MOVING THROUGH LANGUAGE:

A BEHAVIOURAL AND LINGUISTIC ANALYSIS OF SPATIAL MENTAL MODEL CONSTRUCTION

Fabio Parente, BA, MRes

Thesis submitted to the University of Nottingham for the degree
of Doctor of Philosophy

December 2016

Total Word Count: 108,956

ABSTRACT

Over the past few decades, our understanding of the cognitive processes underpinning our navigational abilities has expanded considerably. Models have been constructed that attempt to explain various key aspects of our wayfinding abilities, from the selection of salient features in environments to the processes involved in updating our position with respect to those features during movement. However, there remain several key open questions. Much of the research in spatial cognition has investigated visuospatial performance on the basis of sensory input (predominantly vision, but also sound, hapsis, and kinaesthesia), and while language production has been the subject of extensive research in psycholinguistics and cognitive linguistics, many aspects of language encoding remain unexplored.

The research presented in this thesis aimed to explore outstanding issues in spatial language processing, tying together conceptual ends from different fields that have the potential to greatly inform each other, but focused specifically on how landmark information and spatial reference frames are encoded in mental representations characterised by different spatial reference frames. The first five experiments introduce a paradigm in which subjects encode skeletal route descriptions containing egocentric ("left/right") or allocentric (cardinal) relational terms, while they also intentionally maintain an imagined egocentric or allocentric viewpoint. By testing participants' spatial knowledge either in an allocentric (Experiments 1-3) or in an egocentric task (Experiments 4 and 5) this research exploits the facilitation produced by encoding-test congruence to clarify the contribution of mental imagery during spatial language processing and spatial tasks. Additionally, Experiments 1-3 adopted an eye-tracking methodology to study the allocation of attention to landmarks in descriptions and sketch maps as a function of linguistic reference frame and imagined perspective, while also recording subjective self-reports of participants' phenomenal experiences. Key findings include evidence that egocentric and allocentric relational terms may not map directly onto egocentric and allocentric imagined perspectives, calling into question a common assumptions of psycholinguistic studies of spatial language. A novel way to establish experimental control over mental representations is presented, together with evidence that specific eye gaze patterns on landmark words or landmark regions of maps can be diagnostic of different imagined spatial perspectives.

Experiments 4 and 5 adopted the same key manipulations to the study of spatial updating and bearing estimation following encoding of short, aurally-presented route descriptions. By employing two different response modes in this triangle completion task, Experiments 4 and 5 attempted to address key issues of experimental control that may have caused the conflicting results found in the literature on spatial updating during mental navigation and visuospatial imagery. The

impact of encoding manipulations and of differences in response modality on embodiment and task performance were explored.

Experiments 6-8 subsequently attempted to determine the developmental trajectory for the ability to discriminate between navigationally salient and non-salient landmarks, and to translate spatial relations between different reference frames. In these developmental studies, children and young adolescents were presented with videos portraying journeys through virtual environments from an egocentric perspective, and tested their ability to translate the resulting representations in order to perform allocentric spatial tasks. No clear facilitation effect of decision-point landmarks was observed or any strong indication that salient navigational features are more strongly represented in memory within the age range we tested (four to 11 years of age). Possible reasons for this are discussed in light of the relevant literature and methodological differences.

Globally, the results presented indicate a functional role of imagery during language processing, pointing to the importance of introspection and accurate task analyses when interpreting behavioural results. Additionally, the study of implicit measures of attention such as eye tracking measures has the potential to improve our understanding mental representations, and of how they mediate between perception, action, and language. Lastly, these results also suggest that synergy between seemingly distinct research areas may be key in better characterising the nature of mental imagery in its different forms, and that the phenomenology of imagery content will be an essential part of this and future research.

ACKNOWLEDGMENTS

I am grateful to my supervisors, Alastair D. Smith and Ruth Filik, for their support and guidance over the past few years. They have kept me grounded during my frequent flights of fancy, challenged me, and helped me navigate (some degree of pun intended) a very complex research area. I would also like to extend my gratitude to the University of Nottingham and those of its students, both undergraduate and postgraduate, who have agreed to take part in my experiments over the years. Additionally, I would like to thank the parents and the children who have attended Summer Scientist Week, and who have contributed to making it an exceptional opportunity both for data collection and for science outreach within the local community. A special thank you to Alastair for tirelessly coordinating the organisation of the event.

On a more personal note, I want to thank the many people whose presence in my life has made all the difference both during and before my doctorate. Alexandra, Beerelim, and Lawrence, and all the other PhD students and postdocs (Dominic and Prasannah deserve a special mention and a pint) – for being there and sharing in all the good, bad, and frustrating times, and for the countless stimulating conversations.

