Connecting perceptual and procedural abstractions in physical assembly

Will McCarthy
Dept. of Cognitive Science
UC San Diego
wmccarthy@ucsd.edu

Marcelo Mattar
Dept. of Cognitive Science
UC San Diego
mmattar@ucsd.edu

David Kirsh
Dept. of Cognitive Science
UC San Diego
kirsh@ucsd.edu

Judith Fan
Dept. of Psychology
UC San Diego
jefan@ucsd.edu

Abstract

Compositionality is a core feature of human cognition and behavior. People readily decompose visual objects into parts and complex procedures into subtasks. Here we investigate how these two abilities interact to support learning in a block-tower assembly experiment. We measured the way participants segmented these towers based on shape information alone, and asked how well the resulting parts explained the procedures other participants used to build them. We found that people decomposed these shapes in consistent ways and the most common parts appeared especially frequently as subroutines in the assembly experiment. Moreover, we found that the subroutines participants used converged over time, reflecting shared biases toward certain ways of reconstructing each tower. More broadly, our findings suggest important similarities between the perceptual and procedural abstractions humans use to perceive and interact with objects in their environment.

Keywords: planning; visual reasoning; task decomposition; abstraction; chunking

Human environments are filled with objects and structures that members of our species have made. The variety of these objects demonstrates our capability of learning how to construct new kinds of entity. Making things is easy for us– we are able to pick up a new piece of furniture from IKEA and put it together ourselves, usually on the first try. Furthermore, the presence of a new kind of object is rarely an enigma– our visual systems are extremely good at breaking down objects into component parts, providing a representation we can use to understand what something is and how it works. How are the ability to create and the ability to perceive complex objects related? Here we explore the relationship between procedural and perceptual abilities through two kinds of abstraction– the composition of multiple actions into *procedural abstractions*, and the decomposition of visual scenes through *visual decomposition*. While these two phenomena have been studied separately, we present a first attempt at studying their relationship in an explicitly compositional task– physical assembly.

Procedural abstractions In tasks that involve making extended sequences of decisions, an effective way of learning is through *procedural* or *temporal* abstraction– the chunking of multiple actions or decisions into a single unit. Humans learn extended sequences of actions (Huys et al., 2015; Xia & Collins, 2020), allowing them to represent task structure hierarchically (Botvinick, Niv, & Barto, 2009; Solway et al., 2014). The use of 'temporal' abstractions (e.g. *options* (Sutton, Precup, & Singh, 1999)) in reinforcement learning drastically improves the ability of agents to learn complex behaviors, particularly in tasks where simulating possible actions with a model has proved useful (Botvinick et al., 2009), as is the case in physical assembly (Bapst et al., 2019; Hamrick, 2019). However, the task of *discovering* useful temporal abstractions is computationally expensive (Botvinick & Weinstein, 2014), so is unlikely to explain how humans rapidly decompose problems into tractable units. Recent work has begun to explore other computational methods for task decomposition in humans (Correa, Ho, Callaway, & Griffiths, 2020), however these approaches have largely focussed on inferring hidden problem structure (Tomov, Yagati, Kumar, Yang, & Gershman, 2020), rather than making decisions about how best to organize behavior in a largely observable world.

Visual decomposition If aggregating actions we perform while making a something is one part of learning how to assemble something, another vital element is understanding the changes in object structure that correspond to those actions. One source of this information is the visual perception, which actively structures our representations of what we see. This process is partly achieved through stable principles and biases (Koffka, 1935; Wertheimer, 1938) that determine the which elements in a scene are likely to be grouped together (such as proximity, similarity, common fate, and good continuation) (Wertheimer, 1938), as well how to segment figure from ground. Other approaches have explored the flexibility of our object representations in response to certain kinds of knowledge, such as an object's category (Schyns, Goldstone, & Thibaut, 1998; Goldstone, 2003). As well as carving up the visual array into distinct units, we group elements together into chunks (Miller, 2020) through principles that determine the hierarchical relationships between elements (Palmer, 1977), as well as by picking up on statistical regularities such as the likelihood of parts occurring together (Fiser & Aslin, 2001, 2005; Orbán, Fiser, Aslin, & Lengyel, 2008). We also use such statistical information to adaptively change the parts (or 'features') of an object we emphasize when making judgements about category or identity (Austerweil & Griffiths, 2011, 2013), and information about an object's parts has been explored as object recognition (Biederman, 1987). Prior work has explored visual grouping due to expertise in settings involving interaction (De Groot, 2014; Chase & Simon, 1973), although less attention has been given to how humans adaptively change representations of more meaningful objects over short timescales. Perceptual decomposition is also rarely explored in the context of generative behaviors that involve interacting with object structure.

