Learning induced illusions: Statistical learning creates false memories

Yu Luo (yuluo@psych.ubc.ca)
Department of Psychology
University of British Columbia

Jiaying Zhao (jiayingz@psych.ubc.ca)
Department of Psychology
Institute for Resources, Environment and Sustainability
University of British Columbia

Abstract

The cognitive system readily extracts regularities in terms of object co-occurrences over space and time through statistical learning. However, how does learning such relationships influence the memory representations of individual objects? Here we used a false memory paradigm to examine the impact of statistical learning on memory representations of individual objects. Observers were exposed to a temporal sequence (Experiment 1) or spatial arrays (Experiment 2) of objects which contained object pairs (e.g., A-B). In a subsequent recognition phase, observers viewed a sequence or an array containing only one member of the original pair, and judged whether either the presented object or the missing object in the original pair was present. We found that statistical learning not only sharpened the detection of the presented object, but also induced a false memory of the missing object. This reveals a novel consequence of statistical learning: learning of regularities can create illusory memories.

Keywords: Statistical learning; false memory; implicit learning; regularities;

Introduction

A remarkable ability of the cognitive system is to detect and learn the relationships among objects in the environment. Statistical learning is one mechanism that extracts the statistical relationships between individual objects in terms of object co-occurrences over space and time (Fiser & Aslin, 2001; Saffran, Aslin, & Newport, 1996). This process occurs incidentally, without conscious intent or explicit awareness, produces knowledge about object associations that people are not explicitly aware of (Turk Browne, Jungé, & Scholl, 2005; Turk-Browne, Scholl, Chun & Johnson, 2009), and can operate in multiple sensory modalities and feature dimensions (Conway & Christiansen, 2005; Fiser & Aslin, 2001; Saffran et al., 1996; Turk-Browne, Isola, Scholl, & Treat, 2008). In addition, several cognitive consequences of statistical learning have been identified, such as the compressing of information (Brady, Konkle, & Alvarez, 2009; Zhao & Yu, 2016), attentional prioritization of co-occurring objects (Yu & Zhao, 2015; Zhao, Al-Aidroos, & Turk-Browne, 2013; Zhao & Luo, 2017), and enhanced memory representation (Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014; Otsuka & Saiki, 2016).

One important but unexplored question is: how does learning statistical associations influence the representation of individual objects? Initial evidence comes from studies on false memories of semantically related objects. One pioneer work by Roediger and McDermott (1995) shows that after memorizing a list of words (e.g., nurse, sick, medicine, etc.) that are highly related to a target word that was never present (e.g., doctor), people falsely remember seeing the target word, and label it as an “old” word of the list in the recognition task. This finding was replicated using a visual paradigm, where participants viewed a stereotypical scene (e.g., classroom), and falsely recalled and recognized a target object that was never present (e.g., chalkboard, Miller & Gazzaniga, 1998). One explanation for this phenomenon is that seeing one object can automatically activate other associated objects based on semantic memory (Roediger, Balota, & Watson, 2001).

Here, we provide a new explanation behind this old phenomenon which focuses on a simpler mechanism: learning the co-occurrences of objects can create the false memories of non-present object when only its partner is present. We propose that the mere statistical co-occurrence of two objects can produce false memory, independent of semantic associations. Thus, the goal of the current study was to examine whether statistical learning alters the representations of individual objects.

Experiment 1

This experiment examined how statistical learning alters the representations of individual objects in a temporal context, by first exposing participants to a sequence of paired objects and then testing them on whether seeing an individual object in the pair can produce the false memory of the non-present object in the pair.

Participants

A total of 120 undergraduates (96 female; mean age=20.6 years, SD=2.8) from University of British Columbia (UBC) participated in the experiment for course credit. Participants reported normal or corrected-to-normal visual acuity and provided informed consent. The protocol was approved by the UBC Research Ethics Board.

Stimuli

The stimuli consisted of eight real-world objects (Fig.1a) which were selected from a stimulus set in a previous study

(Brady, Konkle, Alvarez, & Oliva, 2008). All objects were converted to grayscale, and were adjusted to a mean brightness of 84. Each object subtended 2.8° of visual angle. The eights objects were randomly assigned into four pairs for each participant and remained constant throughout the experiment (Fig.1a). In each pair, the first object was always followed by the second object. The random assignment of objects into pairs ensured that there was no systematic semantic relation between two objects in a pair, but rather the two objects were associated with co-occurrences. Each pair was repeated 50 times to form a single continuous temporal sequence of objects in a pseudorandom order with a constraint where no single pair could repeat back-to-back.

