Rapid policy updating in human physical construction

Will McCarthy [1] **Judith Fan** [2]

Abstract

The ability to build a wide array of physical structures, from sand castles to skyscrapers, is a hallmark of human intelligence. What computational mechanisms enable humans to reason about how such structures are built? Here we conduct an empirical investigation of how people solve challenging physical assembly problems and update their policies across repeated attempts. Participants viewed silhouettes of 8 unique towers in a 2D virtual environment simulating rigid-body physics, and aimed to reconstruct each one using a fixed inventory of rectangular blocks. We found that people learned to build each target tower more accurately across repeated attempts, and that these gains reflect both group-level convergence upon a smaller set of viable policies, as well as error-dependent updating of each individual's policy. Taken together, our study provides a novel benchmark for evaluating how well algorithmic models of physical reasoning and planning correspond to human behavior.

1. Introduction

The ability to build a wide array of physical structures, from sand castles to skyscrapers, is a hallmark of modern human intelligence. What computational mechanisms enable humans to reason about how such structures are built? Here we examine how people learn to "reverse-engineer" existing structures — that is, infer a sequence of actions that can be used to recreate them from simpler components. Specifically, we investigate how people make use of prior experience to update their policies across repeated attempts. Overall, our paper presents a novel benchmark task, human dataset, and set of evaluation metrics for AI construction agents. Our specific contributions are: (1) a web-based task environment for physical assembly, enabling scalable and dense measurement of human construction behavior; (2) a dataset containing 2,520 construction attempts across 105 human participants, including 22,793 actions; and (3) a quantitative evaluation of how humans reason about physical construction problems, update their policies based on prior performance, and converge upon similar solutions over time.

2. Related Work

2.1. Physical reasoning

Our paper builds on prior work on both classic (McCloskey, 1983) and contemporary (Battaglia et al., 2013) work investigating how humans reason about the properties of physical objects and their relationships, a suite of abilities known as intuitive physics. While many tasks in this literature involve passive judgments about physical scenes, a promising new direction is to consider tasks that involve active interventions on physical systems to achieve various goals (Allen et al., 2019; Hamrick et al., 2018). Our specific approach draws inspiration from prior work in cognitive science (Cortesa et al., 2018) and AI (Bapst et al., 2019; Jones et al., 2019) that has examined physical construction behavior. In particular, our approach takes inspiration from prior work investigating how such active interventions can be beneficial for learning (Gureckis & Markant, 2012), but in the context of physical reasoning tasks.

2.2. Planning

Our paper is also informed by recent advances in theories of human planning that highlight the pervasive role of mental simulation in guiding human sequential decision making (Solway & Botvinick, 2015; 2012; Daw et al., 2011), combined with reasonable assumptions about the cognitive costs of conducting mental simulations (Callaway et al., 2018; Hamrick et al., 2015). However, the generalizability of classic theories of planning to construction behavior may be limited by the historically narrow focus on tasks with low state-space complexity (van Opheusden et al., 2017) and abstract action spaces (Solway & Botvinick, 2015) far removed from the physical environment. Moreover, these theories do not address our core question of how people make efficient use of prior task experience to quickly update suboptimal plans (Hamrick et al., 2020).

[1] Department of Cognitive Science [2] Department of Psychology, UC San Diego, La Jolla, California, USA. Correspondence to: Will McCarthy <wmccarthy@ucsd.edu>, Judith Fan <jefan@ucsd.edu>.

Workshop on Object-Oriented Learning at ICML 2020. Copyright 2020 by the author(s).