Generative behaviors and perception The way people perceive objects and the procedures uses to generate them have typically been studied separately, however a growing body of evidence points to a mutually supportive relationship between the two. A number of approaches have explored how generative behavior can lead to more robust perceptual skills, for example in hand-writing (James & Gauthier, 2006; James, 2017; Zemlock, Vinci-Booher, & James, 2018), and in drawing (Fan, Yamins, & Turk-Browne, 2018). The "vision as inverse graphics" (Yildirim, Kulkarni, Freiwald, & Tenenbaum, 2015; Kulkarni, Whitney, Kohli, & Tenenbaum, 2015) also suggests that generative *representational* formats, a kind that might be learned during assembly experience, can underlie effective recognition (Lake, Salakhutdinov, & Tenenbaum, 2015). Conversely, perceptual abstractions may also support effective decision making in complex tasks. The integration of state abstractions mirroring human visual perception provided a breakthrough in achieving human-level results in video games (Mnih et al., 2013, 2015), and the object-oriented nature of human cognition has lead researchers to explore the computational benefits of graph-based state representations that highlight relations rather than features (Battaglia et al., 2018; Bear et al., 2020). While work in this area has started to explore physical assembly (Bapst et al., 2019), whether state abstractions could support rapid discovery of procedural abstractions remains unclear.

Physical assembly In this paper we explore perceptual decomposition as an explanation for people's ability to successfully recreate complex objects with little to no experience. Physical assembly is a natural behavior that requires a continuous back and forth between perception and the active transformation of the thing you are perceiving. Prior work investigating the cognitive mechanisms supporting physical assembly analyzed the people's actions as they attempted to accurately reconstruct 2D block towers, given only an ambiguous silhouette of the tower as reference (McCarthy, Kirsh, & Fan, 2020). In a web-based building environment participants were supplied with a fixed set of rectangular blocks, and aimed to recreate the same 8 towers over multiple attempts (*Figure 1 B, C*. The authors found a great deal of overlap in the sequences of block placements made by participants, even the first time they tried to build each tower. A potential explanation for this is that participants identified many of the same subtasks– namely, specific parts of the silhouette that they reconstructed over a sequence of actions. Here we explore a specific reason for why this might be the case– that participants perceptually decomposed silhouettes in consistent ways, and that they treated these perceptual decompositions as subtasks. To explore this possibility, we present an novel experiment in which a new set of participants were asked to partition the same ambiguous block-tower silhouettes into what they considered were its 'natural subparts' (Palmer, 1977). To assess how these parts relate to procedural abstractions, we directly compare them to common subroutines of participants who assembled the same silhouettes out of blocks. In sum, our paper represents a first attempt at delineating the relationship between perceptual and procedural abstractions during physical assembly.

Methods

We first needed a measure of how people perceptually decomposed silhouette stimuli. Although segmentation-like processes likely occur at multiple points during perception, we were most interested in those that directly influence decisions about how to consrtuct an object. We thus asked participants to identify the "natural subparts" of an object using a different generative task: coloring-in.

Stimuli Participants were presented with the 8 block-tower silhouettes from the *silhouette assembly experiment* (McCarthy et al., 2020) (*Figure 1B,C*). Towers from the assembly experiment were created from a fixed set of rectangular blocks, designed to fit within an 8*x*8 grid-world environment. In our (*decomposition experiment*), this grid was superimposed onto each silhouette, partitioning it into squares that could be filled with a color to delineate the "parts" of the shape (*Figure 1A*). Unlike the assembly experiment, silhouettes in the decomposition experiment were centered in a uniform grey box, and participants were not told that the "shapes" presented were silhouettes of stable towers, or that they were generated from a specific set of blocks.

Task Participants were presented with all 8 silhouettes in a randomized order. They were told that there was "no correct way" of decomposing them, but to "color them in according to how you see the parts of the shape", and were encouraged to provide decompositions that "felt natural". Squares were selected by clicking on them individually or, to encourage selection of contiguous regions, by dragging the mouse to select multiple squares. We assessed comprehension with a sequence of practice trials, which involved a range of rectangular and non-rectangular "parts" to avoid biasing participants towards selecting regions of any particular shape. Participants could use up to 9 different colors, and filled the entire silhouette before progressing to the next trial.

