Verbal interference suppresses object-scene binding in visual long-term memory

Zhisen J. Urgolites[1]✉, Timothy F. Brady[1], and Justin N. Wood[2]

[1]Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093
[2]Department of Informatics; Department of Cognitive Science; Center for the Integrative Study of Animal Behavior School: School of Informatics, Computing, & Engineering, Indiana University, 700 N Woodlawn Ave, Bloomington, IN 47408

Building a unified representation of an event requires binding object and scene information in visual long-term memory (VLTM). While previous studies have examined how humans remember individual objects and scenes, little is known about the mechanisms that support object-scene binding. In this study, we examined whether language plays a role in binding objects and scenes in VLTM. Participants studied a large number of object-scene pairs, either while performing no concurrent task, a concurrent verbal shadowing task, or a concurrent rhythmic shadowing task. Participants were then tested on their memory for the individual objects and scenes (entity memory) or their memory for which objects were displayed in which scenes (object-scene binding). We found that (1) the rhythmic load and verbal load impaired memory for objects and scenes to a similar extent, but (2) the verbal load impaired object-scene binding significantly more than the rhythmic load. Thus, suppressing verbal resources during encoding selectively disrupts object-scene binding in long-term memory. We conclude that language networks play an important role in object-scene binding in VLTM.

visual long-term memory | object-scene binding | language | attention | binding

Correspondence: ***zurgolites@ucsd.edu***

Introduction

To build a unified representation of an event, we must bind object and scene information in visual long-term memory (VLTM). For example, remembering the location of your keys requires binding an object representation of 'your keys' with a scene representation of the place where you last left your keys, and then storing that bound object-scene set in VLTM. Similarly, to provide accurate eyewitness testimony, an observer must encode visual information about the perpetrator with information about the scene in which the crime took place. The present study examined the cognitive mechanisms that support object-scene binding in VLTM.

Existing work has found that people are capable of binding visual entities in VLTM. We can remember the identities and locations of objects within scenes (Hollingworth, 2005, 2006, 2010; Hollingworth & Henderson, 2002), and remember which agents performed which actions (Earles et al., 2008; Kersten & Earles, 2010). However, we are not perfect at binding visual entities into unified memories. For instance, participants often falsely believe they have seen an object from a particular viewpoint if both the scene viewpoint and object are independently familiar, even if they had never been seen them together (Varakin & Loschky, 2009). Similarly, unconscious transference errors — in which an eyewitness mistakenly identifies someone seen in a non-criminal context as a perpetrator in a criminal context — occur frequently in eyewitness testimony (Loftus, 1976; Perfect & Harris, 2003; Ross et al., 1994). These errors can be easily triggered: When witnesses are exposed to mug shots of a suspect, then they are more likely to subsequently identify that suspect as a perpetrator (Brown, Deffenbacher, & Sturgill, 1977; Deffenbacher, Bornstein, & Penrod, 2006; Deffenbacher, Carr, & Leu, 1981; Perfect & Harris, 2003). According to the dual-process theory (Atkinson & Juola, 1973, 1974; Mandler, 1980; Jacoby, 1991), these errors are typical cases of successful familiarity-based item memory with failed recollection-based contextual memory. On the one hand, people easily remember the individual entities that they have seen, consistent with findings that VLTM stores accurate representations of objects, actions, and scenes (Brady et al., 2008; Urgolites & Wood, 2013a; Konkle et al., 2010; Varakin & Loschky, 2009). On the other hand, people have difficulty remembering the associations between these entities in VLTM (i.e., which entities were seen together in an event).

The difficulty in binding objects and scenes in long-term memory might occur because object information and spatial/scene information are supported by separate processing systems (Tresch et al., 1993) and separate neural substrates (Epstein, 2008; Kanwisher, 2010; Kravitz, Saleem, Baker, & Mishkin, 2011; Moscovitch et al., 1995). A critical question is: how are the distributed representations for objects and scenes bound in memory? It has been proposed that the hippocampus is the critical neural substrate in the early stage of memory encoding and consolidation for forming associative long-term memories (Eichenbaum, 2004; Frankland & Bontempi, 2005). In the case of object-scene binding, the hippocampus receives input about object information from the nearby perirhinal cortex and input about scene information from the parahippocampal cortex (Davachi, 2006). The hippocampus encodes and holds information about the object-scene associations first, and then over time, through interaction with the neocortex, the associative memory is stored in neocortical regions including medial prefrontal cortex and angular gyrus (Takashima et al., 2009; Bonnici et al., 2016). Synchronized firing of neurons at different regions of the brain might also support binding of information stored in distributed regions (Engel et al., 1997). The distributed neural

representations of object vs. scene information, as well as the difficulty in binding across such distinct brain regions, is an important reason to believe that object-scene binding might be particularly challenging for our memory system. So far, there has been little research on the cognitive processes and mechanisms that underlie binding of different visual entities into event memory. In this study, we focused on examining the role of language in binding objects and scenes in long-term memory.

