Semantic elaboration in auditory and visual spatial memory

Meghan Taevs[1], Louisa Dahmani[1], Robert J. Zatorre[2] and Véronique D. Bohbot[1]*

[1] Department of Psychiatry, Douglas Mental Health University Institute, McGill University, Verdun, QC, Canada
[2] Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

The aim of this study was to investigate the hypothesis that semantic information facilitates auditory and visual spatial learning and memory. An auditory spatial task was administered, whereby healthy participants were placed in the center of a semi-circle that contained an array of speakers where the locations of nameable and non-nameable sounds were learned. In the visual spatial task, locations of pictures of abstract art intermixed with nameable objects were learned by presenting these items in specific locations on a computer screen. Participants took part in both the auditory and visual spatial tasks, which were counterbalanced for order and were learned at the same rate. Results showed that learning and memory for the spatial locations of nameable sounds and pictures was significantly better than for non-nameable stimuli. Interestingly, there was a cross-modal learning effect such that the auditory task facilitated learning of the visual task and vice versa. In conclusion, our results support the hypotheses that the semantic representation of items, as well as the presentation of items in different modalities, facilitate spatial learning and memory.

Keywords: audition, vision, hippocampus, spatial memory, cognitive map

INTRODUCTION

Spatial memory is based on the formation of a cognitive map, i.e., a mental representation of the spatial relationships among various elements in the environment. It has been shown to be critically dependent on the hippocampus (Scoville and Milner, 1957; O'Keefe, 1978; Maguire et al., 1998; Bohbot et al., 2004). It is allocentric, meaning that the relationship between environmental elements or landmarks is constructed independently of the position of the observer. However, it is less well understood whether a cognitive map can be formed of abstract sensory features as readily as a cognitive map based on semantically meaningful elements. In other words, does prior semantic knowledge of elements facilitate the learning of their spatial relationships? Here, we ask whether semantic elaboration has an impact on spatial memory based on a cognitive map.

Previous research in the area of spatial memory and cognition has revealed several factors that influence location memory, such as emotional valence (Crawford and Cacioppo, 2002) and generation and mental rehearsal of words (Slamecka and Graf, 1978; Greene, 1992; Mulligan, 2001; Marsh, 2006). Additionally, object location memory was previously found to be better in women (Silverman and Eals, 1992; Eals and Silverman, 1994; James and Kimura, 1997; Barnfield, 1999), but Choi and L'Hirondelle (2005) suggested that the female advantage was due to superior verbal memory ability, and that this effect disappeared when objects are abstract or unfamiliar. The findings of the above mentioned studies are very intriguing and reflect the current state of object location memory research. Here, we extend this research by raising specific questions pertaining to semantic elaboration and object location memory.

Semantic elaboration can be defined as the process of rehearsal of a stimulus representation in words. It is well known that for many memory tasks, semantic elaboration during learning leads to better recall than does learning without semantic elaboration (Hyde and Jenkins, 1969; Craik and Tulving, 1975; Belmore, 1981; Mennemeier et al., 1992; Brown and Lloyd-Jones, 2006). This effect is often referred to as the levels-of-processing effect (Craik and Lockhart, 1972) and has been shown in several types of verbal recall tasks, usually involving lists of words. It has also been tested for recalling details about nameable pictures (Marks, 1989) and the recall of faces (Anderson and Reder, 1979; Bruce and Young, 1986; Schooler et al., 1996; Brown and Lloyd-Jones, 2006), but it has not yet been investigated for auditory stimuli. It has also not yet been investigated how simply the ability to name a stimulus, which by definition involves more semantic elaboration than perceiving a stimulus without assigning it a name, may aid in storing and retrieving spatial memories. The purpose of this study is to investigate how naming a sound or visual object might lead to better spatial memory, and if it does, whether it is pervasive across modalities.

A study by Marks (1989) examined the degree to which elaborative processing of picture names affects retention of the names and content of the pictures. In this experiment, participants always had initial semantic access by naming the picture, because all the pictures of familiar objects used were easily nameable. It was found that further semantic elaboration, by rehearsal of the name of the picture in a sentence, leads to better name recall and name recognition. Semantic elaboration did not, on the other hand, benefit picture-recognition performance, measured as the participants' ability to remember specific details about the picture. This suggests that while semantic elaboration facilitates memory for picture names, it does not aid memory for perceptual details of the pictures. This study confirms that semantic elaboration does help recall and suggests that labeled pictures are encoded in two separate ways: semantic access to the picture's name and perceptual details of the picture itself. The semantic aspect is aided by semantic elaboration, while memory for the perceptual

details is not. This implies that visual stimuli that can be named may be easier to remember because there are two different routes for encoding.

Klatzky et al. (2002) investigated whether multiple locations of stimuli could be learned from spatial language (a verbal description of the locations) as easily as from auditory and visual perception. The stimuli were all names of objects, either spoken, or written. Stimuli were presented sequentially through a head-mounted virtual reality display for the visual condition, in which the object labels appeared on simulated cards in a particular virtual direction relative to the participant. In the auditory condition, stimuli were presented from loudspeakers at target azimuths. In a third condition, the locations of these same stimuli were described using spatial language. Recall of directions was tested in all groups by using objects' names as probes. The experimenters found that sets of five stimulus locations were learned more slowly using spatial language than using either visual or auditory perceptual cues. The authors suggest that this difference arises because the semantic representation of a place must be converted into a spatial representation. This is different from Marks (1989) because in the latter study, semantic elaboration was used in addition to perceptual representation, whereas in Klatzky et al. (2002) the two representations were presented separately. Nevertheless, results of both studies suggest that there are two possible paths to encoding and recall: the actual stimulus paired with the location (perceptual), and the name of that stimulus paired with the description of the stimulus or the location (semantic). Taken together, these studies suggest that using both pathways is superior to using either one of them, and also that the perceptual pathway alone is better than the semantic pathway alone for spatial learning.

While there is evidence that semantic elaboration plays a role in spatial memory, previous studies assessed it using verbal material. In this study, both verbal and non-verbal material was intermixed into one session in order to directly contrast the two learning conditions. In addition, we replicated our study design with two independent modalities. The aim of this study was to investigate whether naming stimuli facilitates the learning of their locations in both the auditory and visual modalities. We hypothesized that, in both audition and vision, stimuli that are semantically meaningful, i.e., stimuli that can be named in words, would have better spatial encoding and recall than non-semantically meaningful stimuli, i.e., stimuli that cannot be named. We further hypothesized that the advantage of nameable stimuli over non-nameable stimuli would remain despite a practice effect over two sessions with different stimuli in both the auditory and visual modalities.

MATERIALS AND METHODS
PARTICIPANTS
Twenty young healthy participants (12 women, 8 men) with no known vision or hearing problems were recruited. Ages ranged from 20 to 35 (mean = 23.9). Participants were tested in either English or French. Each volunteer participated in two auditory spatial memory sessions and two visual spatial memory sessions, counterbalanced for order of presentation within and across modalities as well as for stimulus set (**Table 1**). Informed consent was obtained from all participants and the experiment was approved by the local ethics committee.

Table 1 | Experimental design.

