The Relation between Gist and Item Memory Over a Month

Haiyun (Tima) Zeng (hzen@sas.upenn.edu)
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Alexa Tompary (atompary@sas.upenn.edu)
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Anna C. Schapiro (aschapir@sas.upenn.edu)
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Sharon L. Thompson-Schill (sschill@psych.upenn.edu)
Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, United States

Abstract

Memory requires both individuation of specific episodes as well as extraction of gist across related experiences. This study developed a spatial memory paradigm to track changes in item memory (memory for specific locations) and gist memory (estimate of the center of the locations) across a period of a month, and to measure the relation between these two forms of memory. We found that item memories decayed compared to gist memory after a month, yet there was a positive relationship between the two forms of memory that persisted. Moreover, item memories were biased towards gist memory only after a month. These findings together indicate that gist memory, initially extracted from item memories, gradually develops into a stable representation that can guide item memory retrieval over longer durations.

Keywords: Consolidation; gist memory

Introduction

How learners extract summary information across individual items is a fundamental question to psychology. Much human and animal work in long-term memory has demonstrated that this 'gist' memory persists or even improves over time while memory for the individual items from which it is built fades (Posner & Keele, 1970; Richards et al., 2014).

How the representation of gist is computed and preserved over time as memory for individual items fades remains unclear. On the one hand, gist memory may become independent of item memory. A persisting gist memory with less accurate item memory is often considered as evidence that gist becomes independent of individual item representations as it is abstracted during encoding (Posner & Keele, 1970) or through consolidation (Richards et al., 2014). Under this account, item memories no longer have an impact on gist representation.

On the other hand, it has been argued that lower accuracy for item memories is not sufficient evidence for the independence of gist (Alvarez, 2011; Squire, Genzel, Wixted, & Morris, 2015). Even when item memories become noisy and less accurate, they still can retain some information that can support a relatively intact gist memory at retrieval.

Disentangling the two possibilities using existing evidence is hard, since previous paradigms often cannot directly measure the information item memories can contribute to gist memory. In this work, we aimed to develop a paradigm to test item memory, gist memory, and a 'predicted' gist memory based on item memory. Ensemble perception research inspired us in developing such a paradigm. Studies of rapid perception of complex visual arrays reveal precise representations of gist (i.e., ensemble statistics) with less accurate item memory retrieval in working memory (Ariely, 2001). Whether item representations are still necessary to estimate the gist at test is also of critical interest for perception researchers (Alvarez, 2011). Ensemble perception paradigms often operationalize the gist as the average representation across instances, and this 'predicted' gist can be computed based on performance on individual items to compare with participants' reported gist memory. In our paradigm, we adapted this approach to study the relation between gist memories and item memories over the course of long-term memory consolidation, using participants' reported gist and an estimate of their 'predicted' gist based on the average, or center, of their item memories. The accuracy of predicted gist may thus reveal the accuracy of gist information preserved in item memories. A positive correlation between predicted and reported gist accuracy could mean that participants' reported gist is still supported by individual item memories, or that the reported gist is influencing the retrieval of items.

To probe the direction of this relationship, we developed a gist-based bias measurement, an approach borrowed from research on hierarchical clustering models that reveals how much gist memory influenced memory for specific items (Brady & Alvarez, 2015). Gist memory has been shown to have an influence on item memory retrieval both in long-term memory of existing semantic categories (Huttenlocher, Hedges, & Vevea, 2000; Tompary & Thompson-schill, 2019) and in ensemble perception (Brady & Alvarez, 2011). Theories suggest that this influence reveals a reconstructive memory retrieval process (Brady, Schacter, & Alvarez, 2015; Hemmer & Steyvers, 2009; Schacter, Guerin, & Jacques, 2011) that depends on the relative strength of item and gist memories (Tompary, Zhou, & Davachi, 2020). Consistent with this theory, as the strength of gist memory preserves or improves and/or that of item memory decreases, gist memory has a larger influence on item memory retrieval (Tompary, Zhou, & Davachi, 2020; Richter, Bays, Jeyarathnarajah, & Simons, 2019). Therefore, an increasing bias of item memories towards the gist would be strong evidence for the increasing strength of gist representation over item memories.

