Title

How does assembling an object affect memory for it?

Permalink

https://escholarship.org/uc/item/71h690dm

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

McCarthy, William P
Anderson, Sean P
Fan, Judith E.

Publication Date

2024

How does assembling an object affect memory for it?

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

Sean P. Anderson
Department of Psychology
Stanford University
seanpaul@stanford.edu

Judith Fan
Department of Psychology
Stanford University
jefan@stanford.edu

Abstract

What impacts what we remember about objects we have just encountered? Influential theories of learning suggest that more active engagement leads to stronger memories than passive observation. However, it is not clear which aspects of interaction lead to stronger memories, nor what kinds of memories are supported by active engagement. Here we conduct several experiments to investigate the impact of assembling an object on subsequent recognition and recall performance. We found that reconstructing a block tower by copying it part-by-part could *impair* subsequent memory for that tower, compared to passively viewing that tower. By contrast, when participants initially encoded each tower by building it from working memory, their subsequent recall was enhanced relative to when they held the tower in working memory without building it. Together our results suggest a complex relationship between the nature of our interactions with objects and our subsequent memories of them.

Keywords: memory; working memory; construction; active learning; encoding specificity; procedural memory

Introduction

To interact with the world in complex ways, we need to remember things about the objects we have interacted with. Sometimes, all we need to remember about an object is whether or not we have seen it before (Brady, Konkle, Alvarez, & Oliva, 2008; Standing, 1973). Other time, we need to remember specific details about our prior interactions. What determines the kinds of information we remember about objects we encounter, and what about our interactions with objects determines how well we remember them?

A substantial body of prior work had found that more *active* forms of encoding, in contrast to more passive observation, lead to stronger memories (Craik & Lockhart, 1972; Bonwell & Eison, 1991; Chi, 2009; Markant, Ruggeri, Gureckis, & Xu, 2016). These findings suggest that people will remember more about objects they actively manipulate, compared to those they just see. Indeed, actively rotating 3D objects does lead to better recognition of those objects compared to passively viewing the same sequence of images (Harman, Humphrey, & Goodale, 1999). Some forms of interaction may be particularly beneficial to memory. Many memory researchers have identified strong mnemonic benefits of *generation*: people are more likely to remember words (Slamecka & Graf, 1978; Bertsch, Pesta, Wiscott, & McDaniel, 2007) and numbers (Crutcher & Healy, 1989) when they have generated them as answers to questions, compared to when those same answers are given to them. These findings suggests that visual memory might also benefit from generative processes, such as altering an object's appearance, or even constructing an object from scratch. Moreover, production of visual objects (i.e. drawing) has been shown to support memory for depicted words and concepts (Fernandes, Wammes, & Meade, 2018; Wammes, Meade, & Fernandes, 2016), however, whether constructing a visual object strengthens memory of the object itself is less clear.

The experience of constructing an object is a complex physical and cognitive act that could impact memory in various ways, from providing more visual exposure, to "deeper" or embodied processing through multiple sensory channels (Craik & Lockhart, 1972; Limata et al., 2023), to practice "retrieving" objects from memory (Schuetze, Eglington, & Kang, 2019; Roediger III & Karpicke, 2006; Rowland, 2014). A unique but perhaps critical aspect of construction is the sequence of transformative actions performed. The procedural learning (Ryle & Tanney, 2009; McCarthy, Kirsh, & Fan, 2020) that occurs during this process may be intimately related to how we visually represent objects (Lake, Salakhutdinov, & Tenenbaum, 2015; Yildirim, Belledonne, Freiwald, & Tenenbaum, 2020). On the other hand, our memory of how an object looks might be entirely independent of our memory of how to build an object, which we may only observe in decoding contexts that leverage that information.