Earlier studies have demonstrated that language plays an important role in many of our seemingly non-verbal cognitive processes. These processes include visual search (Spivey et al., 2001), spatial cognition (Hermer-Vazquez et al., 1999; Pyers et al., 2010), nonverbal false belief reasoning (Newton & de Villiers, 2007), categorical perception (Winawer et al., 2007), numerical representations (Frank et al., 2012), and labeling for familiar objects in memory (Lupyan, 2008). The general findings are that availability of language supports efficient processing, whereas either lack of language for coding the information or the temporary inaccessibility of language undermines the cognitive processes (Frank et al., 2012), particularly the ones that require connecting multiple distributed representations. For example, in one study (Dassalegn & Landau, 2008), four-year olds saw a split square with red on the left and green on the right, and later were asked to find the target in an array that included the target, its reflection (red on the right and green on the left), and a square with a different geometric split. Children were more likely to make errors with the reflections, indicating binding errors between color and location. The four-year olds performed significantly better when the targets were accompanied by sentences specifying the relationship between color and location (e.g., "the red is on the left") but not when the sentences specified a non-directional relationship (e.g., "the red is touching the green"). For adults, verbal shadowing drove their performance down to the level of the 4-year-olds on the same task (again, due predominantly to reflection errors). Thus, language appears to be important for binding features to locations.

We addressed the question of whether language is important for object-scene binding by using a dual-task paradigm. Specifically, participants viewed a large number of object-scene pairs in the study session. While studying the pairs, they performed either (1) no concurrent task, (2) a concurrent verbal shadowing task, or (3) a concurrent rhythmic shadowing task. Afterwards, they were tested on their memory for the individual entities (entity memory: what objects and what scenes had been studied) or for the binding between the entities (binding memory: which objects were displayed in which scenes). Note that a verbal shadowing task occupies language resources as well as attentional resources. To separate the effect of language from the effect of attention, we included the rhythmic shadowing condition which loaded attention without loading language resources. Similar to earlier studies (Newton & de Villiers, 2007; Hermer-Vazquez et al. 1999; Dungan & Saxe, 2012), we made sure that the rhythmic shadowing task and the verbal shadowing task were matched in difficulty (see results). By contrasting memory performance between the verbal shadowing condition and the rhythmic shadowing condition, we could investigate the role of language in object-scene binding memory without the confounding effect of attention distraction.

Methods

Participants. We tested 120 participants with normal or corrected-to-normal vision (females = 99, males = 21; mean age = 20 years, SD = 1.38). The data from two additional participants were excluded from the final analyses because they failed to maintain continuous verbal shadowing. Forty participants were randomly assigned to each of three encoding conditions (i.e., the baseline condition, the rhythmic shadowing condition, and the verbal shadowing condition). Within each encoding condition, eight participants were randomly assigned to each of the five testing conditions (described below). Thus, all conditions (3 encoding conditions crossed with 5 testing conditions) were conducted across participants. Informed consent was obtained.

Stimuli. The stimuli consisted of images of 80 objects and 80 scenes. The objects were selected from The Object Databank, which consists of realistic 3-D objects (website: http://wiki.cnbc.cmu.edu/Objects). The scenes were selected from a previous study that tested VLTM for scenes (Konkle et al., 2010). Each object-scene set displayed one object in the center of one scene (see Figure 1). The object and scene presented in each set were randomly selected. The scenes subtended 17.5° x 14.6° of viewing angle in the center of a 17" computer screen. To ensure that the objects were perceived as individual objects rather than features of the scenes, the objects rotated within the scenes. Specifically, the objects rotated continuously, completing a full 360° rotation during the 1,000-ms presentation of each object-scene set. The scene stimuli set included images of oceans, forests, cities, mountains, and open countries.

Procedure. The experiment contained three encoding conditions: the baseline condition, the rhythmic shadowing condition, and the verbal shadowing condition. In the baseline condition, participants studied the object-scene sets without performing a concurrent task. In the rhythmic and verbal shadowing conditions, participants performed a concurrent rhythmic or verbal shadowing task (described below) while studying object-scene sets. For each encoding condition, a test phase followed the study phase. A test phase had either an entity memory test in which memory for studied items were tested (in 3 testing conditions) or a binding memory test in which memory about which object was displayed in which scene was tested (in 2 testing conditions).

In the study phase, participants viewed 40 randomly selected object-scene sets. Each trial began with a black screen (1,000-ms), followed by an object-scene set (1,000-ms), which was in turn followed by a black screen (1,000-ms). During the presentation of the object-scene set, the object rotated 360°. To maintain their attention, participants performed a repeat-detection task during the study phase.

