UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Abstract strategy learning underlies flexible transfer in physical problem solving

Permalink

https://escholarship.org/uc/item/1jn8k4hg

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Allen, Kelsey R.
Smith, Kevin A.
Piterbarg, Ulyana
et al.

Publication Date

2020

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Abstract strategy learning underlies flexible transfer in physical problem solving

Kelsey R. Allen
krallen@mit.edu
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Kevin A. Smith
k2smith@mit.edu
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Ulyana Piterbarg
upiter@mit.edu
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Robert Chen
robertcc@mit.edu
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Joshua B. Tenenbaum
jbt@mit.edu
Brain and Cognitive Sciences
Massachusetts Institute of Technology

Abstract

What do people learn when they repeatedly try to solve a set of related problems? In a set of three different exploratory physical problem solving experiments, participants consistently learn strategies rather than generically better world models. Participants selectively transferred these strategies when the crucial context and preconditions of the strategy were met, such as needing to "catapult", "support", "launch" or "destabilize" an object in the scene to accomplish their goals. We show that these strategies are parameterized: people can adjust their strategies to account for new object weights despite no direct interaction experience with these objects. Taken together, these results suggest that people can make use of limited experience to learn abstract strategies that go beyond simple model-free policies and are instead object-oriented, adaptable, and can be parameterized by model-based variables such as weight.

Keywords: problem solving; intuitive physics; tool use

Introduction

Imagine that you are attending a conference dinner buffet with a container of delicious-looking fried rice. As you go to scoop the rice onto your plate, you realize there is no serving spoon. The container is too hot to pick up, and being too polite to scoop the rice with your hands, what do you do? You might notice a stack of paper plates near the container that you can fold into a large scoop to pick up the rice. Having solved your rice problem, you could apply the same solution to other dishes without serving utensils. At future buffets with missing utensils, you would know to apply the same strategy, even if the plates were a different shape or made of a somewhat more rigid material, though you might have to adapt how you fold and scoop. And you might even transfer this strategy to scenarios that are nothing like buffets superficially, like using a soft Frisbee to dig a hole at the beach.

How are people able to learn and generalize these object oriented strategies so efficiently? Learning a simple, stereotyped scene-action response would not let you generalize this strategy beyond buffets with specific paper plates. Having general knowledge that paper plates are malleable does not tell you how you might use them to solve your problem. Instead it requires some insight about what the right kind of thing to do is, and how to adapt that idea to the new objects, scenes, and tasks you encounter. This applies not just to making novel paper utensils, but also to learning other kinds of abstract physical strategies such as levering heavy objects that could not otherwise be lifted, or placing an object under a wobbly table to stabilize it.

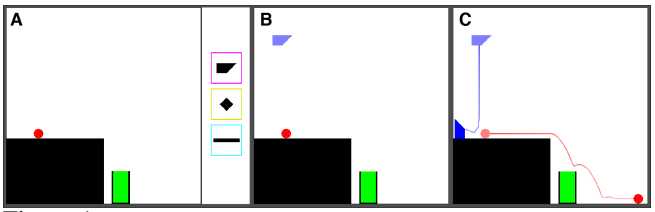


Figure 1: Diagram of a trial. (A) Participants attempt to get a red object into the green goal using one of the three tools on the right. (B) Participants choose a tool and where to place it. (C) Upon placing the tool, physics is turned on and participants see an animation of the scene unfolding. Figure credit: Allen et al. (2019).

Trial and error learning of this form is often modeled using either model-based or model-free reinforcement learning (Gershman, Markman, & Otto, 2014). Model-based learning assumes that an agent is learning more about the dynamics of the world through its interactions, which allows it to form better plans in the future, e.g., learning more about the flexibility of the plates. But even with a perfect model of the world, there is a near-infinite set of actions one might take – how would you know that folding a plate is the correct thing to do, as opposed to searching elsewhere for utensils?