	First task	Second task
Group 1	Visual	Auditory
	Session 1 (set 1)	Session 1 (set 1)
	Session 2 (set 2)	Session 2 (set 2)
Group 2	Visual	Auditory
	Session 1 (set 3)	Session 1 (set 3)
	Session 2 (set 4)	Session 2 (set 4)
Group 3	Auditory	Visual
	Session 1 (set 2)	Session 1 (set 2)
	Session 2 (set 1)	Session 2 (set 1)
Group 4	Auditory	Visual
	Session 1 (set 4)	Session 1 (set 4)
	Session 2 (set 3)	Session 2 (set 3)

MATERIALS AND APPARATUS
Non-semantically meaningful sounds
Sounds of 1 s duration were used. They had been previously assessed to not be easily nameable in a pilot study (for example chains grinding, baboon call). Three non-semantic stimuli were used for each session, intermixed with three semantic stimuli. Two out of four sets of three stimuli were used for each participant, and the order of presentation was counterbalanced.

Semantically meaningful sounds
Sounds of familiar objects, 1 s in duration, previously demonstrated in a pilot study to be easily nameable (for example bird call, telephone ring) were used. Three semantic stimuli were used for each session, intermixed with three non-semantic stimuli. Two out of four sets of three stimuli were used for each participant, and the order of presentation was counterbalanced.

Non-semantically meaningful pictures
Rectangular abstract colored pictures previously demonstrated not to be readily nameable were used. Three non-semantic stimuli were used for each session, intermixed with three semantic stimuli. Two out of four sets of three non-semantic stimuli were used for each participant, and the order of presentation was counterbalanced. All three non-semantic pictures per set had similar colors.

Semantically meaningful pictures
Black and white pictures of familiar objects (for example apple, acorn) were used. Three semantic stimuli were used per session, intermixed with three non-semantic stimuli. Two out of four sets of three semantic stimuli were used for each participant, and the order of presentation was counterbalanced.

A large auditory array, 2.3 m in diameter, was used to present the stimuli. The array formed a 180° semi-circle with the listener at the center and the speakers evenly spaced out (**Figure 1**). This array consisted of 13 speakers, one of which was placed in the center and the others 15° apart, arranged in one plane at the level of the listener's head. The speakers were wired such that any sound being played by the computer could be directed through a switchboard to a particular speaker. Only six of the 13 possible speaker locations were used per session. The entire array was covered with black fabric

to conceal its three supporting legs and the positions of the speakers from the participants. To indicate both perceived and remembered sound locations, participants pointed with a laser. Location was measured in reference to a discreet paper lining the bottom of the array approximately at the participants' shoulder level, with tick marks every 5° and without any other inscriptions of any kind.

All pictures were presented on a computer monitor using PowerPoint (Microsoft). For the encoding, the screen was effectively divided into 16 separate sections where pictures might appear. For each stimulus set, a PowerPoint presentation was created in which each slide contained one picture. The first six slides showed six pictures in their specific locations (for example, the apple appeared in the bottom right corner of the screen, **Figure 2**). The experimenter scrolled through the frames for a stimulus presentation of approximately 1 s per picture. Following an instruction slide which informed participants that the upcoming trial was a recall trial, six slides were presented containing the same pictures but in a new arrangement, all in the center of the screen. In recall trials, participants used the computer mouse to drag the pictures to the places on the screen where they remembered having seen them during encoding. Encoding slides and recall slides alternated 12 times, always with pictures presented in a different arrangement in each PowerPoint presentation. The two types of trials were separated each time by an instruction slide that indicated whether the following trial would be a "Learning Trial" or "Recall Trial" and that repeated the instructions for that particular type of trial.
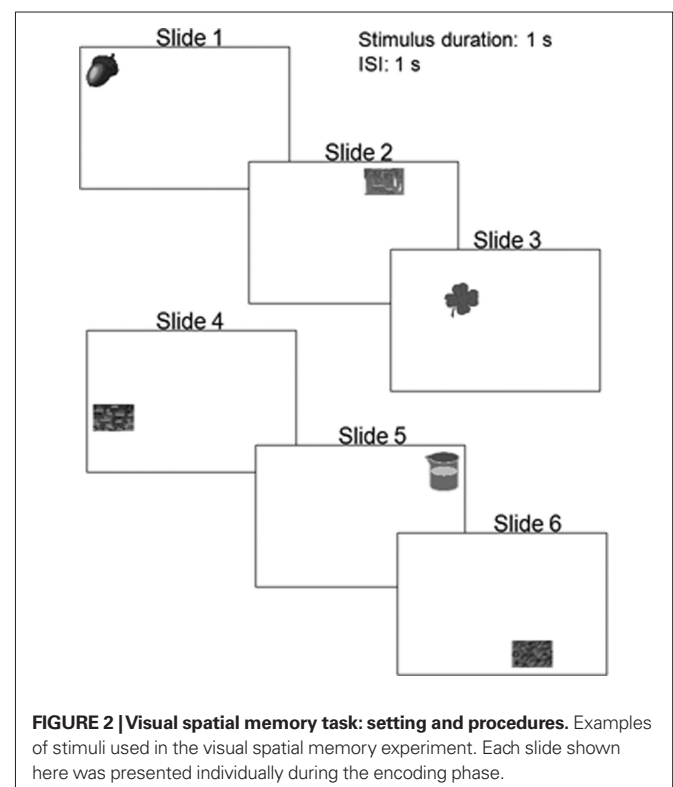
PROCEDURE

Participants first took part in a practice session which included semantic and non-semantic stimuli that were not used in the experimental tasks. In the auditory practice task, participants heard the practice sound in four different locations to get accustomed to localizing sounds, turning their head toward the sounds, and pointing to the locations with the laser. In the visual practice task,
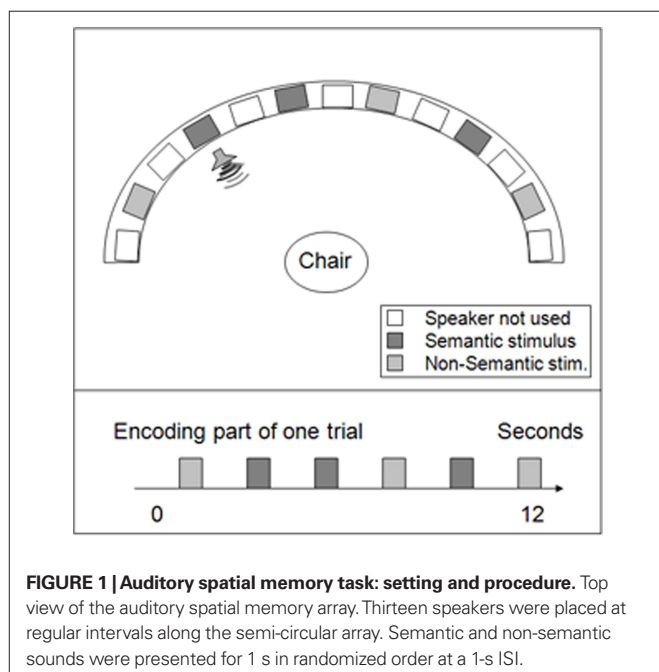
participants saw four practice pictures in the center of the screen to get accustomed to localizing the pictures on the monitor and dragging and dropping icons with the mouse. Following the practice session, participants were administered two experimental sessions in one modality followed by two experimental sessions in the other modality. Each session consisted of 12 trials where participants had to learn the location of the stimuli. Each trial consisted of an encoding and a recall segment. Each session was given with a different set of stimuli. In the auditory-first and visual-first groups, 10 participants heard stimulus sets 1 and 2 (three non-semantic, three semantic for each set) for Sessions 1 and 2. The other 10 participants heard stimulus sets 3 and 4 (three non-semantic, three semantic for each set) for Sessions 1 and 2. Each stimulus set had different stimuli presented to different locations. The order of presentation was counterbalanced. For example, five of the participants hearing stimulus sets 1 and 2 heard set 1 in the first session and set 2 in the second session, and the other five heard set 2 first and then set 1. The same was true for sets 3 and 4.