The biggest thank you, and all my love, goes to my family for their unwavering support and patience during the most challenging, exciting, beautiful, and exhausting four years of my life. Your ability to believe in me even when I cannot is my greatest source of strength, and I hope I can make you as proud as you make me.

Last, but not least, Ms Binotti and Ms Neri – pillars of my high-school experience – deserve recognition and the accolade of Most Patient Teachers in the History of Teachers. I drove you nuts, yet you never failed to push me to do better. If I am taking these first steps in hope of becoming a capable researcher, it is only because you were able to keep my love for science alive even when school tried to kill it. Whatever mess I make in academia from now on is on you.

Author's declaration

I hereby declare that the work contained within this thesis was carried out in accordance with the Regulations of the University of Nottingham. The work is original except where indicated by special reference in the text, and no part of this work has been submitted for examination as part of any other degree, in the United Kingdom or overseas. All views expressed in this thesis are those of the author and in no way represent those of the University of Nottingham.

Signed: **Fabio Parente** Date: **17/10/2016**

CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

I dedicate this thesis to my grandmother, Pina. You unknowingly led me here even as you were losing yourself, and gave me the strength and motivation to pursue a dream I had cast aside. I will be forever grateful.

CHAPTER 1
Navigation and Mental Imagery

1.1. Overview

Humans are inherently spatial creatures, and our survival is ultimately conditional on our ability to interact meaningfully and efficiently with our surroundings. Whether we are reaching for a glass of water located on the desk next to our laptop, walking across the room to reach a pile of papers in a bookcase, or walking into town to run an errand, this crucial ability involves a complex interplay of different cognitive mechanisms. These include the identification of salient features in the environment (e.g. a bright red post box), and the construction of mental representations of environmental space that can merge object identity with location and distance information (e.g. a bright red post box located at a T junction, a few hundred metres down the road from our house). Finally, active navigation requires the planning, execution, and online monitoring of motor behaviour (e.g. which way do we turn to get to the T junction, and which way do we turn there with respect to the post box to reach our destination?).

However, much of our usual navigational behaviour can often involve heading towards more distal goals that lie beyond our immediate perceptual field. In such situations, the spatial representations that we require to plan our motor behaviour must be informed by our long-term knowledge of the environment in which we are operating. Planning a route between two buildings located on opposite sides of one's university campus, for example, requires an understanding of the relative spatial positions of the two locations and of the potentially salient navigational landmarks that might be located between them, as well as knowledge of the network of roads connecting them. There are, however, situations in which even our long-term memory cannot be relied upon to guide navigation. The exploration of a novel environment, such as a town we are not familiar with, might require us to operate on the basis of information provided to us via linguistic propositions. Whether we are asking a stranger for directions or reading a series of route directions on the Web, we will need to extract navigational information from the linguistic content and generate on that basis an appropriate mental model of an environment we cannot directly perceive. Depending upon the type and richness of the information provided, the resulting mental model may display more or less detailed visuospatial properties and may be perceived as phenomenologically analogous to the active exploration of the real environment. The research presented in this thesis aims to explore, at least in part, the nature of and interactions between the various cognitive processes involved in the construction of said spatial mental models and visual mental images from linguistic input when allothetic (i.e. optic flow) and idiothetic (i.e. proprioception) cues generated by active motion during navigation are not available. This body of work is principally concerned with exploring the way in which linguistic manipulations and imagery manipulations interact with each other both during the encoding of spatial linguistic content and during subsequent performance of spatial tasks.

Chapters 1 and 2 will lay the theoretical foundations for the work presented in this thesis, discussing and introducing a number of notions necessary to ground and interpret the following experiments and central to an understanding of spatial cognition. These are: spatial reference frames (Section 1.2), landmarks, and landmark salience (Sections 1.3). Subsequently, I will discuss how these components might be integrated within mental representations that can support navigation (Sections 1.4-1-10), and what factors (individual differences and environmental factors) might drive the selection of certain representations over others (Section 1.11). In Chapter 2, this body of work will be framed within the context of the processes that might underlie the transfer of information between external representations (e.g. maps, or linguistic descriptions of space) and internal representations thereof (Section 2.1). Accordingly, research will be reviewed that has explored the interaction between the human language faculty and navigational abilities (Section 2.2), providing a theoretical motivation for studying this interaction. This will then be followed by a discussion of the factors (cognitive and linguistic) that can influence encoding processes during language processing during the construction of mental representations (Section 2.3). Similarly, the factors influencing the production of external representations will be discussed, particularly with respect to the role played by representational congruency between encoding and test (Section 2.5). Last but far from least, eye movements will also be discussed as potential windows into the construction of mental representations of space during the processing of spatial language, and on mental imagery in general (Section 2.6). The information presented in this section will be paramount for a complete understanding of Experiments 1-3, presented in Chapter 3, and will additionally introduce elements that are central to the broader theoretical framework of this thesis, such as the susceptibility of eye movements to top-down effects during reading and scene processing.