In contrast, one variant of model-free reinforcement learning assumes that learning guides an agent towards promising actions by updating a policy (which action or sequence of actions you should take from the current state to maximize reward). In this way the agent learns exactly what action to take in any given situation, and so does not have to consider the outcomes of large sets of actions. However traditional representations of policies are relatively brittle: common methods create look-up tables that represent one-to-one mappings of states and their features directly to values or actions (Collins & Frank, 2013; Gershman et al., 2014). Such policy representations will not necessarily be flexible enough for an agent to transfer the paper plate skill to a new kind of paper, a new shape of plate, or possibly even a new buffet.

Allen et al. (2019) found that people instead use a hybrid learning system for solving physical problems: they have pre-existing models of dynamics that they can use to assess the outcome of their actions (Battaglia, Hamrick, & Tenenbaum, 2013), but learn what actions are useful in a given context by combining the output of that model with observations from their own actions. However, while Allen et al. (2019) found that people learn *within* a problem context, they found no evidence that people generalize *across* problems, perhaps be-

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

cause the physics puzzles they used in their study required a diverse set of strategies.

Here we ask whether people naturally transfer abstract policies (which we refer to as "strategies") for using objects across different contexts, and if so, what the representation of those strategies are. This is an important question because previous work focuses on how to reuse past experience by directly copying policies (Collins & Frank, 2013), or learning separate dynamics and reward functions from training tasks (Franklin & Frank, 2018) by clustering *context* variables (Xia & Collins, 2020; Momennejad, 2020). A natural question is then: what are these context variables, and by which criteria should they be clustered together?

Through a set of three exploratory experiments, we examine what people learn when they are repeatedly exposed to similar physical problems, and what contexts they transfer that learning to. In Experiment 1, we show that people learn strategies, not generically better world models, that are selectively applied to problems that appear similar to the ones they encounter during training. In Experiment 2, we test the generalization and adaptability of these learned strategies by seeing whether people can apply them to scenes that look visually distinct from, but require the same physical concepts of, the training problems. Finally, in Experiment 3, we test whether learned strategies can be parameterized by model-based variables such as an object's weight. People trained on medium weight objects were able to immediately adapt their strategies to account for objects which were heavier or lighter than those experienced during training despite no direct interaction experience with the new objects. Together, these results suggest that people learn strategies that are object-oriented, adaptable, and can be partially parameterized by learned world dynamics.

Experiments

We perform three exploratory experiments to examine whether participants would learn strategies, and if so, how those strategies might be represented. We focus on the Virtual Tools game proposed by Allen et al. (2019), shown in Figure 1. Each problem is presented as an initial scene (with physics turned off), a goal description, and three 'tool' objects to choose from (Fig. 1A). Participants must accomplish the stated goal objective by selecting one of the 'tool' objects and clicking to place it somewhere in the scene (Fig. 1B). After a single 'tool' is placed, physics is turned on using the Chipmunk 2D physics engine, and participants can see the resulting trajectories of all objects in the scene (Fig. 1C). Participants were given three attempts for each problem to accomplish the objective. We recorded all attempted actions including which tool was used and where it was placed.

We followed a similar experimental protocol to Allen et al. (2019) in order to familiarize participants with the task. First, participants were given a set of initial instructions explaining the kinds of objects that might exist in the different problems. They then interacted with a 'playground' level without a goal, and were required to make 3 placements before moving on. Finally, participants were given two simple practice levels that they had to solve before moving on; these were not analyzed. Participants then continued on to the main body of the experiment described in the sections below.

Experiment 1: Learning strategies

Allen et al. (2019) found no evidence of learning when people played 14 unrelated levels of the Virtual Tools game. They proposed that people were likely not learning better models of the world while they interacted with it, but instead learned a policy which allowed people to make better use of a model in individual trials. Since the trials differed substantially across training, reusing these policies in new scenes would be maladaptive. Instead, it could be that participants will only reuse learned policies when those policies would be useful in accomplishing their goals.

Procedure To test whether people could learn strategies, we designed four types of levels: *Catapulting*, *Tipping*, *Blocking* and *Tabling*. We then created 10 random variations of each level type where different aspects of the scene were varied, including the sizes and positions of each object and the shapes of each tool (see Fig 2A). Each participant was assigned to one of the four level types, and played all 10 random levels of that type. Participants were only given three attempts to solve each trial; if they did not succeed within those attempts, the level was marked as unsolved and they moved on to the next training level. The trial order was randomized across participants. After the training phase, all participants were given the same six testing problems: one from each of the four basic categories, and two chosen as controls to ensure that certain training conditions did not allow participants to get generically better at the game (Fig 2B). The two control levels were taken from Allen et al. (2019): *Chaining* was relatively difficult for participants, while *Unbox* was relatively easy. Just as in the training levels, participants were only allowed three attempts to solve the test levels.