Apparatus

Participants in all experiments were seated 50cm from a computer monitor (refresh rate=60 Hz). Stimuli were presented using MATLAB and PsychophysicsToolbox (http://psychtoolbox.org).

Procedure

The experiment consisted of two conditions. In the structured condition, the eight objects were grouped into four pairs. In the random condition, the eight objects appeared in a random order in the sequence. Participants were randomly assigned to one of the two conditions (N=60 in each). The experiment contained three phases: exposure phase, recognition phase, and test phase. During the exposure phase, one object appeared at the center of the screen for 500ms followed by a 500ms inter-stimulus interval (ISI) in each trial (Fig.1b). Participants performed a 1-back task where they judged as quickly and accurately as possible whether the current object was the same as or different from the previous object (by pressing the "/" or "z" key for same or different, respectively, key assignment counterbalanced). For the 1-back task, each object had a 20% chance of repeating the previous object, producing 480 trials in total. This 1-back task served as a cover task which was irrelevant to learning the object pairs, in order to conceal the true purpose of the study. This ensured that learning of the object pairs was incidental. Participants were not told anything about the object pairs.

After exposure, participants performed a recognition phase (Fig.1c). In each trial, participants viewed a continuous sequence of objects first and then judged whether a certain object was present in the sequence. In the structured condition, there were three types of trials. The first type was missing trials: the sequence contained all four pairs, except for one pair, one member was missing, and observers judged whether the missing member was present in the sequence. The missing trials measured the false alarm rate for the missing object. The second type was presented trials: the sequence contained all four pairs, except for one pair, one member was missing, but this time observers judged whether the presented member was present in the sequence. The presented trials measured the hit rate for the presented object. The third type was baseline trials: the sequence contained all four pairs, and observers judged whether one member in a pair was present in the sequence. The baseline trials measured the hit rate for the presented object. In the missing trials, each member of an original pair was missing for once, resulting in 8 trials. In the presented trials, the presented object of an original pair was tested once, resulting in 8 trials. In the baseline trials, each member of a pair was tested once, resulting in 8 trials. The 24 trials were repeated twice, producing 48 trials in total (order of the trials was randomized). In the random condition, the trials were the same, except the objects in the sequence appeared in a random order, so the sequence contained no pairs. Each object was presented for 500ms followed by a 500ms ISI. After the sequence was presented, a 3000ms blank screen followed. After the blank screen, an object was presented on the screen as a probe, and participants judged whether the object was presented in the previous sequence (by pressing the "1" or "0" key for "yes" or "no", respectively). The object remained on the screen until response.

