University of Cyprus

DEPARTMENT OF PSYCHOLOGY

INTEGRATION OF SPATIAL INFORMATION ACROSS VISION AND LANGUAGE

DOCTOR OF PHILOSOPHY DISSERTATION

STEPHANIE N. PANTELIDES

2014

DEPARTMENT OF PSYCHOLOGY

INTEGRATION OF SPATIAL INFORMATION ACROSS VISION AND LANGUAGE

STEPHANIE N. PANTELIDES

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2014

VALIDATION PAGE

Doctoral Candidate: Stephanie N. Pantelides

Doctoral Thesis Title: Integration of spatial information across vision and language

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the ***Department of Psychology*** *and was approved on the 22/05/2014 by the members of the* ***Examination Committee****.*

Examination Committee:

Research Supervisor: ____

Committee Member: ____

Committee Member: ____

Committee Member: ____

Committee Member: ____

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

....

....

ΠΕΡΙΛΗΨΗ

Τα επιστημονικά ευρήματα από ένα μεγάλο αριθμό ερευνών έδειξαν ότι οι χωροταξικές πληροφορίες οι οποίες κωδικοποιούνται μέσω διαφόρων αισθήσεων μπορούν να ενοποιηθούν σε μια κοινή χωροταξική αναπαράσταση στη μνήμη. Εντούτοις, καμιά επιστημονική μελέτη μέχρι την παρούσα στιγμή δεν έχει διερευνήσει κατά πόσο αισθητηριακές πληροφορίες και πληροφορίες οι οποίες κωδικοποιούνται μέσω έμμεσων μορφών χωροταξικής μάθησης όπως είναι η λεκτική κωδικοποίηση, μπορούν να ενοποιηθούν σε μια κοινή χωροταξική αναπαράσταση. Τα πειράματα από την παρούσα ερευνητική μελέτη προέβαλαν αποτελέσματα τα οποία δείχνουν ότι οι άνθρωποι μπορούν να ενοποιήσουν οπτικές και λεκτικές πληροφορίες τις οποίες αποκτήσαν σε διαφορετικές χρονικές στιγμές σε μια κοινή χωροταξική αναπαράσταση κατά την ώρα της εκμάθησης. Εντούτοις, όταν η ενοποίηση είναι δύσκολο να επιτευχθεί κατά την ώρα της εκμάθησης, οι άνθρωποι διατηρούν στη μνήμη τους ξεχωριστές αναπαραστάσεις για τις οπτικές και λεκτικές πληροφορίες μέχρι τη στιγμή της ανάκτησης των πληροφοριών από τη μνήμη κι εφόσον επίδοση σε ένα χωροταξικό έργο επωφελείται από την ύπαρξη μιας ενιαίας αναπαράστασης.

ABSTRACT

A wide body of studies has provided evidence that spatial memories derived through multiple sensory modalities can be integrated to a single memory representation during encoding. However, it is not yet known whether spatial memories acquired through direct perception (e.g., vision) and indirect methods of spatial learning (e.g., language) can also be integrated into a single memory representation. The experiments reported here provided evidence that people can readily integrate visual and verbal spatial memories acquired at distinct learning experiences into a single memory representation during learning. However, when integrating during learning is difficult, as when learning takes place from different perspectives, people keep separate representations until the time of retrieval when an integrated representation can facilitate spatial reasoning.

ACKNOWLEDGMENTS

Foremost, I express my sincere gratitude to my advisor, Dr. Marios N. Avraamides. His guidance, support, motivation, and expertise have been priceless for the completion of this dissertation. I am deeply convinced that I could have never asked for a better advisor and mentor with whom to begin my career. Furthermore, I thank my really good friend Dr. Anna Zapiti for her support, encouragement, and stimulating discussions. Last but not least, I thank my family from the bottom of my heart. I am grateful to my mother Georgia, my father Nicolas, and my brother Marios for their unconditional love, support, and understanding throughout this research project.

DEDICATION

This dissertation is dedicated to my loving family: to my mother Georgia, to my father Nicolas, and to my brother Marios. Their great and unconditional love, support, and encouragement have been priceless for the achievement of this goal.

TABLE OF CONTENTS

ΠΕΙΛΗΨΗ.......... iii

ABSTRACT.......... iv

ACKNOWLEDMENTS.......... v

DEDICATION.......... vi

LIST OF FIGURES.......... ix

LIST OF TABLES.......... xi

CHAPTER 1 - INTRODUCTION.......... 1

Integration of Spatial Information across Vision and Language.......... 1

Goals of the Study and Overview of the Methodology.......... 14

Hypotheses.......... 15

Significance of the study.......... 16

CHAPTER 2.......... 17

Experiment 1.......... 17

Method.......... 19

Results.......... 22

Discussion.......... 23

Experiment 2.......... 24

Method.......... 25

Results.......... 25

Discussion.......... 27

Experiment 3 28

Method 29

Results 29

Discussion 31

Interim General Discussion 32

CHAPTER 3 34

Experiment 4 34

Method 35

Results 35

Discussion 38

CHAPTER 4 – GENERAL DISCUSSION 39

Conclusions 45

Limitations 45

Future directions 46

References 47

Appendix 53

Figure captions 57

LIST OF FIGURES

Figure 1: The layout used in all experiments..

Figure 2: The pointer used for pointing responses in all experiments..

Figure 3: Pointing error as a function of orientation modality and target modality in Experiment 1...

Figure 4: Pointing error as a function of learning order and orientation modality in Experiment 1...

Figure 5: Pointing error as a function of learning order and target modality in Experiment 1.....

Figure 6: Reaction time as a function of orientation modality and target modality in Experiment 1...

Figure 7: Pointing error as a function of learning order, orientation modality, and target modality in Experiment 2..

Figure 8: Pointing error as a function of orientation modality and target modality in Experiment 2..

Figure 9: Reaction time as a function of orientation modality and target modality in Experiment 2...

Figure 10: The layout used in Experiment 3 in the 5-object layouts condition.............................

Figure 11: Pointing error as a function of orientation modality and target modality for 3-object layouts in Experiment 3...

Figure 12: Pointing error as a function of orientation modality and target modality for 5-object layouts in Experiment 3...

Figure 13: Reaction time as a function of orientation modality and target modality for 3-object layouts in Experiment 3...

Figure 14: Reaction time as a function of orientation modality and target modality for 5-object layouts in Experiment 3...

Figure 15: Pointing error as a function of orientation modality and target modality in Experiment 4..

Figure 16: Reaction time as a function of orientation modality and target modality in Experiment 4..

Figure 17: Pointing error as a function of orientation modality and target modality in testing blocks 1 and 2 in Experiment 4...

Figure 18: Pointing error as a function of orientation modality and target modality in testing blocks 3 and 4 in Experiment 4...

Figure 19: Reaction time as a function of orientation modality and target modality in testing blocks 1 and 2 in Experiment 4...

Figure 20: Reaction time as a function of orientation modality and target modality in testing blocks 3 and 4 in Experiment 4...

LIST OF TABLES

Table 1. Summary of the results

Table 2. Sample descriptives using t-test for equality of means

CHAPTER 1 - INTRODUCTION

Integration of Spatial Information across Vision and Language

Often in everyday life people experience at different points in time different locations within the same environment. For instance, when a new student visits the campus for the first few times he/she may notice different buildings each time. A question of interest is whether at the time of encoding new spatial information people integrate it with the information that is presumably maintained in an existing spatial representation. Such integration could occur in two ways: (1) people may assimilate new information into the existing spatial representation, or (2) they can create a new spatial representation in which they recode the already existing spatial information to accommodate newly acquired spatial information (Greenauer, Mello, Kelly, & Avraamides, 2013). However, it could be the case that, in at least some cases, people neither accommodate nor assimilate spatial information during learning but keep instead each encoding episode distinct as a separate memory representation. In this case they could either coordinate spatial information across representations when the task requires doing so or merge the two representations into one at the time of retrieval, once they realize that the task entails using information across representations.

A number of previous studies have provided indirect evidence that people readily integrate individual locations into spatial representations during learning. This evidence comes from the comparison of findings from studies with external scenes where people study the locations of objects that are visible at once (e.g., Shelton & McNamara, 2001) with those from studies with internal scenes where people encode locations sequentially (e.g., Kelly et al., 2007). Studies with external scenes require participants to study object locations from a viewpoint that is external to the layout and provide evidence that people maintain all object locations in a single spatial representation that is held in memory from a preferred orientation. For instance, Shelton and McNamara (2001) had participants view objects that were placed on a square mat within a rectangular room. Participants studied objects from a perspective that was aligned (0°) and from a perspective that was misaligned (135°) with the mat and the walls of the room and were later tested with a series of Judgments of Relative Direction (JRD's; statements of the form "Imagine standing at x, facing y, point to z", where x, y, and z represent objects from the studied layout). Results showed that for both learning conditions participants' performance was superior for the aligned (0°) than for the misaligned (135°) perspective suggesting that during learning, participants organized their memory using a preferred orientation that was aligned with 0°. Many other studies using external scenes have provided results supporting the conclusion that spatial memories are orientation dependent (e.g., Diwadkar & McNamara, 1997; Mou & McNamara, 2002; Shelton & McNamara, 2001; Yamamoto & Shelton, 2009). Consistent with these findings are those from studies with internal scenes where participants study the locations of objects that are placed around them and are therefore not all visible at once (e.g., Kelly et al., 2007). In such experiments, participants study locations by rotating either their head or their whole body to view each object of the layout. In the study by Kelly et al. (2007) participants studied the locations of objects within a round virtual environment and were then tested with JRD's either in the room in which they learned the objects or after they were guided to an adjacent room. In both conditions, participants performed better for judgments in which the imagined perspective was aligned, rather than misaligned, with their initial facing orientation during learning. This finding suggests that in this study as well participants formed a single spatial representation from a preferred orientation to organize their spatial memory despite the fact that they did not experience directly object-to-object relations. These results suggest that when people encode locations sequentially they form similar orientation-dependent representations to those they construct when they experience locations simultaneously (e.g., Mou & McNamara, 2002). Comparable results from studies with external and internal scenes indicate that people have no difficulty integrating into a single representation locations studied sequentially regardless of the presence or absence of temporal segregation in the encoding of each location (e.g., Mou & McNamara, 2002; Shelton & McNamara, 2001; Kelly et al., 2007).

Although the comparable representations from studies with external and internal scenes indicate that people readily integrate individual locations into spatial representations at the time of encoding, evidence from other studies suggest that at least in some cases people maintain a separate representation for each encoding episode.

For example, in one study, Wang and Brockmole (2003a) assessed whether people update with movement the spatial relations of immediate laboratory objects and the distal locations of landmarks in the campus in which the laboratory was situated. Participants learned the locations of objects in the laboratory and they were then asked to point to these objects and to different familiar locations in the campus. Results showed larger configuration errors (i.e., greater standard deviations of the signed pointing errors) when pointing to campus locations compared to laboratory objects. This result suggests that participants must have stored information about the locations of laboratory objects in a distinct representation from the one they had about the campus (Wang & Brockmole, 2003a; Experiment 1). However, as this study only measured the accuracy of pointing, there still exists the possibility that participants integrated laboratory objects and campus landmarks in a single representation but the relations among the campus landmarks were stored with less precision. It could be that participants in the study of Wang and Brockmole (2003a; Experiment 1) relied on an online, transient memory when they pointed to laboratory locations and to on an offline, enduring memory when they pointed to campus landmarks. The transient system of spatial representation represents spatial memories more accurate and with higher precision while the enduring one, maintains more imprecise spatial memories, even in cases at which it represents a structured and familiar environment (e.g., Easton & Sholl, 1995; McNamara, 2003; Mou et al., 2004; Wang, 2000; Wang & Spelke, 2000). Transient and enduring representations work together as people interact with the environment, and people can switch from a transient representation to an enduring one (Waller & Hodgson, 2006; Experiments 3 & 4). Even if participants integrated laboratory and campus locations in a single representation, they could have been less accurate to point to the campus landmarks rather than laboratory locations because the former originated from a coarser enduring representation.

In another study, Wang and Brockmole (2003b) also had participants learn several laboratory object locations and then point to them or to various other locations on the campus while blindfolded. During learning, participants were asked, depending on condition, to physically rotate on a swivel chair to face either a laboratory object or a campus location. After completing a session of rotations they were asked to point to objects from both environments. Results revealed that participants were faster to point to laboratory objects when they turned relative to those objects rather than to campus locations. However, participants' performance for pointing to campus locations and laboratory objects did not differ when they rotated relative to campus locations. This discrepancy of findings for pointing to the two sets of objects most likely indicates that participants maintained distinct representations for laboratory objects and campus locations. However, the similar latencies for pointing to locations in the two environments when participants turned relative to the campus may indicate that, at least for this condition, participants may have relied on an integrated representation they created probably by assimilating the newly learned laboratory locations to an existing spatial memory about the campus. If this was the case, it remains to be explained why participants did not use this integrated representation when they oriented relative to laboratory objects. One explanation is that in the case of nested environments such as those used by Wang and Brockmole (2003a, 2003b), people organize their memories hierarchically (Hirtle & Jonides, 1985; Greenauer & Waller, 2010). Turning to face laboratory objects could have activated only the subordinate representation for the lab causing participants to respond faster for locations maintained in this representation. On the other hand, rotating relative to campus landmarks may have activated the superordinate representation of the campus that also contained the laboratory objects causing participants to respond equally to campus landmarks and laboratory objects. Supporting this possibility are findings by Rieser, Garing, and Young (1994), Kelly et al, (2007; Experiment 3), and Avraamides and Kelly (2010) who provided evidence that physical movement coupled with instructions to visualize a remote environment allows for updating remote locations.

Although the studies by Wang and Brockmole (2003a; 2003b) allow inferences on whether people integrate representations at learning or not, a number of studies examined integration directly by comparing performance for within- vs. between-layout judgments (e.g., Giudice et al., 2009; Montello & Pick, 1993). The typical paradigm used in these studies requires that people learn two spatial layouts and then carry out spatial judgments that entail using spatial relations from within a layout or across the two layouts. Their rationale is that if integration occurs during learning either by assimilation or accommodation then performance should be equally fast and accurate for within- and between-layout judgments because in such a case information will be retrieved from a single spatial representation that includes all locations. However, if distinct representations are maintained until the time of retrieval then between-layout judgments should produce higher errors and longer latencies compared to within-layout judgments due to the need to coordinate the two representations to compute the response. Findings from most studies using within- vs. between-layout comparisons revealed that people perform better at within- compared to between-layout judgments, a finding which suggests that they keep each layout as a separate representation in memory (e.g., Giudice et al, 2009; Experiment 1; Montello & Pick, 1993). For instance, in a study by Giudice et al. (2009) participants encoded 3 objects by vision and 30 seconds later another 3 objects through touch while blindfolded (or vice versa; Experiment 1). Then, they were asked to imagine facing an object and to point to another. The target object could have been encoded from either the same or different modality as the orientation object allowing trials to be categorized as intramodal or intermodal. Results from the experiment showed that participants were faster and more accurate with intramodal judgments compared to intermodal judgments suggesting that they maintained the visual and the haptic targets in separate representations. Notably, when participants experienced, in a follow-up experiment, the visual and haptic objects as a single spatial layout by alternating the encoding of visual and haptic objects, performance was equal across intramodal and intermodal judgments. Overall, the findings of Giudice et al. (2009) indicate that when encoding consists of two learning experiences that are temporally segregated from each other, people are likely to keep a distinct representation for each experience. In contrast, when learning can be viewed as a single experience, then all objects are maintained in a single representation.

The finding that information from different modalities can co-exist in a single representation is consistent with neuroimaging data showing that a common brain area is responsible for the processing of information acquired through multiple senses (e.g., Andersen et al., 1997). Particularly, findings suggest that sensory inputs and copy signals from motor movements converge in the posterior parietal cortex. In this brain region, the spatial locations of goals for movement are coded. According to Andersen et al. (1997), a specific mechanism promotes the integration of these signals into a single spatial representation. For example, in an experiment by Stricanne, Andersen, & Mazzoni (1996) monkeys experienced auditory signals and then performed eye saccadic movements in darkness to the remembered locations of auditory signals. Brain activity was measured in the delay between the offset of the auditory signal and the offset of the saccadic movement triggered by a fixation light. Results showed that during the delay period cells responding to auditory signals in the lateral intraparietal area (LIP; a cortical area included in the posterior parietal cortex) were found to code the location of the auditory signal in eye-centered coordinates, indicating that memory fields of auditory signals were activated during eye saccades. These findings suggest that this group of neurons in LIP shares a single spatial representation for visual and auditory inputs (Striccane et al., 1996) and provide further support that the human brain binds spatial information acquired through different sensory modalities. Similarly, Wolbers, Loomis, Klatzky, Wutte, & Giudice (2011) provided neuroimaging evidence for the same activation / representation of scenes in the PPA that were either encoded by vision or touch, suggesting a common neural region for the spatial processing of scenes.