153 participants were recruited using Amazon Mechanical Turk. We only analyze data from participants who succeeded on at least one of their training levels – criteria excluded four participants.

Results We first test whether there is any learning during training, and find that accuracy (whether a participant succeeded within 3 attempts) does improve over this phase (for all conditions, all $\chi^2(1) > 4.4$, all $ps < 0.036$; Fig. 3A), thus suggesting that training was effective.

During testing, we find no evidence of difference in performance on the control trials based on training condition (*Unboxing*: $\chi^2(3) = 2.29,\ p = 0.51$, *Chaining*: $\chi^2(3) = 6.16,\ p = 0.10$ respectively). This suggests that participants were not differentially learning more about the game in general, similar to how Allen et al. (2019) found no transfer learning across different types of levels.

Instead, people behave differently in test levels depending

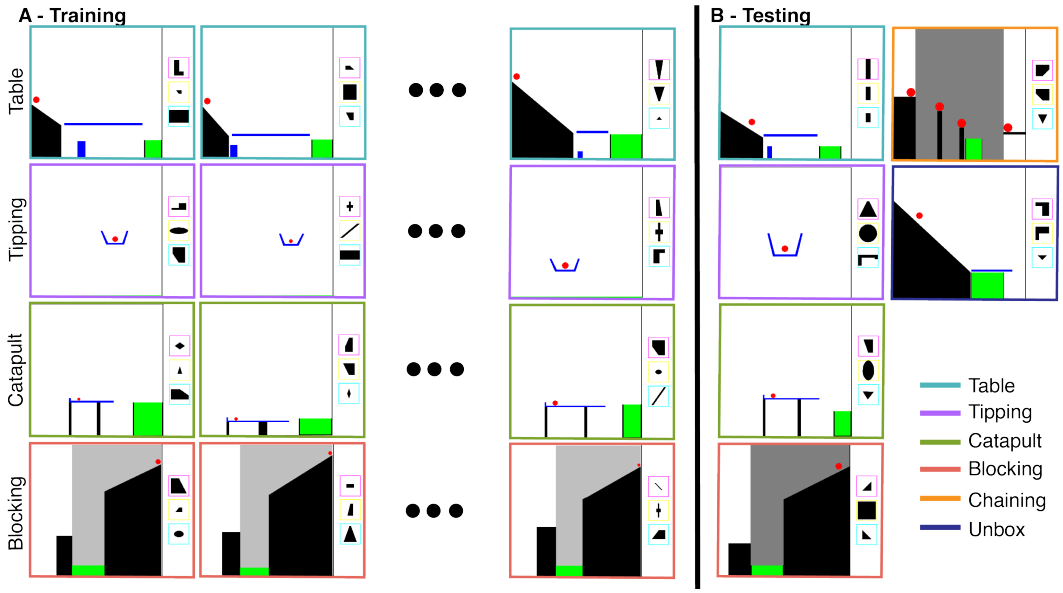


Figure 2: Diagram of Experiment 1. (A) Participants start with a learning phase where they play 10 related levels with random tool options with only three attempts per level. (B) Participants play the same set of 6 levels in random order during the testing phase. See `bit.ly/toolgame` to play the test levels without limits on the number of attempts.