Encoding phase

In the auditory task, the three non-semantic and three semantic sounds, randomly interspersed, were presented to specific speakers in the array, one after the other, with a 1 s inter-stimulus interval. Participants were instructed to try to remember the locations of sounds as precisely as possible, but not to pay attention to the order of presentation. They turned their heads toward each sound as it was being played and pointed the laser to its location. After each sound, participants returned head and laser to the forward position. Localization of the stimuli was recorded on the first learning trial only and was used to assess recall performance. The rationale for

FIGURE 1 | Auditory spatial memory task: setting and procedure. Top view of the auditory spatial memory array. Thirteen speakers were placed at regular intervals along the semi-circular array. Semantic and non-semantic sounds were presented for 1 s in randomized order at a 1-s ISI.

FIGURE 2 | Visual spatial memory task: setting and procedures. Examples of stimuli used in the visual spatial memory experiment. Each slide shown here was presented individually during the encoding phase.

doing so was that the introduction of a localization session within the training session would have introduced further uncontrolled opportunities to encode the location information. Furthermore, should localization have gotten better with practice, this would have no effect on the learning curves due to the criteria used to assess correct performance and consequently this would have no effect on the data reported in this paper. In the visual task, both non-semantic and semantic pictures, randomly interspersed, were presented at specific locations on the monitor. Participants were instructed to try to remember the locations of the pictures as precisely as possible and to ignore the order of presentation.

Recall phase

In the auditory task, the same stimuli as in the encoding phase were presented in a different randomized order through headphones. Participants turned their heads and used the laser to indicate the location from which they had previously heard the sound come. In the visual task, the stimuli were presented in a different (randomized) order in the center of the computer screen. Participants indicated remembered locations by clicking with the mouse and dragging the pictures to where they remembered them appearing in the previous encoding trial. Encoding and recall trials were alternated for each participant until recall was correct for all sounds on two trials in a row or until 12 trials were completed.

QUESTIONNAIRE

After completion of both the visual and auditory components of the experiment, participants were given a questionnaire in which they indicated what kind of strategies they used in order to remember locations for both auditory and visual stimuli. They were also asked whether they named any stimuli in either modality, and if so, which ones they named. After filling out the questionnaire, participants were debriefed with a written explanation of the experiment and the opportunity to ask any questions they may have had.

DEPENDENT VARIABLES

Two dependent variables were used as a measure of recall: trials to criterion (TTC) and the number of correct locations on trial 3 (T3). Criterion was reached when all stimuli were recalled correctly on two trials in a row (the mean number of TTC is 6), with a maximum of 12 trials per session. The sound locations were judged to be correct if the participant pointed anywhere between the actual location and the perceived location (recorded on trial 1 during the localization of the stimuli), plus 5° on either side. For the pictures, locations were judged to be correct when placed within 0.25″ from the actual location in any direction on the PowerPoint ruler. The number of correct locations on T3 was used as a dependent variable because it was a mid-point to criterion.

ANALYSIS

A mixed model ANOVA was conducted for both the visual and auditory task in order to investigate whether participants who started with one modality task or the other differed in performance, taking into account practice and the semantic property of stimuli. Session (Session 1 vs. Session 2) and Semantic Property (semantic vs. non-semantic stimuli) were considered within-subjects factors and First Task (auditory-first vs. visual-first) was considered

a between-subjects factor in the analysis. Effect sizes for within- and between-subjects factors were calculated using the within- and between-variances as denominators, respectively, as suggested in Salkind (2010). We will consider semantic elaboration and cross-modal effects separately.
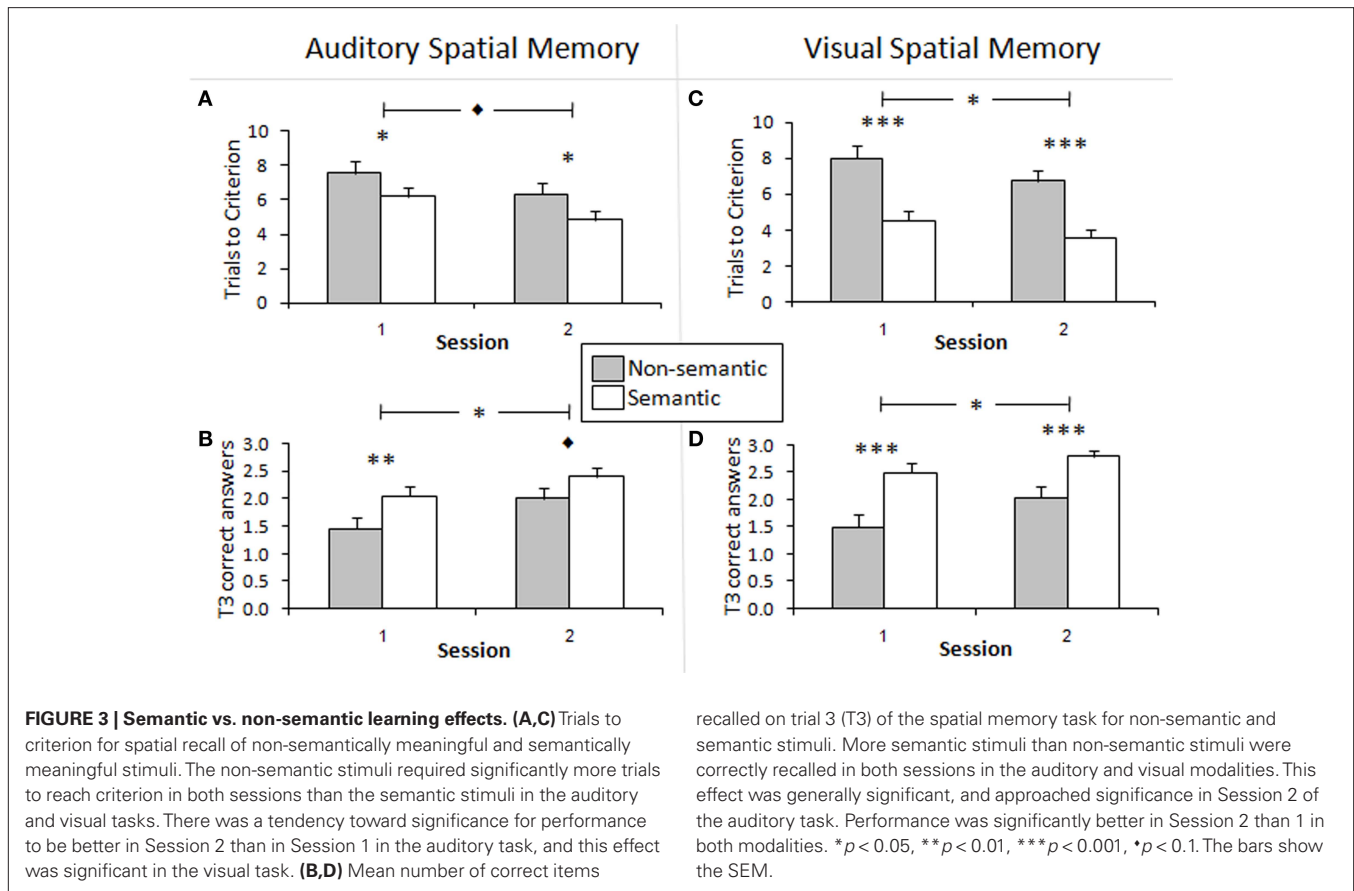
RESULTS