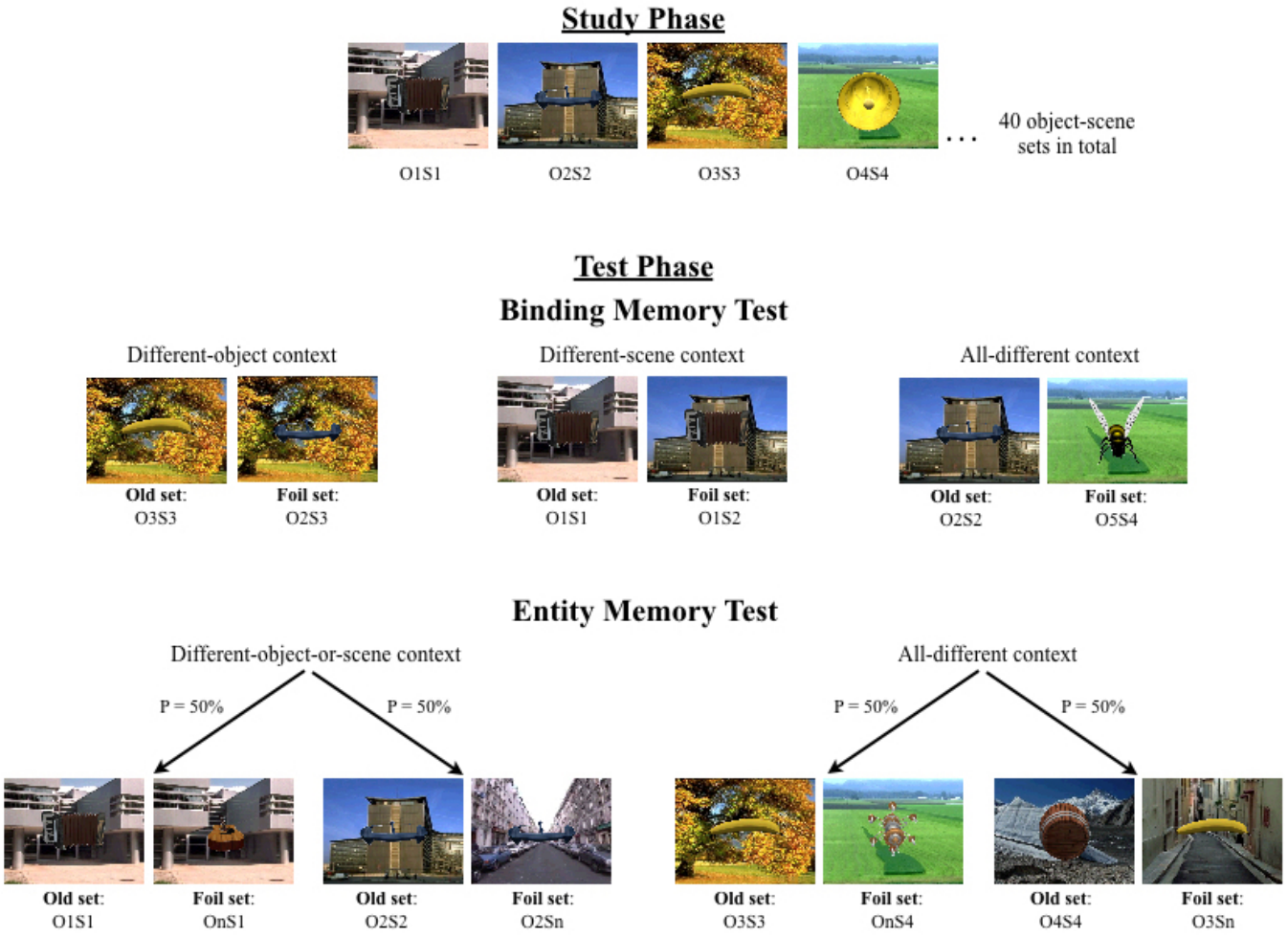

Fig. 1. Schematic of the experimental design. In the study phase, participants studied 40 object-scene sets and concurrently performed no other task, a rhythmic shadowing task, or a verbal shadowing task. Participants from each study condition were then tested for either binding or entity memory. In the binding memory test, participants needed to remember which objects were displayed in which scenes. In the entity memory test, participants needed to remember the individual objects and individual scenes, but did not need to remember which objects where displayed in which scenes. The images show the scenes with the starting position of the objects before rotation. "O1S1" indicates that the set contained Object 1 in Scene 1, both items having been studied together in the study phase; "O2S2" indicates that the set contained Object 2 in Scene 2, both items also having been studied together in the study phase, and so forth. Likewise, "O1Sn" indicates that the set contained Object 1 from the study phase which was now displayed in a novel scene, whereas "OnS1" indicates that the set contained a novel object displayed in Scene 1 from the study phase.

Ten object-scene sets were presented on two separate trials in the study phase, such that 0-4 trials intervened between the first presentation of the set and the repeat presentation of the set. After each set was presented, the phrase "Old or New" appeared on the screen, prompting participants to indicate whether the object-scene set was a repeat presentation of a set or a new set. Participants responded without time pressure.

In the baseline condition, participants performed no concurrent shadowing task. In the verbal and rhythmic shadowing conditions, participants performed the concurrent shadowing task for the entirety of the study phase. We used the rhythmic and verbal shadowing tasks described by Newton and de Villiers (2007) and Hermer-Vazquez et al. (1999). The rhythmic shadowing task required the participant to repeat short rhythmic patterns. Participants listened to a 4/4 measure of beats, and then tapped the rhythm during a 4/4 measure of silence. Afterward, a new rhythmic measure played and participants listened before they tapped during the next silent measure. The rhythmic measures averaged 5-6 notes per measure. Participants tapped on the desk surface with their left hand. Participants were trained in advance to ensure that they could tap the varying rhythms correctly. The verbal shadowing task entailed constant verbal shadowing of English sentences. Participants were trained until they were sufficiently fluent to shadow continuously for 1 minute without pausing for 1s (Hermer-Vazquez et al., 1999). The test phase began approximately one minute after the study phase. Each participant received 40 test trials. On each trial, one old (previously seen) object-scene set and one new object-scene set were presented on the screen sequentially. Specifically, participants were shown a black screen (1,000 ms), followed by an object-scene set on the left side of the screen (1,000 ms), which was then replaced by another object-scene set on the right side of the screen. The old set was presented an equal number of times on the left and right sides of the screen.

After viewing the second set, participants indicated which set (left or right) was the old set, by pressing one of two keys on the keyboard, with no time pressure. Either binding memory or entity memory was tested in the test phase.