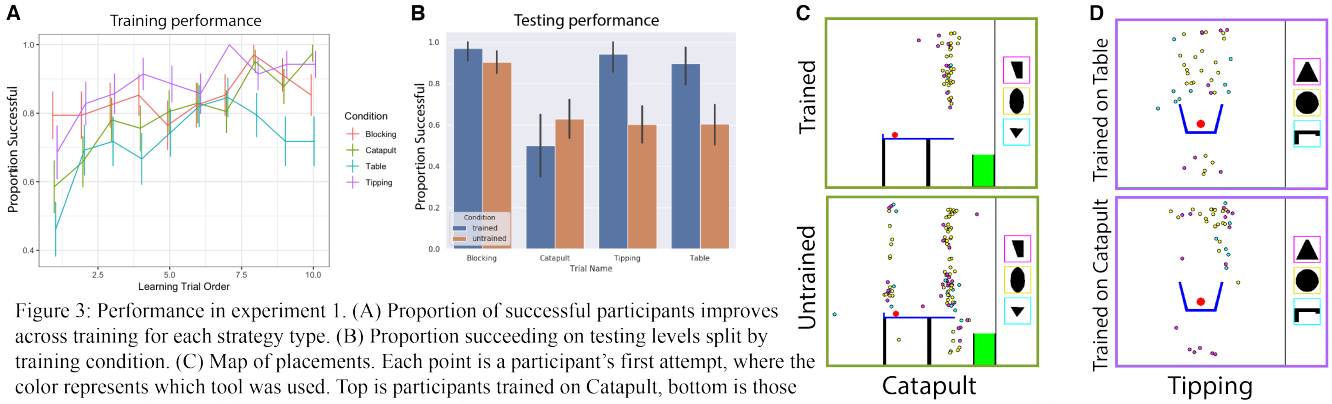


Figure 3: Performance in experiment 1. (A) Proportion of successful participants improves across training for each strategy type. (B) Proportion succeeding on testing levels split by training condition. (C) Map of placements. Each point is a participant's first attempt, where the color represents which tool was used. Top is participants trained on Catapult, bottom is those trained on all other conditions. (D) First placements for Tipping level. Top is participants trained on Table, bottom is participants trained on Catapult. There is no difference in above/below placement proportions, although there may be addditional structure in the above placements for those trained on Catapult.

on the level type they were trained on. Participants were reliably more accurate in the *Tipping* (94% trained vs. 61% untrained; $\chi^2(1) = 17.5$, $p < 0.001$, odds-ratio $= 10.8$) and *Table* conditions, (90% trained vs. 61% untrained; $\chi^2(1) = 13.0$, $p < 0.001$, odds-ratio $= 5.7$), but not in the *Catapult* (50% trained vs. 63% untrained; $\chi^2(1) = 2.01$, $p = 0.16$, odds-ratio $= 0.59$) and *Blocking* (97% trained vs. 90% untrained; $\chi^2(1) = 1.93$, $p = 0.17$, odds-ratio $= 3.52$) conditions (Fig. 3B). While the *Blocking* condition can be explained by ceiling effects (all participants find this level easy), *Catapult* requires a more detailed analysis.

To understand why trained participants are not more accurate on the *Catapult* testing level, we examine the detailed placements people tried (Fig. 3C). For this level there are 2 types of solutions: one solution type involves catapulting the ball into the goal, while the other involves hitting the ball directly and launching it into the goal. Participants in the *Catapult* training condition always try to use the catapulting strategy. However, in this particular level, that strategy is more difficult than directly hitting the ball. Participants who were not trained on *Catapult* hit the ball directly more often, and therefore had a slight advantage, leading to similar overall performance in terms of accuracy but easily distinguishable behavior.

What is learned? While we suggest that people are learning flexible strategies over objects and scenes, an alternate explanation could be that people are instead learning simple associations of where to place objects. This association must be more complex than simply placing objects in particular spatial locations, as the particular places that lead to success vary across each level. Instead, to improve in performance, they would need to learn an object-oriented strategy such as putting an object beneath the platform in *Table*, or above the lever in *Catapult*.

However, this does not rule out learning a simple object-oriented spatial prior on actions to take – e.g., placing the tool under or over the key object. We can examine whether

a simple over/under prior is being learned by looking at how participants perform on *Tipping* – a test level that could be solved by placing the tool over or under an object – depending on whether they were trained on levels that require dropping a tool from above (*Catapult*) or placing a tool underneath an object (*Table*). If this simple strategy is learned, we would expect participants in the *Catapult* training condition to be more likely to place a tool above, whereas participants trained on *Table* levels would be more likely to place a tool below. However, we find no evidence for any differences: the same proportion of participants placed the tool above regardless of whether they were trained on *Catapult* (82%) or *Table* (80%; $\chi^2(1) = 0,\ p = 1$; Fig. 3D). Thus we find evidence that the strategies participants are learning are context specific.