In general, the way in which we probe different kinds of memory may have a critical effect on the results we observe. The standard measure of visual recognition memory– asking whether or not someone has seen a stimulus before– may reveal whether someone has stored some aspect of a stimulus in memory, but not which aspects of the stimulus were used to make those judgements (Brady et al., 2008). Theories of verbal and concept memory distinguish between *recognition* (or "familiarity") and *recall* (Yonelinas, 2002), tests of which are able to provide richer readouts of memory. This had led some researchers to explore *visual production* (i.e. drawing) to provide more detailed insight into the *contents* of visual memory (Bainbridge, Hall, & Baker, 2019). These generative readouts may be especially sensitive to memories formed during construction, by providing a decoding context that is consistent with how the objects are encoded (Godden & Baddeley, 1975; Tulving & Thomson, 1973).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

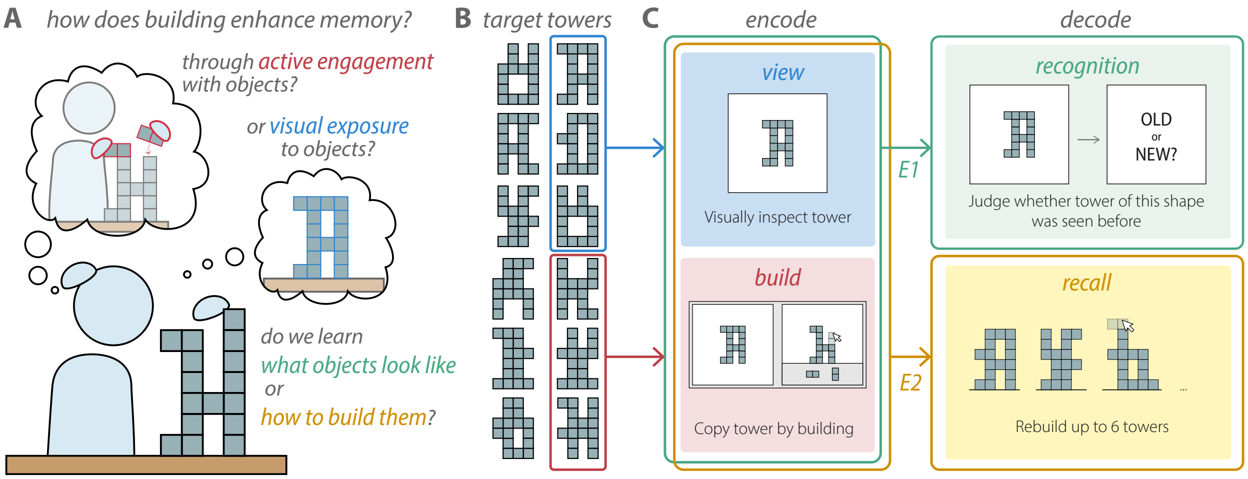

Figure 1: Building might impact memory simply by being more "active", but might also require existing memories to strengthen or elaborate. It could impact our ability to recognize the things we build, or our memories of how to build them (A). Target block towers can be built from 8 blocks (B). 3 towers were assigned to each encoding task (C left). In the *View* task, participants inspected the tower for 15 seconds. In the *Build* task, participants rebuilt the tower. We tested recognition (Experiment 1) by asking participants if they had seen each tower before (top-right); we tested recall (Experiment 2) by asking participants to rebuild each tower from memory (bottom-right).

General Methods

In this paper, we present a series of 4 experiments designed to assess the impact of generative visual encoding tasks on subsequent memory of objects. We use a task domain with objects that can themselves be constructed– 2D block towers– allowing us to compare the impact of generative experience on recognition as well as recall. All experiments reported consisted of an **encoding phase** and **decoding phase**. In each **encoding phase**, each participant viewed 6 block towers that were randomly split between two encoding conditions, *View* and *Build*. Encoding tasks for each condition varied slightly across experiments. In each **decoding phase**, memory of these towers was tested with an assessment of recognition or recall.

Stimuli To design a set of visually homogeneous stimuli that could be generated with distinct sequences of actions, we generated a set of 2D block towers (Fig 1B). Each tower was constructed out of 8 dominoes, 4 horizontal and 4 vertical (i.e. *2x1* and *1x2* blocks), and fit within a *4x6* grid.