1.2. Spatial Reference Frames

A fundamental requirement of successful navigation is the ability to encode the position of objects and environmental features within cognitive structures that can both support immediate navigation and the formation of enduring spatial representations in long-term memory. This ability relies on the use of spatial coordinate systems onto which spatial locations can be anchored.

Traditionally, the spatial cognition literature has distinguished between egocentric (or body-centred) and allocentric (or geocentric) reference frames. This distinction has hinged largely on three aspects: the type of input required to generate them, the type of cognitive processes and spatial tasks they support, and the developmental and cognitive hierarchy in which they are structured. Early developmental models (Piaget & Inhelder, 1967) postulated that the ontogenesis of spatial abilities in children followed a set of sequential milestones, with early reliance on egocentric representations and a qualitative shift towards more complex allocentric representations upon onset of independent locomotion. This model was later expanded into a more general model of spatial microgenesis (Siegel & White, 1975), which assumed a stepwise acquisition of three categories of environmental knowledge. Landmark knowledge concerns the identity of salient and stable environmental features, or discrete object, and is based on egocentric reference frames. Route knowledge involves an egocentric understanding of the paths connecting the various landmarks, and of the sensorimotor sequences that allow navigation between them. It is initially non-metric and improves with repeated exposures to the environment (Ishikawa & Montello, 2006; Montello, 1998). Survey knowledge is a map-like, allocentric representation of the global environment that can support the plotting of alternative routes and shortcuts.

An egocentric frame of reference codes spatial relations on a coordinate system centred on the organism itself. This type of reference frame is thought to be the one most readily constructed on the basis of sensory input during active navigation. Due to this, the encoding (and, consequently, the recall) of visuospatial information within an egocentric frame of reference is also orientation-specific and viewpoint-dependent. They are considered less flexible and primarily used to support perception-driven navigation in near or peripersonal space. As we walk towards and reach for an object, for example, we must construct a motor program that will first direct our legs to move in its general direction, and then our arm and hand towards it.

However, not all navigational (or, more generally, visuospatial) behaviour relies on the processing of spatial relations within a body-centred frame of reference. Updating self-object spatial relations during movement on the basis of idiothetic input (a process known as egocentric path integration) is thought to be subject to cumulative error over increasing distances (Burgess, 2008). During instances of navigation in larger environments and over longer distances, allocentric representations are usually preferred. An allocentric reference frame encodes the position of objects in an environment not with respect to the navigator's body, but with reference to each other or to other stable environmental features on a set of coordinates centred on the global environment itself. This type of spatial relation coding is fundamental, for example, in the process of maintaining a stable heading while moving towards a more distal location in extrapersonal space, and it is central to many models of visuospatial long-term memory (Burgess, 2006; 2008).

Over time, evidence has emerged to challenge the assumption of a stepwise acquisition and hierarchical organisation of reference frames, both in microgenetic and ontogenetic terms (discussed in more detail in Chapter 5, in which three developmental studies are presented). First of all, the egocentric perceptual experience we perceive as unitary is already the result of the synthesis of sensory input originating in different intrinsic reference frames (Galati, Pelle, Berthoz & Committeri, 2010). In order to code and update the relative position of our body and limbs during motion with respect to the reference object we must rely on predominantly egocentric sensory experiences, such as visual, somatosensory, vestibular, and auditory input. At the lowest level, these sensory inputs are acquired within slightly different body-centred coordinate sets. Optic flow input is first used to plot the necessary spatial relations within retinotopic coordinates (i.e. the object in question might appear in the lower-right quadrant of our visual field) (Török, Nguyen, Kolozsvári, Buchanan & Nadasdy, 2014). Auditory and vestibular input is acquired in head-centred coordinates, and proprioceptive information in body-centred coordinates. In a series of processing stages, these various inputs must be integrated by shifting and merging the receptive fields of different neuronal populations into a single coherent reference frame that can support navigation (Avillac, Denève, Olivier, Pouget & Duhamel, 2005; Fogassi & Lupino, 2005) and that we perceive as egocentric.