In this experiment, we aimed to understand the relation between item and gist memory over the course of a month. We trained three groups of participants on spatial locations of six landmarks, and we measured the change in memory of these items as well as the estimated average location (i.e., the gist of the encoded information) at three delay periods: 24 hours, 1 week and 1 month. We predicted that gist memory would persist or improve despite item memories becoming less accurate over a month, consistent with prior work. We extend prior observations by including two new measures—predicted gist and gist-based bias—in order to explore how the representations of item memory and gist memory, and the relation between them, change over the course of one month.

Methods

Stimuli

Item memories were operationalized as six "landmarks", i.e. dots of different locations associated with landmark names on a laptop screen (Figure 1). The gist was defined as the center of these landmarks.

The gist was never presented to participants and the individual landmarks were never presented together at the same time. The locations were the same for each participant, but the mapping between the location and landmark name was randomized.

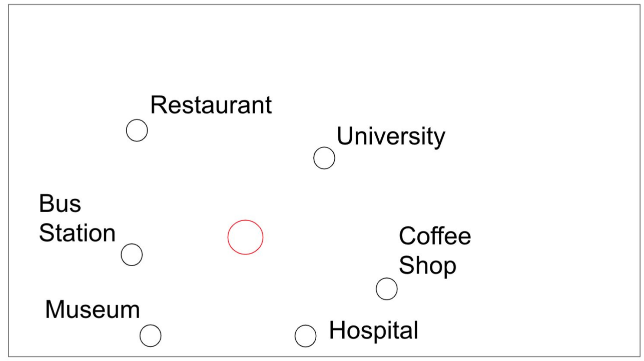

Figure 1. An illustration of the 6 landmarks. Black circles indicate individual landmark locations. Red circle indicates the 'gist' location, i.e. the center of all landmarks.

Participants

We recruited 130 members of the University of Pennsylvania community (18-30 years old; normal or corrected to normal vision) to participate in the experiment for monetary compensation. Participants signed up for a second session that followed their first session by either 24 hours (N = 44), one week (N = 43), or one month (N = 43). We excluded data from 17 participants because of either low performance on Session 1 (i.e. reported gist was out of the scope of the learned landmarks) or individual and gist performance of any sessions lower than 3 *SD* below average.

Experimental Procedure

The experimental procedure is displayed in Figure 2. All participants completed Session 1 and 2; the only difference between groups was the time delay between sessions.

During Session 1, participants were trained to retrieve six landmark locations consecutively on a laptop until their retrieval error for each landmark was fewer than 50 pixels in any direction. After participants reached the training criterion, they completed ten unrelated arithmetic tasks, in order to eliminate potential influences from working memory. Finally, participants were tested for their memory of the locations: They indicated their guess about the center of the landmarks (gist memory test) and then separately recalled each landmark location (individual memory test).

After 24 hours/1 week/1 month, participants returned for Session 2. Session 2 was identical to the testing phase of Session 1, with a gist test followed by an item memory test. This testing order was chosen to reduce the influence of item memories on gist estimation.

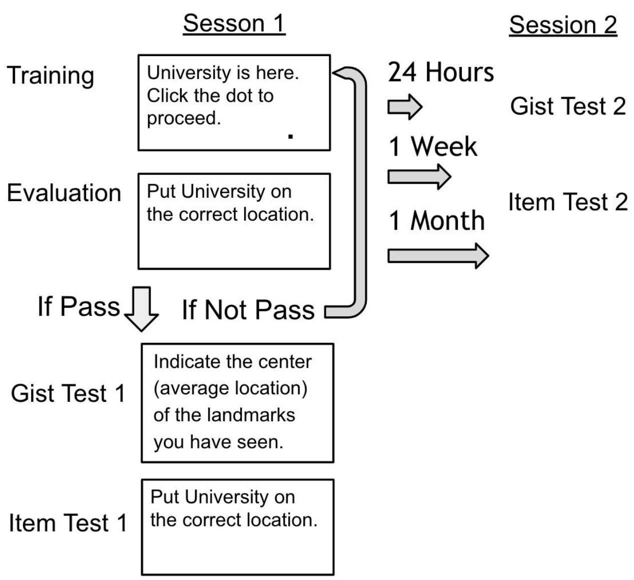

Figure 2. The procedure of the experiment.