Participants Participants (18+ years, from USA and UK) were recruited online using the Prolific platform and were paid approximately $16 per hour for their time (20-30 minutes). For E1 and E2, we recruited participants until 50 participants completed each study without meeting any of our pre-defined exclusion criteria. For E3 and E4, we recruited participants until 50 participants in each group completed the study.

Experiment 1: Impact of building objects on visual recognition

We manually selected a subset of 12 block towers to be shown to all participants (Fig 1B). For each participant, the 12 towers were randomly divided into sets of 6 *target towers* and 6 *foils*. The 6 target towers were randomly split between two conditions– *Build* and *View*– and were all presented in the same color.

Encoding Participants were informed that their memory for the shape of each tower would be tested later in the experiment. All 6 target towers were presented in a psuedorandom order. *View* towers were displayed on screen for 15000ms, and participants were instructed to "study the shape of the tower" for the entire time it is on screen (Fig 1C, upper-left). *Build* towers were presented alongside a building interface: a gridworld environment where blocks could be picked up and placed on any supporting surface by clicking with the mouse. Participants were instructed to "copy the tower" by building it in the environment. Blocks could not be moved once placed, however, the building environment could be reset at any time, and undo/ redo was available. When the participants had perfectly reconstructed the target tower, they automatically proceeded to the next trial (Fig 1C, lower-left).

Decoding Visual recognition memory was measured with an **old-new** task (Fig 1C, upper-right). Participants were presented with the target towers one-by-one, randomly interleaved with foils, and asked to indicate whether they had seen the presented tower in the previous phase by keypress.

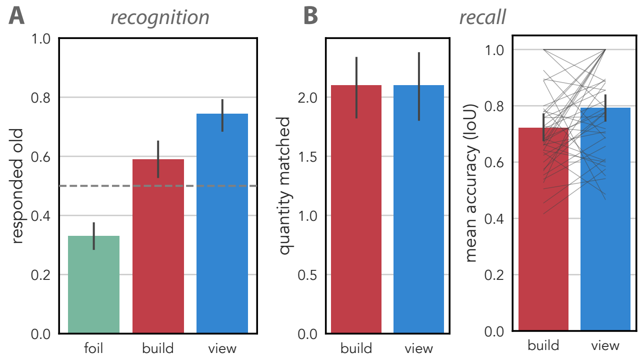

Figure 2: Participants correctly responded 'old' to *View* stimuli more often than to *Build* (A). Participants recalled roughly the same amount of *Build* and *View* towers, and those they did recall were of roughly the same accuracy. Error bars in all plots represent 95% CI.

Results

We excluded 8 participants for incomplete data. To determine whether participants had any ability to discriminate between old and new stimuli, we created bootstrapped distributions of the number of times participants responded "old," separately for target towers and foils (Fig 2A). Distributions and confidence intervals were created by resampling over 1000 iterations; in each bootstrap iteration we sampled participants with replacement and included all data from a participant every time they were sampled. We found that participants responded "old" more often to target towers (0.667, 95% CI : [0.62, 0.708]) than to foils (0.33, 95% CI : [0.283, 0.377]) ($p = 0$), confirming that they could, in general, discriminate between towers they had seen and those they had not.

We also found that participants were more likely to respond "old" to *View* towers (0.743, 95% CI : [0.683, 0.793]) than to *Build* towers (0.59, 95% CI : [0.527, 0.653]) ($p = 0$). This was particularly surprising given that participants took on average 61.1s (95% CI : [60.8, 61.3]) to complete each *Build* trial, far longer than the 15s exposure in the *View* trials. Primarily, however, it conflicts with the prediction that the more active task, building, would lead to stronger memories than the viewing task, which required no overt activity at all.