Participants 50 US- and UK-based participants were recruited from Prolific. 1 participant was excluded for not following instructions.

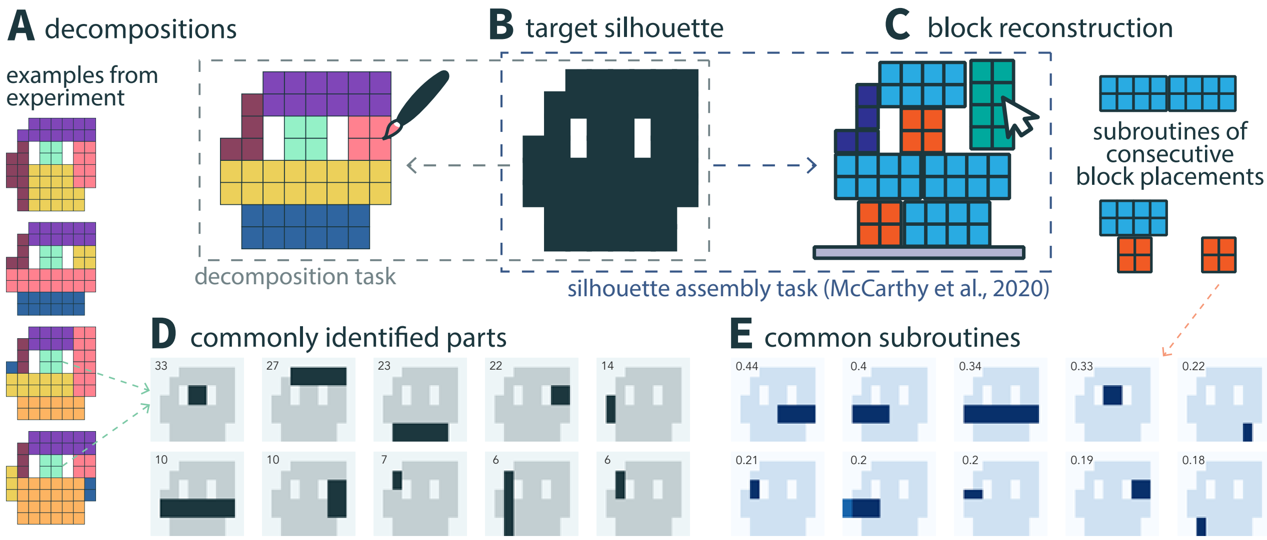

Figure 1: By coloring-in squares (A), participants were asked to segment silhouettes of 2D block towers (B). In a separate assembly experiment, the same silhouettes were reconstructed by placing blocks (C). We identified the most common parts selected by participants in the decomposition experiment (D) (top-left: number of participants who identified part), and compared these with common subroutines used by participants who constructed towers (E) (top-left: proportion of assembly participants who built part over consecutive actions during first assembly attempt).

Results

Participants decompose silhouettes into consistent parts

Participants produced a single decomposition of each silhouette into a set parts. 63.8% (95% CI: $[43.6, 83.9]$) of the decompositions for each silhouette were unique, confirming that each could be segmented in a variety of ways. Some silhouettes evoked more varied decompositions than others, ranging from one that yielded just 14 unique decompositions (28.6%), to another that was decomposed differently by all 49 participants. The number of parts highlighted in each tower was also highly variable ($M = 5.39, SD = 0.947$), with silhouettes that were identified as having more parts generally receiving a more varied set of decompositions.

Despite the large variety in entire decompositions, there was a great deal of overlap in the parts that participants identified (*Figure 1 D*). Participants seemed largely biased towards rectangular parts, which made up (82.9%, 95% CI: $[81.4, 84.6]$) of all parts identified, and were particularly well represented among the most frequently identified parts– of the 8 most commonly identified across all 8 silhouettes, only two were not rectangular. Still, a substantial portion of the *unique* parts identified were non-rectangular 39.8% (95% CI: $[28.5, 51.1]$). These were typically shared across fewer decompositions– 2.24 (95% CI: $[1.11, 3.36]$) compared to 7.11 (95% CI: $[4.86, 9.35]$) for rectangular parts– suggesting that participants shared some common strategies for decomposing silhouettes into rectangles among more idiosyncratic ways of decomposing the silhouettes into less regular shapes, in line with prior work suggesting a bias towards geometric simplicity in perceptual organization. This may explain another feature of the parts identified: that a large proportion of the them (22.4% (95% CI: $[18.7, 26.0]$)) were of the same shape as one of the 5 kinds of blocks used to generate the towers. This may be surprising, given that participants in the decomposition task did not know that the silhouettes were generated from rectangular blocks, let alone the dimensions of those blocks, and may suggest that perceptual decomposition provides relatively accurate clues about the underlying structure of towers in the assembly experiment.