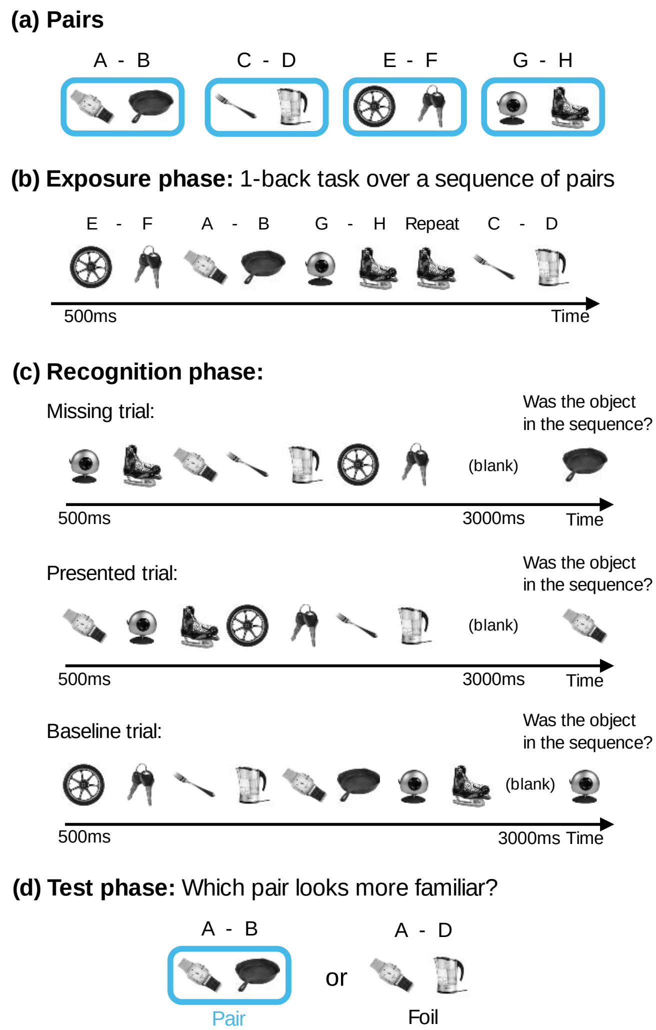

Figure 1. Experiment 1. (a) Four object pairs (e.g., A-B). (b) Exposure phase: 1-back task. (c) Recognition phase: missing trials, presented trials, and baseline trials. (d) Test phase: two-alternative forced-choice task.

To examine whether they had successfully learned the object pairs, participants in the structured condition completed a surprise two-alternative forced choice test phase following the recognition phase (Fig.1d). In each trial, participants viewed two sequences of objects. Each object appeared for 500ms followed by a 500ms ISI, and each sequence was separated by a 1000ms pause. Participants judged whether the first or second sequence looked more familiar based on what they saw in the exposure phase (by pressing the "1" or "0" key for "first sequence" or "second sequence", respectively). One sequence was a pair (e.g., A-B), and the other was a "foil" (e.g., A-D) composed of one object from an original pair (e.g., A-B), and the other from a different pair (e.g., C-D), while preserving the temporal positions in the pairs (Fig. 1d). Each pair was tested against each foil twice, which resulted in 16 trials in total (4 pairs × 2 foils × 2 repetitions). Importantly, each pair and foil were presented the same number of times at test. Thus, to discriminate the pair from the foil, participants needed to know which two specific objects followed each other. The order of the trials was randomized, and whether the pair or foil appeared first was counterbalanced across trials. Participants in the random condition was not tested, since there were no pairs in the sequence.

A debriefing session was conducted at the end of the experiment, where participants were asked if they had noticed any objects that appeared one after another. For those who responded yes, we further asked them to specify which objects followed each other.

Results and Discussion

At the test phase, pairs were chosen over foils on 60.0% (SD=19.3%) of the time, which was reliably above chance (50%) [$t(59)=4.01$, $p<.001$, $d=0.52$]. Thus, learning of the object pairs was successful. During debriefing, 12 participants reported noticing the pairs, but none could correctly report which specific objects followed each other. This suggests that participants had no explicit awareness of the object pairs.

The false alarm rate (FA) and the hit rate in the recognition phase were presented in Fig.2 and analyzed with a 2 (condition: structured vs. random; between-subjects) × 3 (trial type: missing vs. presented vs. baseline; within-subjects) mixed-effects ANOVA.

There was a main effect of condition [$F(1,118)=12.38$, $p<.001$, $\eta_p^2=.09$] and trial type [$F(2,236)=425.06$, $p<.001$, $\eta_p^2=.78$], but no reliable interaction between condition and trial type [$F(2,236)=0.73$, $p=.48$, $\eta_p^2=.006$]. Tukey's HSD post-hoc test showed that the FA rate of the missing trials was reliably higher in structured condition (27.9%) than in random condition (20.4%), $p=.03$; the hit rate of the presented trials was reliably higher in structured condition (72.5%) than in random condition (62.60%), $p<.001$; and a marginal difference in the hit rate of baseline trials between structured (72.5%) and random condition (65.9%), $p=.09$.

These findings suggest that statistical learning not only sharpens the memory of the object within the pairs, but also induces the false memory of the missing object.