Experiment 2: Impact of building objects on visual recall

We had several hypotheses about why building a tower might lead to worse memories, however we first sought to establish whether this phenomena was isolated to visual recognition, or extended to other forms of memory. Recall– the ability to bring an item to mind without a related cue– provides an opportunity for participants to share contents of memory, even if it does not reach threshold for visual recognition. Our task domain provides a natural way of testing visual recall: asking participants to build block towers from memory. Furthermore, this decoding context is highly consistent with the context of encoding (i.e. building towers) (Godden & Baddeley, 1975; Tulving & Thomson, 1973), which may give participants the best chance of leveraging kinds of representations learned during building.

Encoding The encoding phase was identical to that of Experiment 1, except that participants performed 2 repetitions of each encoding trial. We increased the number of repetitions as we found in piloting that many participants struggled to recall any towers after a single encoding trial, consistent with prior findings that visual recall demands a stronger memory signal than recognition (Yonelinas, 2002).

Decoding Participants were presented with a building environment almost identical to the one available to them in the *Build* encoding task, minus the target tower. Participants were asked to reconstruct as many towers as they could remember from the previous part of the study, in any order (Fig 1C, lower-right). The experiment ended when a participant submitted 6 towers, or pressed a button indicating that they could not remember any more towers.

Results

We excluded 11 participants for incomplete data. After removing duplicate submissions of towers, participants submitted an average of 4.2 towers (95% CI : [3.7, 4.64]). On average, 1.46 (95% CI : [1.06, 1.84]) of these towers were perfect reconstructions of a target tower, suggesting that accurately recalling towers of this complexity was a difficult task. Fewer *Build* towers (0.56, 95% CI : [0.34, 0.78]) were perfectly recalled than *View* towers (0.9, 95% CI : [0.62, 1.22]) ($p = 0.020$), providing initial evidence that building did not benefit recall memory.

To measure accuracy of the imperfect reconstructions we calculated the "Intersection Over Union" (IoU): the area of overlap between target and reconstruction, divided by the total area covered by both, allowing for horizontal translation. Imperfect reconstructions present a challenge for analysis: how should we identify which target towers participants were attempting to reconstruct? We made an assumption– that each unique tower built in the recall phase corresponded to a genuinely recalled target tower. To map these recalled towers to their intended targets, we calculated the IoU between every reconstruction and target, then found the mapping that maximizes the mean score. We found no reliable difference between the number of towers paired to targets from the *Build* (2.1, 95% CI : [1.82, 2.34]) and *View* (2.1, 95% CI : [1.8, 2.38]) conditions ($p = 0.440$) (Fig 2B). However, we did find that participants who recalled towers from both conditions generally built more accurate reconstructions of *View* condition towers ($p = 0.0208$, Cohen's $d = 0.433$), revealed by a paired t-test between reconstruction means in each condition (Fig 2C).

In sum, these results point to a moderate recall advantage for towers in the *View* condition, compared to *Build*, which is

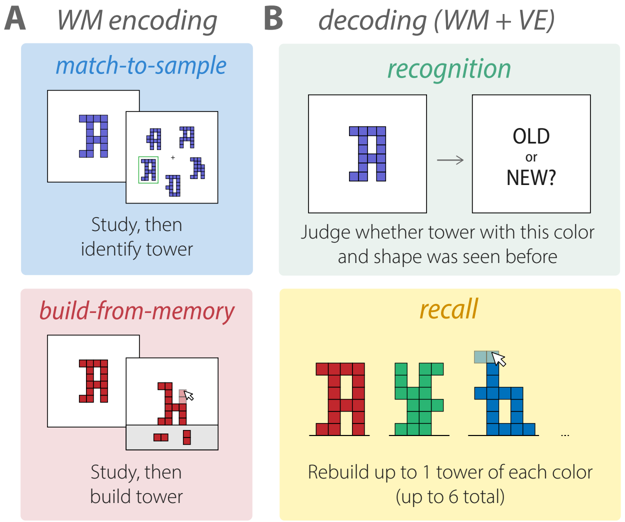

Figure 3: Experiments 3 and 4 compared the effect of our original encoding tasks with two new encoding tasks (A). In both tasks, participants studied a tower until it disappeared. They were then asked to either identify the tower from a group (top) or to rebuild the tower (bottom). Different colors were used for each tower, allowing us to measure recognition and recall of specific towers (B).

also at odds with the prediction that more active engagement leads to stronger memories. This also happened despite the highly similar encoding and decoding contexts in the *Build* condition, suggesting that if generative encoding can actually benefit visual memory, something about the generative experience our participants are engaging in is failing to induce this effect, or is interfering with memory in some way.