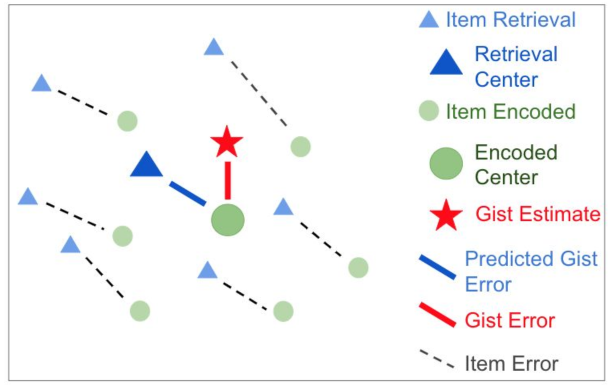

Figure 3. Error Measurements. Dashed lines indicate item memory errors. Each participant's item memory error is operationalized as the average of the dashed lines. Blue solid line (retrieval center - encoded center) indicates error of the predicted gist. Red solid line (gist estimate - encoded center) indicates error of the reported gist memory.

Analysis

Error

In order to test the accuracy for item memory (memory for each landmark), gist memory (reported memory for the center of the landmarks), and predicted gist (estimated center of the landmarks using item memory), we developed three error measurements as follows.

Item Memory Error. The error for item memory was defined as the Euclidean distance between the retrieved location for each landmark and its encoded location (dashed lines in Figure 3). Each participant's item memory error was computed as the average error for the six landmarks. Greater error indicates lower accuracy.

Gist Memory Error. The error in gist memory was defined as the Euclidean distance between the participant's reported center estimate and the true center of all the encoded landmark locations, that is, where the real center should be (red line in Figure 3). Greater error indicates lower accuracy of gist estimate.

Predicted Gist Memory Error. The error for what we are calling the predicted gist was defined to be the Euclidean distance between the center of each participant's retrieved locations and the center of all encoded locations (blue line in Figure 3). In other words, the predicted gist can be thought of as an estimate of what the participant's gist memory would be if that memory is directly computed by averaging across all retrieved locations.

Statistics. In order to test how the measures described above changed over time, we subtracted the Session 1 values from the Session 2 values for each error measure. To test whether gist memory persisted when memory for items faded over time, we entered the change in error for item and gist memory for each group (i.e. 24 hours, 1 week, 1 month) into an aligned ranks transformation ANOVA (group X memory type) and used two-tailed Wilcox tests for between-group comparisons, since change in error was not normal as determined by a Shapiro-Wilk test.

Item - Gist Memory Relationship

In order to test the possible dependence of gist memory and item memory at retrieval, we conducted a Spearman correlation between the predicted gist and reported gist error for Session 2 of the three delay groups (i.e. 24 hours, 1 week, 1 month). Since the accuracy of predicted gist is a measurement for the accuracy of gist information preserved in item memories, a positive correlation between predicted and reported gist error could mean that participants' reported gist is still supported by individual item memories, or that the reported gist is influencing the retrieval of items.

Bias Controlling for Error

In order to examine the influence of gist on item memory, we developed a bias measurement as follows. Since the error analysis revealed an increase in gist memory error after a month, we initially used the reported gist as the center for bias analysis. However, we also report bias using the true gist (center of encoded items) for consistency with common practices in ensemble perception research (Brady & Alvarez, 2011; Lew & Vul, 2015).

True Bias. The bias towards the gist for each retrieved landmark was defined as the relative difference in distance between a participant's gist estimate and each landmark's encoded location versus each landmark's retrieval location. This relative difference was then divided by the error for that landmark: (Encoded - Gist) - (Retrieval - Gist) / (Encoded - Retrieval) (Figure 4).

Bias > 0 indicates that item memory is biased towards the gist while Bias < 0 indicates that individual retrieval is biased away from gist. Each participant's bias was computed as the average across the biases of the 6 landmarks.