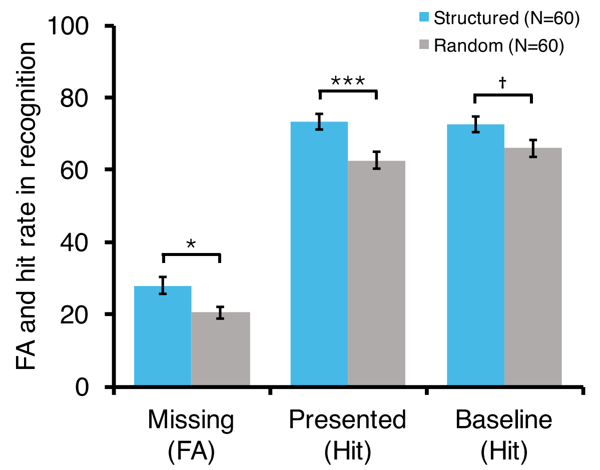

Figure 2: The false alarm (FA) rate and the hit rate in recognition phase (error bars reflect ±1 SEM; †$p<.1$, *$p<.05$, ***$p<.001$).

To examine the relationship between statistical learning and recognition performance, we found that there were no correlations between learning of the pairs at the test phase and the FA rate or the hit rate. However, in structured condition there was a weak correlation between the FA rate in the missing trials and the hit rate in the presented trials. There was a moderate correlation between the FA rate in the missing trials and the hit rate in the baseline trials, but no correlations in random condition (Table 1).

Table 1: Correlations between learning of the pairs at the test phase and the false alarm rate or the hit rate

Condition	Correlation	Correlation results
Structured	Learning vs. missing	$r(58)=-.15$, $p=.26$
(N=60)	Learning vs. presented	$r(58)=.21$, $p=.11$
	Learning vs. baseline	$r(58)=-.03$, $p=.79$
	Missing vs. presented	$r(58)=.28$, $p=.03$
	Missing vs. baseline	$r(58)=.31$, $p=.02$
	Presented vs. baseline	$r(58)=.48$, $p<.001$
Random	Missing vs. presented	$r(58)=.17$, $p=.21$
(N=60)	Missing vs. baseline	$r(58)=.20$, $p=.12$
	Presented vs. baseline	$r(58)=.63$, $p<.001$

Experiment 2

This experiment aimed to generalize the findings in Experiment 1 from the temporal context to a spatial context.

Participants

A new group of 68 undergraduates (51 female, mean age=20.2 years, SD=2.4) from UBC participated in the experiment for course credit.

Stimuli

The stimuli were identical to those in Experiment 1, except that in structured condition the four pairs were grouped into horizontal, vertical, and diagonal spatial configurations (Fig.3a). Each array contained all four pairs, and was placed on an invisible 4 × 4 grid (subtending 8.2° × 8.2°) with the constraint that one pair was adjacent to at least another pair. This was to prevent participants from learning the pairs based on spatial segmentation cues other than object co-occurrences. In random condition, the eight objects were randomly assigned to one of cell on the grid, with the constraint that each object neighbored at least one other object. Thus, the only difference between structured and random condition was the presence or absence of the pairs.

Procedure

As in Experiment 1, there were two conditions (i.e., structured vs. random, *N*=34 in each) and three phases (i.e., exposure, recognition, and test). In the exposure phase, participants in both conditions viewed arrays of objects, and performed a duplicate detection task where they judged as quickly and accurately as possible whether there were two identical objects in a single array (by pressing the "/" or "z" key for yes or no, respectively, key assignment counterbalanced, Fig. 3b). This duplicate detection task served as a cover task irrelevant to statistical learning, to ensure that learning of the object pairs was incidental. Participants were not told anything about the object pairs. Each array was presented on the screen for 1000ms followed by a 1000ms ISI in each trial. There were 480 trials in total, and 20% of the trials (80 trials) contained a duplicate object in the array.

The recognition phase was identical to that in Experiment 1, except that objects were presented all at once on the screen. In each trial, participants viewed an array for 800ms followed by a 3000ms pause, and judged whether the probe object was presented in the array. The display time was increased to 800ms, as it required more time for participants to view all eight objects at once. As before, there were three types of trials: (1) missing trials, where one member in the pair was missing, and the missing object was tested; (2) presented trials, where one member in the pair was missing, but the presented object in the pair was tested; and (3) baseline trials, where all pairs were presented, and one object was tested (Fig. 3c).