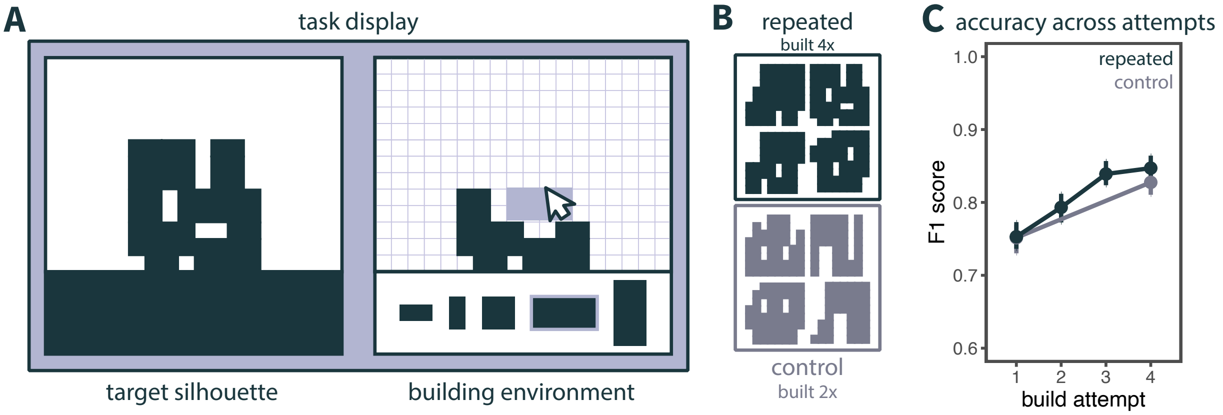

Figure 1. (A) Schematic of task display. The left window contained a target silhouette, and the right contained a building environment. (B) For each participant the 8 target towers were randomly assigned to the repeated and control conditions. (C) Reconstruction accuracy across build attempts.

3. Approach

3.1. Task environment

We developed a web-based gridworld environment (15x13) in which people could construct various block towers under simulated rigid-body physics (`Matter.js`). On each trial of our *Silhouette* task, participants aimed to reconstruct specific target towers in less than 60 seconds using a fixed inventory of rectangular blocks. Only the silhouettes of the towers were provided, requiring participants to infer which blocks to use, where to place them, and in what order (Fig. 1A). After the placement of each block, participants' towers became subject to gravity. Thus, if their tower was not sufficiently stable, single blocks or even the entire tower could fall over.

3.2. *Silhouette* dataset and experimental design

We randomly sampled a large number of stable configurations of 8-16 blocks, then manually selected 8 of these towers to create silhouettes with many valid solutions. For each participant, the 8 towers were randomly split into 2 sets of 4: a *repeated* set and a *control* set (Figure 1B). Each tower in the repeated set was attempted 4 times, interleaved among other towers. Each tower in the control set was attempted 2 times, once at the beginning and once at the end of the experiment. In subsequent comparisons between the *first* and *final* attempt for each tower, we combine data from repeated and control sets. In analyses of fine-grained changes in behavior across successive attempts, we restrict our analysis to repeated sets.

3.3. Human participants

105 participants, who provided informed consent and were recruited via Amazon Mechanical Turk, completed the task.

3.4. Representing states and actions

We define the current *state* as the silhouette of the current reconstruction. Under this definition, reconstructions that are composed of different blocks but share the same silhouette are treated as occupying the same state. A *state trajectory* consists of the sequence of all states a particular participant visited between the start and end of their reconstruction. We define *actions* as individual block placements, represented by 4-vectors $[x, y, w, h]$, where $0 \leq x \leq 15$, $0 \leq y \leq 13$ represents the coordinates of the bottom-left corner of the current block and where $(w, h) \in \{(1,2),(2,1),(2,2),(2,4),(4,2)\}$ represent its width and height, respectively.

4. Experiment

4.1. Improvement in human reconstruction accuracy

We used the F_1 score to measure reconstruction accuracy:

$$F_1 = \frac{2}{(recall^{-1} + precision^{-1})}$$

which reflects how well participants' reconstructions coincided with the target silhouette, and lies in the range $[0, 1]$. To evaluate changes in accuracy between the first and final attempts, we fit a linear mixed-effects model predicting F_1 score from build attempt (first, final) and condition (repeated, control) as fixed effects, including random intercepts for participant and tower. This analysis revealed a strong effect of build attempt ($b = 0.0759$, $t = 6.994$, $p < 0.001$), showing that participants learned to reconstruct towers more accurately over time (Figure 1C). We found no effect of condition, and no interaction between attempt and condition, suggesting that the improvement primarily reflected task-general, rather than tower-specific learning.

