* distribution.

problems.

One reason for that might be the opportunity to propose or make use of composite tiles to assemble recurring substructures, which were much more common in the *Highly structured* conditions. We observed that participants who proposed the most frequent tile for their set completed assembly rounds faster than other participants. ($\delta_{\text{Highly structured}} = -45.9$ seconds, $F_{1,197} = 54.4$, $p < 0.001$, see Figure 3B). These participants could have benefited from having a useful tile available during assembly, or from enhancements in planning as a result of explicitly considering what composite tiles might be useful before use. Future work could investigate this by directly manipulating what composite tiles are available during assembly, or whether participants get the opportunity to propose or use composite tiles at all.

Discussion

We asked how experience solving assembly problems with recurring patterns affects the "shortcuts" people propose to aid their future performance. Our findings suggest that the decision to invent a shortcut, and the actual shortcuts that people propose, are closely related to statistical structure that people have learned over tasks in several senses. In general, people choose to invest in a shortcut that reflects statistically common sequences of actions in the problems they have observed so far. Unlike just practicing a subroutine or learning to recognize statistically likely patterns, shortcut invention intuitively involves optimization: participants should seek to invent shortcuts that would be maximally useful to them in the future. Our findings suggest that, indeed, while people propose shortcuts based on statistically common patterns in prior problems, they tend to favor the *most likely* patterns for candidate shortcuts. Finally, our results suggest that the degree of statistical structure in prior experience also shapes peoples' uncertainty about shortcut invention. Less obvious structure, learned across limited problem solving experience, yields greater uncertainty about what shortcuts might be useful at all in the future.

Together, these findings suggest that people can leverage statistical structure in problems to invent shortcuts for future ones. However, the current experimental design and findings raise several open questions that will guide future work. One concerns the relationship between structure in existing problems and peoples' reasoning about future ones: do people not always choose the single most common subroutine for shortcuts because they are imperfect statistical learners, or because they are uncertain about the payoff of any one shortcut across future problems? Future experiments could distinguish between these by further varying the degree of structure in problems, the cost of shortcut creation, and participants' expectations about the number and similarity of future tasks to current ones. A second question concerns the causal role of shortcut invention in problem solving: do people get better *because* they have invented shortcuts, or have the people who invent better shortcuts already learned more from the tasks? Future experiments could compare our current results to participants who simply complete a pattern recognition memory probe (but invent no shortcut), and those who are given shortcut actions that they did not propose themselves.

One exciting area for future study would be exploring candidate mechanisms of how people generate shortcuts from assembly experience. One class of algorithms, library-learning program synthesis models (Ellis et al., 2021), can capture human-like reasoning and generalization in generative domains like handwriting (Lake et al., 2015). In some cases, these algorithms exhibit curriculum effects; the explicit shortcuts or strategies they propose depend on the order of problems they see during training (Poesia & Goodman, 2023). Future work could investigate the alignment of learning trajectories between models and people.

People learn from solving problems, inventing new ways to get better based on what they have tried before. While we studied how people develop shortcuts that directly capture parts of problems that repeat, real problem solving–and domain expertise–can be far more creative. People discover much more abstract similarities and over much longer timescales (like noticing duplicates, or that they've been always building animals), and invent entirely new tools that break new ground on the range and kinds of problem solving available. Our work takes a step in this direction, suggesting that general mechanisms for learning from experience help guide the way that people plan ahead to perform later tasks.

Acknowledgments

We thank members of the Cognitive Tools Lab at Stanford University for their input and feedback. This work was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-2146755) to S.P.A. J.Y. is supported by an NDSEG fellowship. J.E.F. is supported by NSF CAREER Award #2436199, NSF DRL #2400471, and awards from the Stanford Human-Centered AI Institute (HAI) and Stanford Accelerator for Learning.

References

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

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55–81.

Chi, M. T., Glaser, R., & Rees, E. (1981). Expertise in problem solving.

Elffers, J., & Hollingdale, R. J. (1976). Tangram: The ancient chinese shapes game. *(No Title)*.

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., & Tenenbaum, J. B. (2021). Dreamcoder: Bootstrapping inductive program synthesis with wake-sleep library learning. *Proceedings of the 42nd acm sigplan international conference on programming language design and implementation*, 835–850.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lancia, G. L., Eluchans, M., D'Alessandro, M., Spiers, H. J., & Pezzulo, G. (2023). Humans account for cognitive costs when finding shortcuts: An information-theoretic analysis of navigation. *PLOS Computational Biology*, *19*(1), e1010829.

Poesia, G., & Goodman, N. D. (2023). Peano: Learning formal mathematical reasoning. *Philosophical Transactions of the Royal Society A*, *381*(2251), 20220044.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological review*, *82*(4), 225.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, *331*(6022), 1279–1285.

Terrace, H. S. (2001). Chunking and serially organized behavior in pigeons, monkeys and humans. *Avian visual cognition*.

Tian, L., Ellis, K., Kryven, M., & Tenenbaum, J. (2020). Learning abstract structure for drawing by efficient motor program induction. *Advances in Neural Information Processing Systems*, *33*, 2686–2697.

Wong, C., McCarthy, W. P., Grand, G., Friedman, Y., Tenenbaum, J. B., Andreas, J., Hawkins, R. D., & Fan, J. E. (2022). Identifying concept libraries from language about object structure. *arXiv preprint arXiv:2205.05666*.