How well do commonly identified parts explain assembly behavior?

We suspected that such clues may have helped participants accurately reconstruct towers in the assembly experiment (McCarthy et al., 2020). In particular, we hypothesized that perceptual decomposition defined regions of the silhouettes that participants would be likely to construct over a sequence of block placements. We therefore inspected the actions made by participants in the assembly experiment for *contiguous subsequences of block placements* that generated the parts identified in the decomposition experiment. Because this is a between-subject comparison, and the less frequently identified parts from the decomposition experiment seemed to reflect more idiosyncratic strategies, we limited our search to the k most common parts from each tower, where k was selected to include 85% of all parts identified, giving us up to $k = 28$ parts for each tower.

More common parts built more frequently 31.9% (95% CI: $[21.9, 41.9]$) of the *common parts* did not appear reconstructions at all. Qualitative analysis revealed that many of these parts would have been impossible to build with the set of blocks available in the building study, for example because they contain the wrong number of grid squares or would be physically unstable. In other words, some of the ways participants segmented the silhouettes do not correspond to valid task decompositions for the assembly experiment, and additional work is needed to assess the extent to which systematic errors in the assembly experiment can be explained by these structural parses. The majority of the common parts (68.1% 95% CI: $[58.1, 78.1]$) did appear in building procedures, consistent with perceptual decomposition playing some role in identifying subtasks in the assembly experiment.

To better understand how parts from the decomposition experiment relate to assembly behavior, we analyzed how often parts were built at different points in the assembly experiment. Each participant in the assembly experiment attempted to recreate half of the silhouettes 2 times, and the other half 4 times. As prior work revealed no substantial differences between final reconstructions of towers attempted 2 times and those attempted 4 times, we collapsed across conditions and compared *first* and *final* building attempts from both groups. We fit a linear mixed effects model predicting the proportion of tower reconstructions that contained each part, limiting our search to those parts that were built at least once. We included fixed effects for attempt (first vs. final) and the number of times the part was identified in the perceptual experiment, and random intercepts and slopes for each tower. Also consistent with our hypothesis that perceptual decomposition contributes to subtask selection, we found that parts that were more frequently identified in the decomposition experiment were also built more frequently in the assembly experiment ($b = 0.01074$, $t = 9.87$, $p < 0.001$).

Change in prevalence of parts with assembly practice
We also found that the parts identified in the decomposition experiment were built more frequently in *final* attempts than in first attempts ($b = 0.0449$, $t = 2.18$, $p = 0.0302$) of the assembly experiment. Two prior findings may explain this result. Firstly, participants in the assembly experiment built more complete and accurate towers in their final attempts. As parts from the decomposition experiment are necessarily regions *within* the silhouette, the more blocks that are placed within a silhouette, the more likely it is for a part to be built. Secondly, participants' used increasingly consistent sets of actions throughout the experiment, which may be the due to a set of subroutines becoming more common with practice. We thus decided to analyze the subroutines that participants used in the assembly experiment, first to assess whether they do become more consistent throughout the experiment, and second to see how they relate to the parts identified in the decomposition experiment.

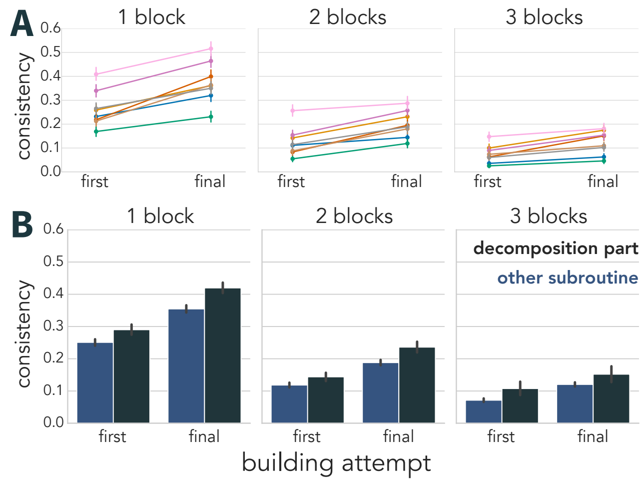

Figure 2: (A) Common subroutines occurred across more participants' reconstructions in final attempts, compared to first (lines are different towers). (B) Those subroutines that were also commonly identified parts were more strongly shared across participants.