Experiment 3: Impact of building from working memory on visual recognition

Why did participants not remember the towers they built better than the ones they viewed? Much of the prior work demonstrating mnemonic benefits of "generation" investigates processes of reconstructing or generating an example or word from memory or from an internal thought process. Retrieval from an internal representation may serve to reinforce prior representations through retrieval (Schuetze et al., 2019; Roediger III & Karpicke, 2006; Rowland, 2014; Fan & Turk-Browne, 2013), or link these representations to novel experiences. Our building task, in contrast, asks people to copy an object that already exists in the world, meaning it could in principle be completed without any holistic representation of the object. If, for example, participants reconstructed towers by iteratively determining which one block should be placed next, they may have never associated their actions with a representation of what any particular tower looked like. Moreover, if participants learned that they only needed to attend to individual blocks, they may have stopped attending to the entire the tower.

In Experiments 3 and 4, we aimed to test whether a pre-existing visual memory is a prerequisite for a mnemonic advantage of building. We introduced two new encoding tasks that each required participants to hold a representation of an entire tower in working memory before performing some an adapted *Build* or *View* task. Similarly to Experiments 1 and 2, Experiment 3 tests visual recognition and Experiment 4 tests recall.

Stimuli For each of the target towers used in Experiments 1 and 2, we derived a set of 5 *distractors* by performing the following transformations: horizontal flip, vertical flip, 180 degree rotation, lower half swapped with upper half, and left half swapped with right half. We sampled one of these distractors to act as the *foil* in the old-new decoding task. The remaining 5 became distractors in the *match-to-sample* encoding task, described below. We randomly sampled sets of 6 target towers until all of the target towers and derived distractors were distinct, and presented this set to all participants in Experiments 3 and 4. Each target tower and its corresponding distractors were assigned one of six colors. As with Experiments 1 and 2, target towers were randomly split between *Build* and *View* conditions for each participant.

Encoding Participants in the **Visual Exposure** group performed the same *Build* and *View* tasks from Experiments 1 and 2. Participants in the **Working Memory** group performed *modified Build* and *View* tasks that required participants to visually encode each tower before responding. Prior to each Working Memory task, the target tower was displayed on screen for 8000ms and participants were prompted to "study" the shape of the tower. Then, for towers in the *View* condition, participants performed a **match-to-sample task**: they were presented with a centered fixation cross, followed by a circular array of 5 towers– the 4 sampled distractor towers plus the target tower. Participants were instructed to select the tower they had just studied by clicking on it, after which they received feedback. For towers in the *Build* condition, participants performed a **build-from-memory task**: they were presented with an empty building environment, with blocks in the same color as the tower they had just viewed, and prompted to build the target tower from memory. They could submit a tower once they had placed 8 blocks. They received feedback after submission (correct or incorrect), and the target tower was revealed in an adjacent window to allow comparison with their reconstruction.

Decoding Experiment 3 used the same old-new task from Experiment 1, except that participants saw two trials of each color: one *target* and the randomly sampled *foil* generated from that target.

Results

We excluded 11 participants for failing to complete all trials, leaving 50 in each group. We first analyze performance in the

Working Memory encoding phase. In the *match-to-sample* task, participants correctly selected the target tower from the 5 distractors on 91.5% of trials (95% CI : [86.3, 95.8]), suggesting that they successfully encoded the target towers in working memory. In the *build-from-memory* task, participants perfectly reconstructed the target tower on 73.3% of trials (95% CI : [0.688, 0.774]), consistent with this being a more difficult task.