After the recognition phase, participants in the structured condition completed the surprise two-alternative forced choice test phase to see whether they had successfully learned the object pairs (Fig. 3d). In each trial, one set of objects was presented on the left and another on the right side of the screen for 1000ms. Participants judged whether the left or right set of objects looked more familiar based on what they saw in the exposure phase (by pressing the "1" or "0" key for "left" or "right", respectively). The foils were created following the same logic as in Experiment 1. Participants in the random condition was not tested, since there were no pairs in the array during exposure.

A debriefing session was conducted after test, where participants were asked if they had noticed any objects that appeared with one another. For those who responded yes, we further asked them to specify which objects appeared adjacent to each other.

Results and Discussion

At the test phase, pairs were chosen over foils on 52.2% (SD=10.3%) of the time, which was not reliably above chance (50%) [$t(33)=1.25$, $p=.22$, $d=0.21$]. This suggests that participants failed to learn the spatial co-occurrences between the two objects in the pairs. During debriefing, four participants reported noticing the pairs, but none could correctly report which specific objects appeared with each other. This suggests that participants had no explicit awareness of the pairs.

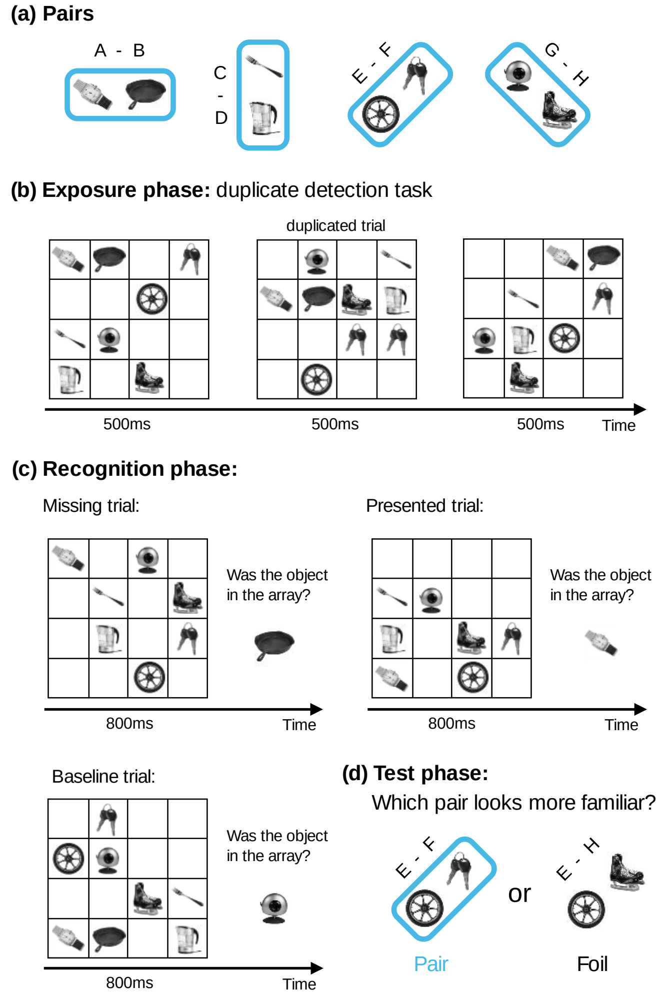

Figure 3. Experiment 2: (a) Four pairs in four different spatial configurations. (b) Exposure phase: duplicate detection task. (c) Recognition phase: missing trials, presented trials, and baseline trials. (d) Test phase: two-alternative forced-choice task.

The FA rate and the hit rate in the recognition phase were analyzed with a 2 (condition: structured vs. random; between-subjects) × 3 (trial type: missing vs. presented vs. baseline; within-subjects) mixed-effects ANOVA. There was a main effect of trial type [$F(2,132)=137.32$, $p<.001$, $\eta_p^2=.68$], but no main effect of condition [$F(1,66)=0.004$, $p=.95$, $\eta_p^2=.00$] and no significant interaction between condition and trial type [$F(2,132)=0.45$, $p=.63$, $\eta_p^2=.007$]. Tukey's HSD post-hoc test showed that the FA rate of the missing trials was not different between the structured condition (30.7%) and the random condition (30.5%), $p=.99$, the hit rate of the presented trials was not different between the structured condition (57.5%) and the random condition (59.4%), $p=.99$, and no difference in the hit rate of baseline trials between the structured (66.7%) and the random conditions (64.3%), $p=.97$ (Fig. 4).