Experiment 2: Generalizing strategies

How context specific are the learned strategies? In Experiment 2 we tested whether the context depends on the presence of more abstract physical concepts even when the problem appears visually distinct. Participants were randomly assigned to three training conditions (*Catapult*, *Tipping* and *Table*) and trained using the same 10 levels from Experiment 1. Each participant then played 6 testing levels: 3 matched testing levels from Experiment 1, and 3 "transfer" levels designed for each level type (Fig 4C).

The transfer levels were designed to maintain the same physical concept as the training levels but appear visually distinct. For example, in the *Catapult Transfer* level, one must drop a tool on the left side of the plank (it is always on the right during training) in order to catapult the ball so that it hits a smaller ball which will then roll into the goal. In the *Table Transfer* level, succeeding requires supporting two planks simultaneously by putting a large block underneath them. In the *Tipping Transfer* level, participants need to place an object underneath the platform to destabilize it, thus tipping the ball into the container. If trained participants solve their transfer level more efficiently, this suggests that the representation of strategies and when to apply them is based on abstract physical concepts rather than simple visual similarity.

Sixty-two participants were recruited using Amazon Mechanical Turk. We only analyze data from participants who succeeded on at least one of their training levels – this caused four participants to be filtered from our analysis.

Results We broadly replicate the results of Experiment 1, first finding a significant improvement in accuracy over the course of training (for all conditions, all $\chi^2(1) > 11.9$, all $ps < 0.001$). The important test for Experiment 2 involves looking at the test accuracy in the *generalization* conditions (Figure 4.). Here, in both the *Catapult Transfer* and *Table Transfer* levels, we see significantly better accuracy for participants trained on that level type (*Catapult Transfer*: 76% trained vs. 32% untrained; $\chi^2 = 9.58$, $p = 0.002$, odds-ratio $= 6.75$, *Table Transfer*: 100% trained vs. 63% untrained; $\chi^2 = 10.9$, $p < 0.001$). We do not find evidence of improvement for the *Tipping Transfer* level (39% trained vs. 52% untrained; odds-ratio $= 0.59$). We can again look at the fine-grained participant behavior to understand why this is.

During training on the *Tipping* levels, participants could solve each level using two distinct strategies. One strategy involves tipping the container from *above* with a precise placement on the side of the container to flip it over. The other strategy involves tipping the container from *below* by putting any object beneath the container to destabilize it. In Experiment 1, almost all participants found and used the "tipping from below" strategy since it is more robust. We expected a similar effect in Experiment 2, and so the *Tipping Transfer* level is *only* solvable by destabilizing the platform from below. However, in Experiment 2 a significant proportion of our participants found and consistently used the "tipping from above" strategy: 55% of participants in the *Tipping* condition never tried tipping the container from below. Instead, they became adept at tipping from above.[1]

In Figure 5, we split participants by those who "tip from below" and those who "tip from above", and compared them to participants who were not trained in this condition. It is clear that (a) participants who learned to tip from above develop an exceptionally precise strategy compared to those who were untrained in this condition, and (b) if participants discover "tipping from below" they use it consistently and can generalize it to the transfer level reasonably well (with a mean accuracy of 70% compared to 0% for those "tipping from above" and 50% for those who were untrained).

Broadly, the successful transfer of strategies to different levels which maintain an abstract strategy concept but change the composition of the scene suggests that people flexibly adapt learned skills to new settings. While evidence from Experiments 1 and 2 suggests that people do not form generic spatial relation priors such as "putting something below", they do transfer more abstract fine-grained strategies, such as "supporting an object from below", "tipping an object from below", or "catapulting" an object.

Experiment 3: Composing strategies and models

Based on Experiments 1 and 2, we established that people can learn strategies which shape how they interact with a new problem. Previous work has found that people can learn about object properties such as weight by observing how they interact with other objects (Schwettmann, Tenenbaum, & Kanwisher, 2019; Yildirim, Smith, Belledonne, Wu, & Tenenbaum, 2018), but we do not know whether this kind of object property learning can be combined with learned strategies.