As with Experiment 1, the Visual Exposure group responded "old" to target towers (0.807, 95% CI : [0.76, 0.853]) more often than to foils (0.29, 95% CI : [0.243, 0.34]) ($p = 0$). However, while *View* towers (0.833, 95% CI : [0.753, 0.9]) were remembered marginally more often than *Build* (0.78, 95% CI : [0.713, 0.847]) ($p = 0.173$), we did not see a reliable difference between responses (Fig 4A left). Convergence between conditions may have been driven by ceiling effects, as the introduction of colors and increased number of repetitions did appear to result in stronger recognition performance overall (75.8% correct, 95% CI : [71.8, 79.7]), relative to Experiment 1.

This explanation is supported by the fact that in the Working Memory condition, where participants' responses were marginally more accurate again (80.1% correct, 95% CI : [76.3, 83.8]), the difference in responses between *Build* (0.88, 95% CI : [0.827, 0.927]) and *View* (0.873, 95% CI : [0.827, 0.92]) was even less distinct ($p = 0.565$) (Fig 4A right). In sum, we find no evidence that building from working memory reliably led to better or worse recognition.

Experiment 4: Impact of building from working memory on visual recall

Finally, we asked whether building from memory impacts visual recall. We introduced block towers of different colors to provide a way of inferring which towers participants were attempting to recall, as well as provide an additional channel by which participants could discriminate between towers.

Encoding The encoding phase was identical to the encoding phase in Experiment 3.

Decoding As in Experiment 2, participants were presented with an empty building environment, and asked to recall as many towers as they could remember from the encoding phase. This time, however, participants first had to select the color of the tower they wanted to build. Once they had placed 8 blocks of their chosen color, then pressed a button to submit their tower and remove that color as an option. The experiment ended when a participant submitted towers of all 6 colors, or pressed a button indicating that they could not remember any more towers.

Results

We excluded 6 participants for failing to complete all trials, and 1 failing to start the decoding task within 10 minutes of finishing the encoding task, leaving 50 participants in each group.

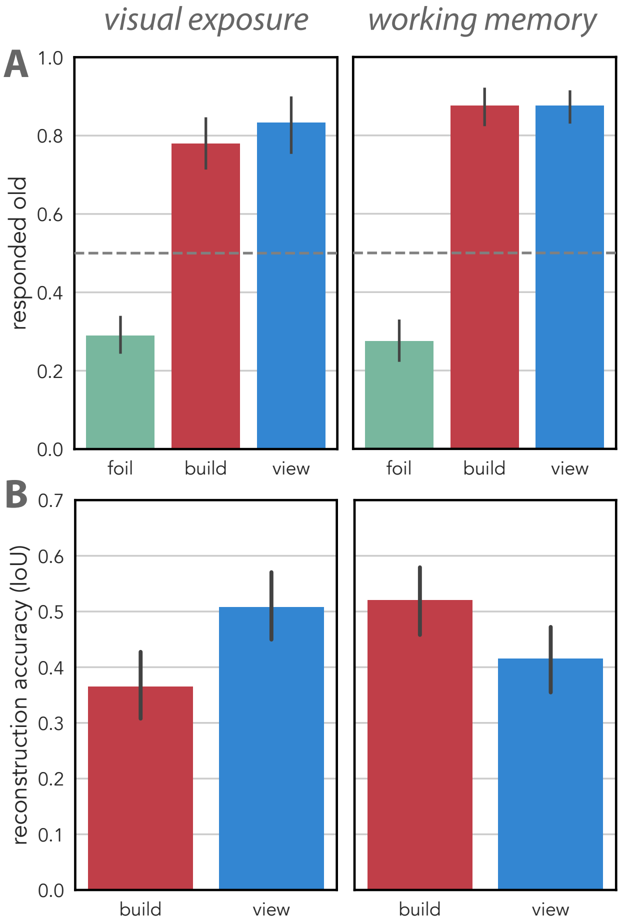

Figure 4: Recognition performance was similar for *Build* and *View* (A left), regardless of whether the tower was encoded in working memory (A right). As in Experiment 1, participants *recalled* towers they viewed more accurately than towers they built (B left), unless those towers were first encoded in working memory (B right).