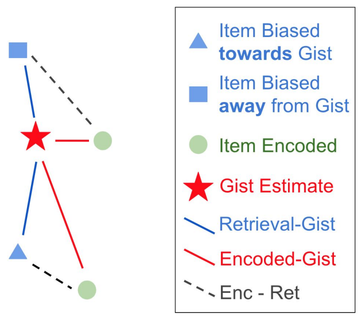

Figure 4. Bias Measurement. The true bias for each location is (red - blue) / dashed. The blue square is an example of an item biased away from gist and the blue triangle is an example of item biased towards gist.

Baseline Bias. When item memory error increases, the increased error has a higher chance of producing a negative bias value despite no real bias away from the center of the landmarks (see Figure 5 for an illustration).

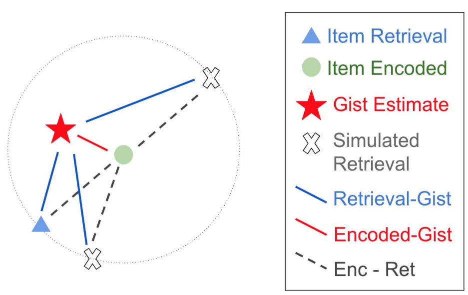

Figure 5. An illustration of a large error causing negative bias for a retrieved location that is not meaningfully retrieved farther from the center relative to its encoded location. Since bias = (red - blue) / dashed line for each location, the bias value will likely be negative, even when retrieval is not biased away from the gist (the blue triangle).

Therefore, to control for the possible influence of error magnitude on our estimate of bias, we conducted a simulation to generate a baseline. We generated 1000 simulated participants for each participant. Each simulation consisted of six simulated retrievals, corresponding to the six landmark locations. For each location, we randomly generated a retrieved location based on the participant's true error for this specific location, allowing angle to vary randomly across the simulations (Figure 5; gray cross). If a simulated location fell out of the boundaries of the screen, the algorithm generated a new location. The bias value for each of the 1000 simulations was the average across each simulation's six retrievals. The baseline bias for each participant was the average across the 1000 simulations.

Bias = True Bias - Baseline Bias. We subtracted the simulated bias from participants' true bias, which resulted in a bias measure that controls for error. Bias > 0 indicates that item retrieval is biased towards the gist, controlling for item memory error.

Statistics. In order to test how the bias described above changed over time, we subtracted the Session 1 values from the Session 2 values for the bias error measure and entered the change in error for each group (i.e. 24 hours, 1 week, 1 month) into an ANOVA, and used two-tailed t-tests for between-group comparisons.

Results

Accuracy of item and gist memory over time. To test the change of accuracy in item memory compared to gist memory, we conducted a 3 (group: 24-hour, 1-week, and 1-month) X 2 (memory type: item, gist) ANOVA. This test revealed a main effect of group, $F(2, 252) = 36.29$, $p < .001$, memory type, $F(1, 252) = 63.16$, $p < .001$, and an interaction between group and memory type, $F(2, 252) = 18.02$, $p < .001$. This interaction indicates that item memory decreased more over time compared to gist memory (Figure 6). Specifically, whereas each pairwise comparison between groups was significant for item memory (all p's < .01), the only reliable group difference for gist memory was between the 24-hour and 1 month groups ($Z = 685$, $p = .026$; uncorrected for multiple comparisons).

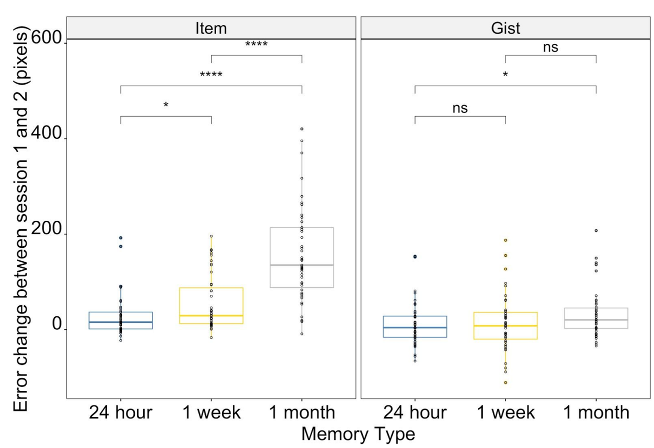

Figure 6. Change in error by group and memory type. Greater values indicate an increase in error in Session 2 over Session 1. * indicates $p < .05$ and **** $p < .0001$.