Similar findings to those of Giudice et al. (2009; Experiment 1) were provided by studies in which participants experienced large scale environments by vision. For example, in a study by Ishikawa and Montello (2006) participants were guided along two routes in a real environment for ten times each. After completing the first three sessions, participants experienced a connecting route between the two routes. At the end of each session, participants were asked to name the four landmarks on each route in the order they experienced them and to estimate directions and route-distances between the four landmarks. After the completion of the fourth session, participants were asked to estimate directions and straight-line distances within- and between the four landmarks. Results revealed that participants' performance (in terms of speed and accuracy) improved over the sessions indicating that they were able to construct spatial relations between the two routes. However, their performance was better for within- compared to between-route landmarks suggesting that participants did not integrate the two routes in a single spatial representation but maintained distinct representations for each route. Other studies in which two environments were learned one after the other also showed that although people are able to relate spatial information between two layouts that were learned at distinct experiences, they perform better at within- compared to between-layout judgments (e.g., Adamou et al., 2013; Montello & Pick, 1993). Such findings may be due to people's exerting cognitive effort to relate spatial information across representations to compute the response at the time of retrieval.

In contrast, results by other studies suggest that temporal segregation does not always lead to better performance for within-layout responses (e.g., Moar & Carleton, 1982; Holding & Holding, 1989). For example, in a study by Holding and Holding (1982), participants experienced pairs of intersecting suburban routes depicted on slides. Each route was experienced either one or three times. In testing, slides with route pairs were presented and the participants' task was to make judgments about the direct distance between the two routes. Their performance across within- and between-route judgments was equally fast and accurate indicating that participants might have integrated the two route networks into a single spatial representation at the time of learning (either by assimilation or accommodation).

These results allow drawing conclusions that besides temporal segregation other factors, such as the learning conditions, may encourage people to integrate spatial locations studied at separate temporal events at the time of learning. For instance, it is likely that when participants experienced pairs of route networks (i.e., Holding and Holding, 1982) they considered that information from the two routes were both relevant to the task because they viewed them as pairs of routes. Therefore, they might have chosen to integrate the two routes in a single spatial representation during learning because they had predicted that it would facilitate their performance. In the study by Ishikawa and Montello (2006), participants experienced the two routes at different times. Therefore, they might have not consider it necessary to integrate the two routes in a single representation prior to retrieval and therefore maintained separate representations until the task required coordinating spatial information across representations.

Overall, findings from these studies indicate that, in most cases, locations that have been encoded with temporal and/or spatial separation are maintained in distinct spatial representations in memory (e.g., Giudice et al., 2009; Ishikawa & Montello, 2006; Montello & Pick, 1993). This may be the case because spatial information experienced much later or is regarded as not being part of the same environment might not be deemed relevant to the task; therefore, people may choose to store it in a separate representation. Thus, it might be the case that people only integrate information when it is necessary. In fact, this might support cognitive economy as by organizing locations in distinct representations people store fewer object-to-object relations in total. Another possibility is that spatial representations are automatically tagged with the time they were formed and the space they referred to and are thus stored in memory by default as distinct episodic memories.

Notably, all previous findings on integration come from studies in which information was encoded through direct perception (e.g., vision, touch). A fundamental characteristic of perception is that when one perceives an object, s/he also has direct experience with its position in space. For example, seeing an object in a room leads to perceiving not only its identity but also where it is exactly. That spatial information may be encoded automatically by the brain is supported by several findings documenting a functional distinction between the dorsal and the ventral pathways in the brain (e.g., Goodale & Milner, 1992; Milner & Goodale, 2008; Schneider, 1969; Ungerleider & Mishkin, 1982). The dorsal pathway ends in the parietal lobe and is characterized as the “where or how pathway” because it is involved in the instant processing of the object’s location relative to the observer for the guidance of human action in space. This pathway is linked to the ventral pathway which moves through to the temporal lobe and is regarded as the “what pathway” because it is involved in object recognition and identification; also, it supplies abstract representations of what people perceive, that can be used at any time later (Goodale & Milner, 1992). Evidence for the distinction between the two visual systems comes from observations of patients with impairments in brain areas associated with object localization or recognition. For instance, patients with visual agnosia who exhibit specific brain damages in the occipitotemporal area usually reveal impairments in recognizing or describing objects or faces but they can efficiently navigate to execute at least, everyday skills (Farah, 1990). In contrast, patients with optic ataxia who exhibit damages in the posterior parietal area reveal impairments in reaching visual targets but they can easily recognize them though (Perenin & Vighetto, 1988). These findings give support to the argument that a visual pathway in the occipitotemporal region is responsible for object identification but not for the processing of spatial location and a distinct visual pathway in the parietal area is responsible for the processing of spatial location but not for object identification (Goodale & Milner, 1992). Milner and Goodale (1995; see also Goodale & Milner, 1992) suggested that although there are distinct processes for visual perception and visual control of actions, the two visual systems interact to produce adaptive behavior. Therefore, given that with direct perception one encodes directly an object’s location in addition to the information s/he gets about its identity, Milner and Goodale’s argument is compatible with the idea that with direct perception spatial information may be encoded automatically.

For some scientists, the automatic encoding of spatial information has an important role in attention and perception. For example, Triesman and Gelade (2000) suggested that one must first encode the spatial location of an object in order to focus attention on it so. Thus, perceiving the location of an object is necessary for identifying its identity. Support for the claims of Treisman and Gelade (2000) comes from earlier studies which provided evidence that people use spatial indexes to which eye movements respond to construct a spatial representation (e.g., Ballard, Hayhoe, Pook, & Rao, 1997). For instance, when perceiving objects in a visual scene people directly get information about the location of each object, which they can use as a cue to direct attention to interesting objects for further processing. Notably, there is empirical evidence which demonstrates that locations are encoded automatically during perception (e.g., Richardson & Spivey, 2000; Spivey & Geng, 2001). For instance, in a study by Spivey and Geng (2001) participants studied four objects presented in the four corners of the display. The display then became blank for a second and came back with only three of the objects. Results showed that when participants were asked what color the missing object was or what direction it was tilted (right or left), they made more eye movements to the blank region where the missing object was previously presented (Spivey & Geng, 2001; Experiment 2). This shows that participants encoded in memory the spatial locations of the four objects and used them during the recall phase, even though the task was not spatial. Similar findings were provided by Richardson and Spivey (2000). In this experiment, participants listened to four auditory messages each originating from a different corner of the computer display. Only the speaker delivering each message was visible to participants. Then they heard a probe audio clip and had to verify its content based on information from one of the four messages they listened to earlier. Results showed that although during the verification phase participants were looking at a blank screen, they made more saccadic movements towards the screen location at which the message containing the relevant information was previously presented (Richardson & Spivey, 2000; Experiment 1). This finding shows that participants encoded in memory the spatial locations of the four messages even when these were irrelevant to the task, and used them as a cue for retrieval, indicating that even with auditory information the locations may be encoded automatically.

Although perception is the primary means of experiencing space, in many occasions in everyday life people are required to reason about space on the basis of information acquired through indirect means, such as language (e.g., when people interpret route directions provided verbally or in text). Behavioral and neuroimaging findings suggest that spatial learning from language is not different from that taking place through direct modalities such as vision and touch (for reviews see Loomis, Klatzky, Avraamides, Lippa, & Golledge, 2007; Loomis, Klatzky, & Giudice, 2013; Avraamides, Mello, & Greenauer, 2013). In fact, a number of theories in spatial cognition claimed that at some point, spatial memories acquired from different modalities are represented in a similar format (e.g., Bryant, 1997; Jackendoff, 1987; Miller & Johnson-laird, 1976). For instance, Bryant (1997) argued that people use a spatial representational system (SRS) to represent object locations encoded from perception and language in either egocentric (i.e., self-to-objects relations are encoded) or allocentric (i.e., object-to-object relations are encoded) coordinate systems. According to Bryant (1997), perceptual and linguistic inputs are primarily processed in systems that are modality-specific, but then information is moved through to the SRS which represents that information in a form that is not necessarily linked to a specific modality (Bryant, 1997). Also, Bryant states that similar processes are involved in the construction of both perceptual and linguistic representations and that perception, speech, or text can support the construction of the final representation, which is referred to as a mental model. The SRS therefore, operates to represent inputs encoded from perception and language that are finally maintained in similar representations (Bryant, 1997).

The claim that people construct similar representations after encoding information from different modalities (i.e., vision, touch, language) is supported by various studies which provided evidence that different modalities lead to functionally equivalent spatial representations (e.g., Avraamides & Kelly, 2010; Giudice, Betty, & Loomis, 2011; Mello, Greenauer, Pantelidou, & Avraamides, under revision; Pantelides, 2010). *Functional equivalence* indicates that performance in a spatial task is independent of the input modality (e.g., Giudice, Betty, & Loomis, 2011; Avraamides & Pantelidou, 2008). Notably, findings by a number of studies revealed that people perform in an equivalent manner in spatial tasks (i.e., JRD's) after they learn locations from vision, touch, and also spatial language (e.g., Avraamides et al., 2004; Giudice et al., 2011). Particularly, two views account for the hypothesis that functional equivalence may result from the convergence of spatial information from different modalities to a single representation (e.g., Loomis, Lippa, Golledge, & Klatzky, 2002). On one hand, the *Common Recoding hypothesis* assumes that equivalence occurs because inputs acquired from different modalities result in representations which are recoded into another modality, most possibly vision (Loomis et al., 2002). On the other hand, the *Amodal hypothesis* states that different encoding modalities lead to the construction of a spatial image that is not linked to any input modality but it's accessible by all modalities (i.e., amodal; see Giudice et al., 2009 for discussion). Recently, Struiksma, Noordzij, and Postma (2009) suggested two ways through which functionally equivalent spatial representations could arise from perception and language; on the one hand, they suggest that functional equivalence may result from combined modal representations that are constructed by similar but distinct processes. On the other hand, they propose that an abstract modal representation that captures both perceptual and linguistic information may be constructed. Struiksma et al. (2009) take side with the second notion and they argue that spatial representations derived from perception and language are supramodal by nature because they capture inputs from both modalities. That is, supramodal representations don't preserve spatial representations specific to each modality but maintain the connections to the input modalities. Thus, the supramodal nature of spatial representations supports that perceptual and linguistic inputs are maintained within a single spatial representation in memory (Struiksma et al., 2009). Consistent with theoretical frameworks (e.g., Struiksma et al., 2009) are findings by neuroimaging studies (e.g., Noordzij, Neggers, Ramsey, & Postma, 2008). In particular, mental scanning studies (e.g., Mellet, Bricogne, Tzourio-Mazoyer, Ghaem, Petit, Zago et al., 2000; Mellet, Bricogne, Crivello, Mazoyer, Denis, & Tzourio-Mazoyer, 2002) provided evidence that when viewing a map or reading a verbal description, common brain areas reveal activation. Also, in an experiment designed by Noordzij et al. (2008), participants viewed a sentence (i.e., verbal context) or a picture (i.e., visual context) in order to verify a spatial sentence. The left supramarginal gyrus (SMG) revealed activation when a spatial sentence was processed either visually or verbally. Hence, Noordzij et al. (2008) suggested that the left SMG is responsible for the construction of a single representation that involves spatial information derived from vision and language (Noordzij et al., 2008).

Regardless of the way functional equivalence arises (i.e., common recoding, amodal representation, supramodal representation), the theory implies that individual locations encoded from different modalities could be easily integrated into the same representation. Moreover, if spatial learning from language is not different from direct modalities such as vision and touch, one should expect similar findings with those from integration studies with perceptual learning (e.g., Giudice et al., 2009; Adamou et al., 2013). However, encoding spatial information from language differs from doing so through direct perception in one fundamental way: in contrast to perception, when people learn information through language they do not experience directly the object's position in the environment. For instance, in cases in which people learn verbal information with statements of the form "There is a ball at 2 o' clock, 3 meters away from you", they get information about the identity (e.g., "there is a ball...) and the location ("...at 2 o' clock, 3 meters away from you) of the object, which they should combine to construct the spatial image of the object occupying a location in the visual scene. To carry out a task that involves the object's location one must first recode the information describing its location from verbal to spatial format. If people choose to translate the verbal code to spatial at the time of learning then they might be more likely to assimilate the verbal information with other information that is already in a spatial format. Alternatively, people may choose to maintain the spatial information acquired through language in a verbal format until they experience additional information through direct perception. In this case, people may choose to recode the already existing verbal memories to spatial and integrate them with the newly learned information (i.e., accommodation; Greenauer et al., 2013). Yet another possibility is that people avoid integrating the information acquired verbally with other spatial information at the time of learning but instead, keep separate modality-specific representations until the time of retrieval when the task will require relating spatial information across representations. In this case, they may either choose to merge the two representations at the time of retrieval or retrieve at each trial the information needed from separate representations specified by the encoding modality. This could especially be the case when integrating information at learning is difficult.

Support for the latter possibility comes from a recent study by Adamou, Avraamides, & Kelly (2013). In two experiments, the authors investigated whether participants would integrate in a single representation two separate sets of objects studied within the same environment but from different orientations or whether they would maintain each layout in a separate representation. In Experiment 1, participants studied two layouts of objects placed around them in a round room. Participants studied each layout facing a different orientation; the two orientations were offset by 150°. Then participants' memory was tested with perspective-taking trials that required participants to imagine facing a target and point to another. Trials involved objects from the same layout (within-layout trials) or different layouts (between-layout trials). Results showed that participants were faster and more accurate at within- than between-layout trials. Notably, for each layout participants' performance was best from the orientation that the layout was encoded. These findings indicate that participants maintained separate orientation-dependent representations for the two layouts and related information across representations at retrieval. It could be that integrating information encoded from different orientations is a difficult task, which encourages participants to maintain separate representations and only to relate information across representations when the task required doing so, at the time of retrieval (Adamou et al., 2013; Experiment 1). In a follow-up experiment in which participants studied the two layouts in a square room with one of the layouts studied from an orientation aligned with the walls of the room, many, but not all participants, performed equally well across within- and between-layout trials. Also, these participants maintained the two layouts in memory from the same preferred orientation suggesting that they might have organized them in a single representation. In contrast, participants who exhibited a performance advantage for within- compared to between-layout trials also showed a different preferred orientation for each layout, matching its learning orientation. Overall, the results of Adamou et al. suggest that people typically maintain separate representations for spatial information learned at different times and only relate information across representations when needed. However, the presence of a stable environmental reference frame may cause some people to integrate all spatial relations into a single representation; this may suggest that integration takes place strategically by taking into account the availability of environmental, and possibly other, cues.

The goal of the present study is to examine these hypotheses in a series of experiments that involve encoding spatial information from perception and language. As in many occasions people need to reason about spatial relations using both visually- and verbally-encoded information (e.g., when people follow driving instructions while driving), knowing whether and under which conditions people integrate spatial information acquired through vision and language will provide insight about the spatial representations that underlie the execution of spatial tasks. So far no research has assessed integration using layouts involving indirect methods of spatial learning such as language.

Goals of the Study and Overview of the Methodology

The current study involves experiments with spatial layouts encoded visually or-verbally to investigate spatial integration across modalities. In particular, the study examines whether people integrate visual and verbal spatial information at the time of learning or whether they maintain separate representations during learning but relate spatial information across representations only when the task requires doing so or merge the two representations into one at the time of retrieval, when they realize that the task entails using information across representations.

All experiments assessed performance with pointing to target objects from imagined perspectives. Depending on condition, the orienting object (i.e., the object defining the imagined perspective) and the target object were encoded as part of the same (i.e., visual-visual, verbal-verbal) or a different layout (i.e., visual-verbal, verbal-visual). Most studies on integration that used the within- vs. between-layout paradigm were based on measures of both pointing accuracy and latency (e.g., Giudice et al., 2009; Ishikawa & Montello, 2006). Pointing error shows the accuracy with which each target is localized in the environment but not the access to spatial relations within- and across representations. Therefore, one could argue that is not a direct measure of integration but instead, a measure which indicates the precision with which people maintain spatial memories. Latency seems to be a more sensitive measure of integration because when a response must be computed by coordinating information across distinct representations, a switching cost should arise. That is, if people directly retrieve locations from a single spatial representation then it shouldn't matter in terms of pointing latency if these relations come from the same or a different layout. For this reason, the current study does not regard accuracy as a measure of integration and relies on latency to evaluate integration. If integration occurs at the time of learning then pointing latency should be similar for within- and between-layout judgments because retrieval will occur from a single spatial representation that includes all spatial locations. This hypothesis predicts similar performance (i.e., in terms of speed of pointing) across within- and between-layout judgments. Notably, although one could integrate in a single spatial representation at the time of learning, the precision for within- and between-layout spatial relations may differ because participants will learn the visual and the verbal layouts at different times. If participants will maintain in separate representations the visual and the verbal targets then overall performance (in terms of latency) is expected to be better for within- compared to between-layout judgments. Between-layout judgments should require extra time because participants will need to switch from one representation to the other in order to retrieve the information needed for the trial. If participants keep separate representations for the targets encoded at different times but they relate spatial information across representations early at retrieval then performance is expected to be superior (in terms of speed and accuracy) for within- compared to between-layout judgments. However, in this case it is expected that although participants will be faster for within- compared to between-layout judgments their latency will most probably improve over testing trials. This means that at the early phases of testing there will be significant performance differences between within- and between-layout judgments that will be reduced or completely disappear over the course of testing.