Use of subroutines in assembly behavior

To measure the use of subroutines throughout the experiment, we identified the parts of each structure that were most likely to be built over a contiguous sequence of actions, and obtained a reliable measure of their consistency across participants using leave-one-out (LOO) cross validation. We found the 12 most common parts generated by sequences of 1 to 5 block placements, from the reconstructions of $N - 1 = 48$ assembly experiment participants, and recorded whether or not each of these common parts was also constructed (across any sequence of block placements) by the remaining 'left-out' participant. As the frequency of subroutines could possibly change throughout the experiment, common subparts were identified separately for first and final attempts. This process was repeated for all participants, yielding a set of common *subroutines* along with a proportion of left-out participants that used them (e.g. *Figure 1 E*), allowing us to calculate an aggregate measure of the consistency of subroutines used in an attempt, as well as identify which of the most common subroutines were also parts identified in the decomposition task.

To assess whether participants' final reconstructions contained more consistent subroutines than their initial ones, we fit a logistic linear mixed-effects model predicting whether each common subroutine appeared in the building procedure of the left-out participant. We included fixed effects for attempt (i.e., first, final) and number of block placements, an interaction term between the two, random intercepts and slopes for each tower, as well as a random intercept for participant. We found that larger subroutines were less likely to be shared by multiple participants than smaller subroutines ($b = -0.699$, $t = -53.4$, $p < 0.001$), as expected. Nevertheless, participants shared a modest number of subroutines in their initial building attempts: single-block subroutines appeared in 19.8% (95% CI [14.5 26.4]), two-block subroutines in 10.9% (95% CI [7.77 15.1]), and three-block subroutines in 5.74% (95% CI [4.02 8.13]) of attempts (*Figure 2 A*). These results indicate that participants do reconstruct towers by building some of the same parts, even when building the tower for the first time. More revealingly, the subroutines used by different participants were more consistent in final attempts than in first attempts ($b = 0.407$, $t = 4.43$, $p < 0.001$) (*Figure 2 A*). This suggests that the parts identified in the decomposition experiment are built more frequently across the assembly experiment because they make up some of the subroutines that participants build more consistently throughout- a possibility we explore more in the next section. We also found a reliable interaction between number of blocks and attempt ($b = 0.0754$, $t = 4.54$, $p < 0.001$), which may reflect a greater increase in agreement about how to reconstruct larger parts of the towers.

How do common subroutines relate to decomposition behavior?

Commonly identified parts built more than other subroutines To see whether participants were more likely to discover subroutines specifically for constructing parts identified during decomposition, we augmented the subroutine model from above with a binary variable indicating whether or not each subroutine reconstructed a commonly identified part in the decomposition experiment. This version of the model yielded a similar pattern of results as the previous model (attempt ($b = 0.365$, $t = 0.0952$, $p < 0.001$), number of blocks ($b = -0.670$, $t = -49.7$, $p < 0.001$), interaction ($b = 0.0839$, $t = 4.89$, $p < 0.001$)). We found a slight, but statistically insignificant, interaction between attempt and being a common part ($b = 0.0985$, $t = 1.85$, $p = 0.0634$), suggesting that the convergence of subroutines was not primarily driven by convergence to these parts. However, subroutines that were parts were built by more participants across the board ($b = 0.376$, $t = 9.38$, $p < 0.001$) (*Figure 2 B*), suggesting that the way people visually decompose silhouettes plays an important but stable role in the selection of subroutines throughout.

Common subroutines form part of accurate reconstructions Finally, we sought to better characterize the subroutines that become more and less consistent with practice. We visualized subroutines with the greatest increases and decreases in prevelance from first to final attempt *Figure 3*. Several of the greatest decreases in consistency were from subroutines that extended beyond the silhouettes, supporting the idea that convergence in subroutines is partly achieved by participants pruning systematic errors from the space of actions they consider. Furthermore, we found that many of the largest increases in consistency came from subroutines representing individual block placements (rectangles of `1x2`, `2x1`, `2x2`, `2x4`, and `4x2` grid squares), many of which were required to perfectly reconstruct the tower they were attempting.