Binding Memory Test. The participants tested in the binding memory task were told at the beginning of study phase that they would see a large number of object-scene sets and then be tested on their ability to remember which objects occurred within which scenes. In the binding memory test, the old object-scene set contained an object and a scene that were previously presented together in the study phase (i.e., old combination). The new object-scene set consisted of a new combination of an object and a scene that were previously presented in different object-scene sets in the study phase (i.e. recombination). Thus, to succeed, participants needed to remember which objects occurred within which scenes and then select the sets that were the old combinations of previously seen objects and scenes.

To ensure that the results would generalize across different arrangements of the test stimuli, we probed binding memory in three testing contexts (see Figure 1). In the 'different-object' context, the two object-scene sets that were displayed on the same trial had the same scene and different objects. In the 'different-scene' context, the two object-scene sets had the same object and different scenes. And in the 'all-different' context, the two object-scene sets had different objects and different scenes.

Entity Memory Test. The participants tested in the entity memory task were told at the beginning of the study phase that they would see a large number of object-scene sets and then be tested on their ability to remember the objects and scenes that were presented in the study phase. In the entity memory test, the old object-scene set contained an object and a scene that were previously presented together in the study phase (i.e., old combination), similar to the binding memory test. However, the new object-scene set contained one new item, either a new object in a previously studied scene (half of the trials) or a previously studied object in a new scene (half of the trials). The two types of trials were randomly intermixed with each other. Critically, because each new set could contain either a new object or a new scene, participants needed to try to remember both the objects and the scenes presented in the study phase to succeed in the test. They did not need to remember which objects were displayed in which scenes. We also probed entity memory in multiple ways to ensure that the results would generalize across different arrangements of test stimuli. Our design included two testing contexts (see Figure 1). In the 'different-object-or-scene' context, the old object-scene set and the new object-scene set had the same old scene on one-half of the trials and the same old object on the other one-half of the trials. In the 'all-different' context, both object-scene sets had different objects and different scenes.

Verbal instructions were provided before the experiment. Written instructions were also provided on the screen before each phase of the experiment.

Results

Performance was good on the repeat detection task in the study phase across all encoding and testing conditions (all means > 89%, all SEM < 2%). As expected, participants were more accurate in detecting repeated trials when they did not perform a concurrent task than when they performed a concurrent shadowing task (M = 97%, SEM = 1.5% for the baseline condition, M = 93%, SEM = 1.5% for rhythmic shadowing condition, and M = 91%, SEM = 1.4% for the verbal shadowing condition; $F(2,117) = 9.05$, $p < .001$, $\eta^2 = .13$). Crucially, participants' accuracies in detecting repeated trials were similar whether they performed a concurrent rhythmic or verbal shadowing task during the study phase ($t(98) = 0.12$, $p = 0.60$), providing one piece of evidence that the two shadowing tasks distracted attention to the same level (see below for additional evidence from entity memory performance).

Performance in the test phase was similar for the object-scene sets that had, or had not, been repeated as part of the repeat detection task for all testing conditions (e.g., ps > .17), with the exception of the 'all-different' testing condition for the binding memory test in the rhythmic shadowing condition (repeated object-scene sets were remembered better than non-repeated object-scene sets, p = .02). Thus, for this one testing condition, only the data from the non-repeated object-scene sets (30 out of 40 object-scene sets) were used in further analyses. For all other testing conditions, we used the data from both the repeated and non-repeated object-scenes sets.

For all three encoding conditions (i.e., baseline, rhythmic, or verbal shadowing condition), performance in the test phase did not differ across the three testing contexts for binding memory (ps > .50) or the two testing contexts for entity memory (ps > .50). The multiple testing contexts were designed only to ensure we tested binding and entity memory with a broad range of possible probes. We thus pooled the data for the binding test contexts and entity test contexts for the remaining analyses. (Note that for the 'all-different' context for binding memory test in the rhythmic shadowing condition, performance did not differ from performance in the other testing contexts whether or not the score for the repeated object-scene sets were included in the analyses, ps > .19)

Analysis of Performance Within Conditions. The accuracy rates from the test phase are depicted in Figure 2. Accuracy rates in all conditions for entity and binding memory were higher than chance level (50%) (ps < .01).

In the baseline condition, participants correctly recognized the old object-scene set on 91% (SEM = 2%) of the entity memory trials (objects: 95%, SEM = 2%; scenes: 86%, SEM = 3%) and 78% (SEM = 2%) of the binding memory trials. The accuracy rates from the binding memory test were significantly lower than average entity memory ($t(48) = 3.88$, $p < .001$), object memory alone ($t(48) = 5.34$, $p < .001$), and scene memory alone ($t(48) = 2.24$, $p = .03$) from the entity memory tests.