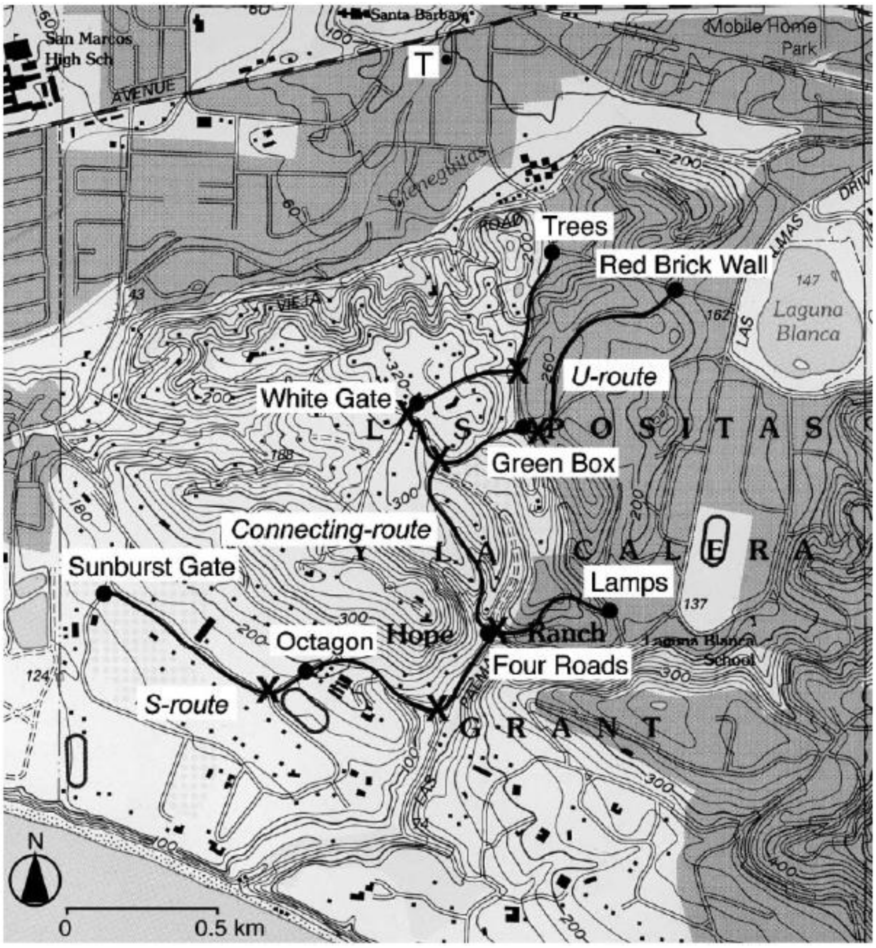


Figure 1.1 - The two routes travelled by Ishikawa and Montello's (2006) participants.

Additionally, the idea that our spatial understanding of a novel environment is initially purely egocentrically constrained has also failed to stand up to further scrutiny. Montello (1998) proposed an alternative theoretical framework according to which spatial knowledge acquisition follows a continuous trend. Rather than progressing through qualitatively different stages, knowledge of distances between environmental locations should be above chance already after early exposures and increase continuously as a function of experience with the environment. However, the extent and accuracy of said spatial knowledge (as well as the rate of improvement over time) will be a function of individual differences. Additionally, this continuous framework posited that integrating spatial knowledge of separate environments acquired during distinct navigational events into a single allocentric knowledge structure represents the only real qualitative step during spatial microgenesis.

Ishikawa and Montello (2006) tested this framework by exposing a sample of university students to two routes in unfamiliar neighbourhoods (Figure 1.1) over 10 weekly sessions. Participants were driven along the routes, and along a shorter path connecting them. During the first three sessions, they wore blindfolds while travelling circuitously between the two test routes. Starting from the fourth session, they were driven along a direct connecting route without blindfolds, in order to allow them to integrate their knowledge of the two routes into a single mental representation. After each session, participants carried out direction and straight-line distance estimation tasks between pairs of landmarks, and after every other session they drew sketch maps of the routes, including their shapes, the spatial relation between them, and the four landmarks encountered on each one. Following exposure to the connecting route, participants were probed on both within- and between-route direction and distance estimates. Additionally, participants took the Santa Barbara Sense of Direction (SBSOD) self-report scale (Hegarty, Richardson, Montello, Lovelace & Subbiah, 2002). Results showed that already after a first exposure, participants were able to acquire landmark, route, and survey knowledge that included above-chance awareness of metric knowledge (understood as quantitative but approximate knowledge of distances between locations), confirming one of the predictions of Montello's (1998) framework.

Very little group-level improvement was observed between the first and the 10th session for within-route tasks, whereas participants' understanding of the connection between the two test routes (as evidenced by the maps drawn after the fourth session) showed more evidence of improvement (and was reported by participants to be more challenging than other tasks). Analyses of individual participants' data, however, showed evidence of considerable between-subject variability, also consistent with predictions. Good and poor performers were found to be such consistently already from the first session, and approximately half of the participants showed slight evidence of improvement over time. Interestingly, participants' SBSOD scores were found to positively correlate with their performance in direction and distance estimates, and in the map drawing tasks. However, this was only the case following exposure to the more complex U-route (see Figure 1.1), which involved multiple changes in heading, and for between-route direction estimates which required the construction of a more complex survey representation of the environment containing both routes. This confirmed multiple predictions of the alternative framework: that route complexity and individual differences would modulate performance, and that integrating spatial representations of distinct routes into a single one would represent a qualitative step in the acquisition of spatial knowledge.