Hypotheses

Hypothesis 1: Because people must recode the verbal code to spatial at some point, participants are likely to do so at the time of learning and integrate as well the verbal targets with the visual targets (which should be in a spatial format). In the case that participants encode the verbal targets second then they could recode them into spatial format and assimilate them with the existing representation that maintains the visual targets. When they encode the verbal targets first they could either transform them to spatial format during learning or they could do so when they subsequently study the visual objects. In either case, latency for responding to different-layout pairs should be as short as that for responding to same-layout pairs. However, because participants will learn the visual and the verbal layouts at different times it is not unlikely that the precision for allocentric relations is different for same layout and different layout pairs. For example, it could be that there will be more noise about the exact location of objects learned first due to memory decay. If this is the case and as this noise may influence the pattern of latencies as well, of particular interest would be to examine if the learning order also affects the pattern of latencies.

Hypothesis 2: Because participants will learn the visual and the verbal objects at different times they are likely to maintain the visual and verbal objects in separate representations. This outcome should definitely take place if participants choose not to translate the verbal code to spatial but instead maintain a verbal memory for the objects encoded linguistically. This hypothesis predicts longer latencies for between- compared to within-layout judgments due to switching costs in retrieving the visual and the verbal targets from distinct representations and coordinating them to compute the response. As with the previous hypothesis, greater error for between- than within-layout judgments might be observed.

Hypothesis 3: Participants will keep separate representations at the time of learning but they could integrate the visual and the verbal targets in a single representation during the first stages of task execution. This hypothesis assumes inferior performance in terms of latency for between- compared to within-layout judgments initially but equivalent performance later in the task. As with the two previous hypotheses, lower error rates for within- compared to between-layout judgments might be evident.

Significance of the study

The expected findings will extend current findings on multimodal integration. Obtaining findings that favor integration of vision and language will provide further evidence that the efficiency of spatial reasoning is independent of the input modality. Such a conclusion will have important implications for applications in which people are required to reason about their environment after binding spatial information derived through vision and language. For instance, Global Positioning Systems (GPS) often use both visual (i.e., a map) and linguistic information (i.e., verbal descriptions) to communicate navigational information. Through a combination of this information people construct a single representation that includes inputs from both modalities that could result in effortless, and maybe more effective, navigation and wayfinding rather than coordinating spatial information across multiple representations when needed at the time of retrieval. The present study is the first attempt to directly and systematically assess whether and under which conditions visual and linguistic spatial information is integrated in a single representation. Findings from the current study will further our understanding about how the human brain deals with spatial information derived from different modalities.

CHAPTER 2

Experiment 1

Experiment 1 was designed to assess the integration of visual and verbal spatial information that is spatially congruent but temporally segregated during learning by comparing performance for within- and between-layout spatial judgments. Also, Experiment 1 examined whether explicit instructions given by the experimenter to visualize as part of the same environment locations encoded separately would influence the pattern of findings.

According to Giudice et al. (2009), integration relies on spatial and/or temporal congruence of object locations at the time of encoding. Giudice et al. (2009) showed that when participants studied layouts through vision and touch at separate times, but within the same general space, they were faster and more accurate to respond to trials testing within- compared to between-layout relations; this finding suggests that they maintained the two layouts in distinct representations (Experiment 1). However, when visual and haptic locations were intermixed, participants responded equally well for the two types of judgments (Experiment 2); this result indicates that when layouts were not temporally segregated, participants integrated them into a single spatial representation at the time of learning.

The argument by Giudice et al. (2009) refers to different modalities in general but no research so far has assessed whether it applies to spatial learning from non-sensory modalities, such as language. Spatial learning through language is by nature more difficult compared to learning from sensory modalities (e.g., Avraamides et al., 2004). Also, in order to carry out actions in space (e.g., when people follow driving instructions) people must put effort, either at the time of encoding or later, to translate the verbal code to spatial in order to use it. If people perform this translation at the time of encoding then a question that arises is whether at that point they also integrate the verbal information with other information that is already spatial at the time of learning (either through assimilation or accommodation) or whether they maintain it in spatial format but as a distinct episodic memory. An alternative possibility is that people maintain the information learned through language in a verbal format and carry out the translation to spatial format and relate information across representations during retrieval, when the task requires doing so.

The present experiment aimed to investigate these hypotheses by using spatial layouts encoded through vision and language. Objects from the two layouts were presented within the same enclosing space but in two different study sessions separated by a 30-second delay. Thus, while they were spatially congruent, the objects were temporally segregated. If translation to spatial format takes place early, then because of the extra demands that language poses in translating the verbal code to spatial, it might be that participants will be more willing to integrate during learning. If this is the case, similar performance in the latency across within- and between-layout judgments should be expected. But, if spatial learning from language is no different from learning through perceptual modalities then similar findings to those provided by studies assessing integration with layouts encoded through sensory modalities (e.g., Adamou et al., 2013; Giudice et al., 2009) should be obtained; that is, faster performance for within-compared to between-layout judgments documenting the presence of distinct representations (e.g., Giudice et al., 2009; Experiment 1) should be observed. Greater latency for between- than within-layout judgments will reflect the cost in performance associated with retrieving the visual and the verbal targets from two distinct representations and coordinating them to compute the response. However, if people maintain separate representations for the visual and the verbal objects during learning but integrate them in a single representation after some testing (i.e., when they realize that the task requires coordinating information across layouts), then any initial differences in performance between within- and between-layout judgments should be reduced or eliminated as the experiment progresses.

Notably, regardless of whether participants integrate information into a single representation or they maintain separate representations, they may still exhibit poorer precision for relations across layouts due to learning the visual and verbal objects with larger temporal segregation. Therefore, differences in accuracy for between- compared to within-layout judgments may be observed regardless of the pattern of latencies.

Moreover, if integration is strategic then it may be influenced by top-down factors such as following explicit instructions to integrate. Previous studies in the field of multisensory perception have provided evidence that instructions given by the experimenter to consider a visual target (e.g., a face) and an auditory signal (e.g., a voice) perceived at different times as having a common origin or not influence integration of multiple perceptual experiences of a stimulus (i.e., visual, auditory) in a single perceptual experience (e.g., Arnold, Johnston, & Nishida, 2005; Jack & Thurlow, 1973; Warren, Welch, & McCarthy, 1981). For instance, findings by Warren, Welch, & McCarthy (1981) showed that when participants were given instructions to integrate the visual target and the auditory signal in a single percept, they could easily do so. However, when they were given instructions to consider the visual target and the auditory signal as two different events, they maintained separate perceptual experiences for the visual and the auditory presentation of the stimulus. These findings indicate that instructions given to participants most possibly encouraged them either to integrate different perceptual experiences of a stimulus in a single event or maintain separate perceptual experiences. Therefore, such top-down factors may also influence integration of spatial locations encoded through different modalities in a single memory representation. It might be the case that explicit instructions given to participants to consider the visual and the verbal objects encoded at separate times as spatial locations of a single layout would encourage them to integrate them in a single memory representation. To test this hypothesis, in the present experiment half of the participants were given instructions to consider the visual and the verbal objects as locations of a single layout while the remaining participants were not given any instructions. If the case is that integration of locations in spatial memory is influenced by similar factors to those that influence integration in perception then similar performance in the latency across within- and between-layout judgments should be observed when participants receive explicit instructions to integrate. Additionally, longer latencies for within- compared to between-layout judgments should be observed when participants do not receive any instructions. Such a finding would indicate that integration is under volitional control with instructions influencing whether participants integrate the visual and the verbal objects in a single memory representation.

Method

Participants

Forty-four students (ages: 20-33, 23 male) from the University of Cyprus participated in this experiment in exchange of a small monetary compensation (€10). All participants signed an informed consent form before the experiment.

Materials

Participants were tested individually. Objects were learned in a common lab room and verbal descriptions were given by the experimenter. Each layout (i.e., visual and verbal) had three objects so that a total of six objects (ball, wine, pot, shoe, vase, and mug; see Figure 1 in the Appendix) were learned. Participants responded using the mouse on a computerized pointer (Figure 2). Pointing trials were controlled using a Python script within the Panda 3D software.

Design

The experiment followed a mixed-factorial design with learning order (visual layout first vs. verbal layout first) and integration instructions (given vs. not given) as the between-subjects factors and orientation modality (visual vs. verbal) and target modality (visual vs. verbal) as the within-subjects factors. The combination of levels for the orientation and target modalities (i.e., Imagine facing x point to y; x = orientation, y = target) yielded four conditions: visual-visual, verbal-verbal, visual-verbal, and verbal-visual. Each participant carried out a total of 120 trials presented in a different randomized order. After learning the two layouts, half of the participants were instructed to visualize the 6 objects together as a single layout, while the other half were given no instructions after learning.

Procedure

Before the beginning of experimental trials participants completed a short *practice phase* using laboratory objects in order to become familiarized with the nature of the pointing trials. Particularly, they were asked to respond to statements of the form “Imagine facing x, point to y”, where “x” and “y” represented lab objects). For practice trials, participants received corrective feedback. Participants continued to the learning phase as soon as they pointed correctly in the first ten practice trials. Some of the participants needed to make some extra practice trials (2 or 3) in order to reach the pointing criterion of the ten correct practice trials.

For the *learning phase* participants stood in the center of the lab. When participants learned the visual layout first, they were asked to view three objects that were placed around them and memorize their locations. After memorizing the layout participants were asked to point to each object with their eyes closed from the learning standpoint, after rotating 90º to their right or left, and after rotating 135º to their right or left, with the direction of turn counterbalanced across participants. This was done to ensure that participants knew where objects were located in space while they changed their facing direction, and therefore memorized them in their correct locations. After pointing correctly to all objects of each layout from the different perspectives they were guided to an adjacent laboratory and their memory was tested with *memory verification trials.* These trials required participants to point egocentrically to the remembered object locations (e.g., Point to the mug) as if they were standing in the learning standpoint using a computerized pointer (Figure 2). Participants continued pointing until they responded within ±30° for all memorized objects. Then participants were guided back to the learning room to learn the verbal layout. For the verbal layout the experimenter provided the object locations with statements using clock hours, e.g., such as "Imagine that a mug is placed at 1 o'clock". Participants were encouraged to visualize objects encoded verbally in their described positions. After memorizing the layout participants were asked to point to the remembered object locations from the different perspectives and then executed *memory verification trials*. The procedure was reversed when the verbal layout was learned first. The order in which layouts were learned (i.e., visual first vs. verbal first) was manipulated across participants. Before proceeding to testing trials, half of the participants were given instructions to imagine the 6 objects as part of a single layout. The other half were given no instructions. Importantly, participants were not aware during learning of the nature of the testing trials.

Afterwards, participants continued with the *testing phase*. In this phase, participants executed pointing judgments (i.e., "Imagine facing x point to y") on a computer screen. Each participant carried out 4 blocks of 30 trials. Pointing error and response time were logged and used as the dependent measures. At the end of the experiment participants were debriefed and were questioned about their strategies on memorizing the layouts. They were specifically asked to report whether they visualized objects as two distinct layouts or whether they merged the objects into a single layout during learning or at any other point in the experiment.

Results

Data for pointing error and latency were analyzed using a repeated-measures Analysis of Variance (ANOVA) with learning order (visual first vs. verbal first) and integration instructions (given vs. not given) as the between-subjects factors and orientation modality (visual vs. verbal) and target modality (visual vs. verbal) as the within-subjects factors.

The analysis of pointing error showed a significant orientation modality x target modality interaction, $F(1, 40) = 7.23$, $p = .01$, $\eta^2 = .15$. As seen in Figure 3 (see Appendix), when participants oriented to a visual object they were more accurate to point to a visual object rather than to a verbal object, p = .02. In the case that participants oriented to a verbal object, they were more accurate to point to a verbal object rather than to a visual object, p = .04.

The analysis also revealed that the orientation modality x learning order (Figure 4) and the target modality x learning order (Figure 5) interactions were significant, $F(1, 40) = 6.45$, $p = .01$, $\eta^2 = .13$ and $F(1, 40) = 5.06$, $p = .03$, $\eta^2 = .11$ respectively. When participants learned the visual layout first they were more accurate when the orientation object was verbal rather than visual. In the case they learned the verbal layout first they were more accurate when the orientation object was visual rather than verbal. The learning order x target modality interaction revealed that when the layout studied first was visual participants were more accurate to point to a verbal target rather than to a visual target. When the layout studied first was verbal, they were more accurate to point to a visual target rather than to a verbal target.

Although the main effect of learning order was marginally significant, $F(1, 40) = 3.28$, $p = .07$, $\eta^2 = .07$, it revealed that participants' overall performance (in terms of accuracy) was better when the layout studied first was verbal compared to when it was visual. The main effect of integration instructions was not significant, $F(1, 40) = .00$, $p = .96$, $\eta^2 = .00$. The main effects of orientation modality and target modality were also not significant, $F(1, 40) = .06$, $p = .79$, $\eta^2 = .00$ and $F(1, 40) = .14$, $p = .70$, $\eta^2 = .00$ respectively. All other interactions were not significant.

For the latency measure, the analysis revealed significant main effects of orientation modality and target modality, $F(1, 40) = 14.77$, $p = .00$, $\eta^2 = .27$ and $F(1, 40) = 4.17$, $p = .04$,

$\eta^2 = .09$ respectively. Participants were faster to respond when either the orientation object or the target object was visual rather than verbal. The orientation modality x target modality interaction did not approach significance (Figure 6), $F(1, 40) = .30$, $p = .58$, $\eta^2 = .00$. The main effects of learning order and integration instructions were not significant, $F(1, 40) = .01$, $p = .90$, $\eta^2 = .00$ and $F(1, 40) = .18$, $p = .67$, $\eta^2 = .00$ respectively. All other interactions were not significant.

To investigate whether any significant performance differences between within- and between-layout judgments were evident early at retrieval but not afterwards separate analyses were conducted on pointing error and latency by splitting the data into two phases; the first phase included the first two testing blocks (block 1, block 2) and the second phase included the two last testing blocks (block 3, block 4). Data from these analyses showed that, in both phases, participants were similarly fast across within- and between-layout judgments and more accurate for within- compared to between-layout judgments. Also, they were more accurate when the orientation or the target object was encoded in the layout studied first.

Further analyses were conducted comparing performance across within- and between-layout judgments after collapsing data across modalities. The results from these analyses are shown in Table 2 (see Appendix).

Finally, data from participants' reports in the informal interview that took place after testing showed that 37 out of 44 participants reported that they merged the visual and the verbal objects in a single representation at the time of learning. Twenty-seven of those participants studied the visual layout first and 10 participants studied the verbal layout first. The remaining 7 participants reported that they kept the visual and the verbal objects in distinct representations (5 participants studied the visual layout first and 2 participants studied the verbal layout first).

Discussion

Results from Experiment 1 revealed that participants pointed equally fast across within- and between-layout judgments. This was the case regardless of whether they were given explicit instructions to integrate or not. However, they were more accurate at within- compared to between-layout judgments. Also, participants were more accurate when the orientation object or the target object was encoded in the layout presented second, regardless of whether it was visual or verbal, suggesting that precision for an object's location deteriorated with time. In terms of latency, participants were faster when either the orientation object or the target object was visual. Such a finding is compatible with earlier experiments demonstrating that spatial memories from direct perception are generally more vivid, and therefore more easily accessible, than those derived from other modalities and language (e.g., Avraamides et al., 2004). What is more, in the informal interview most participants reported that they integrated the objects from the two layouts in a single memory representation at the time of learning. Corroborating the findings from the informal interview, the similar latency for accessing spatial relations within and across layouts suggests that participants retrieved the visual and the verbal objects from a single spatial representation which they constructed during learning. This was the case regardless of whether participants received instructions to integrate or not, indicating that instructions did not influence their performance, most likely because participants had already integrated the visual and the verbal objects in a single memory representation at the time of encoding. Notably, the spatial relations between objects that were encoded at distinct learning experiences were less precise compared to those for objects encoded together and participants were more accurate for both orienting and pointing to objects encoded in the layout presented second, indicating that integration was not perfect. It might be that the overall lower accuracy for between- compared to within-layout judgments was a result of the lower accuracy for the objects encoded in the layout presented first. This finding may be driven at least in part by memory decay of locations from the first studied layout. Experiment 2 was designed to assess this by refreshing participants' memory for the layout studied first.