In the rhythmic shadowing condition, participants cor-

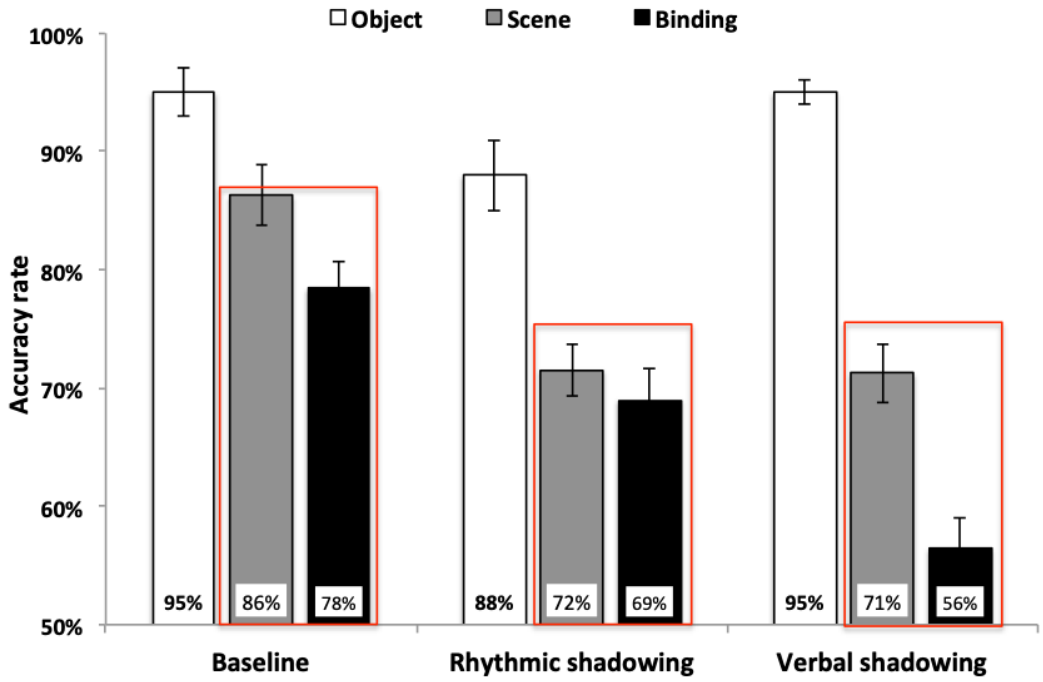

Fig. 2. Accuracy rates in recognizing previously observed object-scene sets in the entity memory test and in the binding memory test across the baseline, rhythmic shadowing, and verbal shadowing conditions. Accuracy rates from the entity memory test are depicted separately as object memory alone (white bars) and scene memory alone (grey bars) (see text for the overall entity memory performance for objects and scenes). Accuracy rates from the binding memory test are depicted by the black bars. Red boxes enclose the two variables (i.e., scene memory and binding memory) that were compared when contrasting entity memory vs. binding memory performance (see text for more explanation). Error bars denote standard errors.

rectly recognized the old object-scene set on 80% (SEM = 2%) of the entity memory trials (objects: 88%, SEM = 3%; scenes: 72%, SEM = 2%) and on 69% (SEM = 3%) of the binding memory trials. The accuracy rate for binding memory was significantly lower than the accuracy rates for object memory ($t(48) = 4.82$, $p < .001$), but not lower than the accuracy rates for scene memory ($t(48) = .57$, $p = .57$) or average object-scene memory ($t(48) = 2.87$, $p = .24$) from the entity memory test.

In the verbal shadowing condition, participants correctly recognized the old object-scene set on 83% (SEM = 1%) of the entity memory trials (objects: 95% SEM = 1%; scenes: 71%, SEM = 2%) and on 56% (SEM = 3%) of the binding memory trials. The accuracy rate from the binding memory test was significantly lower than the accuracy rates of the average object-scene memory ($t(48) = 7.72$, $p < .001$), object memory alone ($t(48) = 11.36$, $p < .001$), or scene memory alone ($t(48) = 3.98$, $p < .001$) in the entity memory condition.

Note that for entity memory in each condition, the accuracy rates for object memory were significantly higher than the accuracy rates for scene memory (ps < .001), potentially because (1) the objects were more readily namable than the scenes, and/or (2) the objects moved. Since an accurate memory of a bound object-scene set requires remembering both the object and the scene, binding accuracy should be limited by entity memory for the more difficult entity type (Urgolites & Wood, 2013b). To provide a fair comparison between entity and binding memory performance, we thus used the lower of the two accuracy rates in the entity memory test (i.e., the accuracy rates for remembering scenes) to represent entity memory and compared those values to binding memory performance (see red boxes in Figure 2) in the analysis of performance across conditions (Model 1).

Analysis of Performance Across Conditions.

Model 1. In this first model of across-condition analysis, we carried out a 2 × 3 ANOVA with the factors of Memory Test (binding memory vs. scene-entity memory) and Encoding Condition (baseline vs. rhythmic shadowing vs. verbal shadowing). The analysis revealed a significant main effect of Memory Test ($F(1,114) = 15.55$, $p < .001$, $\eta^2 = .12$), a significant main effect of Encoding Condition ($F(2,114) = 26.85$, $p < .001$, $\eta^2 = .32$), and a significant interaction between the two factors ($F(2,114) = 3.10$, $p = .049$, $\eta^2 = .05$). Post hoc analyses showed that rhythmic shadowing and verbal shadowing reduced the accuracy rate of entity memory to a similar level ($p = .94$), providing further support that these two shadowing tasks imposed similar levels of impact on entity memory. In contrast, verbal shadowing reduced the accuracy of binding memory to a significantly lower level than rhythmic shadowing did ($p < .001$), suggesting that verbal shadowing has a selective effect on binding memory that rhythmic shadowing does not have. A 2 × 2 ANOVA examining entity and binding memory for only the rhythmic and verbal shadowing conditions also found the significant interaction between Memory Test (entity memory vs. binding memory) and Condition (rhythmic shadowing vs. verbal shadowing) ($F(1,76) = 4.91$, $p = .03$, $\eta^2 = .06$, Figure 3), confirming the finding that