In a more recent study, Ishikawa (2013) presented participants with a video of an urban route containing five turns and five landmarks (counterbalanced between turn and non-turn locations). After watching the video, participants were then tested on four spatial tasks. In a landmark memory task, participants had to list the names of the five landmarks encountered, in order of appearance. In a route-choice task, participants were shown five egocentric snapshots of intersections and asked to whether they remember turning at that location during encoding, and in what direction. In the direction estimation task, participants estimated the spatial relationships between the five landmarks encountered (for a total of 10 pairs). In the map-sketching task, participants were asked to draw as accurate a map of the learned route as possible. Additionally, half of participants repeated the spatial tasks after 2 weeks, and half after 3 months from exposure, in both cases without watching the video a second time.

Results showed differential patterns of memory decay for landmark, route, and survey knowledge as a function of sense of direction (as measured by participants' SBSOD scores). More specifically, the two groups displayed comparable rates of rapid decay of landmark name recall and topological route knowledge, but individuals with a better self-reported sense of direction showed a significantly lower rate of decay of survey knowledge. It therefore appears that the use of allocentric representations of space is per se no more effortful than the construction and processing of egocentric representations. During active navigation, the sensory input acquired via different modalities can be merged and form the basis of viewpoint-dependent egocentric snapshots of events and locations. These are action-oriented representations of self-object spatial relations (Burgess, 2006), and the automatic use of visual, vestibular, and kinaesthetic input allows them to support spatial updating over short distances (Riecke, Cunningham & Bülthoff, 2007). Allocentric representations can also be generated based on sensory input after relatively short exposures to novel environments, and a tendency to favour either spatial reference frame is the result of a complex interplay of disparate factors. These include environmental features, the degree of motion involved, task demands, neurodevelopmental characteristics, sociogeographical differences, age, and others (see Section 1.11).

Egocentric and allocentric representations, however, are not only constructed in parallel but are also inherently interactive. This means that the way in which humans initially experience an environment can also influence the resulting long-term representations of that space. McNamara, Rump and Werner (2003) had participants learn the locations of eight objects located at the intersections of two paths encircling a large, rectangular building, and in the vicinity of a salient environmental landmark (a lake). One of the paths was aligned to the walls of the building, while the other was out of alignment by 45 degrees. Subjects subsequently had to inspect their mental representations of the environment and point to the target objects from imagined vantage points and headings. Pointing accuracy was greater after experiencing the environment from the aligned path compared to the misaligned one, indicating the fundamentally allocentric nature of participants' representations. However, imagined headings aligned with the salient landmark also led to increased pointing accuracy, and this was taken as indication that the geocentric features used to construct intrinsic reference frames are selected on the basis of egocentric experience. Additionally, the results provided evidence of orientation-dependent alignment effects in otherwise allocentric spatial memories.

This deeply interactive system of parallel reference frames raises several important questions that are relevant to the current research. Namely, what processes mediate the construction of spatial representations based on linguistic input and how do egocentric and allocentric reference frames interact within this domain? Are both egocentric and allocentric representations constructed in parallel based on linguistic input? And, if that is the case, can experimental paradigms be developed that will allow to determine, on the basis of dependent measures of linguistic encoding and visuospatial performance, the type of reference frame adopted in the construction of the underlying spatial representations? However, before addressing these questions, other fundamental notions must be discussed in more detail. Among them is the idea of landmark, which will be covered in the next section.

1.3. Landmarks and Spatial Learning

Previous research (Newcombe & Huttenlocher, 2003; Newcombe, Huttenlocher, Drummey & Wiley, 1998) has categorised these spatial coding systems we use to encode the locations and relative positions of entities in environments on the basis of the reference frame upon which they rely, of the type of spatial relations they encode, and of the behavioural complexity they can support. More specifically, spatial coding systems can be classified depending on whether they code spatial relations with respect to the self (and within an egocentric frame of reference) or with respect to external landmarks (and within an allocentric frame of reference).

The systems known as *response learning* and *dead reckoning* fall within the first group and require a constant awareness of one's own position in space. The former involves the re-enactment of motor sequences whose accuracy in reaching a target depends on a constant starting point (e.g. reaching for the right-hand drawer when seated on one particular side of the desk), whereas the latter is a more complex system involving the integration of optic flow, vestibular, and kinaesthetic information in order to update one's position.