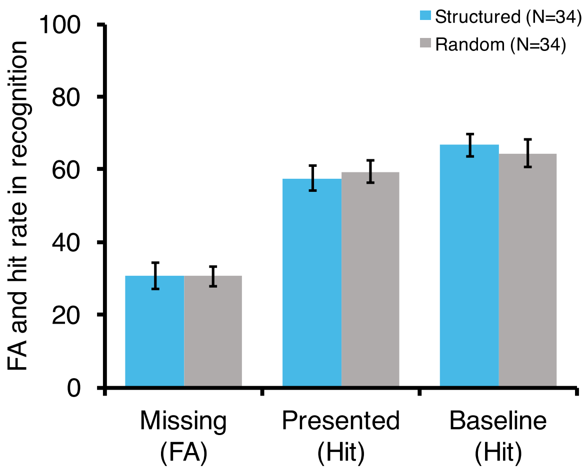

Figure 4: The false alarm rate and the hit rate in recognition phase (error bars reflect ±1 SEM).

We found no correlation between learning of the pairs at the test phase and the FA rate or the hit rate. But in both the structured and random condition, there was a correlation between the FA rate, and the hit rate in the presented trials and in the baseline trials (Table 2).

Table 2: Correlations between learning of the pairs at the test phase and the false alarm rate or the hit rate

Condition	Correlation	Correlation results
Structured (N=34)	Learning vs. missing	$r(32)=.24$, $p=.17$
	Learning vs. presented	$r(32)=.22$, $p=.22$
	Learning vs. baseline	$r(32)=.16$, $p=.35$
	Missing vs. presented	$r(32)=.66$, $p<.001$
	Missing vs. baseline	$r(32)=.57$, $p<.001$
	Presented vs. baseline	$r(32)=.40$, $p=.02$
Random (N=34)	Missing vs. presented	$r(32)=.44$, $p=.009$
	Missing vs. baseline	$r(32)=.53$, $p<.001$
	Presented vs. baseline	$r(32)=.62$, $p<.001$

The lack of memory difference between the structured condition and the random condition could be due to the lack of learning of object pairs in the spatial context.

To further explore whether learning of spatial pairs changed the representation of individual objects in the pairs, we separated participants who successfully learned the pairs (those who chose pairs over foil above chance, N=15), and those who failed to learn the pairs (those who chose pairs over foil at or below chance, N=19) in the structured condition. Among participants who showed learning, pairs were chosen over foils on 61.2% (SD=6.8%) of the time, which was reliably above chance (50%) [$t(14)=6.44$, $p<.001$, $d=1.66$]. Only one participant reported noticing the pairs, but could not correctly report which specific objects appeared with each other. A 2 (group: learners vs. non-learners; between-subjects) × 3 (trial type: missing vs. presented vs. baseline; within-subjects) mixed-effects ANOVA revealed a main effect of trial type [$F(2,64)=68.40$, $p<.001$, $\eta_p^2=.68$], but no main effect of group [$F(1,32)=1.25$, $p=.27$, $\eta_p^2=.04$] and no significant interaction between group and trial type [$F(2,64)=0.42$, $p=.66$, $\eta_p^2=.01$]. Although the results were not reliably different between the two groups, the learners consistently showed numerically greater FA rate and hit rate than the non-learners (Fig.5), a pattern that was consistent with the findings in Experiment 1.

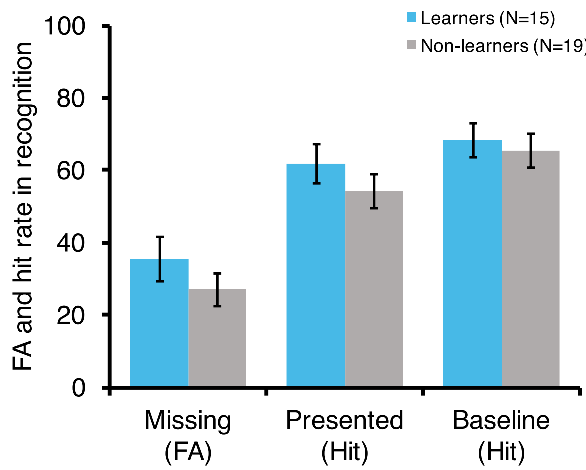

Figure 5: The false alarm rate and the hit rate of learners and non-learners in recognition phase in the structured condition (error bars reflect ±1 SEM).