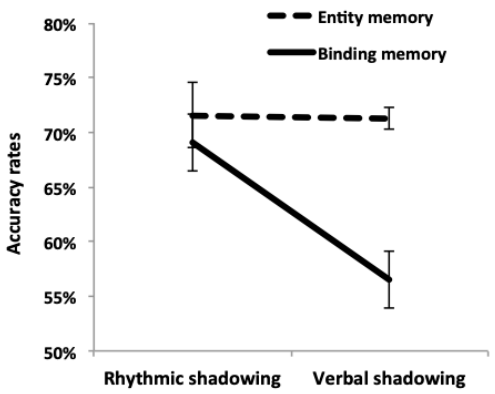

Fig. 3. Interaction between Memory Test (entity memory vs. binding memory) and Encoding Condition (rhythmic shadowing vs. verbal shadowing). Error bars denote standard errors.

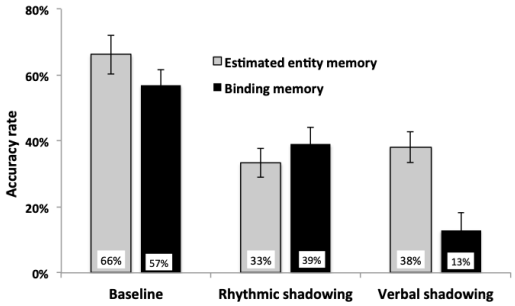

Fig. 4. Estimated probability in remembering both objects and scenes in the entity memory test (estimated entity memory) and the probability of remembering bound object-scene sets for the baseline, rhythmic shadowing, and verbal shadowing conditions. Error bars denote standard errors.

reducing verbal resources selectively impairs binding memory.

Model 2. In the first model of across-condition analysis, the accuracy rate of scene memory was compared with that of binding memory, because scene memory is the lower of the two separate entity memories (lower than object memory). The reasoning is that if objects and scenes are perfectly correlated (such that if the scene is remembered, the object is also remembered), scenes should be the limiting factor for binding memory, and, if binding is perfect, binding memory performance should be equal to scene memory. An alternative way to estimate the overall entity memory level is to calculate the product of the percent of objects remembered and the percent of scenes remembered, which assumes the particular scenes and objects that are remembered are independent. In this sense, if a participant remembered 75% of previously studied objects and 90% of previously studied scenes, then the probability that this participant remembered both objects and scenes would be 68% (75% × 90% = 68%). Because chance guessing in a two-alternative forced-choice test can lead to a 50% accuracy rate, to correct for chance guessing, we follow the equation of (accuracy rate – chance)/(1 - chance) to arrive at the probability of remembering from percent correct (Standing et al., 1970; Brady et al., 2008). For each participant, we calculated the probability of remembering objects, the probability of remembering scenes, and the probability of remembering bound object-scene sets based on the participant's accuracy rates in the three categories. We then calculated the estimated probability of remembering both a given object and scene (i.e., estimated overall entity memory) by computing the product of the probability of remembering objects and the probability of remembering scenes, as described above. The results are shown in Figure 4. Thus, in the second model of across-condition analysis, we compared the estimated overall probability of remembering objects and scenes with the probability of remembering bound object-scene sets.

In the baseline condition, the estimated entity memory was 66U+0025 (SEM = 6%). This performance rate was similar to performance from the binding memory condition (mean = 57%, SEM = 5%; t(48) = 1.3, p = .22), indicating that, when there was no concurrent task during encoding, the probability that both objects and scenes were remembered closely predicted the probability in remembering which objects were displayed in which scenes. In the rhythmic shadowing condition, the estimated entity memory (mean = 33%, SEM = 4%) also did not differ from the binding memory in this condition (mean = 39%, SEM = 5%; t(48) = .78, p = .44), indicating that, when participants performed a concurrent rhythmic shadowing task during encoding of object-scene sets, the probability that both objects and scenes were remembered also closely predicted the probability in remembering which objects were displayed in which scenes. In contrast, in the verbal shadowing condition, estimated entity memory (mean = 38%, SEM = 5%) was significantly higher than binding memory (mean = 13%, SEM = 5%; t(48) = 3.4, p < .01), indicating that binding memory was significantly impaired when participants performed a concurrent verbal shadowing task during encoding of object-scene sets.

We also carried out a 2 × 3 ANOVA with factors of Memory Type (estimated entity memory vs. binding memory) and Encoding Condition (baseline vs. rhythmic shadowing vs. verbal shadowing). The analysis revealed a significant main effect of Memory Type (F(1,114) = 5.1, p < .05, η^2 = .04), a significant main effect of Encoding Condition (F(2,114) = 25.4, p < .01, η^2 = .31), and a significant interaction between the two factors (F(2,114) = 4.4, p < .05, η^2 = .07). Post hoc analyses showed that rhythmic shadowing and verbal shadowing reduced the estimated entity memory to similar levels (p = .46), suggesting that the two shadowing tasks had a similar impact on entity memory. In contrast, verbal shadowing reduced binding memory to a significantly lower level than rhythmic shadowing (p < .001), indicating that there is selective effect that verbal shadowing has on binding memory which rhythmic shadowing does not have. A 2 × 2 ANOVA for estimated entity memory and binding memory in only the rhythmic and verbal shadowing conditions revealed a significant interaction between Memory Type (estimated entity memory vs. binding memory) and Encoding Condition (rhythmic shadowing vs. verbal shadowing) (F(1,76) = 8.8, p < .01, η^2 = .10), confirming the selective effect of the verbal shadowing task on binding memory. The other two 2 × 2

ANOVAs that broke down the 2 × 3 ANOVA (i.e., comparing estimated entity and binding memory between baseline and rhythmic shadowing or between baseline and verbal shadowing) did not yield any significant interactions.