If people learn strategies that are more akin to *habits*, these new object types should not affect the kinds of actions that people take until they have acquired significant interaction experience with the new object type. In the more extreme case, if people learn about the object properties purely through observation with no direct interaction, their strategies

[1] This could be due to the difference in the number of participants recruited: Experiment 1 had 40 participants in the *Tipping* condition while Experiment 2 only had 19.

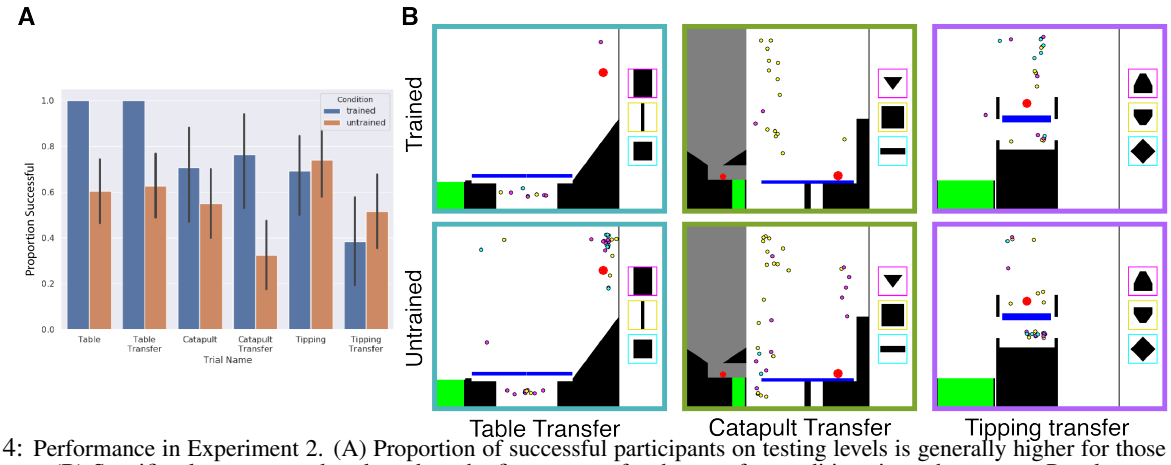


Figure 4: Performance in Experiment 2. (A) Proportion of successful participants on testing levels is generally higher for those trained on the strategy. (B) Specific placements and tool used on the first attempt for the transfer conditions in each strategy. People were trained on levels from Experiment 1.

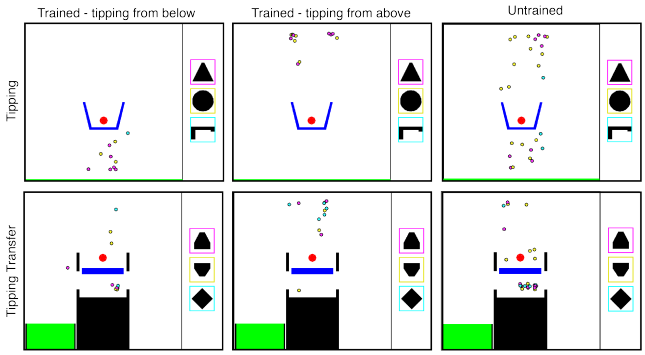


Figure 5: First placements of participants in the *Tipping* condition split into those who learned to tip from above, those who learned to tip from below, and those who were untrained.

could not be updated using most standard RL learning procedures. Therefore, if people do compose observed models with learned strategies without any direct interaction experience, it suggests that people are using strategies to guide a model-based sampling process with an updated model, or that the representations of the strategies themselves are parameterized by a model. This presents a unique and complementary perspective on model-based and model-free decision making in humans, as prior work generally assumes models are learned by interacting with the world, not via passive observation.

Stimuli We designed two new types of levels to test model-strategy composition, *Push* and *Launch*, and included a subset of the *Catapult* levels from Experiments 1 and 2. In each level type, the ball size is kept constant so that participants cannot learn strategies that implicitly depend on the weight of the ball through its size.