General Discussion

The goal of this experiment was to examine whether statistical learning alters the memory representations of individual objects. We found that after learning the temporal co-occurrences of objects, participants showed a reliably higher false alarm rate of seeing a missing object, and a reliably higher hit rate of seeing a presented object (Experiment 1). When the objects co-occurred over space, participants did not successfully express learning of pairs, and therefore did not show differential false alarm and hit rates (Experiment 2). However, with a more detailed analysis, participants who successfully learned the spatial pairs showed numerically higher false alarm rate of the missing object and numerically higher hit rate of the presented object than those who failed to learn the pairs. The current findings suggest that statistical learning not only sharpens the detection of the objects within the pairs, but also induces a false memory of the missing object.

Induced false memory of the missing object can be explained by the automatic statistical association between the missing object and the presented object in the pair. Once the pairs were learned over repeated exposures even implicitly, one member in the pair could serve as a cue to signal the presence of its partner (Turk-Browne, et al., 2009). Thus, participants may have automatically brought the missing object to mind when seeing its partner in the sequence, thus false recalling that the missing object was present. This suggests that the automatic activation of the missing object was possible by merely co-occurring with its partner previously.

Alternatively, the two co-occurring objects may be unitized after learning. Previous studies have demonstrated that regularities compress information (Brady et al., 2009) and reduce perceived numerosity of the objects (Zhao & Yu, 2016), which suggests that the co-occurring objects could be grouped and encoded as one single unit. Seeing a member of the unit could trigger the illusion that the entire unit was presented, and therefore inducing the false memory of the missing partner.

The enhanced hit rate of the presented member in the pair could be due to the possibility that statistical regularities automatically draw attention (Zhao et al., 2013). Given that attention plays an important factor in the recognition task, participants in the structured condition may have prioritized processing of the paired objects, and therefore showed a better hit rate compared to the random condition.

Another account for the enhanced memory is that it may be easier to memorize the objects that were present in the sequence, because statistical learning increases the working memory capacity to encode objects (Brady et al., 2009). The better memory performance of the paired objects in the baseline condition was consistent with previous finding that statistical learning enhances memory of structured objects (Otsuka & Saiki, 2016).

In conclusion, we discovered a novel consequence of statistical learning: it not only enhances the detection of the object within the regularities, but also creates a false memory of the missing object.

Acknowledgement

This work was supported by NSERC Discovery Grant (RGPIN-2014-05617 to JZ), the Canada Research Chairs program (to JZ), the Leaders Opportunity Fund from the Canadian Foundation for Innovation (F14-05370 to JZ), and Elizabeth Young Lacey Graduate Scholarship (to YL).

References

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138,* 487-502.

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

Conway, C. M., & Christiansen, M. H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24-39.

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

Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences, 111*, 8997-9002.

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia, 36*, 513-520.

Otsuka, S., & Saiki, J. (2016). Gift from statistical learning: Visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. *Cognition, 147*, 113-126.

Roediger III, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. The nature of remembering: *Essays in honor of Robert G. Crowder,* 95-115.

Roediger III, H. L., & McDermott, K. (1995). Creating false memories - remembering words not presented in lists. *Journal of Experimental Psychology-Learning Memory and Cognition, 21*, 803-814.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 399–407.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*, 1934–1945.

Yu, R., & Zhao, J. (2015). The persistence of attentional bias to regularities in a changing environment. *Attention, Perception, & Psychophysics, 77,* 2217-2228.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science, 24*, 667–677.

Zhao, J., & Luo, Y. (2017). Statistical regularities guide the spatial scale of attention. *Attention, Perception, & Psychophysics, 79,* 24-30.

Zhao, J., & Yu, R. (2016). Statistical regularities reduce perceived numerosity. *Cognition, 146,* 217-222.