In summary, the results from the two models indicate that: (1) the rhythmic load and verbal load impaired memory for objects and scenes to a similar extent, but (2) the verbal load impaired object-scene binding significantly more than the rhythmic load. Thus, suppressing verbal resources during encoding selectively disrupts object-scene binding in long-term memory.

Discussion

This study investigated the impact of verbal interference on the binding of objects and scenes in visual long-term memory. Participants observed 40 object-scene pairs while concurrently performing a rhythmic shadowing task, a verbal shadowing task, or no shadowing task. We then tested memory for the objects and scenes (entity memory) and memory for which objects were displayed in which scenes (binding memory). We found that (1) the rhythmic and verbal shadowing tasks impaired entity memory and repeat detection performance to similar levels, but (2) the verbal shadowing task selectively impaired binding memory. We conclude that verbal interference during encoding suppresses binding of objects and scenes in long-term memory.

Note that the rhythmic shadowing task and verbal shadowing task were matched in their impact on attention. Specifically, across the two shadowing conditions, performance was similar on both the repeat detection task and the entity memory task. The crucial difference between the two tasks is that verbal resources are required for one task, but not the other. Thus, these data suggest that language networks are important for binding objects and scenes in long-term memory.

An alternative explanation is that the binding memory task was more difficult than the entity memory task, and accordingly, binding memory suffered more than entity memory when observers were distracted by a secondary task. If this explanation were correct, then binding memory should also have been lower than entity memory when the memory tasks were performed with the rhythmic shadowing task. However, this was not the case. Binding memory and entity memory were nearly identical in the rhythmic shadowing condition (Figure 2, rhythmic shadowing panel), indicating that binding memory tasks are not always more difficult than entity memory tasks in dual-task settings. The impairment for binding memory occurred only when participants performed a concurrent verbal shadowing task.

Why did the verbal shadowing task impair binding memory? One possibility is that verbal shadowing impaired participants' ability to label the entities (e.g., when seeing a banana in a forest, one could remember the words "banana" and "forest"). While this type of verbal labeling process could be helpful for remembering stimuli, our results indicate that participants did not use this strategy. In particular, in the entity memory task, performance for the objects and scenes was similar in the verbal shadowing condition and the rhythmic shadowing condition (Figure 2). Participants were equally good at remembering objects and scenes irrespective of whether they could recruit verbal labeling processes for the task. A second possibility is that verbal shadowing impaired participants' ability to label the relation between entities (e.g., when seeing a banana in a forest, one could remember the sentence "banana in a forest"). This interpretation is consistent with our data, as well as with the previous results from Dassalegn and Landau (2008), who showed related results in location-feature memory. Participants may make use of such a strategy to label objects because while visual memory for entities is extremely good (e.g., Brady et al. 2008), source memory for where an object was in a scene is quite challenging and may require a distinct recollection signal (e.g., Mandler, 1980). Thus, participants may use labels to help with this more difficult memory problem. Additional research is needed, however, to determine whether verbal shadowing disrupted verbal labeling processes versus disrupting more general areas of the language network. For example, language production tasks like verbal shadowing depend heavily on association areas of the brain (frontal and parietal lobes), and these same brain areas might also be needed to bind entities in long-term memory. If so, a verbal shadowing task could disrupt binding memory without necessarily implicating language-specific verbal labeling processes.

From an applied perspective, our study provides insights for the reliability of eyewitness testimony. In eyewitness testimony, accurate associative memory is critical for identifying the perpetrator of a crime without making mistakes, such as mistaking a person seen in an innocent context for a perpetrator of a crime. Because the availability of language is crucial for binding objects and scenes in long-term memory, a witness whose verbal resources are not available while observing a criminal scene (e.g., because they are engaged in a conversation) may be less accurate in his/her associative memory than a witness whose verbal resources are available. Note that our data indicate that these two types of observers could have similar levels of accuracy in remembering the individual elements of the event (i.e., remembering who was seen, where it was, or what happened). Critically, however, an observer with unavailable verbal resources might have less reliable associative memories (i.e., remembering who did what, or what happened where). It would be interesting for future studies to examine whether the patterns obtained in the present study obtain in more realistic, real-world contexts that parallel eyewitness testimony.

In summary, our study reveals that when observers are performing a verbal shadowing task, they have considerable difficulty binding objects and scenes together in visual long-term memory. These results add to the growing body of work showing that language plays an important role in a range of fundamental cognitive abilities, including visual search (Spivey et al., 2001), spatial cognition (Hermer-Vazquez et al., 1999; Pyers et al., 2010), nonverbal false belief reasoning (Newton & de Villiers, 2007), categorical perception

(Winawer et al., 2007), numerical cognition (Frank et al., 2012), and labeling of familiar objects in memory (Lupyan, 2008). Our results suggest that language may play a particularly important role in challenging object-scene binding tasks.