SEMANTIC ELABORATION

For both auditory and visual modalities, there was a significant difference on both measures of recall for non-semantic vs. semantic stimuli, with the semantic recall being better (fewer TTC; higher number of correct answers on the third trial). Participants' performance improved in Session 2 relative to Session 1, whereby fewer TTC were required and number of correct answers on the third trial increased. Spatial memory for semantic stimuli remained significantly better than spatial memory for non-semantic stimuli in Sessions 1 and 2 in both modalities.

In the auditory task, the main effects showed that participants had better performance in terms of TTC with semantic rather than with non-semantic stimuli, indicating that the location of semantic sounds was easier to remember $[F(1,18) = 9.43, p < 0.01, \eta^2 = 0.09]$. The same was found in terms of number of correct answers on T3 $[F(1,18) = 9.18, p < 0.01, \eta^2 = 0.12]$. When looking at practice effects, there was a significant difference in performance in terms of correct responses on T3, where participants performed better in Session 2 than in Session 1 $[F(1,18) = 6.13, p < 0.05, \eta^2 = 0.10;$ **Figure 3B**]. In addition, participants performed better in terms of TTC in Session 2 compared to Session 1 (**Figure 3A**), which closely approached statistical significance $[F(1,18) = 3.79, p = 0.067, \eta^2 = 0.07]$. Thus, participants got better with practice in the auditory task.

Paired-wise comparisons using paired samples t-tests were conducted in order to determine the differences in semantic property within each session. In Session 1, there was a significant difference between semantic and non-semantic stimuli in terms of both TTC $[t(19) = 2.17, p < 0.05;$ **Figure 3A**] and correct answers on T3 $[t(19) = -2.70, p < 0.01;$ **Figure 3B**]. In Session 2, there was also a significant difference between the two types of stimuli in terms of TTC $[t(19) = 1.78, p < 0.05;$ **Figure 3A**]. This difference approached significance for correct answers on T3 $[t(19) = -1.63, p = 0.06;$ **Figure 3B**].

In the visual task, the results paralleled those of the auditory task. The main effects showed that participants had better performance in terms of TTC with semantic rather than with non-semantic stimuli, indicating that the location of semantic images was easier to remember than that of abstract images $[F(1,18) = 123.67, p < 0.001, \eta^2 = 0.46]$. The same result was found in terms of number of correct answers on T3 $[F(1,18) = 40.39, p < 0.001, \eta^2 = 0.30]$. Thus, it seems that the locations of semantic stimuli, whether they are sounds or images, are recalled better than non-semantic ones. When looking at practice effects, participants performed significantly better in Session 2 compared to Session 1 in terms of both TTC $[F(1,18) = 4.67, p < 0.05, \eta^2 = 0.05;$ **Figure 3C**] and correct answers on T3 $[F(1,18) = 8.10, p < 0.01, \eta^2 = 0.07;$ **Figure 3D**].

Paired-wise comparisons using paired samples t-tests were conducted in order to look at the effect of semantic property within each session. In Session 1, there was a significant difference between semantic and non-semantic stimuli in terms of both TTC

FIGURE 3 | Semantic vs. non-semantic learning effects. (A,C) Trials to criterion for spatial recall of non-semantically meaningful and semantically meaningful stimuli. The non-semantic stimuli required significantly more trials to reach criterion in both sessions than the semantic stimuli in the auditory and visual tasks. There was a tendency toward significance for performance to be better in Session 2 than in Session 1 in the auditory task, and this effect was significant in the visual task. **(B,D)** Mean number of correct items recalled on trial 3 (T3) of the spatial memory task for non-semantic and semantic stimuli. More semantic stimuli than non-semantic stimuli were correctly recalled in both sessions in the auditory and visual modalities. This effect was generally significant, and approached significance in Session 2 of the auditory task. Performance was significantly better in Session 2 than 1 in both modalities. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $\bullet p < 0.1$. The bars show the SEM.