On the other hand, the location of a target object within an environment is often encoded with respect to other stable features (i.e. landmarks) within the environment itself. Given the importance of landmarks in guiding many instances of navigational behaviour, it is important also to construct a taxonomy of functions that they can assume. In this sense, a distinction can be made between landmarks used as associative cues for navigational actions and those used as beacons. Cue learning of spatial locations involves the direct association of a target object or location with a coincident landmark, provided that the association is habitual or otherwise stable over time. For example, one might keep wine glasses in the cupboard right above the sink. The association can also involve a landmark region rather than a landmark object. In this sense, both wine glasses and the sink are associated with a region of space located in one's kitchen. However, in certain situations, no distinctive, coincident landmark may be available that can serve as an associative cue, such as when we are attempting to locate our car in a full parking lot. In such cases, place learning requires that the target object or location be encoded in terms of its distance and relative direction from more distal landmarks. These landmarks, or beacons, are defined as highly visible navigational objects that indicate or are target locations (Chan, Baumann, Bellgrove & Mattingley, 2012), providing highly accurate positional information even from a long distance and from all locations in the environment. A skyscraper or a church's spire would be examples of target landmarks within an urban environment that might act as beacons. In a study aimed at testing the relative advantages and disadvantages of beacon and associative cue navigation, Waller and Lippa (2007) had participants explore a virtual environment composed of 20 rooms in a linear sequence, each of which contained two doors. Only one would allow the participant to progress to the following room, and doors could either be marked by a single landmark placed between them (Associative Cue) or by two landmarks, each placed next to one door (Beacons). Additionally, a "No Landmark" condition was included to test for the facilitating effect of landmark presence. Over the course of several trials, participants navigated through the same environment, allowing the experimenter to record both the number of correct doors selected overall and the increase in accuracy over subsequent trials.

Results revealed that the presence of landmarks led to better performance compared to the No Landmark condition. However, the facilitating effect of landmark presence was modulated by the function of the landmarks, leading to greater increases in accuracy earlier in the experiment when they acted as beacons compared to when they acted as associative cues. That is, accuracy increased more quickly when participants could simply encode the identity of the landmarks to aim for in the various rooms. This, however, also translated into a poorer recall of directional information when landmarks were removed in the last trial, indicating that the need to only encode landmark identity during beacon navigation may lead to weaker consolidation of directional information.

However, a perhaps more fundamental issue than the function of landmarks in navigation is the nature of what constitutes a landmark in the first place. In spite of the central role of landmarks in guiding spatial navigation, no univocal definition of the term has been presented in the literature. This is perhaps an indication of the considerable flexibility with which landmark selection occurs. Stable environmental features are normally selected as navigational aids if they present a higher degree of salience compared to other environmental features. Although the determination of this salience is far from being a simple cognitive task, attempts have been made to determine both its neural and psychological underpinnings.

A number of studies have closed in on the neural circuitry that appears to be involved in responding to navigationally salient features of environments, while also providing behavioural correlates for landmark salience discrimination. Janzen and van Turennout (2004) studied the role of the parahippocampal gyrus (PHG) in encoding landmark objects during navigation. In an fMRI study they presented adult participants with videos of a route through a virtual environment and instructed them to remember both the route and the objects they encountered. These objects could be either toys or objects belonging to other semantic categories, and they could be located either at intersections (decision point objects) or at simple turns (non-decision point objects). Participants were further instructed to pay particular attention to the toys, in order to be able to guide a group of children along the tour. Following route learning, participants engaged in an object recognition task, during which they were shown previously encountered and novel objects of both semantic categories and asked to determine via button press whether they had seen the objects or not. During this phase, the objects were presented from a canonical orientation on a white background, to separate the recall of the object identities from that of the spatial information participants may have encoded during learning.

No significant differences were found in response accuracy rates as a function of semantic category or navigational salience. However, toys were responded to significantly more quickly than non-toys, and toys at decision points significantly more quickly than toys at non-decision points. On the other hand, response times did not differ as a function of navigational salience for objects in the non-toy semantic category, indicating that the navigational salience of landmarks may, to an extent, interact with other task-related top-down demands, such as instructions to attend to specific categories of landmarks. In neural terms, navigational salience and semantic salience were found to be served by distinct neural mechanisms, with stronger activation in the right fusiform gyrus (BA 37) for attended objects (toys) compared to unattended objects (non-toys), and increased activation in the left and right parahippocampal gyri for decision-point objects compared to non-decision point objects. In the right PHG, the increased activation for decision-point objects was also found for forgotten objects (objects that were present in the videos, but that participants had incorrectly judged not to have seen).