Bibliography

Allen, R. J., Hitch, G. J., Mate, J., & Baddeley, A. D. (2012). Feature binding and attention in working memory: A resolution of previous contradictory findings. The Quarterly Journal of Experimental Psychology, 65(12), 2369-2383.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. Attention and performance IV, 583-612.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. WH Freeman.

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal feature integration in the angular gyrus during episodic and semantic retrieval. Journal of Neuroscience, 36(20), 5462-5471.

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, USA, 105, 14325–14329.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. Journal of Applied Psychology, 62(3), 311-318.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. Current opinion in neurobiology, 16(6), 693-700.

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. Law and Human Behavior, 30(3), 287-307.

Deffenbacher, K. A., Carr, T. H., & Leu, J. R. (1981). Memory for words, pictures, and faces: Retroactive interference, forgetting, and reminiscence. Journal of Experimental Psychology: Human Learning and Memory, 7(4), 299-305.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. Psychological Science, 19(2), 189-195.

Dungan, J., & Saxe, R. (2012). Matched false-belief performance during verbal and nonverbal interference. Cognitive Science, 36, 1148-1156.

Earles, J. L., Kersten, A. W., Curtayne, E. S., & Perle, J. G. (2008). That's the man who did it, or was it a woman? Actor similarity and binding errors in event memory. Psychonomic Bulletin & Review, 15, 1185–1189.

Eichenbaum, H. (2004). Hippocampus: cognitive processes and neural representations that underlie declarative memory. Neuron, 44(1), 109-120.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. Trends in Cognitive Sciences, 12(10), 388-396.

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. Cognitive psychology, 64(1), 74-92.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. Nature Reviews Neuroscience, 6(2), 119-130.

Fries, Pascal, et al. "Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry." Proceedings of the National Academy of Sciences 94.23 (1997): 12699-12704.

Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. Cognitive Psychology, 39, 3-36.

Hollingwoth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 396-411.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(1), 58-69.

Hollingworth, A. (2010). Binding objects to locations: The relationship between object files and visual working memory. Journal of Experimental Psychology: Human Perception and Performance, 36(3), 543-564.

Hollingwoth, A., & Henderson, J. M. (2002) Accurate visual memory for previously attended objects in natural scenes. Journal of Experimental Psychology: Human Perception and Performance, 28(1), 113-136.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. Journal of memory and language, 30(5), 513-541.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. Proceedings of the National Academy of Sciences, 107(25), 11163-11170.

Kersten, A. W., & Earles, J. L. (2010). Effects of aging, distraction, and response pressure on the binding of actors and actions. Psychology and Aging, 25, 620–630.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. Psychological Science, 21, 1551–1556.

Kravitz, D. J., Saleem, K. S., Baker, C, I., Mishkin, M. (2011). A new neural framework for visuospatial processing. Nature Reviews. Neuroscience, 12, 217-230.

Lupyan, G. (2008). From Chair to "Chair": A representational shift account of object labeling effects on memory. Journal of Experimental Psychology: General, 137(2): 348-369.

Loftus, E. F. (1976). Unconscious transference. Law & Psychology Review, 2, 93–98.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological review, 87(3), 252.

Moscovitch, C., Kapur, S., Köhler, S., & Houle, S. (1995). Distinct neural correlates of visual long-term memory for spatial location and object identity: a positron emission tomography study in humans. Proceedings of the National Academy of Sciences, 92(9), 3721-3725.

Newton, A. M., & de Villiers, J. G. (2007). Thinking

while talking: Adults fail nonverbal false-belief reasoning. Psychological Science, 18(7), 574-579.

Newton, A. M., & de Villiers, J. G. (2007). Thinking while talking: Adults fail nonverbal false-belief reasoning. Psychological Science, 18(7), 574-579.

Perfect, T. J., & Harris, L. J. (2003). Adult age differences in unconscious transference: Source confusion or identity blending? Memory & Cognition, 31, 570–580.

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. Proceedings of the National Academy of Sciences, 107(27), 12116-12120.

Ross, D. R., Ceci, S. J., Dunning, D., & Toglia, M. P. (1994). Unconscious transference and mistaken identity: When a witness misidentifies a familiar with innocent person. Journal of Applied Psychology, 79, 918–930.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. Psychological Science, 12, 282-286.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. Psychonomic Science, 19(2), 73-74.

Takashima, A., Nieuwenhuis, I. L., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. Journal of Neuroscience, 29(32), 10087-10093.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cognitive Psychology, 12, 97-136.

Tresch, M. C., Sinnamon, H. M., & Seamon, J. G. (1993). Double dissociation of spatial and object visual memory: Evidence from selective interference in intact human subjects. Neuropsychologia, 31(3), 211-219.

Urgolites, Z. J., & Wood, J. N. (2013a). Visual long-term memory stores high fidelity representations for observed actions. Psychological Science, 24(4), 403-411.

Urgolites, Z. J., & Wood, J. N. (2013b). Binding actions and scenes in visual long-term memory. Psychonomic Bulletin & Review. 20(6), 1246-1252.

Varakin, D. A., & Loschky, L. (2009). Object appearance is not integrated with scene viewpoint in long-term memory. Journal of Vision, 9(8), 565-565.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. Proceedings of the National Academy of Sciences, 104(19), 7780-7785.