Relationship between item and gist memory. Even though we found that item memory significantly decreased over time compared to gist memory, it is possible that item memory with lower accuracy could still support a relatively intact gist memory. To explore the possible dependence between item and gist memory, we used a linear model to predict the error in reported gist using the error in predicted gist on Session 2 for 24-hour, 1-week, and 1-month. As a reminder, our measure of predicted gist was computed using the retrieved locations of each landmark, therefore despite item memory becoming inaccurate, the predicted gist from item memory could still have high accuracy.

We found a significant positive relationship between reported gist error and predicted gist error from item memories at 24 hours delay ($rs(42) = .41$, $p = .006$) and at 1 month delay ($rs(41) = .37$, $p = .014$) (Figure 7).

The result indicates a strong relationship between error in the predicted gist—gist calculated from individual item retrieval—and error in directly reported gist over time.

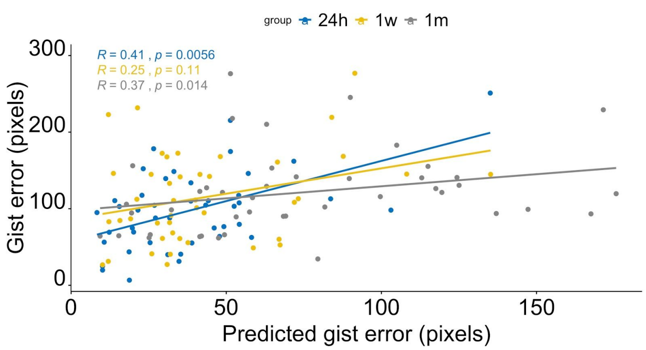

Figure 7. Predicted gist error positively predicted gist memory error 24 hours and 1 month after learning.

Bias towards gist. We found that gist memory continued to correlate with the 'gist' computed from item memory over a long delay of time. The correlation could mean that participants' gist estimate was still supported by individual item memories even after a month, or that an intact, separable gist representation was influencing the retrieval of items.

To examine the direction of the relation between gist and item retrieval, we conducted a 1 (Session 2 - Session 1 change in bias) X 3 (group) ANOVA to examine if participants' item memory became more biased towards gist over time. This test revealed a main effect of group, $F(2, 127) = 5.12$, $p = .007$. This main effect reflected greater change in bias between the 24-hour and 1-month groups, $t(68.77) = 2.266$, $p = 0.027$, and also between the 1-week and 1-month groups $t(70.51) = 2.755$, $p = 0.007$, but not between the 24-hour and 1-week groups, $t(84.72) = -0.706$, $p = .482$ (Figure 8). Furthermore, only the change in bias for the 1 month group was significantly greater than 0, $t(42) = 2.92$, $p = .006$.

To be consistent with common practices in ensemble perception research (Brady & Alvarez, 2011; Lew & Vul, 2015), we then conducted the same bias analysis using the true gist (center of encoded items) and found the same pattern, with a main effect of group, $F(2, 127) = 10.57$, $p < .001$, a greater change in bias between the 24-hour and 1-month groups, $t(71.91) = 3.378$, $p = 0.001$, and also between the 1-week and 1-month groups $t(75.49) = 3.967$, $p < 0.001$, but not between the 24-hour and 1-week groups, $t(84.09) = -0.879$, $p = .382$.

These results suggest that item memory became more biased towards gist memory only after 1 month.