Similarly to Experiment 3, the Working Memory group correctly selected the target tower on 86.7% of *match-to-sample* trials (95% CI : [81.3, 91.7]), and perfectly reconstructed the target tower on 73.9% of *build-from-memory* trials (95% CI : [0.7, 0.778]).

Participants submitted towers on 3.78 towers on average (95% CI : [3.44, 4.11]). The colors of recalled towers provided a mechanism for us to match recalled towers with their intended target stimuli. To compare how different encoding tasks affected recall memory, we fit a mixed-effects logistic regression predicting whether or not a participant submitted a perfect reconstruction of the target tower. We included fixed effects for encoding group (*Visual Exposure* vs. *Working Memory*), encoding context (*Build* vs. *View*), and their interaction; plus random intercepts for participant and tower.

We found no evidence that the *Working Memory* tasks reliable led to a better or worse ability to perfectly recall towers ($b = -0.595$, $z = -1.35$, $p = 0.177$). While we did see evidence for a main effect of encoding context, where *Build* towers were recalled less frequently than *View* ($b = -0.879$, $z = -2.52$, $p = 0.0117$), this effect was small compared to a reliable crossover interaction between encoding task and context ($b = 1.62$, $z = 3.33$, $p < 0.001$): *Build* towers were recalled *more* often than *View* towers when encoded in the *Working Memory* tasks. That is, we see evidence for stronger memories of built towers than viewed towers when building follows prior encoding of the tower.

To verify this finding, we fit a model of the same structure, predicting the *accuracy* of each reconstruction for every target tower, treating towers that were not reconstructed as $IoU = 0$. Again, we found no reliable effect of encoding condition ($b = -0.09261$, $t = -1.58$, $p = 0.116$), a small negative main effect of the *Build* condition ($b = -0.143$, $t = -3.00$, $p = 0.00346$), and a crossover interaction ($b = 0.247$, $t = 3.67$, $p < 0.001$) suggesting that *Build* towers were recalled more accurately than *View* towers in the Working Memory condition (Fig 4 B) (and less in the Visual Exposure condition). Together, these results suggest that building a tower from working memory facilitates visual recall, relative to simply viewing a tower.

Discussion

We asked how generating block towers impacts our subsequent memory of them. We initially compared memory for block towers that participants copied with block towers that they simply viewed on screen, and found that the towers people copied were recognized less frequently and recalled less accurately. We suspected that building block towers while they were still on screen prevented participants from forming holistic representations of them, and that these might be critical for generation to facilitate memory. Consistent with this interpretation, we found that when participants built towers from working memory, they did remember them better later on. Moreover, this relative memory boost was only apparent in visual recall, not visual recognition, suggesting that generative experience impacted some but not all aspects of memory for the object.

Our work has implications for the applicability of active and generative learning to visual memory (Markant et al., 2016; Slamecka & Graf, 1978; Crutcher & Healy, 1989). It suggests that more active engagement does not necessarily translate to better memory of a visual stimulus– that the kind of engagement matters. Our finding that building from memory supports recall but not recognition, as well as hinting at distinct processes underlying these two forms of memory (Yonelinas, 2002), suggests that active engagement differentially affects different kinds of memory. Why is recall prioritized in this way? A possible reason is suggested by theories of situated cognition (Roth & Jornet, 2013), that have long stressed that internal representations do not always present the most efficient solution to a cognitive problem: why remember what something looks like when you can easily check by looking? Actions are not perceivable in this way, making it more worthwhile to dedicate cognitive resources to remembering them.