$[t(19) = 5.67, p < 0.001]$ and correct answers on T3 $[t(19) = -3.82, p < 0.001$; **Figures 3C,D**]. In Session 2, the same results were found in terms of TTC $[t(19) = 6.12, p < 0.001]$ and correct answers on T3 $[t(19) = -3.68, p < 0.001$; **Figures 3C,D**].
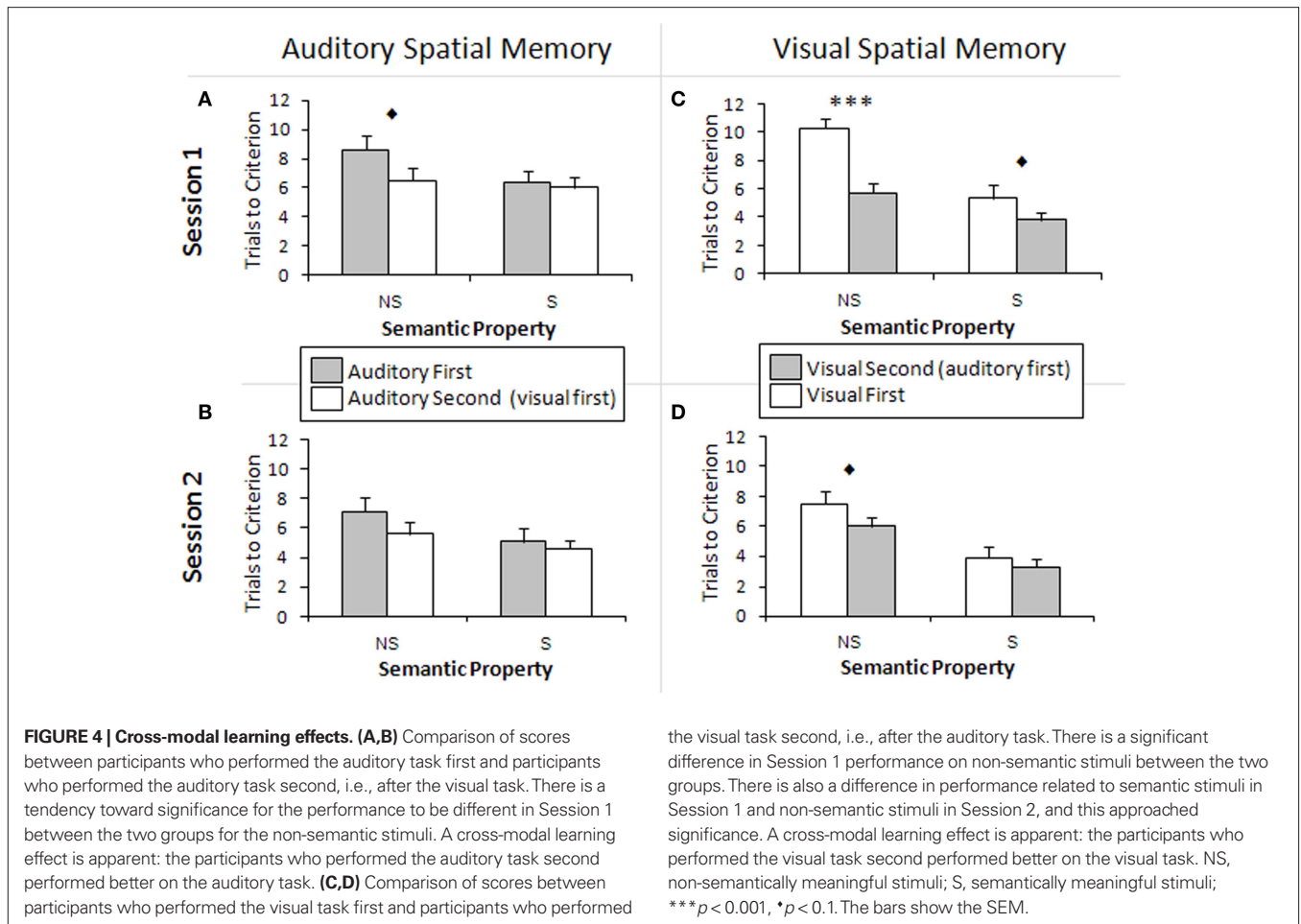
CROSS-MODAL TASK ORDER EFFECTS

A comparison of the auditory scores was made between two groups of participants: those who performed the auditory task first and those who performed the auditory task after the visual task. A comparison of the visual scores was also made between two groups of participants: those who performed the visual task first, and those who performed the visual task after the auditory task (**Figure 4**). In the visual modality, there was a difference in performance between the two groups in both sessions, with a significantly higher number of TTC in the visual-first group than in the visual-second group, showing cross-modal learning effects. Similar findings were observed between the auditory-first group and the auditory-second group. These differences were greater for the visual than the auditory tasks and for the non-semantic stimuli than for the semantic stimuli.

The main effects showed that in the visual modality, participants who performed the visual task after the auditory task required significantly fewer TTC (**Figure 4C**) and obtained more correct answers on T3 than those who performed the visual task first $[$TTC: $F(1,18) = 8.93, p < 0.01, \eta^2 = 0.33$; T3: $F(1,18) = 11.70, p < 0.01, \eta^2 = 0.39]$. Thus, performing the auditory spatial memory task before the visual spatial memory task significantly improved performance on the visual spatial memory task, indicating a cross-modal learning effect. In addition, there was an interaction between Semantic Property and First Task in terms of TTC $[F(1,18) = 11.36, p < 0.01, \eta^2 = 0.04]$, where individuals who performed the visual task second had a smaller difference in TTC between semantic and non-semantic images, compared to the visual-first group. Thus, practice from the same task in a different modality helped bridge the gap between semantic and non-semantic stimuli. There was also a significant interaction between Session and First Task $[F(1,18) = 4.74, p < 0.05, \eta^2 = 0.04]$, where the visual-second group had a smaller difference in T3 correct answers between Sessions 1 and 2 than the visual-first group. This interaction strongly approached significance in terms of TTC $[F(1,18) = 3.86, p = 0.065, \eta^2 = 0.04]$. Thus, the visual-second group benefited less from practice than the visual-first group within the visual task, most likely because they had already benefited from the auditory task. This is supported by the fact that the scores of the visual-second group were still better than those of the visual-first group.

Independent samples t-tests were conducted in order to assess the difference in TTC associated with semantic and non-semantic stimuli between visual-first and second groups. The two groups were found to be significantly different in TTC for non-semantic stimuli in Session 1 $[t(18) = -4.94, p < 0.001$; **Figure 4C**]. This effect approached significance for semantic stimuli in Session 1 $[t(18) = -1.38, p = 0.092$; **Figure 4A**] and non-semantic stimuli in Session 2 $[t(18) = -1.43, p = 0.085$; **Figure 4B**].

FIGURE 4 | Cross-modal learning effects. (A,B) Comparison of scores between participants who performed the auditory task first and participants who performed the auditory task second, i.e., after the visual task. There is a tendency toward significance for the performance to be different in Session 1 between the two groups for the non-semantic stimuli. A cross-modal learning effect is apparent: the participants who performed the auditory task second performed better on the auditory task. **(C,D)** Comparison of scores between participants who performed the visual task first and participants who performed the visual task second, i.e., after the auditory task. There is a significant difference in Session 1 performance on non-semantic stimuli between the two groups. There is also a difference in performance related to semantic stimuli in Session 1 and non-semantic stimuli in Session 2, and this approached significance. A cross-modal learning effect is apparent: the participants who performed the visual task second performed better on the visual task. NS, non-semantically meaningful stimuli; S, semantically meaningful stimuli; ***$p < 0.001$, ♦$p < 0.1$. The bars show the SEM.

In the auditory task, the main effects showed that participants performed better in terms of TTC when they were administered the auditory task after the visual task rather than before, and this effect approached significance [$F(1,18) = 3.39, p = 0.082, \eta^2 = 0.16$]. However, there was no difference in the number of correct answers on T3, $p = 0.842$.

Independent samples t-tests were conducted to assess the difference in TTC related to semantic and non-semantic stimuli between auditory-first and second groups. In Session 1, the auditory-second group performed better with non-semantic stimuli than the auditory-first group (**Figure 4A**) and this effect strongly approached significance [$t(18) = 1.62, p = 0.061$]. The other comparisons were found to be non-significant, $p > 0.05$ (**Figures 4A,B**). This suggests that there is a cross-modal practice effect that transfers from the visual to the auditory task, and that the processing of non-semantic sounds benefits more from this practice than semantic sounds.

QUESTIONNAIRE

In their responses to the questionnaire, all 20 participants reported naming the semantically meaningful sounds, and 10 participants reported naming at least one of the non-semantically meaningful sounds. Many of the names for non-semantic sounds that participants came up with were abstract, e.g., "rustling," "noise," "scaring." The names participants assigned to semantic sounds, when listed, were always correct. The names assigned to non-semantic sounds

were never correct and had no agreement among participants with one exception: two participants reported naming one of the sounds "ocean."