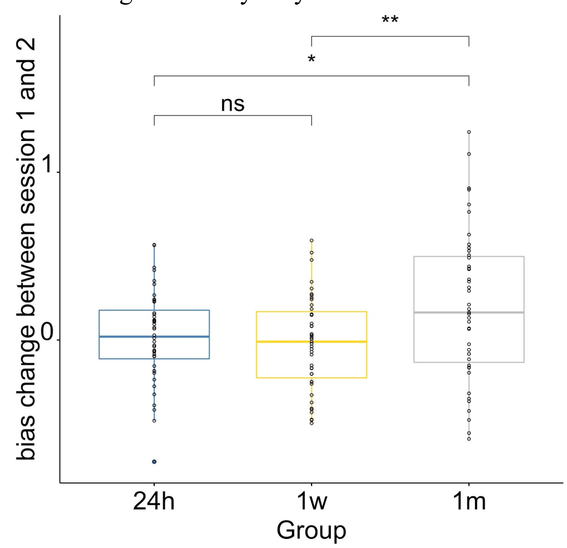

Figure 8. Change in items' bias towards the gist for the 24-hour, 1-week, and 1-month groups. Values > 0 indicate item memory was more biased towards gist in Session 2 relative to Session 1. * $p < 0.05$ and * $p < 0.01$.

General Discussion

We systematically examined the relation between item and gist memory across a month by comparing how memory for six landmarks changed after differing delays between encoding and retrieval: 24 hours, 1 week and 1 month.

Consistent with our prediction (and with prior research; Lutz, Diekelmann, Hinse-Stern, Born, & Rauss, 2017; Posner & Keele, 1970), item memory became less accurate over time while gist memory remained relatively intact. Our results add to this work by showing a gradual decrease from a day to a month.

Surprisingly, the relationship between item memory and the gist persisted across delay periods despite decreased accuracy in item memory. This relationship could have resulted from the influence of gist memory on the retrieval of item memory, from the influence of item memory on gist memory, or both. Although the relationship is correlational, our bias results shed light on the direction of this relationship: Item memory retrieval was biased towards gist only after 1 month, but not after 1 day or 1 week.

Our bias results suggest that gist representation only started to influence item memory retrieval after 1 month. Therefore, at 24 hours, the correlation between predicted and reported gist memory may be because item memory still influences gist representation; in contrast, after 1 month, the correlation appears to be the result of gist memory guiding item memory retrieval.

Although the task explicitly required participants to estimate gist during learning, their gist estimate still seems to be influenced by item memories at retrieval 24 hours after learning. This adds to previous work by raising the possibility that regardless of whether gist takes on an independent representation at encoding, its representation may still depend on item memories at retrieval (Posner & Keele, 1970). In addition, this result is consistent with the theory that the consolidation of gist takes more time compared to the consolidation of individual items (Lutz, Diekelmann, Hinse-Stern, Born, & Rauss, 2017).

Our findings that items are increasingly biased towards the gist as the item memories decrease are consistent with a memory reconstruction framework, which describes memory retrieval as dependent on memories' relative strength and uncertainty (Brady, Schacter, & Alvarez, 2015; Hemmer & Steyvers, 2009; Tompary et al., 2020). This closely parallels models of hierarchical clustering in visual working memory, which also suggest that items are more biased towards their cluster center as uncertainty increases (Brady & Alvarez, 2011; Lew & Vul, 2015; Orhan & Jacobs, 2013). Different from earlier data (Brady & Alvarez, 2011), in our study, the bias towards the gist only appeared at a 1-month delay. This difference may be because the intensive training at session 1 increased the strength of item memories relative to gist memory. In contrast, after 1 month, the uncertainty of item memory increased and therefore its retrieval started to depend on gist memory, which remained stable. Our results enrich the existing work by characterizing the qualitative change of strength of item and gist memories as well as their relationship over time and with consolidation, suggesting an interesting connection between ensemble perception and long term memory consolidation that they might be underpinned by a similar reconstructive mechanism.

The results are consistent with there being a systems consolidation process that results in a qualitatively different representation of these memories by 1 month and the two types of memories interact dynamically (Richards et al., 2014; Winocur & Moscovitch, 2011; Sekeres et al., 2018). An increased reliance on neocortical areas over time would be expected to strengthen gist memory, as neocortex tends to represent information in a 'semanticized' form. The pull of item memories toward the gist may thus reflect the slow establishment or stabilization of such a neocortical trace.