Another key question raised by our study is how building from working memory leads to stronger memories. One possibility is that building from memory requires a large volume of queries of working memory, consolidating any pre-existing representations in longer-term memory through retrieval practice (Schuetze et al., 2019; Roediger III & Karpicke, 2006; Rowland, 2014). Alternatively, generative experience may result in a distinct *kind* of action-based representation, akin to procedural knowledge or “knowledge how” (Ryle & Tanney, 2009; Anderson, 2013). Such representations may elaborate on existing perceptual representations, facilitating processing at a deeper level (Craik & Lockhart, 1972; Bradshaw & Anderson, 1982), or simply constitute a distinct memory trace that can be accessed in future generative contexts. Our results do provide one reason to be skeptical of additional memory formats– a seemingly limited capacity to recall objects. Participants in the Working Memory group did not, in general, recall more towers than the Visual Exposure group, suggesting that the build from memory task served to prioritize memory for certain towers above others, more so than it did to boost memory strength overall.

Our study also raise the question of how goals at encoding time affect memory. We chose not tell participants which Working Memory task they would perform until the stimulus they were encoding had disappeared. However, goals guide visual attention and attention is crucial for determining what gets encoded in memory (Chun & Turk-Browne, 2007). A straightforward way to test whether goals at encoding time impacted memory would be to tell people in advance what task they will perform, potentially cueing different ways of seeing (Goodwin, 2015) and leading to measurable memory effects downstream.

Finally, the hierarchical structure of our stimuli raises the possiblity of relating fine-grained differences in encoding behavior to downstream memory. One well documented strategy for remembering something is to break it down into memorable “chunks” (Miller, 1956; Chase & Simon, 1973; Orbán, Fiser, Aslin, & Lengyel, 2008), a process that may have occurred implicitly as participants built towers. By analyzing the kinds of errors participants made, we may be able to identify subtowers that they did remember, even when they failed to remember the entire tower. Doing so may help to shed light on the structure of the representations used to support visual recognition and recall (Yonelinas, 2002), and tease apart the impact of generative experience on these representations.

Acknowledgements

Thanks to the members of the Cognitive Tools Lab at Stanford University and University of California San Diego for helpful discussion. This work was supported by an NSF CAREER

Award #2047191 to J.E.F.

Data and code availability

All experimental materials, data, and analysis code are available at `https://github.com/cogtoolslab/zipping`.

References

Anderson, J. R. (2013). *The architecture of cognition*. Psychology Press.

Bainbridge, W. A., Hall, E. H., & Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature communications*, *10*(1), 5.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & cognition*, *35*, 201–210.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom. 1991 ashe-eric higher education reports.* ERIC.

Bradshaw, G. L., & Anderson, J. R. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 165–174.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

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

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73–105.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current opinion in neurobiology*, *17*(2), 177–184.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671–684.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 669.

Fan, J. E., & Turk-Browne, N. B. (2013). Internal attention to features in visual short-term memory guides object learning. *Cognition*, *129*(2), 292–308.

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*(5), 302–308.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325–331.

Goodwin, C. (2015). Professional vision. In *Aufmerksamkeit: Geschichte-theorie-empirie* (pp. 387–425). Springer.

Harman, K. L., Humphrey, G. K., & Goodale, M. A. (1999). Active manual control of object views facilitates visual recognition. *Current Biology*, *9*(22), 1315–1318.

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

Limata, T., Bucciarelli, M., Schmidt, S., Tinti, C., Ras, I. N., & Ianì, F. (2023). Action and posture influence the retrieval of memory for objects. *Memory*, *31*(5), 652–664.

Markant, D. B., Ruggeri, A., Gureckis, T. M., & Xu, F. (2016). Enhanced memory as a common effect of active learning. *Mind, Brain, and Education*, *10*(3), 142–152.

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. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

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

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249–255.

Roth, W.-M., & Jornet, A. (2013). Situated cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(5), 463–478.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432.

Ryle, G., & Tanney, J. (2009). *The concept of mind*. Routledge.

Schuetze, B. A., Eglington, L. G., & Kang, S. H. (2019). Retrieval practice benefits memory precision. *Memory*, *27*(8), 1091–1098.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and Memory*, *4*(6), 592.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, *69*(9), 1752–1776.

Yildirim, I., Belledonne, M., Freiwald, W., & Tenenbaum, J. (2020). Efficient inverse graphics in biological face processing. *Science Advances*, *6*(10), eaax5979.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441–517.