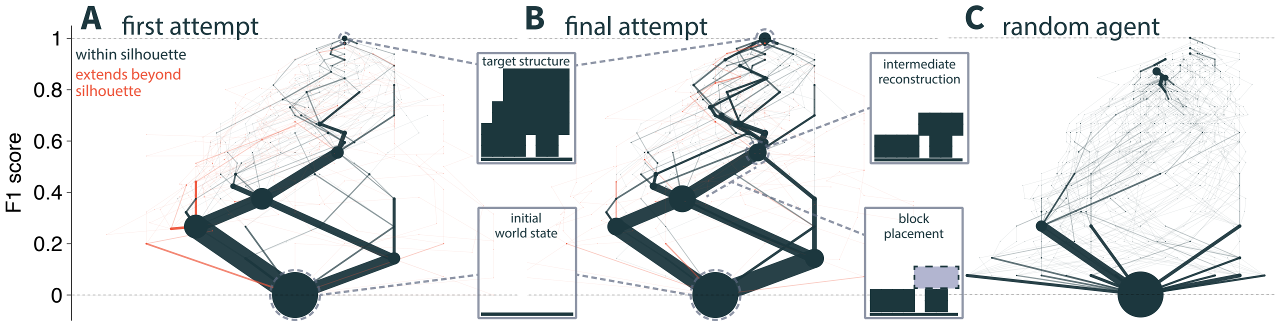

Figure 2. State trajectories for an example target tower for human participants' (A) first attempts, (B) final attempts, and (C) the random-policy agent. Each trajectory consists of a sequence of states (nodes) connected by actions (edges), beginning from the initial world state ($F_1 = 0$) and directed upwards toward complete reconstructions ($F_1 = 1$). Node size represents the number of times a state was visited. Edge thickness represents the number of times a state-state transition was traversed.

4.2. Systematicity in human construction policies

Even in the first attempt, many participants appeared to traverse the same states when reconstructing each target silhouette (Fig. 2A), hinting at broad consistency in the policies humans use to perform this task. To rigorously quantify these systematic biases toward certain states, we computed the Gini index (G) over the frequency of visits to each state across all participants:

$$G = \sum_{i=1}^{n}\sum_{j=1}^{n} |x_i - x_j| * (2\sum_{i=1}^{n}\sum_{j=1}^{n} x_j)^{-1}$$

To estimate how strongly human policies concentrate on the same sequences of states at different timescales, we next extracted n-gram representations for all state trajectories, each defined by n successive states, for $1 \leq n \leq 10$, then calculated G_n for each of these n-gram frequency distributions. To provide a baseline, we also constructed a random-policy agent that samples blocks and viable locations (i.e., within silhouette, maintaining stability) with equal probability. We used this random-policy agent to generate a null distribution of 1000 Gini values, each computed from 105 random-policy agents identified by unique random seeds. When comparing the mean observed G for human trajectories to this null distribution, we found that human state trajectories were reliably more concentrated on fewer n-grams than the random-policy agent, across n-grams of all lengths, for both first attempts (Z-score = 21.6) and final ones (mean Z-score = 42.7; Fig. 3A).

4.3. Policy convergence between individuals over time

Insofar as human participants are biased to discover similar solutions over time, we may expect the Gini index to grow between the first and final attempts. To evaluate this possibility, we fit human Gini coefficients with an LME model including attempt number and linear and quadratic terms for n, as well as random intercepts for target towers and participants (Figure 3A). This analysis revealed that there was lower overall convergence on longer sequences (i.e., larger n) than shorter sequences ($b = -0.0454$, $t = -21.3$, $p < 0.001$), as expected. Importantly, it also revealed a positive effect of attempt number ($b = 0.112$, $t = 6.02$, $p < 0.001$), suggesting that human participants tended to converge on increasingly similar policies over time.