In response to the visual task, 16 of 20 participants reported naming the semantically meaningful pictures, and seven participants reported naming at least one of the non-semantically meaningful pictures. For those who listed them, names of semantic pictures were always correct. The names assigned to the non-semantic pictures usually related to color, e.g., "good green" and "gross green." Only one participant reported naming any of the non-semantic pictures using concrete names of objects ("coleslaw" and "guacamole") rather than color.

DISCUSSION

EFFECT OF SEMANTIC ELABORATION

The locations of semantically meaningful stimuli were easier to learn and remember than those of non-semantically meaningful stimuli when presented in both the auditory and visual modalities. This finding indicates that naming a stimulus, which involves basic semantic elaboration, leads to better spatial memory for sounds and pictures than does merely perceiving a stimulus. One possible explanation is that the representation of semantic objects can be accessed more readily than that for non-semantic objects. This easier access can then help retrieve the location of the object from a cognitive map. Alternatively, the formation of a cognitive map

could be facilitated by the pre-existing semantic information of its individual components. This is supported by a study by Hardt and Nadel (2009), in which the authors showed that people build a cognitive map using concrete cues present in the environment but not abstract paintings, in an adaptation of the Morris Water Maze where both types of cues are available. However, they are able to incorporate the abstract paintings into a cognitive map when asked to do so. Thus, people have a tendency to use concrete nameable cues to construct a cognitive representation of an environment, although they can also use abstract cues.

Half of the participants reported producing names for the non-semantic sounds; fewer named the non-semantic pictures. Participants generally reported naming the non-semantic sounds after they named the semantic sounds. Because familiar objects already have names, they would likely have been learned first. Another possibility is that participants ignored the non-semantic stimuli and began with learning the semantic stimuli because they had pre-existing names. The non-semantic stimuli could be assigned names later, rendering them easier to remember than if a mere perceptual representation was used.

Several factors can contribute to the faster learning of the location of nameable stimuli. For example, as suggested by previous research, naming a stimulus likely provides a secondary pathway for encoding, thus supplementing the direct perceptual pathway. Marks (1989) showed that elaborative processing of picture names aids in later retention of the picture names, but not in recall of the perceptual details of the pictures. Jones (1974) showed that pictorial representations of paired associates during encoding led to significant improvements in recall relative to recall for the words alone, supporting the two-pathway hypothesis for better memory. Klatzky et al. (2002) showed that spatial locations of words can be learned perceptually through vision or audition as well as semantically from a verbal description of the locations. The results of Marks (1989), Jones (1974), and Klatzky et al. (2002) suggest that using two pathways to encode memories is superior to using only one in terms of how quickly information about a stimulus is learned and how well it is later remembered. This experiment provides direct support for this concept. The non-semantic stimuli had only one encoding pathway, which was perceptual, at least at the beginning of the learning phase. The semantic stimuli, which were readily nameable, provided the immediate opportunity to use two pathways: the perceptual pathway and the semantic pathway, both of which would converge onto regions critical for spatial learning. The fact that semantic stimuli were easier to learn than the non-semantic stimuli could be explained by the use of two pathways to encode the stimuli, instead of one.

In addition, the difference between the learning rate of the location of semantic and non-semantic stimuli may be related to the fact that the semantic stimuli are more familiar and consequently may appear more distinct to the participant. Familiarity and distinctiveness have both been shown to lead to a stronger semantic representation and better recognition memory of visual stimuli (Valentine and Bruce, 1986; Gauthier and Tarr, 1997). Conversely, it has also been suggested that elaborative processing leads to more distinctiveness, and distinctiveness alone produces a better memory trace (Marks, 1989). The semantically meaningful stimuli in this study were nameable, but were also more familiar and more distinct than the non-semantically meaningful stimuli. It is therefore possible that it is easier to form a spatial map of familiar stimuli than non-familiar stimuli, which would lead to faster learning of the semantic pictures and sounds. Although the semantic stimuli were pictures of familiar objects or sounds made by familiar objects, the pictures and sounds themselves were new to each participant at the beginning of each session. Familiarity and distinctiveness were not intended to be dissociated from nameability in this study, and all three factors likely contributed to the difference in the learning rate between non-semantic and semantic stimuli.

Alternatively, research participants may not have formed a spatial map of the stimuli, but may have used paired associations between stimulus and location instead. Verbalizing the location could have made it easier to form a paired association between the name of the location and the name of a stimulus than with a perceptual representation of the stimulus, as both the stimulus and its location would be represented in the same way. We know from verbal reports that names were sometimes produced for the non-semantic stimuli. The question is whether participants also used a verbal description of the locations. Based on participants' reports, we found that this was not the case. No participant reported naming the locations in the auditory modality or in the visual modality. This suggests that the verbal label was generated toward developing a semantic representation of the stimuli and was not used to learn the location itself. In fact, all participants reported using purely visual or spatial strategies to remember the locations of the stimuli, and many reported using visual cues on the computer monitor or of folds in the curtain covering the auditory array. Several participants reported forming a "spatial map" or "auditory map" in their heads.

Additionally, we asked whether participants could have confused the various non-semantic stimuli. If they had, we would have expected correct memory for locations and incorrect identification of the objects occupying these locations. However, this was not the case as locations of semantic stimuli were learned very precisely with relative ease, in contrast to the locations of non-semantic stimuli which remained less precise for a greater number of trials in both modalities. In other words, rather than merely swapping the locations of two non-semantic stimuli, participants were less precise in remembering the locations of non-semantic stimuli than those of semantic stimuli. This is important as it implies that the difference between recall for locations of non-semantic and semantic stimuli is not due to difficulties in recognizing the more abstract, less familiar non-semantic stimuli.

Finally, the performance discrepancy between semantic and non-semantic stimuli was smaller in the auditory task than in the visual task. A potential explanation for this effect is that the non-semantically meaningful sounds, although hard to recognize (as in the baboon call and the chains grinding), are not inherently abstract sounds. In the case of non-semantically meaningful pictures, these were fully abstract since they did not represent any kind of object. More participants tried to name non-semantic sounds than non-semantic images. Thus, participants could have identified the non-semantic sounds to a greater extent, which would result in a performance that was closer to the performance related to semantic sounds.

Based on the results of this study, we may conclude that naming stimuli, an elementary form of semantic elaboration, can facilitate spatial memory and the formation of a cognitive map. It can thus improve memory for the specific locations of those stimuli and render this memory superior to that of the locations of non-nameable stimuli.

CROSS-MODAL LEARNING EFFECTS

The order of presentation was counterbalanced, so that half of the participants performed the visual task first (visual-first) while the other half performed the visual task after the auditory task (auditory-first). The results in both modalities represent an average of these two groups, but if the groups are examined separately, a cross-modal practice effect is seen.