Experiment 2

Experiment 2 investigated whether participants' overall lower accuracy for between-layout judgments in Experiment 1 reflects an effect of memory decay for the layout studied first. After learning both layouts participants were asked to point once more to the objects of the layout they studied first (regardless of whether it was visual or verbal) to refresh their memory for that layout.

Method

Participants

Thirty-eight students (ages: 20-25, 8 male) from the University of Cyprus participated in this experiment in exchange of a small monetary compensation (€10). All participants signed an informed consent form before the experiment.

Materials, Design, & Procedure

Everything was identical to Experiment 1 with two notable exceptions. First, participants were not given any instructions to visualize the 6 objects together as results from Experiment 1 revealed that such instructions did not influence performance. Second, after learning both layouts and executing memory verification trials, participants were guided back to the learning room and they were asked to point one more time to the objects of the layout encoded first (visual or verbal depending on the condition they were assigned to) in order to refresh their memory for that layout. This was added to examine whether inferior performance for between- compared to within-layout judgments in Experiment 1 was a result of participants' lower accuracy for the objects of the layout studied first. If this is the case, then refreshing participants' memory for the layout studied first should lead to equivalent performance for within-layout and between-layout judgments in accuracy as well as in latency.

Results

Data for pointing error and latency were analyzed using a repeated-measures ANOVA with learning order (visual first vs. verbal first) as the between-subjects factor and orientation modality (visual vs. verbal) and target modality (visual vs. verbal) as the within-subjects factors.

The analysis of pointing error showed that the main effect of learning order was not significant, $F(1, 36) = .03$, $p = .85$, $\eta^2 = .00$. The main effects of orientation modality and target modality were also not significant, $F(1, 36) = .66$, $p = .42$,$\eta^2 = .01$ and $F(1, 36) = .00$, $p = .99$, $\eta^2 = .00$ respectively. The orientation modality x target modality interaction (Figure 7) was marginally significant, $F(1, 36) = 4.13$, $p = .05$, $\eta^2 = .10$.When participants oriented to a visual object they were numerically more accurate to point to a visual object rather than to a verbal object, but the difference was not significant, $p = .13$. When participants oriented to a verbal object they were numerically more accurate to point to a verbal object rather than to a visual object but the difference was also not significant, $p = .21$.

Importantly, the analysis revealed a significant learning order x orientation modality x target modality interaction, $F(1, 36) = 6.97$, $p = .01$, $\eta^2 = .16$. As shown in Figure 8 (see Appendix), in the case that participants studied the visual layout first, when they oriented to a visual object they were equally accurate to point to a visual and to a verbal object, $p = .83$. Similarly, when participants oriented to a verbal object they were equally accurate to point to a verbal and to a visual object, $p = .71$. However, in the case that participants studied the verbal layout first, when they oriented to a visual object they were more accurate to point to a visual object rather than to a verbal object, $p = .02$. When participants oriented to a verbal object they were more accurate to point to a verbal object rather than to a visual object, $p = .03$.

For the latency measure, the analysis revealed a marginally significant effect of orientation modality, $F(1, 36) = 3.17$, $p = .08$, $\eta^2 = .08$. Participants were faster to respond when the orientation object was visual rather than verbal. The main effect of target modality did not approach significance, $F(1, 36) = 2.69$, $p = .10$, $\eta^2 = .07$. The orientation modality x target modality interaction was not significant (Figure 9), $F(1, 36) = 2.17$, $p = .14$, $\eta^2 = .05$, and so was the main effect of learning order, $F(1, 36) = .03$, $p = .85$, $\eta^2 = .00$. All remaining interactions were non significant.

As in Experiment 1, separate analyses on the first two testing blocks and the last two testing blocks were conducted. Results showed no differences across phases; in both phases participants were similarly fast and accurate across within- and between-layout judgments, but this was the case only when the layout studied first was visual. In the case that the layout studied first was verbal, they were more accurate, but not faster, for within- compared to between-layout judgments. These results indicate that the pattern of participants' performance did not change over the course of testing. Results from further analyses comparing performance across within- and between-layout judgments after collapsing data across modalities are shown in Table 2 (see Appendix).

In the report questionnaires 33 out of 38 participants reported that they integrated the visual and the verbal targets in a single representation at the time of encoding (17 participants studied the visual layout first and 16 participants studied the verbal layout first). The remaining 5 participants reported that they kept the visual and the verbal targets in distinct representations (1 participant studied the visual layout first and 4 participants studied the verbal layout first).

Discussion

Results from Experiment 2 showed comparable accuracy across within- and between-layout judgments but this was the case only when participants studied the visual layout first. In the case that the layout studied first was verbal participants were more accurate at within- compared to between-layout judgments. Latency was similar across the two types of judgments though, indicating that participants integrated the visual and the verbal objects in a single memory representation at the time of learning. This was corroborated by verbal reports where most participants indicated that they had integrated the visual and the verbal objects in a single representation during learning. Finally, participants were faster to respond when the orientation object was visual rather than verbal suggesting overall easier access to the visual objects.

Similar performance in the accuracy across within- and between-layout judgments when the layout studied first was visual suggests that participants maintained equally precise spatial memories for the spatial relations for objects encoded within and across layouts. It could be that the strong and vivid visual memory participants built first allowed for the subsequent precise assimilation of the verbal targets into the existing memory (Greenauer et al., 2013). Notably, when encoding the verbal layout participants had visual access to the environment, which might have helped them integrate more efficiently the described objects into the existing representation containing the visual targets. That is, the availability of extra-layout visual cues during verbal encoding and the precise memory for visual objects could have served as placeholders for determining with high accuracy the described locations of the verbal objects. If this is the case, then precluding visual access to the surrounding environment during verbal learning should result in differences across within- and between-layout judgments even when the visual layout is studied first. This prediction is examined in Experiment 3.

Experiment 3

Findings from Experiments 1 and 2 suggest that people have no difficulty to integrate at the time of learning visual and verbal spatial information learned at different experiences in a single memory representation. Yet, the pattern of accuracy indicates that in the integrated representation between-layout relations are represented with diminished precision.

The results for Experiment 2 showed that although integration across vision and language occurred during learning, the learning order influenced the precision of the resulting integrated representation; differences for within- vs. between-layout judgments in accuracy were observed only when participants studied the verbal layout first. In contrast, performance was comparable across within- and between-layout judgments when participants studied the visual layout first. Experiment 3 examines whether this difference in the pattern of accuracy was an effect of the presence of visual cues during verbal learning. In Experiment 3 participants were blindfolded during verbal learning to eliminate access to visual cues.

An additional goal of Experiment 3 was to examine whether integration depends on the number of objects to be memorized. Experiments 1 and 2 involved learning 6 objects in total, which is fairly easy to do. Perhaps increasing the number of objects per layout would cause participants to keep the representations separate. Although, like the present experiments, the study of Giudice et al. used 3 locations for each layout, the other studies supporting distinct representations typically involved more objects. For example, in the study by Adamou et al. (2013; Experiment 1) participants studied two visual layouts with 4 objects each. Therefore, it could be that with a larger number of objects participants will rather keep the two layouts in memory in distinct representations, as this will reduce the number of allocentric relations to be stored. This hypothesis was tested in Experiment 3 with half of the participants experiencing 5 visual and 5 verbal objects and the other half studying 3 objects per layout as in Experiments 1 and 2.

Method

Participants

Forty-eight students (ages: 20-25, 17 male) from the University of Cyprus participated in this experiment in exchange of a small monetary compensation (€10). All participants signed an informed consent form before the experiment.

Materials, Design, & Procedure

Everything was identical to Experiment 2 with two exceptions. First, participants were blindfolded during verbal learning and when guided back to the learning room to point to the verbally-encoded object locations (i.e., when they studied the verbal layout first). This was done to examine whether the absence of a performance difference in accuracy when the visual layout was studied first was caused by the presence of visual cues during verbal learning. If this is the case, then eliminating visual cues during verbal learning should lead to inferior performance in the accuracy for between- compared to within-layout judgments, regardless of layout learning order. Second, half of the participants studied two layouts with 3 objects each as in Experiments 1 and 2 (Figure 1) while the other half studied two layouts with five objects each (see Figure 10 in the Appendix). If integration is limited to a small number of locations, then increasing the number of the objects in each layout would make participants keep distinct representations for the two layouts. In such a case, inferior performance for between- compared to within-layout judgments in latency should be observed.

Results

Data for pointing error and latency were analyzed using a repeated-measures ANOVA with learning order (visual first vs. verbal first) and layout size (3 objects vs. 5 objects) as the between-subjects factors and orientation modality (visual vs. verbal) and target modality (visual vs. verbal) as the within-subjects factors.

The analysis of pointing error revealed a significant orientation modality x target modality interaction, $F(1, 44) = 19.13$, $p = .00$, $\eta^2 = .30$. When participants oriented to a visual object they were significantly more accurate to point to a visual object rather than to a verbal object, $p < .001$. When participants oriented to a verbal object they were significantly more accurate to point to a verbal object rather than to a visual object, $p < .001$ (Figures11 and 12). The main effect of layout size was marginally significant, $F(1, 44) = 3.65$, $p = .06$, $\eta^2 = .07$. Participants were numerically more accurate when they studied 5 objects in each layout compared to when they studied 3 objects in each layout. Importantly, the layout size x orientation modality x target modality interaction was not significant, $F(1, 44) = 2.63$, $p = .11$, $\eta^2 = .05$. Neither the main effect nor any interactions involving learning order were significant.

For the latency measure, the analysis revealed significant main effects of orientation modality and target modality, $F(1, 44) = 19.64$, $p = .00$, $\eta^2 = .30$ and $F(1, 44) = 9.92$, $p = .00$, $\eta^2 = .18$ respectively. Participants were faster to respond when the orientation object or the target object was visual rather than verbal. The orientation modality x target modality interaction (Figures 13 and 14) was not significant, $F(1, 44) = 1.27$, $p = .26$, $\eta^2 = .02$. The learning order x layout size interaction was marginally significant, $F(1, 44) = 4.04$, $p = .05$, $\eta^2 = .08$. When participants studied the visual layout first they were numerically faster in the case when they studied 5 objects in each layout rather than 3 but the difference was not significant, $p = .24$. When they studied the verbal layout first they were numerically faster in the case when they studied 3 objects in each layout rather than 5 but the difference was again not significant, $p = .10$. No other main effects or interactions were significant.

Given previous findings suggesting that spatial memories derived from vision are more vivid and more easily accessible compared to those derived from other perceptual modalities (e.g., audition, touch) and language (Avraamides et al., 2004; Newell, Woods, Mernagh, & Bülthoff, 2005), it could be that when participants studied the visual layout first they might have constructed a stronger integrated representation during learning which they could access more easily during retrieval, even with more objects in each layout. Instead, when participants studied the verbal layout first they might have constructed a weaker integrated representation during learning which resulted in less efficient access to that representation for layouts with more objects.

As in Experiment 1, data from analyses comparing the first two testing blocks with the last two testing blocks showed the same pattern of findings in the two phases; in both phases, participants were similarly fast across within- and between-layout judgments and more accurate for within- compared to between-layout judgments.

Also, like in the previous experiments, further analyses were conducted comparing performance across within- and between-layout judgments after collapsing data across modalities (Table 2 in Appendix).

Questionnaire data showed that 43 out of 48 participants reported that they integrated the visual and the verbal targets in a single representation at the time of encoding (22 participants studied the visual layout first and 21 participants studied the verbal layout first). The remaining 5 participants reported that they kept the visual and the verbal targets as distinct representations (2 participants studied the visual layout first and 3 participants studied the verbal layout first).

Discussion

Results from Experiment 3 showed that participants were more accurate at within- compared to between-layout judgments while they were equally fast at their responses across the two types of judgments. Taken together with verbal reports, these results suggest that participants integrated the two layouts into a single representation but between-layout relations were represented with lower precision than within-layout relations. Importantly, this result was found regardless of whether participants studied the visual or the verbal layout first and of the number of objects in each of the two layouts. This shows that although participants might have assimilated the verbal targets studied later (in the case that the layout studied first was visual) to their already existing spatial memory for the visual layout, the absence of visual cues during verbal learning most likely resulted in loss of precision for the integrated representation. Such a finding suggests that the similar performance in accuracy across within- and between-layout judgments when participants studied the visual layout first in Experiment 2 was a result of the availability of additional-layout visual cues during verbal learning. Moreover, results from this experiment show that independently of the number of locations per layout, people have no difficulty to integrate visual and verbal spatial information in a single memory representation at the time of learning. It might be that the absence of visual cues during verbal learning encouraged participants to translate the verbal code to spatial at the time of encoding, when they experienced other information that was already in a spatial format and therefore, integrate at that time point.

Notably, participants were numerically more accurate with layouts of 5 objects than with layouts of 3 objects. It could be that a crowded layout forces people to perform more refine parsing of space leading to more precise representation of an object's location. That is, if people maintain categorical (in addition to metric) information about object locations (Newcombe, Huttenlocher, Sandberg, Lie, & Johnson, 1999) then more objects could encourage parsing space into more regions, allowing one to represent more accurately the location of each object. According to the category-adjustment model proposed by Newcombe et al. (1999), spatial coding occurs hierarchically with people estimating relative distances between objects based on combination across different levels of the hierarchy. It could be that more objects might have encouraged participants to estimate the relative distance among the objects in that layout. This could result in a more precise representation of each object's location within and across layouts.

Interim General Discussion

Results from Experiment 1 showed that although participants integrated the visual and the verbal objects in a single memory representation at the time of learning, they maintained more precise spatial memories for within- compared to between-layout spatial relations. Lower accuracy for between-layout judgments might have been caused by the decay of the first layout as it was observed by the decreased accuracy for the locations encoded in the layout studied first, regardless of whether the layout studied first was visual or verbal. Findings from Experiment 2 employed refreshing participants' memory for the layout studied first by pointing to the objects of that layout one more time before testing and showed that when the visual layout was studied first participants were similarly accurate across within- and between-layout judgments. In the case they studied the verbal layout first they were more accurate at within- compared to between-layout judgments. It could be that participants, when learning the visual layout first, could more efficiently relate the verbal targets to their already existing memory for the visual layout because of the presence of extra-layout visual cues during verbal encoding. If this is the case, eliminating visual cues during verbal learning should result in loss of precision for between- compared to within-layout spatial relations.

Experiment 3 showed exactly this; although equal latencies across within- and between-layout judgments suggested that integration occurred during learning, participants maintained more precise spatial memories for within- compared to between-layout spatial relations regardless of whether they studied the visual or the verbal layout first. Importantly, participants seemed to have no difficulty integrating locations across layouts during learning even with an increased number of objects in each layout.

Although Experiments 1, 2, and 3 provide valuable findings suggesting that people can readily integrate visual and verbal spatial information in a single memory representation during learning, a question that remains is under which conditions people may keep separate representations for spatial memories derived from vision and language. In Experiments 1, 2, and 3, participants encoded the visual and the verbal locations from the same viewpoint which was aligned with the geometric structure of the room. Given previous evidence that spatial memories are orientation-dependent and that, in the absence of conflicting environmental cues, people rely on egocentric experience to establish a preferred orientation (e.g., Diwadkar & McNamara, 1997; Mou & McNamara, 2002; Shelton & McNamara, 2001), the common viewpoint might have facilitated integration. Previous studies have provided evidence that people keep separate representations for two layouts studied from different viewpoints (e.g., Adamou et al., 2013; Experiment 1). A possible explanation for this finding could be that the difficulty of relating information studied from different viewpoints at the time of encoding encouraged participants to maintain separate representations. Experiment 4 was designed to examine whether studying the visual and the verbal layouts from different viewpoints would lead to separate representations.

CHAPTER 3

Experiment 4

Experiment 4 investigated whether participants' similar latencies across within- and between-layout judgments in Experiments 1, 2, & 3, which imply retrieval from a common representation, reflect stem from learning the visual and the verbal layouts from the same viewpoint.