In *Push*, a ball rolls down a steep slope and participants must put an object in its path to slow it down such that it remains in the green goal region. When the slope is very steep, solving these levels requires placing an object *far in front* of the goal region to slow the ball. Physically, we measure this as the amount of work that the ball needs to do to the tool in order to end up in the green region. In the training levels, we vary the steepness and length of the slope, as well as the position of the goal region.

In *Launch*, participants must drop a tool onto a ball that will roll into another ball and cause it to fall into the container. Depending on the relative positions and heights of the container and table, succeeding could require hitting the ball very hard or very lightly. Physically, we measure this as the amount of momentum that needs to be applied to the ball in order to succeed. In the training levels, the positions of the balls, height of the table, and position of the goal is varied.

Finally we also include a subset of the *Catapult* levels from Experiments 1 and 2. Depending on the location of the container relative to the ball, succeeding requires applying differing amounts of momentum to the platform to ensure the ball does not under or overshoot the container.

Procedure Participants were trained on 5 levels of either the *Catapult*, *Push* or *Launch* types, with 3 attempts per trial. After training, all participants watched two videos of a blue ball interacting with a pink object, and two videos of it interacting with a purple object. The pink objects were lighter (with a density $0.2\times$ the blue and red objects), while the purple objects were heavier (with a density $2\times$ the blue and red objects). After watching the four videos, participants then played a level of the same type as their training condition, but with either a red, pink or purple object. In this way, we could test whether people composed knowledge from observing dynamics with a learned strategy without requiring any interaction experience. Participants then also played every other level type with a randomly selected object mass.

At the end of the experiment, participants filled out a questionnaire which included the questions: "Which color corresponds to the heaviest objects?" and "Which color corresponds to the lightest objects?". Both questions could be answered with "red", "purple" or "pink".

We recruited a total of 168 participants across 9 conditions (3 category types $\times$ 3 weight conditions). Like in Experiments 1 and 2, we only analyzed data from participants who succeeded on at least one training level. Additionally, since we were interested in whether people composed new knowledge about object properties with learned strategies, we only included participants who either correctly answered which object color was heaviest (purple), or which object

color was lightest (pink) to ensure they had learned relative object weights. Using both filters resulted in 31 participants being eliminated from the analysis (16 who failed because of the success criteria, and 15 because of the questionnaire).

Results Participants showed improvement throughout training in all conditions (all $\chi^2 > 7.8$, all $p < 0.005$). In the *red* ball test trials, there is also an overall effect of training condition ($p = 0.01$), consistent with Experiments 1 and 2. To test model-strategy composition, we examine what people do on their first attempt in the test level when the density of the ball has changed.

In Figure 6A, we show participants' first attempts in the heavy and light testing conditions for each level type. Participants appear to take the weight of the ball into account when choosing where to place an object and which object to place. For both *Catapult* and *Launch*, participants generally choose heavier objects and place them higher when the ball is heavier (purple), while in *Push* they place the tool farther to the left to slow down the ball.

We can quantify this by looking at the momentum applied across weight conditions for *Launch* and *Catapult*, and the work done across weight conditions for *Push* (Figure 6B).

In *Launch*, we find that there is an effect of object weight on how much momentum participants impart with their tool ($F(2,40) = 5.0,\ p = 0.011$, partial $\eta^2 = 0.201$). We do not find a reliable difference for participants trained on *Catapult* ($F(2,43) = 1.6,\ p = 0.22$, partial $\eta^2 = 0.067$), but this appears to be because people in the heavy condition separate into two groups: one group which applies a significant amount of momentum, and one who does not. A subset of participants use the lighter tools on their first attempt, but after observing this single failure, 13/15 participants used the heaviest tool on their second attempt.

Similarly, in Push, there is no statistically significant effect of weight on work done ($F(2,45) = 2.8,\ p = 0.069$, partial $\eta^2 = 0.112$). It again appears that in the heavy condition a subset of participants realized they needed to place the heavy object to the left of the goal, while the rest did not. Of the participants who did not succeed on the first attempt, 10/11 increased the work done in the second attempt.

Discussion

Through three exploratory experiments, we showed that in a physical problem solving task, people learn abstract strategies that are object-oriented (Exp 1), can be transferred to contexts that differ visually but rely on similar principles (Exp 2), and can be parameterized by model-based variables such as weight (Exp 3). These results hint at a picture of rapid trial-and-error learning which focuses on abstract *strategies* that can be used to guide a model-based sampling procedure towards promising regions of the problem space.