Globally, the results suggested that the encoding of navigational salience is automatic (present when participants are instructed to attend to objects based on non-navigational criteria), independent of spatial information requirements during retrieval (when objects are presented in isolation), and even of conscious recall of the object landmarks. The study also specifically implicated the PHG in the acquisition of object-place associations during route learning, and indicated that this form of learning requires limited exposure to the environment, allowing fast and dynamic changes to spatial maps during navigation. This was confirmed in a following study by Janzen, Wagensveld and van Turennout (2007), who exposed participants to different route sequences a different number of times. Results revealed that the number of exposures (one vs three) did not modulate the differential parahippocampal activation for decision-point objects compared to non-decision point objects. The representation of landmark salience was already stable after one exposure to the route, meeting an important requirement for a navigational system capable of quickly acquiring navigationally salient information and of maintaining it over time.

However, Janzen, Jansen and van Turennout (2008) observed that time from exposure and the resulting memory consolidation did influence hippocampal and parahippocampal activity, but that this effect was modulated by navigational ability. In that study, participants were presented with two route sequences through a virtual environment containing landmarks at both decision and non-decision points. As in previous studies, participants were instructed to explicitly attend to a specific class of objects, the toys, rather than the other objects, regardless of their spatial location. One route was presented the evening prior to the fMRI scanning session, and the other immediately before it. An object recognition task was performed during scanning as in previous studies, and participants indicated whether they had seen the presented objects in either of the two routes they had experienced. Participants were divided into good and bad navigators based on their score on the Santa Barbara Sense of Direction Scale (SBSOD), a self-report measure of navigational skills already introduced in Section 1.2 as a correlate of survey spatial abilities.

Behaviourally, accuracy rates revealed higher error rates in response to landmark objects encountered the evening before scanning, and lower error rates for landmarks of the attended semantic category (toys). Additionally, toys at decision points were recalled more accurately than those at non-decision points, but this effect of navigational salience was not present for non-toy objects. Attended objects also elicited faster responses than non-attended ones. An analysis of the fMRI data revealed that objects encountered the night prior to scanning elicited stronger bilateral hippocampal activity. This consolidation effect was positively correlated with participants' SBSOD scores, with good navigators also displaying stronger responses in the PHG to consolidated decision-point landmarks compared to recently encountered ones.

The results of Janzen et al. (2008) pointed to a role of memory consolidation and individual differences in navigational salience perception, and strengthened the view that the PHG is involved in the enduring representation of navigationally salient landmark information. However, the mechanism via which this salience determination is carried out so that only useful information is stored remained to be elucidated. In a following study (Janzen & Jansen, 2010) this mechanism was more closely studied by confronting participants with ambiguous landmark information (i.e. instances in which potentially salient landmarks appear at two different decision points requiring two different directional turns). Participants actively explored a virtual environment containing objects they were explicitly instructed to attend to (toys) and objects belonging to other semantic categories. Each object appeared twice at two different decision points (D-D objects), at two different non-decision points (ND-ND objects), or at one decision and at one non-decision point (D-ND and ND-D objects, also "one-D objects"), for a total of 288 encounters. Active exploration was followed by an object recognition task (during fMRI scanning) that included both previously encountered and novel toys and non-toys. During this task, each object was presented only once, and participants had to judge whether they had encountered it during exploration of the environment or not. Behaviourally, D-D objects were found to elicit the most errors and ND-ND objects the fastest responses. Once again, attended objects yielded lower error rates and faster responses than unattended objects.

An analysis of the fMRI data showed that one-D objects elicited greater parahippocampal activity compared to ND-ND objects, irrespective of the semantic category of the objects and consistent with previous findings (Janzen et al., 2007; Janzen & van Turennout, 2004). On the other hand, D-D objects elicited greater activity than ND-ND objects in the right middle frontal gyrus, a prefrontal region implicated in cognitive control (Miller & Cohen, 2001), spatial working memory (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998), in the selection of contextually relevant information (Ridderinkhof, Ullsperger, Crone, & Niewenhuis, 2004), and the detection of expectation violations (Corlett et al., 2004; Fletcher et al., 2001). Additionally, the middle frontal gyrus was found to respond more strongly to D-D objects associated with different directional turns compared to D-D objects associated with turns in the same direction. Globally, these findings suggest that the determination of navigational salience is a flexible process that is continuously informed by incoming input and that conflicting or misleading information pertaining to navigationally salient regions of a route or environment activates areas involved in executive functions such as cognitive control.

The role of the PHG in the marking of navigationally salient landmarks was further explored by Wegman and Janzen (2011), who studied its resting state connectivity with other brain regions. As in previous studies, participants were shown a video of routes through four sections of a virtual environment containing landmark objects both at decision and non-decision points. Participants were instructed to learn the routes and to pay particular attention to objects of interest to children visiting the environment (i.e. toys). All objects appeared on posters located at decision points and non-decision points, and each section contained the same number of attended and unattended objects located at navigationally salient and non-salient points.