Finally, prior work showed that the more accurately a participant built a tower in one attempt, the less they would update their procedure in the following attempt (McCarthy et al., 2020), suggesting that subroutines that form part of a successful reconstruction may be less affected by assembly practice. We therefore reran the analysis from the previous section on just the perfectly reconstructed towers in the assembly dataset, removing the part indicator variable. We found that common subroutines from these towers were shared by a greater proportion of reconstructions overall– in first attempts single-block subroutines appeared in: 56.0% (95% CI [0.474, 0.642]) of left-out reconstructions; two-block: 41.4 % (95% CI [33.5, 49.8]); and 3-block: 28.3% (95% CI [21.9, 35.6]). Additionally, and consistent with prior work demonstrating that successful procedures are updated less, the subroutines used to create perfect reconstructions did not become more consistent with practice ($b = 0.0363$, $t = 0.270$, $p = 0.788$), and instead remained at high levels of consistency. This suggests that the convergence we observed in the whole set of reconstructions is mainly driven by participants updating their unsuccessful attempts to include those subroutines used throughout by participants who originally reconstructed the tower successfully– namely those that are necessary to accurately reconstruct the towers.

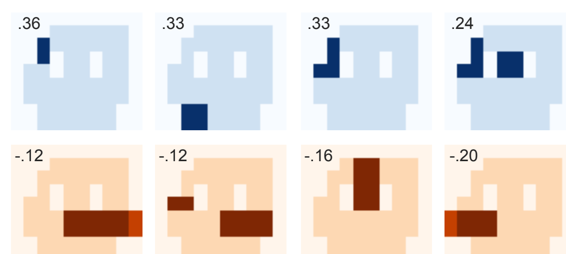

Figure 3: Common subroutines with greatest difference in occurrences between first and final assembly attempts (change in proportion shown top left), for one silhouette. Subroutines that extended beyond the silhouette became less common (orange), and subroutines for accurately recreating the silhouettes became more common (blue).

Discussion

In this paper we presented a paradigm for measuring detailed aspects of decomposition, and explored its relationship to physical assembly behavior. We found a great deal of consistency among the parts identified by participants, but enough variability among entire decompositions to suggest that different decompositions may contribute to differences in reconstruction procedures. We also found that parts that were more frequently identified in the decomposition experiment were more likely to be built by different participants assembling the same towers out of blocks. Moreover, participants assembled towers using increasingly consistent sets of subroutines throughout the experiment and, of these, the most consistent were also identified in the decomposition task.

Our study suggests a relationship between the actions people perform when making an object and the way people perceptually decompose those objects. Illuminating this relationship could have consequences for our understanding of how we rapidly learn to construct new kinds of entities, and more generally for how we utilize the visual system to solve complex problems. The phenomena studied here may also hold implications for algorithmic accounts of abstraction, that have traditionally treated 'state' and 'temporal' abstractions as distinct, as well as raise questions about the content and format of the internal representations that are used to support perception and action.

Our results suggests several avenues for clarifying the underlying relationship between perceptual decomposition and procedural abstraction. Firstly, while we observed a bias towards geometrically simple units that is in line with prior work on perceptual organization, the extent to which our decomposition task is a reliable measure of *perceptual* decompositions is unclear. Evaluating existing models of perceptual segmentation on the decompositions yielded in our task may help disentangle the contributions of perceptual decomposition and the coloring-in task that participants performed. This initial investigation could be augmented to clarify how people represent the parts of an object. For example, where the current study measured decompositions from participants that were naive to the physical structure of the towers, participants that knew that the shapes they were building were subject to gravity, or made from rectangular blocks, may segment them differently. Such studies could reveal the contribution of different kinds of representation on task decomposition. The experiments presented in this paper also provide the tools for exploring how generative experience changes the way people parse and interpret the parts of a physical structure. Such studies may have implications for theories of how perceptual representations and procedural knowledge interact in the mind, as well as the many natural behaviors that rely on a tight coordination between them.

Acknowledgments

Thanks to Felix Binder, and the members of the Cognitive Tools Lab at UC San Diego for helpful discussion. This work was supported by NSF CAREER Award #2047191 to J.E.F.

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

References

Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive psychology*, *63*(4), 173–209.