Performance on the visual task was better for the visual-second group than for the visual-first group in both sessions, especially for the non-semantic stimuli. This indicates that there is an effect of practice from the auditory task, and that the effect transfers from the auditory modality to the visual modality. Similarly, performance on the auditory task was better for the auditory-second group than for the auditory-first group in both sessions, and again there was a greater difference for non-semantic stimuli than for semantic stimuli. This result parallels that of the auditory task and implies that there is an effect of practice on the auditory task that transfers from the auditory to the visual modality. Thus, this practice effect crosses modalities in both directions; auditory practice extends to the visual task, and visual practice extends to the auditory task. However, the cross-modal effect was smaller in the auditory task. This could be explained by the fact that cross-modal practice seems to benefit non-semantic processing more than semantic processing. Taking part in the auditory or visual task second resulted in enhanced spatial localization for non-semantic stimuli, whereas localization accuracy of semantic stimuli generally closely resembled that of the group that was performing the task as their first task. As mentioned above, performance related to non-semantic sounds was more similar to semantic sounds, in comparison to performance associated with non-semantic and semantic images, which was more different. Thus, in the auditory task, there was less room for amelioration, since non-semantic sounds were not completely abstract and were more similar to semantic sounds. In summary, the similarity between semantic and non-semantic sounds limited the emergence of a bigger cross-modal effect, which affects non-semantic stimuli more than semantic stimuli.

There was also an effect of practice within each task. Performance was better in the second session than in the first session. Interestingly, the visual-second group benefited less from practicing the visual task from Session 1 to Session 2 than the vision-first group. Since their overall performance was still better than the visual-first group, it implies that they initially gained greater practice effects from the auditory task. Thus, once they were administered the visual task, they had already improved performance, indicating that the learning curve is steeper early on.

The observed practice effects could be related to non-specific factors, such as an increase in comfort level associated with the testing, or perhaps we have tapped into something more interesting. For example, people may develop a new strategy or technique for

performing the task during their very first session, and this strategy may be applicable in subsequent sessions. This is especially apparent in the large difference between the two groups for the non-semantic stimuli (**Figure 4A**). The fact that there is little difference between memory for locations of non-semantic and semantic sounds in the auditory-second group implies that people develop a specific technique for remembering non-semantic stimuli.

Another potential explanation for these cross-modality effects involves mental imagery, the process of bringing perceptual information to consciousness. Many neuroimaging and lesion studies have reported that the same brain areas are recruited during perception and mental imagery (Kosslyn et al., 2001), whether it be in the auditory (Zatorre and Halpern, 1993; Zatorre et al., 1996; Halpern and Zatorre, 1999) or visual modality (Farah, 1984; Levine et al., 1985; De Vreese, 1991; Young et al., 1994; Chatterjee and Southwood, 1995; O'Craven and Kanwisher, 2000). When remembering the locations of the stimuli, many participants reported using spatial cues and forming maps in their heads, processes that may very well involve imagery. As such, both the auditory and visual tasks may have required mental spatial imagery. A study by Ghaem et al. (1997) supports this hypothesis. In this study, people learned to navigate along landmarks in a real environment. Later, positron emission tomography (PET) was used to image the brain while people imagined walking down the same path, among imagined landmarks. Mental imagery of the navigation experience led to activation of the hippocampus and neighboring medial temporal lobe regions (Ghaem et al., 1997). The brain areas responsible for encoding and retrieving spatial locations of the stimuli, such as the hippocampus, may then respond better to auditory or visual stimuli due to triggered imagery (Amedi et al., 2005) and lead to better memory because stored representations of the stimuli are accessed more readily. Mental imagery may also explain why performance for semantic stimuli was better than that for non-semantic stimuli, as imagining concrete nameable objects is easier than imagining abstract non-nameable objects.

Various processes like object recognition are inherently multisensory. An object can be characterized by its appearance, its sounds, its texture, its smell, its taste, etc., and thus information is taken from diverse sensory modalities to provide a unified perception of an object. Various brain areas process information from different sensory inputs, which converge and are integrated to form this unified representation (Amedi et al., 2005). Thus, an object presented under different modalities can activate the same representation in the brain. Integration occurs in high-order processing areas but also at the primary cortical level (Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006). Importantly, anatomical connections have been shown to exist between the visual and auditory primary cortices (Falchier et al., 2002; Rockland and Ojima, 2003). Other studies (see Schroeder and Foxe, 2005; Ghazanfar and Schroeder, 2006 for reviews) have found that visual processing sometimes takes place in the auditory cortex and that auditory processing sometimes takes place in the visual cortex (Zangenehpour and Zatorre, 2010). In addition, Schneider et al. (2008) have demonstrated that object identification was facilitated by cross-modal priming in the auditory and visual modalities.

Moreover, in the macaque monkey, it was shown that visual, auditory, somatosensory, and multimodal association areas send projections to the entorhinal, perirhinal, and parahippocampal cortices (Jones and Powell, 1970; Van Hoesen and Pandya, 1975; Van Hoesen et al., 1975; Seltzer and Pandya, 1976; Mesulam and Mufson, 1982; Mufson and Mesulam, 1982; Suzuki and Amaral, 1994a). These cortices then relay the sensory inputs to the hippocampus (Van Hoesen and Pandya, 1975; Van Hoesen et al., 1975; Insausti et al., 1987; Suzuki and Amaral, 1994b). In addition, the hippocampus projects back to various association cortices, including the orbitofrontal, medial frontal, anterior temporal, and posterior temporal association cortices (Rosene and Van Hosen, 1987; Van Hoesen, 1995). Some of these outputs project directly to the association cortices, however most are relayed through the entorhinal cortex (Kosel et al., 1982; Van Hoesen, 1995). These relays of information to and from the hippocampus may therefore prime the hippocampus and make it more sensitive to upcoming stimulation. Since object location activates the hippocampus, it is possible that the presentation of an object location task in one modality primes the hippocampus and accelerates processing involved in locating objects in a different modality, therefore accounting for the observed cross-modal effects.

Thus, visual imagery, hippocampal activation, and cross-modal processing in unimodal areas and associative cortex may all serve as priming processes which help produce better "what" and "where" representations of the presented stimuli, as well as their conjunction in a cognitive map. This in turn enhances spatial learning and recall.

CONCLUSION

Naming sounds and pictures, a form of semantic elaboration, makes the locations of sounds and pictures easier to remember than if the stimuli are merely perceived without names. This is true for both the visual and auditory modalities, even after some practice with the task. Additionally, taking part in an object location task in one modality seems to enhance performance of the same task in a different modality, indicating that spatial learning is cross-modal. In summary, these results suggest that the semantic representation of auditory or visual stimuli, in addition to their representation in different modalities, facilitate the formation of a cognitive map. Further research is necessary in order to elucidate the mechanisms involved in the construction of a cognitive map.

REFERENCES

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., and Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res.* 166, 559–571.

Anderson, J. R., and Reder, L. M. (1979). "An elaborative processing explanation of depth of processing," in *Levels of Processing in Human Memory*, eds L. S. Cermak and F. I. M. Craik (Hillsdale, NJ: Lawrence Erlbaum Associates), 385–395.

Barnfield, A. M. (1999). Development of sex differences in spatial memory. *Percept. Mot. Skills* 89, 339–350.

Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnesia for words. *J. Exp. Psychol. Hum. Learn. Mem.* 7, 191–203.

Bohbot, V. D., Iaria, G., and Petrides, M. (2004). Hippocampal function and spatial memory: evidence from functional neuroimaging in healthy participants and performance of patients with medial temporal lobe resections. *Neuropsychology* 18, 418–425.