Although convergence on the same sequences of states is one signature of having similar policies, the above analysis is insensitive to cases where two participants reconstruct a silhouette by placing the same blocks in the same locations, yet only have first and final world states in common. To address this limitation, we defined a measure of *action* dissimilarity that compares sequences of actions while disregarding the states in which they are performed. For a pair of action sequences, we define the "raw" action dissimilarity as the mean Euclidean distance between corresponding pairs of $[x, y, w, h]$ action vectors (Fig. 3C, light). As this measure compares the dissimilarity of sequences on an action-by-action basis, it is brittle with respect to the *sequence* in which they are performed. We therefore also obtained a "transformed" action dissimilarity between *sets* of actions, using the Kuhn-Munkres algorithm to identify the one-to-one mapping between actions minimizing the Euclidean distance between them (Fig. 3C, dark). We found that overall variability in the sets of actions performed was smaller on final attempts than on first attempts, for all target towers ($t(7) = 10.603$, $p < 0.001$; Fig. 3B).

Taken together, these results suggest that human participants converge on similar policies over time, indicating shared biases when reasoning about such construction tasks and updating their strategies.

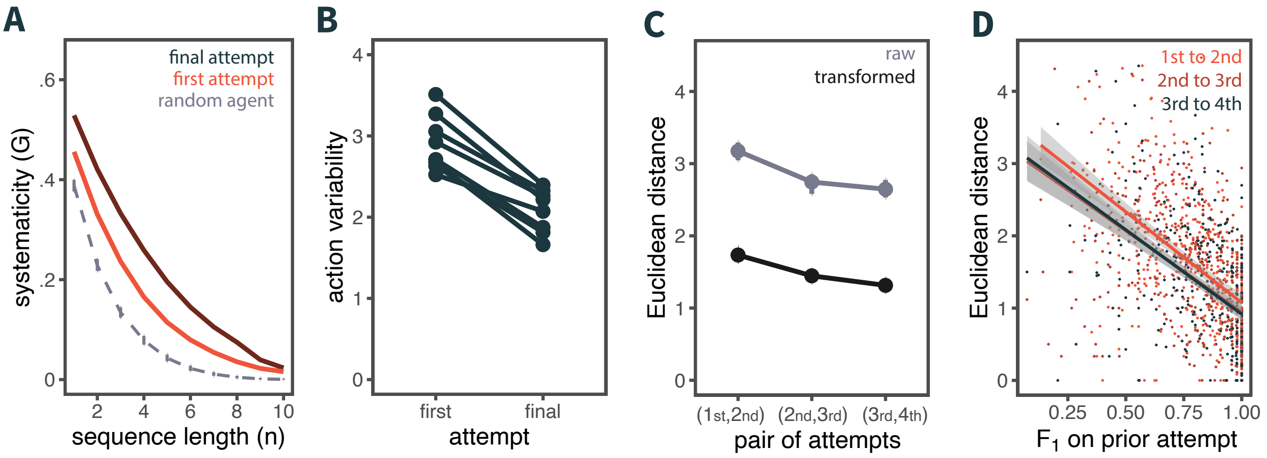

Figure 3. (A) Gini index for state trajectories for human participants in first and final attempts, compared to that of a random-policy agent. (B) Variability between participants in sets of actions performed on first and final attempts. Each line segment represents a different target tower. (C) Magnitude of change in action sequences (raw) and sets of actions (transformed) across successive build attempts. (D) Magnitude of change in sets of actions as a function of accuracy (F_1) on previous attempt, for each pair of successive attempts.

4.4. Policy convergence within individuals over time

If human participants converged upon internally consistent ways to reconstruct each target tower, we would predict that action dissimilarity between successive pairs of attempts (e.g., 1st-2nd) would decrease over time. We fit both raw and transformed action dissimilarities with a LME model including fixed effects for attempt pair, the accuracy of the earlier attempt, and the dissimilarity type (raw or transformed), as well as random intercepts for target tower and participant. Consistent with our prediction, we found that Euclidean distance is negatively related to attempt pair ($b = -0.186$, $t = -7.40$, $p < 0.001$; Figure 3C). We also found that transformed dissimilarities were smaller than raw ones ($b = -0.482$, $t = -2.96$, $p = 0.00315$), suggesting that participants updated policies in a way that achieved a similar outcome, even if they performed actions in a somewhat different sequence across build attempts.