Participants in Experiment 4 studied the temporally-segregated visual and the verbal layouts from different viewpoints. Recent findings by Adamou et al., (2013) showed that when participants studied two visual layouts separated in time and from different viewpoints, they performed better (in terms of speed and accuracy) at within- compared to between-layout judgments, indicating that they maintained the two layouts in separate representations. The authors argued that the difficulty of integrating spatial information from different viewpoints most likely made participants less willing to integrate during learning and instead maintain separate representations and only relate information across representations when needed (Adamou et al., 2013). If this is the case then in the present experiment as well participants should keep the visual and the verbal objects in separate representations during learning. Participants could either coordinate information across representations when performing each trial or integrate the visual and the verbal targets in a single representation when they realize that the task requires relating information across layouts at the beginning of testing. If participants maintain separate representations, inferior performance in the latency for between- compared to within-layout judgments should be observed. If instead participants maintain separate representations during learning but integrate the visual and the verbal targets in a single representation after the first few trials at testing then performance will be faster at within- compared to between-layout judgments, but only in the early stages of testing. Finally, if information is retrieved from two different representations during each testing trial, then slower performance for between-layout judgments should be observed throughout the experiment.

Method

Participants

Twenty-eight students (ages: 20-35, 4 male) from the University of Cyprus participated in this experiment voluntarily. All participants signed an informed consent form before the experiment.

Materials, Design, & Procedure

Materials and procedure were identical to Experiment 3 with three exceptions. First, participants studied the visual layout from the 300° perspective and the verbal layout from the 60° perspective (as indicated in Figure 1); both perspectives were misaligned with the geometry of the lab. If integration of memories formed from different viewpoints is difficult then it is possible that people would maintain separate representations, either throughout or at least until the early stages of testing. Therefore, better performance for within- compared to between-layout judgments in latency should be observed, at least for the first few trials of the experiment. Second, after memorizing the locations of each layout participants were asked to point to each object with eyes closed from the learning standpoint, after rotating to 0°, and after rotating to120° either to their right or left (i.e., right vs. left rotation was counterbalanced between participants). This was done to ensure that participants kept track of object locations when they rotated to the second learning viewpoint. Third, because results from Experiment 3 revealed that integration occurs at the time of learning regardless of the number of objects experienced in each layout, participants in Experiment 4 learned only layouts with 3 objects as shown in Figure 1.

Results

Data for pointing error and latency were analyzed using a repeated-measures ANOVA with learning order (visual first vs. verbal first) as the between-subjects factor and orientation modality (visual vs. verbal) and target modality (visual vs. verbal) as the within-subjects factors.

The analysis of pointing error revealed a significant orientation modality x target modality interaction (Figure 15), $F(1, 26) = 12.53$, $p = .00$, $\eta^2 = .32$. When participants oriented to a visual object they were significantly more accurate to point to a visual object rather than to a verbal object, $p = .01$. When participants oriented to a verbal object they were significantly more accurate to point to a verbal object rather than to a visual object, $p < .001$. The main effect of learning order was not significant, $F(1, 26) = 2.77$, $p = .10$, $\eta^2 = .09$. The main effects of orientation modality and target modality were also not significant, $F(1, 26) = 1.56$, $p = .22$, $\eta^2 = .05$ and $F(1, 26) = 1.36$, $p = .25$, $\eta^2 = .05$ respectively. All other interactions were not significant.

For the latency measure, the analysis revealed a significant orientation modality x target modality interaction (Figure 16), $F(1, 26) = 5.78$, $p = .02$, $\eta^2 = .18$. When participants oriented to a visual object they were significantly faster to point to a visual object rather than to a verbal object, $p < .001$. When participants oriented to a verbal object they were numerically faster to point to a verbal object than to a visual object but the difference was not significant, $p = .28$. The analysis also revealed a significant main effect of orientation modality, $F(1, 26) = 5.48$, $p = .02$, $\eta^2 = .17$. Participants were faster to respond when the orientation object was visual than verbal. The main effect of learning order was not significant, $F(1, 26) = .06$, $p = .80$, $\eta^2 = .00$. All other interactions were not significant.

As in Experiment 1, 2, & 3, data from separate analyses on the first two testing blocks and the last two testing blocks respectively were conducted. Data from these analyses showed that for pointing error, participants were more accurate at within- compared to between-layout judgments in both phases. The analysis of pointing error including only testing blocks 1 and 2 showed a significant orientation modality x target modality interaction (Figure 17), $F(1, 26) = 15.13$, $p = .00$, $\eta^2 = .36$. When participants oriented to a visual object they were marginally more accurate to point to a visual object than to a verbal object, $p = .06$. When participants oriented to a verbal object they were significantly more accurate to point to a verbal object than to a visual object, $p < .001$. The main effect of learning order was marginally significant with participants being more accurate when they studied the visual layout first than when they studied the verbal layout first, $F(1, 26) = 4.19$, $p = .05$, $\eta^2 = .13$. The main effects of orientation modality and target modality were not significant, $F(1, 26) = .84$, $p = .36$, $\eta^2 = .03$ and $F(1, 26) = 3.52$, $p = .07$, $\eta^2 = .11$ respectively. All other interactions were not significant.

The analysis of pointing error for blocks 3 and 4 showed a significant orientation modality x target modality interaction (Figure 18), $F(1, 26) = 7.96$, $p = .00$, $\eta^2 = .23$. When participants oriented to a visual object they were more accurate to point to a visual object than to a verbal object, $p < .001$. When participants oriented to a verbal object they were more accurate to point to a verbal object rather than to a visual object, $p = .05$. The main effect of learning order was not significant, $F(1, 26) = 1.26$, $p = .27$, $\eta^2 = .04$. The main effects of orientation modality and target modality were also not significant, $F(1, 26) = 1.59$, $p = .21$, $\eta^2 = .05$ and $F(1, 26) = .14$, $p = .70$, $\eta^2 = .00$ respectively. All other interactions were not significant.

For latency, the analysis revealed that in the first phase (testing blocks 1 & 2) participants were faster at within- compared to between layout judgments but in the second phase (testing blocks 3 & 4) they were similarly fast across the two types of judgments.

The analysis of latency including only testing blocks 1 and 2 showed a significant orientation modality x target modality interaction (Figure 19), $F(1, 26) = 4.98$, $p = .03$, $\eta^2 = .16$. When participants oriented to a visual object they were significantly more accurate to point to a visual object rather than to a verbal object, $p < .001$. When participants oriented to a verbal object they were numerically more accurate to point to a verbal object rather than to a visual object but the difference was not significant, $p = .76$. The analysis also revealed a significant main effect of orientation modality, $F(1, 26) = 10.27$, $p = .00$, $\eta^2 = .28$. Participants were faster to respond when the orientation object was visual than verbal. The main effect of learning order was not significant, $F(1, 26) = .07$, $p = .78$, $\eta^2 = .00$. The main effect of target modality was not significant, $F(1, 26) = 2.14$, $p = .15$, $\eta^2 = .07$ respectively. All other interactions were not significant.

The analysis of latency including testing blocks 3 and 4 showed that the orientation modality x target modality interaction was not significant (Figure 20), $F(1, 26) = 2.65$, $p = .11$, $\eta^2 = .09$. The main effect of learning order was not significant, $F(1, 26) = .06$, $p = .80$, $\eta^2 = .00$. The main effects of orientation modality and target modality were also not significant, $F(1, 26) = .02$, $p = .87$, $\eta^2 = .00$ and $F(1, 26) = .03$, $p = .85$, $\eta^2 = .00$ respectively. All other interactions were not significant.

Results from analyses comparing performance across within- and between-layout judgments after collapsing data across modalities are reported in Table 2 (see Appendix).

Data from participants' report questionnaires revealed that 23 out of 28 participants maintained the visual and the verbal targets as separate representations, but integrated them into a single representation when the task required doing so at the time of retrieval, after the first few trials at testing (11 participants studied the visual layout first and 12 participants studied the verbal layout first). The remaining 5 participants reported that they kept separate representations for the visual and the verbal targets (3 participants studied the visual layout first and 2 participants studied the verbal layout first). Interestingly, results remained unchanged when analyses were conducted only using the data of participants who reported that they maintained separate representations.

Discussion

Results from Experiment 4 showed that participants were not only more accurate but also faster at within- compared to between-layout judgments (numerically so when orienting to a verbal object). Importantly, the differences in latency were only present for the first two blocks of testing; in the last two blocks pointing latency was equal across within- and between-layout judgments. These results are more in line with the conclusion that participants kept separate representations initially but integrated them during testing. This conclusion is corroborated by participants' verbal reports. Findings from this experiment are compatible with those provided from previous studies assessing integration using perceptual layouts (e.g., Adamou et al., 2013; Giuidice et al., 2009). This suggests that although spatial learning from language may be different from that taking place through direct modalities such as vision and touch, spatial representations that support the execution of spatial tasks after visual and verbal learning are functionally similar (e.g., Avraamides et al., 2004; Avraamides & Pantelidou, 2008; Giudice, Betty, & Loomis, 2011) to that derived after perceptual learning indicating that spatial reasoning is independent of the input modality.

CHAPTER 4 – GENERAL DISCUSSION

Many studies (e.g., Giudice et al., 2009; Ishikawa & Montello, 2006; Moar & Carleton, 1982; Montello & Pick, 1993) have investigated whether people integrate spatial information acquired from different perceptual experiences in a single representation in memory. The paradigm used in these studies relied on comparing performance for judgments involving locations encoded as part of the same perceptual experience (within-layout judgments) or different experiences (between-layout judgments). The current study adopted this paradigm to investigate integration of spatial information across vision and language. In particular, the goal of the present experiments was to investigate whether people integrate visual and verbal spatial information in a single memory representation at the time of learning or whether they maintain separate representations during learning and at the time of retrieval, they either relate information across representations because the task requires doing so, or they integrate the visual and the verbal spatial information in a single memory representation.

Findings from the majority of previous studies with perceptual modalities suggest that people keep in separate representations spatial memories for layouts learned at different times, although they can relate information across representations when the task requires doing so, at the time of retrieval (e.g., Adamou et al., 2013; Giudice, Klatzky, & Loomis, 2009; Experiment 1; Ishikawa & Montello, 2006; McNamara, Halpin, & Hardy, 1992; Montello & Pick, 1993). These findings indicate that temporal and spatial separation is a critical factor that influences whether people integrate newly experienced locations with their already existing spatial memories. Still, findings from a few studies showed that people may integrate in a single spatial representation information experienced at different times at the time of learning (e.g., Holding & Holding, 1989; Moar & Carleton, 1982), either by assimilation or accommodation (Greenauer et al., 2013), despite the temporal segregation in encoding. These findings suggest that besides temporal segregation, the specific learning conditions may influence whether people integrate during learning. For instance, if information studied at different times is considered as related (i.e., pairs of routes; Holding & Holding, 1989), people may be more willing to integrate them in a single memory representation during learning.

The studies that examined integration so far used layouts encoded through direct perception, such as vision and touch (e.g., Adamou et al., 2013; Giudice et al., 2009).

Although people usually experience space through direct perception, many tasks of everyday life require from people to reason about space based on spatial information acquired through indirect means such as language (e.g., when people follow driving instructions). To date, however, no research has assessed integration across spatial memories acquired through sensory modalities and indirect methods of spatial learning such as language. Given the extra demands that language poses for translating the verbal code to spatial at some point to use it in a task it might be that integration across vision and language differs from integration across perceptual modalities in terms of the time point in time at which it occurs (during learning or retrieval) or whether it occurs at all. Providing evidence about whether and under which conditions people integrate spatial information across vision and language will extend our understanding about the representations that support spatial activity.

The present experiments were designed to investigate whether people integrate visual and verbal spatial memories at the time of learning or they keep separate representations during learning but they either relate information across representations when the task requires doing so, or they integrate spatial information across vision and language in a single memory representation at the time of retrieval. Hypothesis 1 stated that because of the requirement to translate the verbal code to spatial in order to carry out the task, people may carry out the translation at the time of learning and also integrate the verbally-encoded locations (either through assimilation or accommodation) with the visual information, which is automatically encoded in a spatial format (Richardson & Spivey, 2000). Hypothesis 2 stated that people may maintain the visual and the verbal spatial memories in separate representations. This should especially be the case if participants prefer to maintain the objects encoded verbally in a verbal format instead of translating the verbal code to spatial. Finally, hypothesis 3 stated that people may keep separate representations during learning but integrate the visual and the verbal objects in a single representation at the very first stages of retrieval, when realizing that the task requires coordinating information across representations. All three hypotheses were cast in terms of latency as it was deemed probable that the precision of maintaining information may differ for locations encoded in close that in far temporal proximity, regardless of integration. Overall, the results of Experiments 1, 2, and 3, which involved studying two layouts from the same perspective, supported Hypothesis 1 whereas results of Experiment 4, in which participants encoded the two layouts from a different perspective, favored Hypothesis 3 (see Table 1 in the Appendix for an overview of results).

In Experiment 1, participants studied a visual and a verbal layout separated by a 30-second delay. Participants were similarly fast at their responses for within- and between-layout spatial judgments suggesting that they most likely retrieved the visual and the verbal objects from a single spatial representation which they constructed at the time of learning. This was corroborated by participants' reports, as most of the participants indicated that they integrated the visual and the verbal objects in a single representation during learning. Results were the same regardless of whether participants received instructions to integrate or not, which further suggests that they had integrated information at the time of learning. Results further indicated that participants lost precision for the spatial information encoded first, regardless of whether it was visual or verbal, an effect that could have been caused by the memory decay of the first layout. Therefore, Experiment 2 assessed whether refreshing participants' memory for the layout studied first would eliminate memory decay and lead to similarly precise spatial memories for within- and between-layout spatial relations. Results from Experiment 2 showed that although participants again seem to have integrated all locations in a single representation at the time of learning (a result that was further supported by their reports), performance differences were present in the accuracy depending on whether they studied the visual or the verbal layout first. Particularly, when the visual layout was studied first participants maintained similarly precise memories for within- and between-layout spatial relations. It could be that participants, when they learned the visual layout first they maintained a memory that was strong enough to allow (likely due to the presence of extra-layout visual cues) easy and accurate assimilation of the verbal targets. However, when the first layout was encoded verbally participants maintained more precise memories for within- compared to between-layout spatial relations. This could suggest that participants maintained a weak verbal memory in which they assimilated the visual targets experienced later. If this is the case then, although they had integrated at the time of learning, they lost precision for the integrated representation as documented by higher error rates for between- than for within-layout spatial judgments.

Experiment 3 verified that that learning order effects in accuracy in Experiment 2 were likely caused by the presence of environmental visual cues. Results from this experiment showed that when visual cues were eliminated during verbal learning, although integration still seems to have occurred, participants had reduced precision for the between- than within-layout relations, regardless of whether they studied the visual or the verbal layout first. Also, results from Experiment 3 support that integration across vision and language occurred during learning independently of the number of locations learned. Notably, participants were overall more accurate with more objects, possibly due to that encoding more objects forces participants to localize more precisely each target relative to the others at the time of encoding.

Overall, findings from Experiments 1, 2, and 3 provide converging evidence that people can easily integrate in a single representation visual and verbal spatial memories acquired at different learning experiences. The finding that no effect of layout learning order was evident, with the exception of Experiment 2, may suggest that both assimilation of new spatial information into an existing representation and accommodation of already existing information into a newly constructed representation (Greenauer et al., 2013) may occur. This might have occurred because as participants studied the visual and the verbal objects from the same orientation, it was easier for them to merge locations in a single representation at the time of learning, despite the temporal separation in encoding. Given previous findings indicating that people use a reference frame they establish at the time of learning to maintain information in spatial memory (Klatzky, 1998) and that, people maintain spatial information using a preferred orientation determined by cues that are available during learning (e.g., the structure of the environment, visual cues; Mou & McNamara, 2002), it might be that the same learning viewpoint facilitated integration by providing a common reference frame to organize the two layouts. Overall, the findings from Experiments 1, 2, and 3 are in line with the results provided from previous studies (e.g., Holding & Holding, 1982; Moar & Carleton, 1982) and suggest that temporal segregation may not be a necessary prerequisite for keeping separate representations during learning.

Experiment 4 demonstrated that the findings from Experiments 1, 2, and 3 were more likely a result of the common viewpoint. In Experiment 4, participants studied a visual and a verbal layout at different times and from different viewpoints. Results showed that participants were faster at within- compared to between-layout spatial judgments, suggesting that they needed extra cognitive effort to retrieve the visual and the verbal targets from separate representations and coordinate them to compute the response. Additional analyses showed that this was the case for half of the testing phase of the experiment (i.e., block 1, block 2). For the other half of the experiment (i.e., block 3, block 4) participants were similarly fast across within- and between-layout judgments, indicating that during retrieval, they most probably integrated the visual and the verbal objects in a single memory representation. This conclusion is corroborated by participants' reports.

Participants in Experiment 4 might have found it difficult to integrate during learning because they studied the two layouts from different orientations (e.g., Adamou et al., 2013; Experiment 1) and therefore, opted for maintaining distinct representations for the visual and the verbal objects during learning. But when they realized that integration would facilitate their spatial performance at the early stages of testing, they merged the visual and the verbal objects in a single memory representation.