Brown, C., and Lloyd-Jones, T. J. (2006). Beneficial effects of verbalization and visual distinctiveness on remembering and knowing faces. *Mem. Cognit.* 34, 277–286.

Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77(Pt 3), 305–327.

Chatterjee, A., and Southwood, M. H. (1995). Cortical blindness and visual imagery. *Neurology* 45, 2189–2195.

Choi, J., and L'Hirondelle, N. (2005). Object location memory: a direct test of the verbal memory hypothesis. *Learn. Individ. Differ.* 15, 237–245.

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684.

Craik, F. I. M., and Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *J. Exp. Psychol.* 104, 268–294.

Crawford, L. E., and Cacioppo, J. T. (2002). Learning where to look for danger: integrating affective and spatial information. *Psychol. Sci.* 13, 449–453.

De Vreese, L. P. (1991). Two systems for colour-naming defects: verbal disconnection vs colour imagery disorder. *Neuropsychologia* 29, 1–18.

Eals, M., and Silverman, I. (1994). The hunter–gatherer theory of spatial sex differences: proximate factors mediating the female advantage in recall of object arrays. *Ethol. Sociobiol.* 15, 95–105.

Falchier, A., Clavagnier, S., Barone, P., and Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.* 22, 5749–5759.

Farah, M. J. (1984). The neurological basis of mental imagery: a componential analysis. *Cognition* 18, 245–272.

Gauthier, I., and Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res.* 37, 1673–1682.

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., and Denis, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport* 8, 739–744.

Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285.

Greene, R. L. (1992). *Human Memory: Paradigms and Paradoxes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Halpern, A. R., and Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex* 9, 697–704.

Hardt, O., and Nadel, L. (2009). Cognitive maps and attention. *Prog. Brain Res.* 176, 181–194.

Hyde, T. S., and Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *J. Exp. Psychol.* 82, 472–481.

Insausti, R., Amaral, D. G., and Cowan, W. M. (1987). The entorhinal cortex of the monkey: II. Cortical afferents. *J. Comp. Neurol.* 264, 356–395.

James, T. W., and Kimura, D. (1997). Sex differences in remembering the locations of objects in an array: location-shifts versus location-exchanges. *Evol. Hum. Behav.* 18, 155–163.

Jones, E. G., and Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820.

Jones, M. K. (1974). Imagery as a mnemonic aid after left temporal lobectomy: contrast between material-specific and generalized memory disorders. *Neuropsychologia* 12, 21–30.

Klatzky, R. L., Lippa, Y., Loomis, J. M., and Golledge, R. G. (2002). Learning directions of objects specified by vision, spatial audition, or auditory spatial language. *Learn. Mem.* 9, 364–367.

Kosel, K. C., Van Hoesen, G. W., and Rosene, D. L. (1982). Non-hippocampal cortical projections from the entorhinal cortex in the rat and rhesus monkey. *Brain Res.* 244, 201–213.

Kosslyn, S. M., Ganis, G., and Thompson, W. L. (2001). Neural foundations of imagery. *Nat. Rev. Neurosci.* 2, 635–642.

Levine, D. N., Warach, J., and Farah, M. (1985). Two visual systems in mental imagery: dissociation of "what" and "where" in imagery disorders due to bilateral posterior cerebral lesions. *Neurology* 35, 1010–1018.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., and O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science* 280, 921–924.

Marks, W. (1989). Elaborative processing of pictures in verbal domains. *Mem. Cognit.* 17, 662–672.

Marsh, E. J. (2006). When does generation enhance memory for location? *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1216–1220.

Mennemeier, M., Fennell, E., Valenstein, E., and Heilman, K. M. (1992). Contributions of the left intralaminar and medial thalamic nuclei to

memory. Comparisons and report of a case. *Arch. Neurol.* 10, 1050–1058.

Mesulam, M. M., and Mufson, E. J. (1982). Insula of the old world monkey. III: efferent cortical output and comments on function. *J. Comp. Neurol.* 212, 38–52.

Mufson, E. J., and Mesulam, M. M. (1982). Insula of the old world monkey. II: afferent cortical input and comments on the claustrum. *J. Comp. Neurol.* 212, 23–37.

Mulligan, N. W. (2001). Generation and hypermnesia. *J. Exp. Psychol. Learn Mem. Cogn.* 27, 436–450.

O'Craven, K. M., and Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci.* 12, 1013–1023.

O'Keefe, J. N. L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Clarendon.

Rockland, K. S., and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Int. J. Psychophysiol.* 50, 19–26.

Rosene, D. L., and Van Hosen, G. W. (1987). "The hippocampal formation of the primate brain: a review of some comparative aspects of cytoarchitecture and connections," in *Cerebral Cortex. Further Aspects of Cortical Function, Including Hippocampus*, Vol. 6, eds E. G. Jones and A. Peters (New York: Plenum Press), 345–456.

Salkind, N. J. (2010). *Encyclopedia of Research Design*, 1st Edn, Vol. 1. Thousand Oaks, CA: Sage Publications.

Schneider, T. R., Engel, A. K., and Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp. Psychol.* 55, 121–132.

Schooler, J. W., Ryan, R. S., and Reder, L. (1996). "The costs and benefits of verbally rehearsing memory for faces," in *Basic and Applied Memory Research, Vol. 2, Practical Applications*, eds D. J. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, and M. K. Johnson (Mahwah, NJ: Lawrence Erlbaum Associates), 51–65.

Schroeder, C. E., and Foxe, J. (2005). Multisensory contributions to low-level, "unisensory" processing. *Curr. Opin. Neurobiol.* 15, 454–458.

Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatr.* 20, 11–21.

Seltzer, B., and Pandya, D. N. (1976). Some cortical projections to the parahippocampal area in the rhesus monkey. *Exp. Neurol.* 50, 146–160.

Silverman, I., and Eals, M. (1992). "Sex differences in spatial abilities: evolutionary theory and data," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, eds J. H. Barkow, L. Kosmides, and J. Tooby (New York: Oxford University Press), 533–549.

Slamecka, N. J., and Graf, P. (1978). The generation effect: delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn.* 4, 592–604.

Suzuki, W. A., and Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J. Comp. Neurol.* 350, 497–533.

Suzuki, W. A., and Amaral, D. G. (1994b). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J. Neurosci.* 14(Pt 2), 1856–1877.

Valentine, T., and Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception* 15, 525–535.

Van Hoesen, G., and Pandya, D. N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res.* 95, 1–24.

Van Hoesen, G., Pandya, D. N., and Butters, N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. II. Frontal lobe afferents. *Brain Res.* 95, 25–38.

Van Hoesen, G. W. (1995). Anatomy of the medial temporal lobe. *Magn. Reson. Imaging* 13, 1047–1055.

Young, A. W., Humphreys, G. W., Riddoch, M. J., Hellawell, D. J., and de Haan, E. H. (1994). Recognition impairments and face imagery. *Neuropsychologia* 32, 693–702.

Zangenehpour, S., and Zatorre, R. J. (2010). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia* 48, 591–600.

Zatorre, R. J., and Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia* 31, 221–232.

Zatorre, R. J., Halpern, A. R., Meyer, E., and Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* 8, 29–46.