4.5. Human policy updating is error dependent

To what extent is the human policy updating sensitive to errors made on previous attempts? Using the same linear model from the previous section, we found a strong negative relationship between accuracy on the most recent build attempt and how much they changed which actions they performed ($b = -0.6426$, $t = -4.054$, $p < 0.001$), such that participants updated their policy more when their previous attempt was less successful. Taken together, these results suggest that human participants make efficient use of prior experience to update their policies accordingly (Figure 3D).

5. Discussion

In this paper, we introduce a novel benchmark task for measuring how humans and AI construction agents reason about challenging physical assembly problems and update their policies across repeated attempts. Our quantitative evaluation of human behavior on this task revealed a large degree of consistency in policies used by different individuals, even on their first attempt. Additionally, we found that people converge on increasingly similar solutions over time, which represent a tiny fraction of all possible solutions, suggesting shared biases on such tasks. Moreover, our data suggest that people rapidly update their policies based on prior performance, even after one or a few attempts.

A key open question concerns the source of the systematicity we see in human strategies for solving these physical reasoning problems. Shared prior experience with a variety of other physical reasoning and planning tasks may play a crucial role, and understanding how humans transfer such broad experiences to new tasks may be critical for developing AI agents that learn as flexibly as humans do.

The highly sample-efficient learning we observed in humans differs starkly from the learning in even the most sophisticated current deep reinforcement learning agents, which require substantial amounts of experience to achieve good performance on similar tasks. Our immediate next steps will be to directly evaluate how well current AI construction agents (Bapst et al., 2019) emulate human behavioral data on the same tasks and metrics. Overall, we hope our study will spur progress at the intersection of cognitive science and AI to advance computational theories of human planning and physical reasoning.

6. Acknowledgements

We would like to thank David Kirsh and members of the Cognitive Tools Lab for helpful discussion.

All code and materials available at:
https://github.com/cogtoolslab/block_construction

References

Allen, K. R., Smith, K. A., and Tenenbaum, J. B. The tools challenge: Rapid trial-and-error learning in physical problem solving. *arXiv preprint arXiv:1907.09620*, 2019.

Bapst, V., Sanchez-Gonzalez, A., Doersch, C., Stachenfeld, K. L., Kohli, P., Battaglia, P. W., and Hamrick, J. B. Structured agents for physical construction. *arXiv preprint arXiv:1904.03177*, 2019.

Battaglia, P. W., Hamrick, J. B., and Tenenbaum, J. B. Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110 (45):18327–18332, 2013.

Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P. M., and Griffiths, T. A resource-rational analysis of human planning. In *CogSci*, 2018.

Cortesa, C. S., Jones, J. D., Hager, G. D., Khudanpur, S., Landau, B., and Shelton, A. L. Constraints and development in children's block construction. In *CogSci*, 2018.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011.

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

Hamrick, J. B., Smith, K. A., Griffiths, T. L., and Vul, E. Think again? the amount of mental simulation tracks uncertainty in the outcome. In *CogSci*. Citeseer, 2015.

Hamrick, J. B., Allen, K. R., Bapst, V., Zhu, T., McKee, K. R., Tenenbaum, J. B., and Battaglia, P. W. Relational inductive bias for physical construction in humans and machines. *arXiv preprint arXiv:1806.01203*, 2018.

Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Pfaff, T., Weber, T., Buesing, L., and Battaglia, P. W. Combining q-learning and search with amortized value estimates. *International Conference on Learning Representations*, 2020.

Jones, J., Hager, G. D., and Khudanpur, S. Toward computer vision systems that understand real-world assembly processes. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 426–434. IEEE, 2019.

McCloskey, M. Intuitive physics. *Scientific American*, 248 (4):122–131, 1983.

Solway, A. and Botvinick, M. M. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological Review*, 119(1):120, 2012.

Solway, A. and Botvinick, M. M. Evidence integration in model-based tree search. *Proceedings of the National Academy of Sciences*, 112(37):11708–11713, 2015.

van Opheusden, B., Galbiati, G., Bnaya, Z., Li, Y., and Ma, W. J. A computational model for decision tree search. In *CogSci*, 2017.