As discussed in the Introduction, the current experiments assessed integration based only on pointing latency. It was hypothesized that even if integration would occur during learning, the precision with which within- and between-layout spatial relations are maintained may be different due to the time delay between learning the visual and the verbal layouts. Results from Experiments 1, 2, 3, and 4 with the exception of Experiment 2 when participants studied the visual layout first, showed that regardless of whether integration occurred during learning (Experiment 1 and 3) or retrieval (Experiment 4), participants maintained more precise spatial memories for the objects encoded in the same layout compared to those encoded in different layouts (regardless of whether they studied the visual or the verbal layout first). Therefore, it seems most likely that the time delay between learning the two layouts indeed affected the precision with which participants maintained within- and between layout spatial relations.

Findings from Experiments 1, 2, and 3 contrast recent findings provided by studies assessing integration with perceptual layouts using the within- vs. between-layout paradigm (e.g., Giudice et al., 2009). For instance, findings provided by Giudice et al. (2009) showed that participants maintained in separate representations visual and haptic spatial locations learned at different times (Experiment 1). In that study, although participants could relate information across representations during retrieval (as evidenced by above-chance performance) they were faster at within- compared to between-layout judgments, documenting the presence of distinct representations. In the present study, only in Experiment 4, in which participants studied the layouts from different viewpoints, were participants faster for within- than between-layout judgments. The difference in representing spatial information acquired through direct perception (e.g., vision, haptics) and indirect methods of spatial learning (e.g., language) could account for the contrasting results between the current experiments and those of studies that used perceptual layouts. Specifically, when people perceive an object in space through vision, audition, and touch, they automatically encode its location in space (e.g., Richardson & Spivey, 2000; Spivey & Geng, 2001). In contrast, when people learn spatial information through language they must semantically process the verbal statement and recode the information described verbally to a spatial code in order to use it in a task. Thus, deciphering the location of an object is an effortful process; when people invest in this process at the time of learning they might be more likely to also contemplate on the most efficient strategy for maintaining the spatial information in memory. In Experiments 1, 2, and 3 in which learning occurred from the same viewpoint (allowing a common reference frame) that strategy might have been to integrate all locations in a single representation. In Experiment 4, in which integration was more demanding due to the discrepant viewpoints, maintaining separate representations was preferred.

Taken together, results from the current experiments are consistent with the claim that at some point (either during learning or later), spatial memories derived from perception and language are represented in a similar format (e.g., Bryant, 1997; Jackendoff, 1987; Miller & Johnson-laird, 1976), although spatial learning though indirect modalities such as language seems to be different from that taking place through direct perception. Such a conclusion is further supported by findings which showed that spatial information acquired through perception and language lead to similar representations that function equivalently as it was observed by similar performance in spatial tasks after perceptual and verbal learning (e.g., Avraamides & Kelly, 2010; Giudice, Betty, & Loomis, 2011; Mello, Greenauer, Pantelidou, & Avraamides, under revision; Pantelides, 2010). Regardless of whether functional equivalence results from the use of spatial images not linked to any modality (i.e., Amodal hypothesis; see Giudice et al., 2009 for discussion), or visual images (i.e., Common recoding hypothesis; Loomis et al., 2002), or supramodal representations that are not linked to the input modalities but maintain connections to them (Struiksma et al., 2009), it could be that spatial information encoded from different modalities could be easily merged into an integrated representation. Consistent with this idea, findings from the current experiments provide converging evidence that at some point (either during learning or retrieval) spatial information encoded through vision and language can be integrated into the same representation.

Conclusions

The present experiments demonstrate that people can readily integrate visual and verbal spatial memories learned at different times at the time of encoding. However, in cases at which integrating during learning is difficult, people maintain separate representations during learning, but they integrate visual and verbal spatial memories in a single representation at the time of retrieval, most probably because it is beneficial for spatial reasoning. Regardless of whether people integrate during learning or retrieval, it seems that they maintain more precise spatial memories for the allocentric relations encoded in the same experience than different learning experiences. These findings support the notion that integration of spatial information across vision and language is task-dependent with people being more willing to translate the verbal code to spatial during learning and integrate the verbal information with the visual that is already in a spatial format when the specific learning conditions facilitate integration. However, when integration is difficult during learning people maintain separate representations and only transform the verbal information to spatial when it must be used in a task; at that point, they integrate it with the visual information that is automatically encoded in a spatial code. Overall, findings from the current research suggest that although spatial learning through indirect modalities such as language is different from perceptual learning, people use a similar code to represent spatial information derived from different modalities at some point, indicating that spatial reasoning is independent of the input modality.

Limitations

A potential limitation of the current study is that most participants in the current study were females and it is possible that integration strategies are different in each gender. Follow-up experiments could include more males in order to investigate any possible gender differences in respect to the strategies used for integration across modalities. Also, another limitation is that conclusions in the three first experiments are based on a combination of a null effect with participants' self-reports, which cannot be verified empirically.

Future directions

The current findings highlight the flexibility of human spatial memory to organize spatial information encoded as different experiences through vision and language. Future studies may aim at identifying under what conditions people will integrate visual and verbal spatial information during learning despite integration being difficult. It could be that if people are aware that the upcoming task will require coordinating information across representations, they might put effort in integrating during learning. Furthermore, future research may aim at assessing individual differences by including other cognitive measures of working memory load, cognitive control and speed of processing. Another possibility for future research is to assess what happens when there is a delay between encoding and retrieval. Will the delay between encoding and retrieval result in memory decay of the integrated representation with people reverting to their distinct episodic memories for each encoding episode? While future research should focus on answering such interesting questions, the present results extend current findings by providing evidence that spatial integration across direct and indirect modalities is possible and provide further understanding about how the human brain deals with spatial memories derived from different modalities that are not necessarily sensory. What remains to be seen is whether the present findings that refer to vision and language can also apply to other combinations of indirect and direct modalities, e.g., language and touch. The study of Giudice et al. (2009) showed that people by default keep visual and haptic objects separate. If indeed things are different with indirect modalities such as language, then it would be interesting to see if verbal and haptic targets are readily integrated into a single representation during encoding. Answers to such questions may be more easily acquired if one collects, in addition to behavioral data, neuroimaging or electrophysiological data that will provide a more direct means of assessing brain functioning during task execution. Such measures may be more likely to provide answers on when and where in the brain integration takes place.

References

Adamou, C., Avraamides, M. N., & Kelly, J. W. (2013). Integration of visuospatial information encoded from different viewpoints. *Psychonomic Bulletin & Review.*

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Reviews in Neuroscience*, *20*, 303-330.

Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research, 45*, 1275-1284.

Avraamides, M. N. & Kelly, J. W. (2010). Multiple Systems of Spatial Memory: Evidence From Described Scenes. *Journal of Experimental Psychology: Learning*, *Memory & Cognition*, *36*, 635-645.

Avraamides, M. N., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2004). Functional Equivalence of Spatial Representations Derived From Vision and Language: Evidence From Allocentric Judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 801-814.

Avraamides, M. N., & Pantelidou, S. (2008). Does body orientation matter when reasoning about depicted or described scenes? Spatial Cognition, 2008-Lecture Notes in Artificial Intelligence, 8-21. Springer-Verlag, Berlin-Heidelberg.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20,* 723-767.

Bryant, D. J. (1997). Representing Space in Language and Perception. *Mind & Language*, *12,* 239-264.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8,* 302-307.

Farah., M. (1990). *Visual Agnosia,* MIT Press.

Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Functional Equivalence of Spatial Images From Touch and Vision: Evidence From Spatial Updating in Blind and Sighted Individuals. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37,* 621-634.

Giudice, N. A., Klatzky, R. L., & Loomis, J. M. (2009). Evidence for amodal representations after bimodal learning: Integration of haptic-visual layouts into a common spatial image. *Spatial Cognition & Computation, 9,* 287-304.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 0-25.

Greenauer, N. A., Mello, C., Kelly, J. W., & Avraamides, M. N. (2013). Integrating spatial information across experiences. *Psychological Research, 77,* 540-554.

Greenauer, N., & Waller, D. (2010). Micro- and macroreference frames: Specifying the relations between spatial categories in memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36,* 938-957.

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition, 13,* 208-217.

Holding, C. S., & Holding, D. H. (1989). Acquisition of route network knowledge by males and females. *Journal of General Psychology, 116,* 29-41.

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned spaces. *Cognitive Psychology, 52,* 93-129.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills, 37,* 967-979.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA.: MIT Press.

Kelly, J. W., Avraamides, M. N., & Loomis, J. M. (2007). Sensorimotor alignment effects in the learning environment and in novel environments. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33,* 1092- 1107.

Loomis, J. M., Klatzky, R. L., Avraamides, M. N., Lippa, Y., & Golledge, R. G. (2007). Functional equivalence of spatial images produced by perception and spatial language. In F. Mast and L. Jancke (Eds.), *Spatial processing in navigation, imagery, and perception* (pp. 29-48). New York: Springer.

Loomis, J.M., Klatzky, R.L., & Giudice, N.A. (2013). Representing 3D space in working memory: Spatial images from vision, touch, hearing, and language. In S. Lacey & R. Lawson (Eds). *Multisensory Imagery: Theory & Applications* (pp. 131-156). New York: Springer.

Loomis, J. M., Lippa, Y., Golledge, R. G., & Klatzky, R. L. (2002). Spatial updating of locations specified by 3-d sound and spatial language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 335–345.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In C. Freska, W. Brauer, C. Habel, & K. F. Wender (Eds.), *Spatial cognition III (Lecture notes in artificial intelligence)* (pp. 174-191). Berlin, Germany: Springer. doi: 10.1007/3-540-45004-1_11

McNamara, T. P., Halpin, J. A., & Hardy, J. K. (1992). Spatial and Temporal Contributions to the Structure of Spatial Memory. *Journal of Experimental psychology: Learning, Memory, and Cognition, 18,* 555-564.

Mellet, E., Bricogne, S., Crivello, F., Mazoyer, B., Denis, M., & Tzourio-Mazoyer, N. (2002). Neural basis of mental scanning of a topographic representation built from a text. *Cerebral Cortex, 12,* 1322-1330.

Mellet, E., Bricogne, S., Tzourio-Mazoyer, N., Ghaem, O., Petit, L., Zago, L., et al. (2002). Neural correlates of topographic mental exploration: The impact of route versus survey perspective learning. *NeuroImage, 12,* 588-600.

Milner, A. D., & Goodale, M. A. (1995). The Visual Brain in Action. Oxford: Oxford University Press.

Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46,* 774-785.

Moar, I., & Carleton, L. R. (1982). Memory for routes. *Quarterly Journal of Experimental Psychology, 34,* 381-394.

Montello, D. R., & Pick, H. L. (1993). Integrating Knowledge of Vertically Aligned Large-Scale Spaces. *Environment and Behavior, 25,* 457-484.

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 162-170.

Newcombe, N., Huttenlocher, J., Sandberg, E., Lie, E., & Johnson, S. (1999). What Do Misestimations and Assymetries in Spatial Judgment Indicate About Spatial Represnetation? *Journal of Experimental psychology: Learning, Memory, and Cognition, 25,* 986-996.

Newell, F. N., Woods, A. T., Mernagh. M., & Bülthoff, H. H. (2005). Visual, haptic and crossmodal recognition of scenes. *Experimental Brain Research, 161,* 233-242.

Noordzij, M. L., Neggerss, S. F., Rasmey, N. F., & Postma, A. (2008). Neural correlates of locative prepositions. *Neuropsychologia, 46,* 1576-1580.

Pantelides, S. (2010). Spatial updating: are described objects automatically updated? Unpublished master's thesis, University of Cyprus, Nicosia, Cyprus.

Perenin, M- T., & Vighetto, A. (1988). Optic Ataxia: A Specific Disruption in Visuomotor Mechanisms: I. Different Aspects of the Deficit in Reaching for Objects. *Brain, 111,* 643-674.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space, and Hollywood Squares: looking at things that aren't there anymore. *Cognition, 76*, 269-295.

Rieser, J. J., Garing, A. E., & Young, M. E. (1994). Imagery, action, and young children's spatial orientation. It's not being there that counts, it's what one has in mind. *Child Development, 65,* 1262-1278.

Schneider, G. E. (1969). Two visual systems. *Science, 163,* 895-902.

Shelton, A. L., & McNamara, T. P. (2001). Visual memories from nonvisual experiences. *Psychological Science, 12,* 343-347.

Spivey, M. J., & Geng, J. J. (2001). Occulomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychological Research, 65,* 235-241.

Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered, head-centered and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology, 76,* 2071-2076.

Struiksma, M. E., Noordzij, M. L., & Postma, A. (2009). What is the link between language and spatial images? Behavioral and neural findings in blind and sighted individuals. *Acta Psychologica, 132,* 145-156.

Treisman, A. M., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psycholoy, 12,* 97-36.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & Mansfield, R. J. W. (Eds.) *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32*, 867-882.

Wang, R. F. (2000). Representing a stable environment by egocentric updating and invariant representations. *Spatial Cognition & Computation, 1,* 431-445.

Wang, R. F., & Brockmole, J. R. (2003a). Human navigation in nested environments. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 398-404.

Wang, R. F., & Brockmole, J. R. (2003b). Simultaneous spatial updating in nested environments. *Psychonomic Bulletin & Review, 10,* 981-986.

Wang, R. F., & Spelke, E. F. (2000). Updating egocentric representations in human navigation. *Cognition, 77*, 215-250.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory ‘compellingness’ in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perceptual Psychophysics, 30,* 557–564.

Wolbers, T., Loomis, J. M., Klatzky, R. L., Wutte, M., & Giudice, N. A. (2011). Modality Independent Coding of Spatial Layout in the Human Brain. *Current Biology. 21,* 984-989.

Yamamoto, N., & Shelton, A. L. (2009). Orientation dependence of spatial memory acquired from auditory experience. *Psychonomic Bulletin & Review*, *16*, 301–305.

Appendix

Table 1

Summary of the Results

Experiment	Number of objects	Learning orientation	Eyes	Integration report	Accuracy	RT
1	3	Same	Open	At the time of encoding	Orientation x target Orientation x learning order Target x learning order	No significant interaction Main effect of orientation modality Main effect of target modality
2	3	Same	Open	At the time of encoding	Orientation x target x learning order	No significant interaction Main effect of orientation modality
3	3 or 5	Same	Closed	At the time of encoding	Orientation x target	No significant interaction Main effect of orientation modality Main effect of target modality
4	3	Different	Closed	At the time of retrieval	Orientation x target	Orientation x target Main effect of orientation modality

Experiment	Acc_Testing block 1 & 2	RT_Testing block 1 & 2	Acc_Testing block 3&4	RT_Testing block 3&4
1	Orientation x target Orientation x learning order Target x learning order	Main effect of orientation modality Main effect of target modality	Orientation x target Orientation x learning order Target x learning order	Main effect of orientation modality
2	Orientation x target Orientation x target x learning order	Main effect of orientation modality	Orientation x target Orientation x target x learning order	Main effect of target modality
3	Orientation x target	Main effect of orientation modality Main effect of target modality	Orientation x target	No significant main effects and/or interactions
4	Orientation x target	Orientation x target Main effect of orientation modality	Orientation x target	No significant main effects and/or interactions

Table 2

Sample descriptives using t-test for equality of means

	Within-layout		Between-layout		
	M	SD	M	SD	t-test
Exp.1_Acc (overall)	14.93	9.42	17.80	12.13	-2.67**
Exp.1_RT (overall)	12.66	3.40	12.56	3.55	.56
Exp. 1_Acc (block 1 & 2)	16.72	11.39	20.03	14.13	-2.22*
Exp.1_Acc (block 3 & 4)	13.20	9.10	15.75	11.27	-2.14*
Exp.1_RT (block 1 & 2)	14.80	4.35	14.49	4.19	1.28
Exp.1_RT (block 3 & 4)	10.71	2.99	10.80	3.41	-.40
Exp.2_Acc (overall)	24.66	10.56	26.42	10.70	-1.88
Exp.2_RT (overall)	13.55	3.35	13.26	3.12	1.49
Exp. 2_Acc (block 1 & 2)	28.99	12.14	29.83	11.97	-.72
Exp.2_Acc (block 3 & 4)	20.37	11.43	23.15	10.89	-2.22*
Exp.2_RT (block 1 & 2)	15.20	3.79	15.17	3.85	.14
Exp.2_RT (block 3 & 4)	11.93	3.32	11.43	2.92	2.07*
Exp.3_Acc (overall)	12.92	8.82	15.12	10.58	-4.26**
Exp.3_RT (overall)	13.64	4.43	13.39	4.21	1.12
Exp. 3_Acc. (block 1 & 2)	21.35	22.08	21.51	20.70	-.22
Exp.3_Acc (block 3 & 4)	18.91	22.61	19.50	20.82	-.75

Exp.3_RT (block 1 & 2)	15.44	5.03	15.05	4.66	1.23
Exp.3_RT (block 3 & 4)	11.54	3.89	11.54	3.92	-.00
Exp.4_Acc (overall)	21.78	10.17	27.15	11.50	-3.49*
Exp.4_RT (overall)	11.52	4.05	12.10	4.40	-2.43*
Exp. 4_Acc. (block 1 & 2)	22.95	10.88	28.86	13.34	-3.80**
Exp.4_Acc (block 3 & 4)	20.62	10.19	25.54	10.95	-2.81*
Exp.4_RT (block 1 & 2)	12.93	4.59	13.60	4.85	-2.24*
Exp.4_RT (block 3 & 4)	10.14	3.86	10.67	4.16	-1.59

Note. * = $p < .05$, *** = $p < .01$.