One limitation is that the testing order of gist before item memory might encourage the recalls of the items to be consistent with the gist (Tversky & Kahneman, 1974; Mutluturk & Boduroglu, 2014). We chose this order because we wanted to minimize the influence of item memories on gist estimation. Indeed it is possible that this order influences the item retrieval. Nevertheless, since the order is the same for three different time delays, the changes in error and bias across delay groups are likely not due to the order of testing.

Another potential limitation is that participants selected which delay group to join, as opposed to random assignment into three delay conditions. A difference in expectation may meaningfully influence their learning and consolidation, although the performance of item and gist memories did not differ between groups at initial learning. In addition, there could be other biases inherent in a spatial memory task, for example, bias towards the vertical or the horizontal axis (Huttenlocher, Hedges, & Duncan, 1991). Future analyses and studies with randomized testing order and delay conditions are needed to address these issues.

It will be interesting to explore if our results, which considered gist memory as a spatial average, can generalize to a broader definition "gist" memory, such as gist-like memory for events (Moscovitch, Cabeza, Winocur, Nadel, 2016). For example, when first learning what a "birthday party" is from attending a few, the "gist" representation of "birthday party" may be dominated by memory for a few recent parties, but over time the "gist" becomes a more independent representation that can influence retrieval of those specific birthday party events.

Overall, our work shows that memory for individual items and memory for the gist of a set of items changed differently across 30 days. We propose that gist memory that is initially extracted from individual items gradually develops into an independent, stable representation that eventually guides item retrieval.

Acknowledgments

This research was supported by NIH R01 DC009209 awarded to STS. The authors gratefully acknowledge the data collected by Siqi Lin, Jennifer Nazario, Emily Potter and the helpful discussions with colleagues in Penn Psychology.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Ariely, D. (2001). Seeing Sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual Items.

Psychological Science, *22*(3), 384–392.
Brady, T., Schacter, D., & Alvarez, G. (2015). The adaptive nature of false memories is revealed by gist-based distortion of true memories. *Preprint*, *15*(12), 948.
Emin Orhan, A., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review*, *120*(2), 297–328. https://doi.org/10.1037/a0031541
Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202.
Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.
Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220–241.
Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, *15*(4), 1–14. https://doi.org/10.1167/15.4.10
Lutz, N. D., Diekelmann, S., Hinse-Stern, P., Born, J., & Rauss, K. (2017). Sleep supports the slow abstraction of gist from visual perceptual memories. *Scientific Reports*, *7*(1), 1–9.
Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annual review of psychology*, *67*, 105-134.
Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2p1), 304–308.
Richards, B. A., & Frankland, P. W. (2017, June 21). The persistence and transience of memory. *Neuron*, Vol. 94, pp. 1071–1084.
Richards, B. A., Xia, F., Santoro, A., Husse, J., Woodin, M. A., Josselyn, S. A., & Frankland, P. W. (2014). Patterns across multiple memories are identified over time. *Nature Neuroscience*, *17*(7), 981–986.
Richter, F. R., Bays, P. M., Jeyarathnarajah, P., & Simons, J. S. (2019). Flexible updating of dynamic knowledge structures. *Scientific Reports*, *9*(1), 1–15.
Schacter, D. L., Guerin, S. A., & St. Jacques, P. L. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474.
Sekeres, M. J., Winocur, G., & Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, *680*(May), 39–53. https://doi.org/10.1016/j.neulet.2018.05.006
Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology*, *72*(4), 499–503.
Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766.
Sweegers, C. C. G., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex*, *59*, 49–61.
Tompary, A., & Thompson-schill, S. L. (2019). Semantic influences on episodic memory distortions. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 1124–1130.
Tompary, A., Zhou, W., & Davachi, L. (2020). Schematic memories develop quickly, but are not expressed unless necessary. Retrieved from https://doi.org/10.31234/osf.io/k4fea
Tversky, A., & Kahneman, D. (2019). Judgment under uncertainty. *Foundations of Cognitive Psychology*, *185*. https://doi.org/10.7551/mitpress/3080.003.0038
Winocur, G., & Moscovitch, M. (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society*, *17*(5), 766-780.