Austerweil, J. L., & Griffiths, T. L. (2013). A nonparametric bayesian framework for constructing flexible feature representations. *Psychological review*, *120*(4), 817.

Bapst, V., Sanchez-Gonzalez, A., Doersch, C., Stachenfeld, K. L., Kohli, P., Battaglia, P. W., & Hamrick, J. B. (2019). Structured agents for physical construction. *36th International Conference on Machine Learning, ICML 2019*, *2019-June*, 722–732.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., ... Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. , 1–40. Retrieved from `http://arxiv.org/abs/1806.01261`

Bear, D. M., Fan, C., Mrowca, D., Li, Y., Alter, S., Nayebi, A., ... others (2020). Learning physical graph representations from visual scenes. *arXiv preprint arXiv:2006.12373*.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, *94*(2), 115.

Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, *113*(3), 262–280.

Botvinick, M. M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655). doi: 10.1098/rstb.2013.0480

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

Correa, C. G., Ho, M. K., Callaway, F., & Griffiths, T. L. (2020). Resource-rational Task Decomposition to Minimize Planning Costs. *arXiv*. Retrieved from `http://arxiv.org/abs/2007.13862`

De Groot, A. D. (2014). *Thought and choice in chess* (Vol. 4). Walter de Gruyter GmbH & Co KG.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive science*, *42*(8), 2670–2698.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504. doi: 10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: Statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521–537. doi: 10.1037/0096-3445.134.4.521

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. *Perceptual organization in vision: Behavioral and neural perspectives*, *233278*.

Hamrick, J. B. (2019). Analogues of mental simulation and imagination in deep learning. *Current Opinion in Behavioral Sciences*, *29*, 8–16. doi: 10.1016/j.cobeha.2018.12

.011

Huys, Q. J., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., ... Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3098–3103. doi: 10.1073/pnas.1414219112

James, K. H. (2017). The importance of handwriting experience on the development of the literate brain. *Current Directions in Psychological Science*, *26*(6), 502–508.

James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory–motor brain network. *Neuropsychologia*, *44*(14), 2937–2949.

Koffka, K. (1935). *Principles of Gestalt psychology.* Oxford, England: Harcourt, Brace.

Kulkarni, T. D., Whitney, W., Kohli, P., & Tenenbaum, J. B. (2015). Deep convolutional inverse graphics network. *arXiv preprint arXiv:1503.03167*.

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

McCarthy, W., Kirsh, D., & Fan, J. (2020). Learning to build physical structures better over time. *Proc. 42nd Annu. Meet. Cogn. Sci. Soc.*

Miller, G. A. (2020). 11. the magical number seven, plus-or-minus two or some limits on our capacity for processing information. In C. R. Evans & A. D. J. Robertson (Eds.), *Brain physiology and psychology* (pp. 175–202). University of California Press. Retrieved from https://doi.org/10.1525/9780520318267-011 doi: doi:10.1525/9780520318267-011

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. Retrieved from http://dx.doi.org/10.1038/nature14236 doi: 10.1038/nature14236

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2745–2750. doi: 10.1073/pnas.0708424105

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*(4), 441 - 474. Retrieved from http://www.sciencedirect.com/science/article/pii/0010028577900160 doi: https://doi.org/10.1016/0010-0285(77)90016-0

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and brain Sciences*, *21*(1), 1–17.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. M. (2014). Optimal Behavioral Hierarchy. *PLoS Computational Biology*, *10*(8). doi: 10.1371/journal.pcbi.1003779

Sutton, R. S., Precup, D., & Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, *112*(1-2), 181–211.

Tomov, M. S., Yagati, S., Kumar, A., Yang, W., & Gershman, S. J. (2020). *Discovery of hierarchical representations for efficient planning* (Vol. 16) (No. 4). Retrieved from http://dx.doi.org/10.1371/journal.pcbi.1007594 doi: 10.1371/journal.pcbi.1007594

Wertheimer, M. (1938). Laws of organization in perceptual forms.

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

Yildirim, I., Kulkarni, T. D., Freiwald, W. A., & Tenenbaum, J. B. (2015). Efficient analysis-by-synthesis in vision: A computational framework, behavioral tests, and comparison with neural representations. In *Thirty-seventh annual conference of the cognitive science society* (Vol. 4).

Zemlock, D., Vinci-Booher, S., & James, K. H. (2018). Visual–motor symbol production facilitates letter recognition in young children. *Reading and Writing*, *31*(6), 1255–1271.