Figure captions

Figure 1. The layout used in all experiments.

Figure 2. The pointer used for pointing responses in all experiments.

Figure 3. Pointing error as a function of orientation modality and target modality in Experiment 1. Error bars are standard errors from the ANOVA.

Figure 4. Pointing error as a function of learning order and orientation modality in Experiment 1. Error bars are standard errors from the ANOVA.

Figure 5. Pointing error as a function of learning order and target modality in Experiment 1. Error bars are standard errors from the ANOVA.

Figure 6. Reaction time as a function of orientation modality and target modality in Experiment 1. Error bars are standard errors from the ANOVA.

Figure 7. Pointing error as a function of orientation modality and target modality in Experiment 2. Error bars are standard errors from the ANOVA.

Figure 8. Pointing error as a function of learning order, orientation modality, and target modality in Experiment 2. Error bars are standard errors from the ANOVA.

Figure 9. Reaction time as a function of orientation modality and target modality in Experiment 2. Error bars are standard errors from the ANOVA.

Figure 10. The layout used in Experiment 3 in the 5-object layouts condition.

Figure 11. Pointing error as a function of orientation modality and target modality for 3-object layouts in Experiment 3. Error bars are standard errors from the ANOVA.

Figure 12. Pointing error as a function of orientation modality and target modality for 5-object layouts in Experiment 3. Error bars are standard errors from the ANOVA.

Figure 13. Reaction time as a function of orientation modality and target modality for 3-object layouts in Experiment 3. Error bars are standard errors from the ANOVA.

Figure 14. Reaction time as a function of orientation modality and target modality for 5-object layouts in Experiment 3. Error bars are standard errors from the ANOVA.

Figure 15. Pointing error as a function of orientation modality and target modality in Experiment 4. Error bars are standard errors from the ANOVA.

Figure 16. Reaction time as a function of orientation modality and target modality in Experiment 4. Error bars are standard errors from the ANOVA.

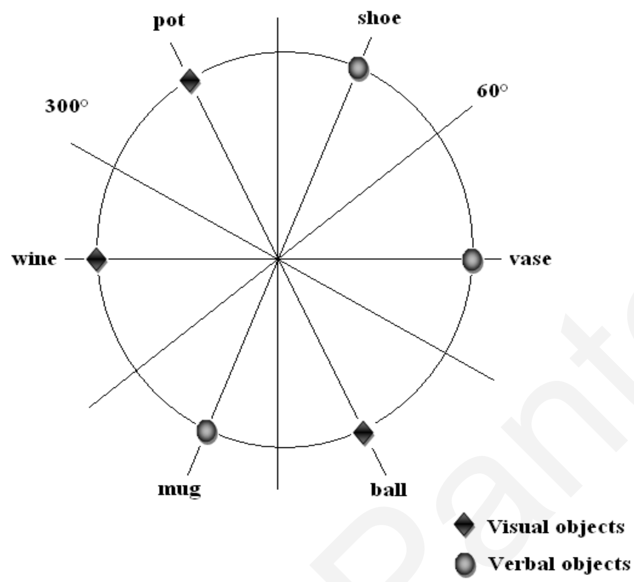
Figure 17. Pointing error as a function of orientation modality and target modality in testing blocks 1 and 2 in Experiment 4. Error bars are standard errors from the ANOVA.

Figure 18. Pointing error as a function of orientation modality and target modality in testing block 3 and 4 in Experiment 4. Error bars are standard errors from the ANOVA.

Figure 19. Reaction time as a function of orientation modality and target modality in testing blocks 1 and 2 in Experiment 4. Error bars are standard errors from the ANOVA.

Figure 20. Reaction time as a function of orientation modality and target modality in testing blocks 3 and 4 in Experiment 4. Error bars are standard errors from the ANOVA.

Figure 1



pot
shoe
300°
60°
wine
vase
mug
ball
Visual objects
Verbal objects

Figure 2

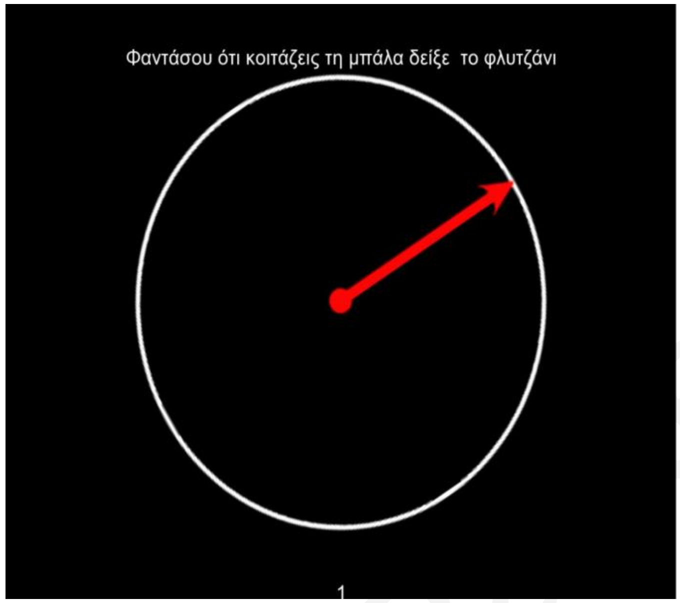
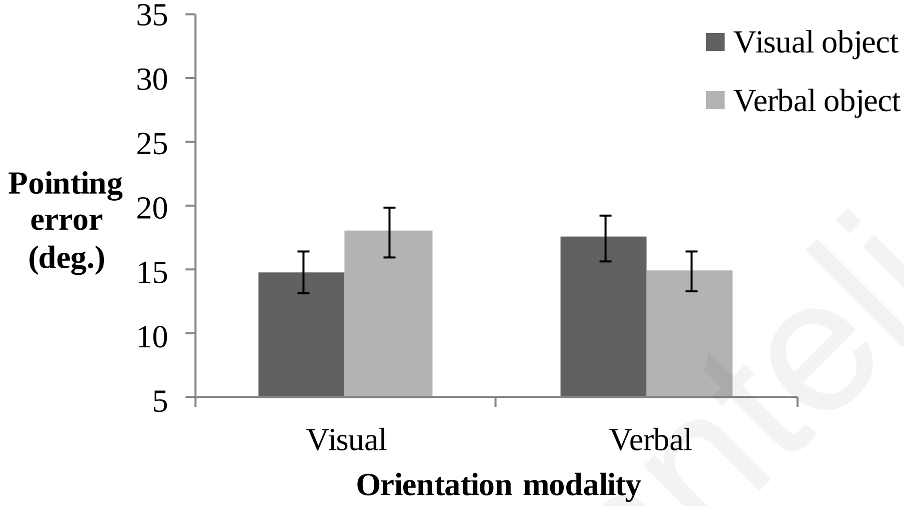


Figure 3



35
30
25
20
15
10
5
Pointing error (deg.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 4

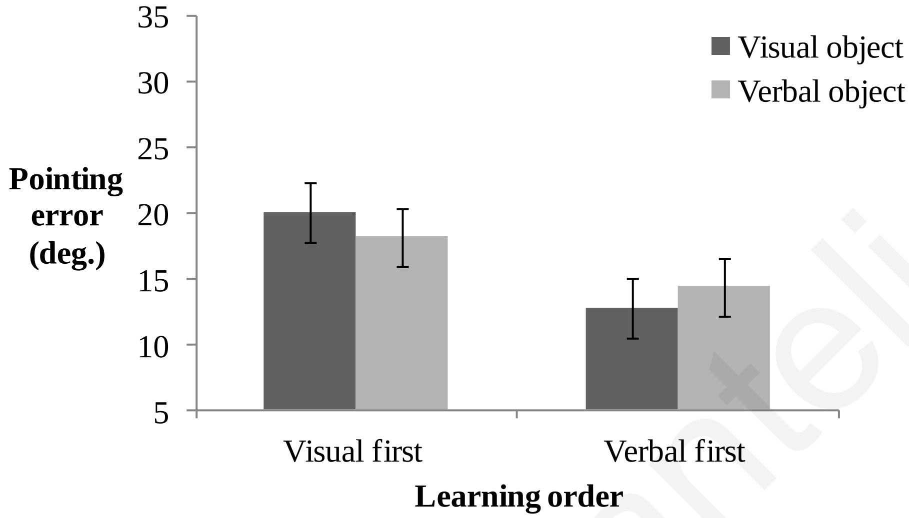

35
30
25
20
15
10
5
Pointing error (deg.)
Visual object
Verbal object
Visual first
Verbal first
Learning order

Figure 5

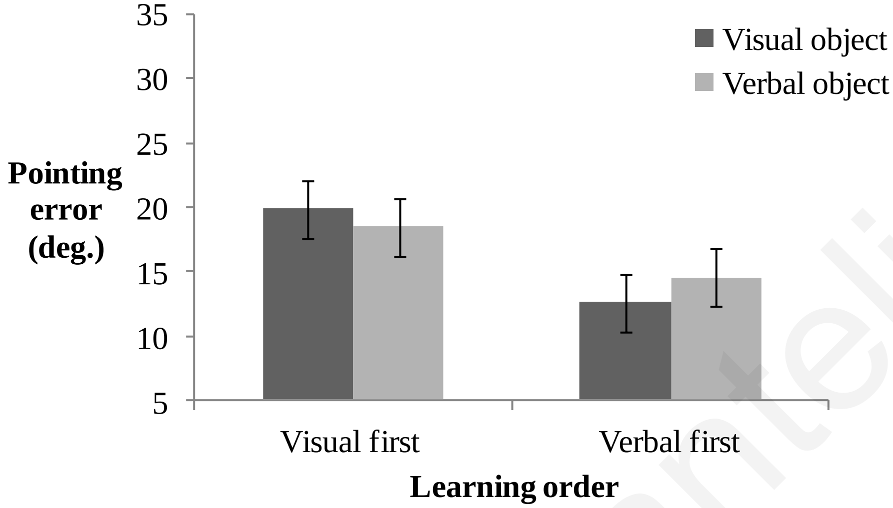
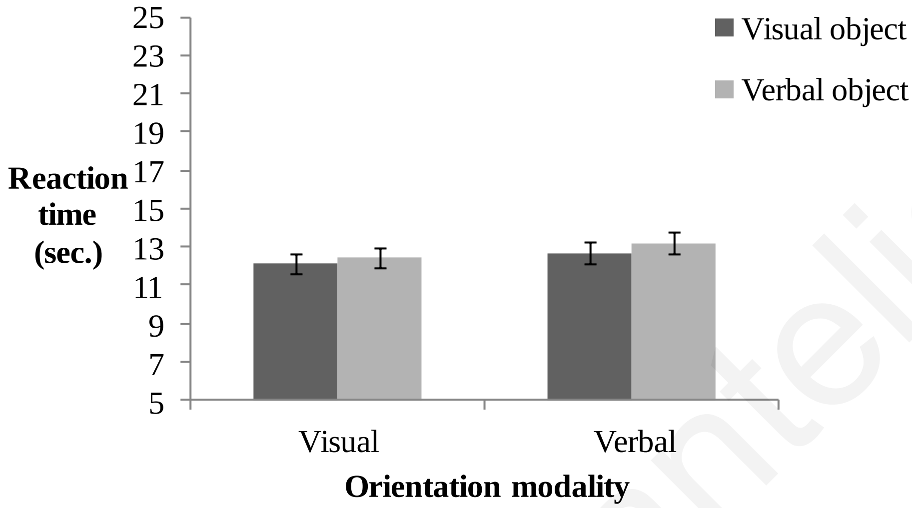

35
30
25
20
15
10
5
Pointing error (deg.)
Visual object
Verbal object
Visual first
Verbal first
Learning order

Figure 6



25
23
21
19
17
15
13
11
9
7
5
Reaction time (sec.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 7

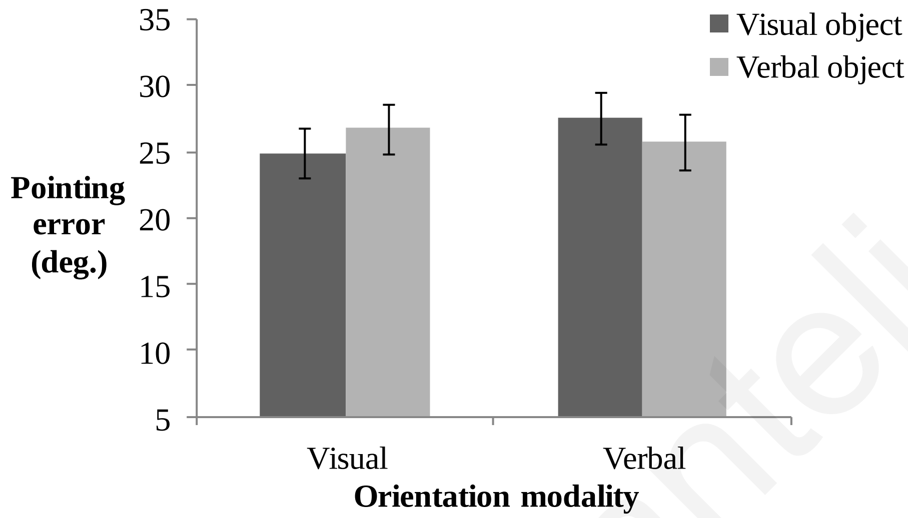
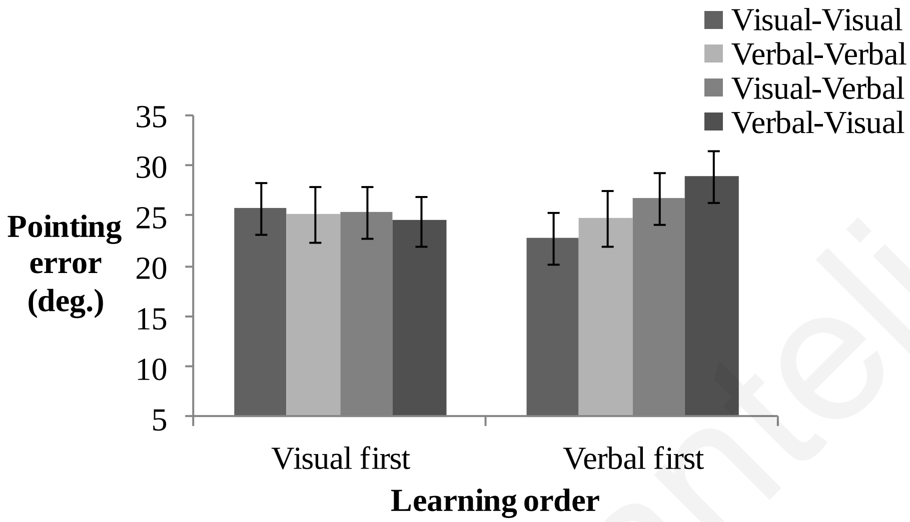
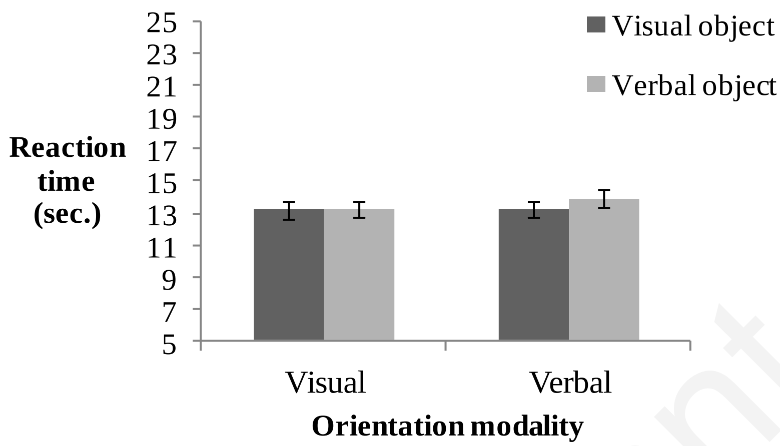