There remain several open questions for future work to better understand the underlying representation of these strategies and how they are connected to model-based reasoning. Specifically, the kinds of strategies learned here are not the same as more traditional object-specific affordances that have been a popular framework for thinking about tools in cognitive psychology (Gibson, 1979). This was by design – in our experiments we explicitly randomized the tools available for each problem, and therefore the learned strategies had to be agnostic to any specific tool. However, we want to emphasize that we consider the objects here to still be tools, just unfamiliar ones. We see this as akin to a paper plate that can be re-purposed as a scoop, or a broom that can be re-purposed as a table leg. In future experiments, we would like to investigate whether people can learn object-specific strategies that dictate how a particular object can be used to accomplish a new objective.

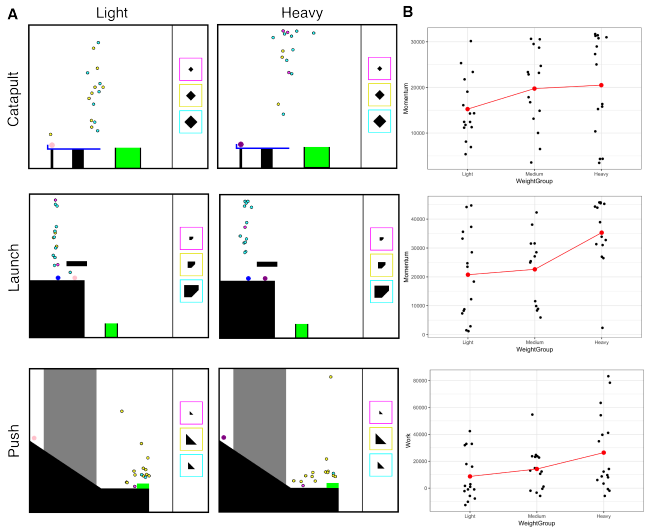


Figure 6: Momentum applied in the *Catapult* and *Launch* conditions, and the amount of work that would need to be done to the placed object to succeed in the *Push* condition.

In Exp 3, we saw a mixture of people who composed strategies and models without any experience, and those who required a single interaction with the scene to adjust their strategy. This suggests that even those who do not immediately generalize learn a strategy that is abstract enough to allow rapid updates. Further work could study the form of this representation and why individual differences exist.

Our experiments suggest that people's trial-and-error physical problem solving does not fit neatly into model-based reinforcement learning or model-free reinforcement learning. Instead, the content of people's strategies and how they know to apply them might be based on more abstract principles, allowing for broader generalization than previous studies of decision making would imply. However, these experiments are exploratory, and many open questions remain. Does the applicability of a strategy really depend on physical concepts, or is it the recognition of key objects and object-object relationships that appeared in a training level? How diverse can the training levels be? Can people learn multiple strategies simultaneously? Such questions present a wide and exciting space for future work to better understand the representations of physical problem solving strategies and how they are applied to new scenarios.

Acknowledgments

KRA, KAS, and JBT are supported by CBMM funded by NSF STC award CCF-1231216. KAS and JBT are supported by ONR grant N00014-13-1-0333.

References

Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2019). Rapid trial-and-error learning in physical problem solving. *Proceedings of the Cognitive Science Society*.

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

Collins, A. G., & Frank, M. J. (2013). Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological review*, *120*(1), 190.

Franklin, N. T., & Frank, M. J. (2018). Compositional clustering in task structure learning. *PLoS computational biology*, *14*(4), e1006116.

Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, *143*(1), 182.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin Co.

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166.

Schwettmann, S., Tenenbaum, J. B., & Kanwisher, N. (2019). Invariant representations of mass in the human brain. *eLife*, *8*, e46619.

Xia, L., & Collins, A. G. E. (2020). Temporal and state abstractions for efficient learning, transfer and composition in humans. *bioRxiv*.

Yildirim, I., Smith, K. A., Belledonne, M., Wu, J., & Tenenbaum, J. B. (2018). Neurocomputational modeling of human physical scene understanding..