35
30
25
20
15
10
5
Visual object
Verbal object
Pointing error (deg.)
Visual
Verbal
Orientation modality

Figure 8



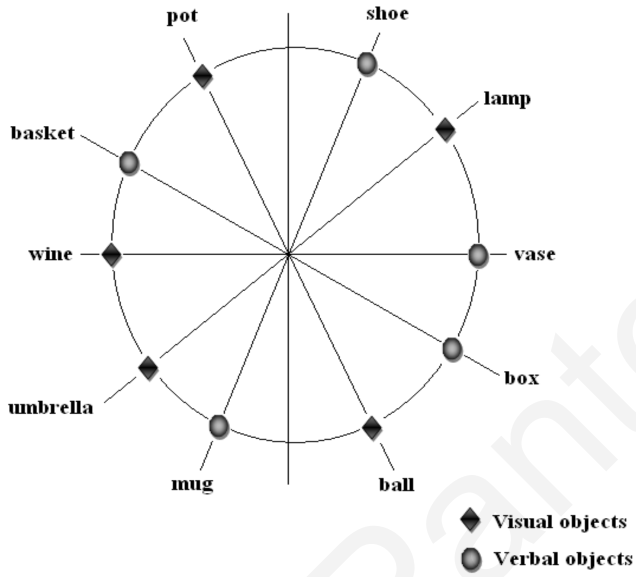
Visual-Visual
Verbal-Verbal
Visual-Verbal
Verbal-Visual
35
30
25
20
15
10
5
Pointing error (deg.)
Visual first
Verbal first
Learning order

Figure 9



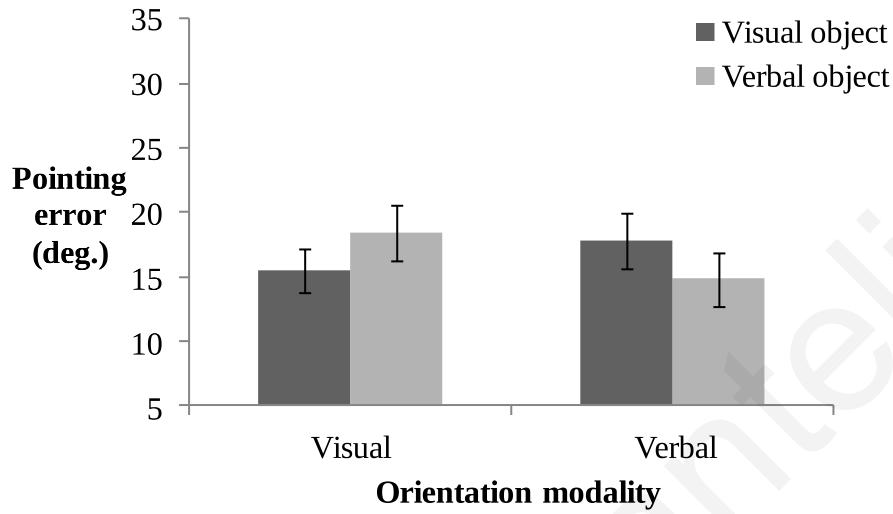
25
23
21
19
17
15
13
11
9
7
5
Reaction time (sec.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 10



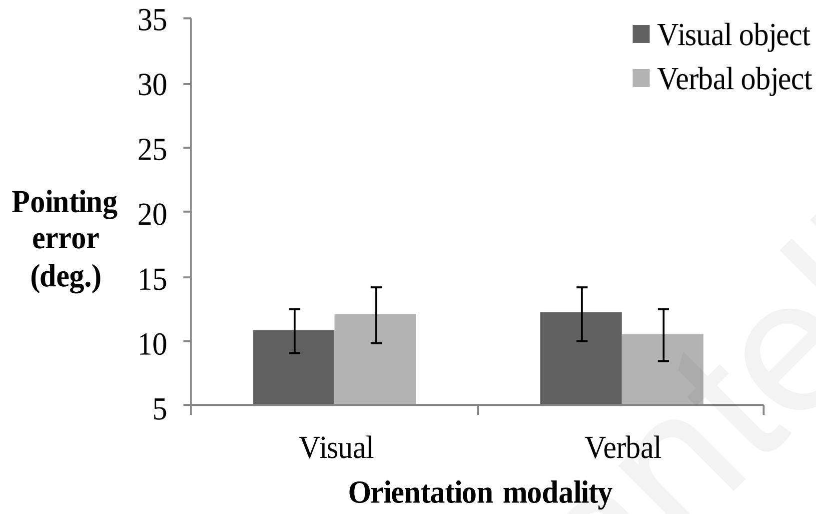
pot
shoe
lamp
basket
wine
vase
box
umbrella
mug
ball
Visual objects
Verbal objects

Figure 11



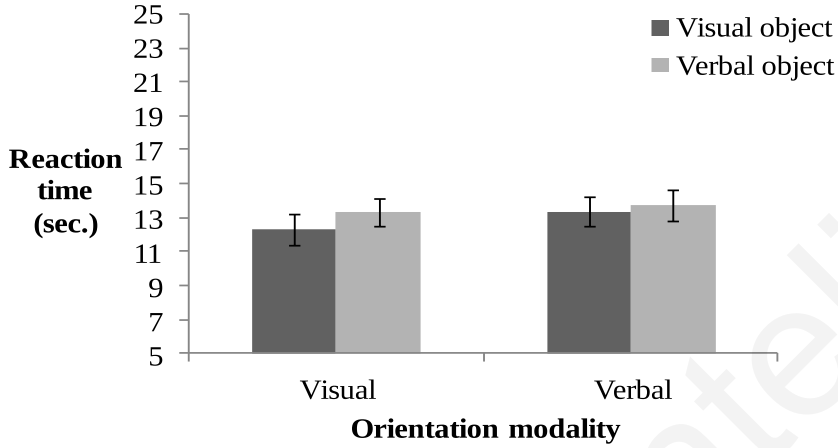
35
30
25
20
15
10
5
Pointing error (deg.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 12



35
30
25
20
15
10
5
Pointing error (deg.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 13



Visual object
Verbal object
Reaction time (sec.)
25
23
21
19
17
15
13
11
9
7
5
Visual
Verbal
Orientation modality

Figure 14

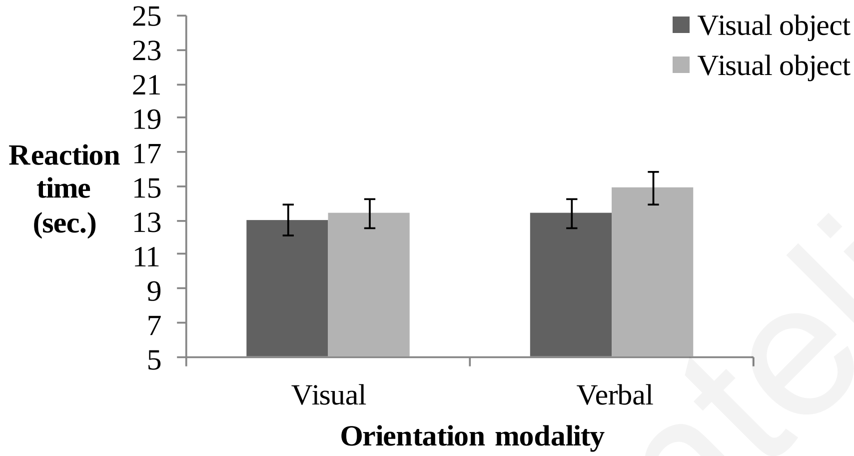
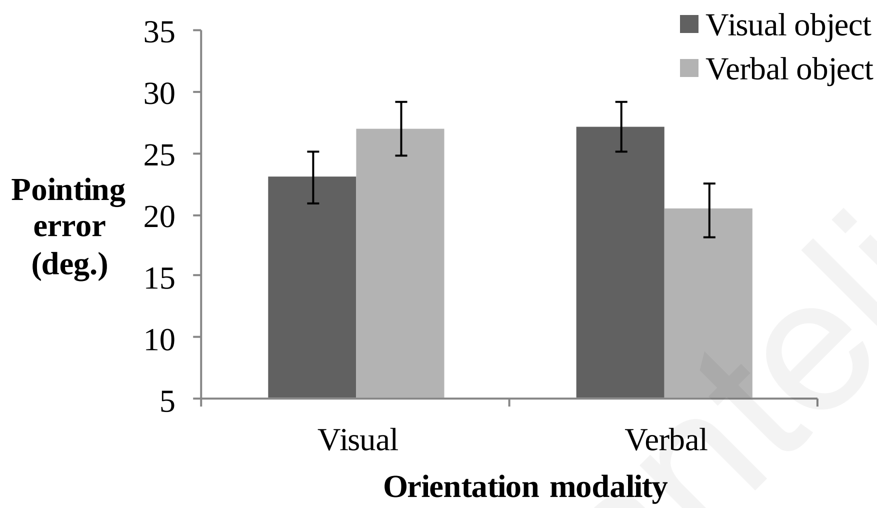

Visual object
Visual object
Reaction time (sec.)
25
23
21
19
17
15
13
11
9
7
5
Visual
Verbal
Orientation modality

Figure 15



Visual object
Verbal object
35
30
25
20
15
10
5
Pointing error (deg.)
Visual
Verbal
Orientation modality

Figure 16

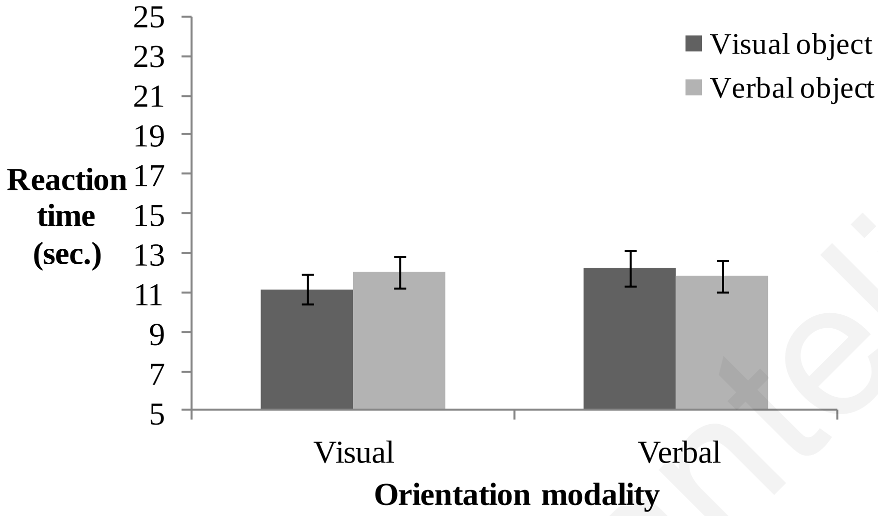
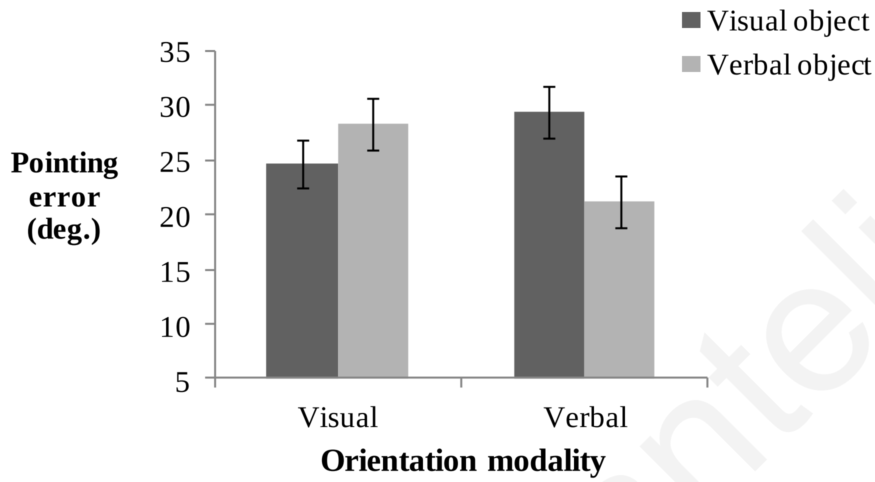
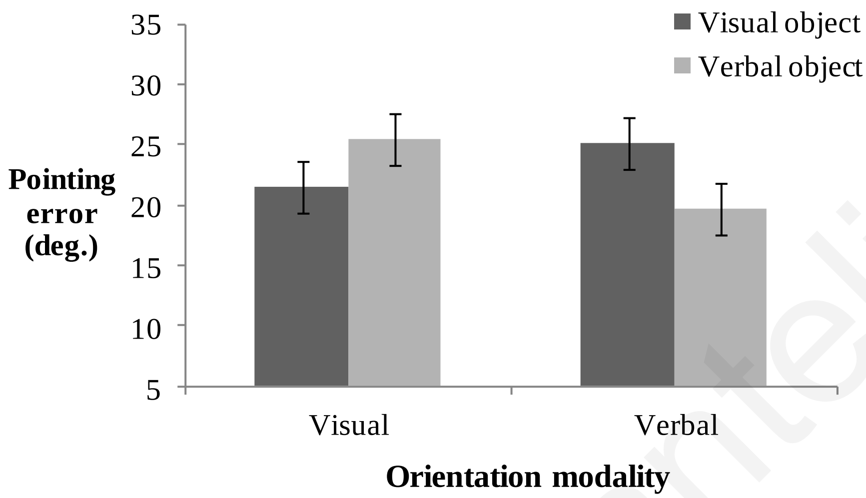

25
23
21
19
17
15
13
11
9
7
5
Reaction time (sec.)
Visual object
Verbal object
Visual
Verbal
Orientation modality

Figure 17



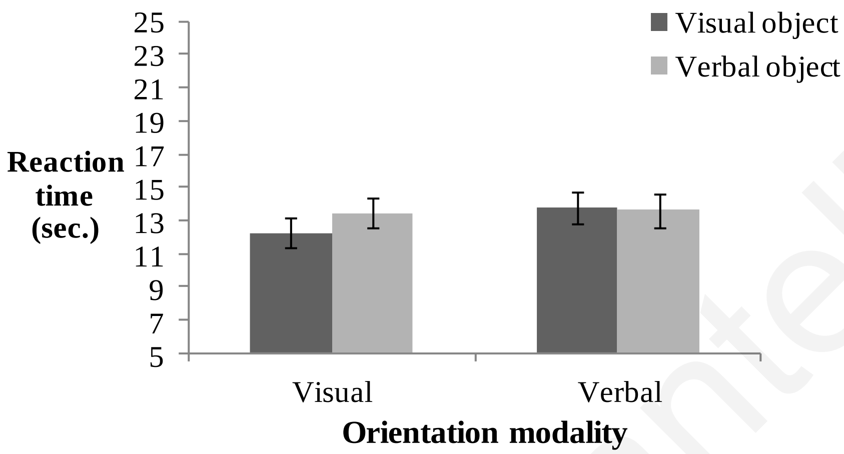
Visual object
Verbal object
35
30
25
20
15
10
5
Pointing error (deg.)
Visual
Verbal
Orientation modality

Figure 18



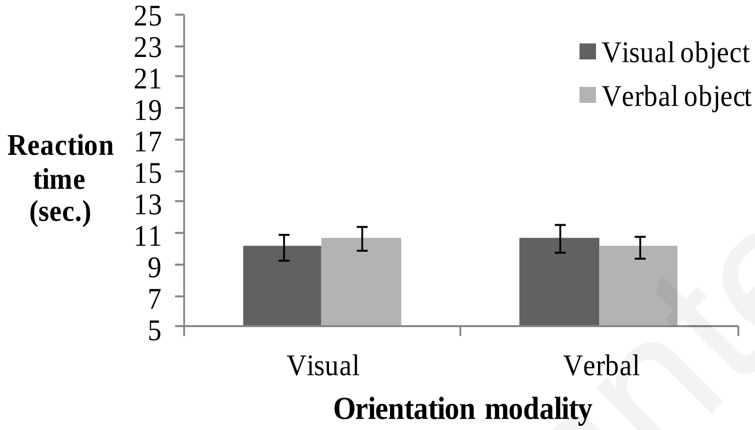
Visual object
Verbal object
35
30
25
20
15
10
5
Pointing error (deg.)
Visual
Verbal
Orientation modality

Figure 19



Visual object
Verbal object
Reaction time (sec.)
25
23
21
19
17
15
13
11
9
7
5
Visual
Verbal
Orientation modality

Figure 20


Reaction time (sec.)
25
23
21
19
17
15
13
11
9
7
5
Visual object
Verbal object
Visual
Verbal
Orientation modality