Unlike in previous studies, participants' eye movements were recorded during the learning phase. These data were used to segment sections of the fMRI recordings that corresponded to object viewing period, defined as the number of consecutive frames participants' eye gaze was on the object's coordinates. For each object, the video frame in which the object was no longer visible was taken as the offset of the object viewing trial. However, eye gaze data also provided a measure of attention allocation. They revealed that participants spent longer looking at toys compared to objects belonging to other semantic categories, but also that toys located at non-decision points were fixated for longer than toys at decision points, and toys at non-decision points for longer than non-toys at non-decision points.

In this study, fMRI recordings were made during route learning, and while participants performed a landmark recognition task. During learning, first fixations on decision-point landmarks were found to result in increased activity in the PHG compared to fixations on non-decision point landmarks. Relatedly, periods of looking at screen locations with objects corresponded to periods of increased activity in the PHG compared to looking at regions without objects, and increased PHG activation for an object was predictive of its successful recall during object verification. Additionally, decision points without landmark objects also resulted in higher PHG activation compared to empty non-decision points, indicating that this region is sensitive to the navigational salience of a decision point within a route, irrespective of the concurrent presence of a landmark object.

Furthermore, resting state functional connectivity scans were performed before and after the learning phase. This was intended to investigate how spatial learning alters the connectivity between the PHG and the rest of the brain. More specifically, changes in functional connectivity were investigated between the PHG and regions involved in egocentric and allocentric navigation respectively: the caudate nucleus and the hippocampus (Hartley, Maguire, Spiers, & Burgess, 2003;

Voermans et al., 2004). The functional connectivity analysis revealed changes in connectivity between pre- and post-learning that correlated with participants' self-reported navigational abilities as measured by the SBSOD. More specifically, SBSOD scores were found to positively correlate with the rate of post-learning connectivity increase between the PHG and the right hippocampus, but negatively with the rate of post-learning connectivity increase between the PHG and the right caudate nucleus. This finding is consistent with the idea that higher self-reported navigational abilities correlate with a preference for allocentric navigational strategies which rely on hippocampal regions. As discussed in Section 1.2., both egocentric and allocentric spatial reference frames can be computed in parallel, but such ability is susceptible to considerable between-subject variability. Accordingly, Wegman and Janzen suggest that an individual's propensity to employ an allocentric or egocentric navigational strategy might be a function of the degree to which landmark information is transmitted from the PHG to the hippocampus or the right caudate nucleus respectively.

The post-learning resting state scan was followed by an object recognition task akin to those used in previous studies. Recognition performance was found to be higher for toys compared to non-toys, and response times were found to be faster for toys compared to non-toys. An analysis of BOLD responses to D and ND objects during the recognition task revealed higher bilateral PHG and bilateral middle occipital gyrus activation for the former. Additionally, toys resulted in higher activation in the fusiform gyrus bilaterally, right middle temporal gyrus, and right superior occipital gyrus. Non-toys, however, resulted in greater activity in the left fusiform gyrus.

While the studies presented in this thesis do not contain brain-imaging components, the studies by Janzen and colleagues provide a theoretical foundation to explore the processing of the navigational salience of landmarks. Their results constitute evidence of a network of brain regions involved in the extraction of navigational salience information during egocentric route learning and landmark recall. The results indicate that the perception of navigational salience is fast and automatic, and that decision points in a route are perceived as inherently salient by the human navigational system even in the absence of landmarks. Furthermore, certain behavioural and neurophysiological measures of landmark salience (e.g. eye tracking measures of viewing time, or fusiform gyrus activity) were also found to be modulated by factors such as task demands (e.g. the requirement to focus on specific semantic classes of objects).

This might suggest that the determination of landmark salience is, despite its speed, a complex and multifactorial process integrating different types of bottom-up and top-down information, and that the interactions between these different factors must be better understood in order to correctly model landmark salience perception in its various forms. In one such model, Caduff and Timpf (2008) have proposed that landmark salience can be described as the vector product of three individual vectors representing Perceptual, Cognitive, and Contextual Salience. Perceptual Salience (PS) models the bottom-up allocation of attentional resources to features detected in the stream of sensory input. In the visual modality, Caduff and Timpf identify Location- and Object-based Attention (LA and OA), and Scene Context (SC) as the fundamental units of attention. LA involves the processing of visual stimuli from the entire visual field and their decomposition into feature maps that extract colour, intensity, and texture orientation information based on discontinuity, and their subsequent recombination into global saliency maps (Itti, Koch & Niebur, 1998) (Figure 1.2). OA can single out individual objects in a scene based on their structure and geometric features, such as size, shape, and orientation. SC operates at the global scene level, and integrates the other two components of perceptual salience with relevant contextual information. This component can allow the differential salience weighting and disambiguation of otherwise perceptually identical objects owing to their different spatial locations and spatial relations within the scene.

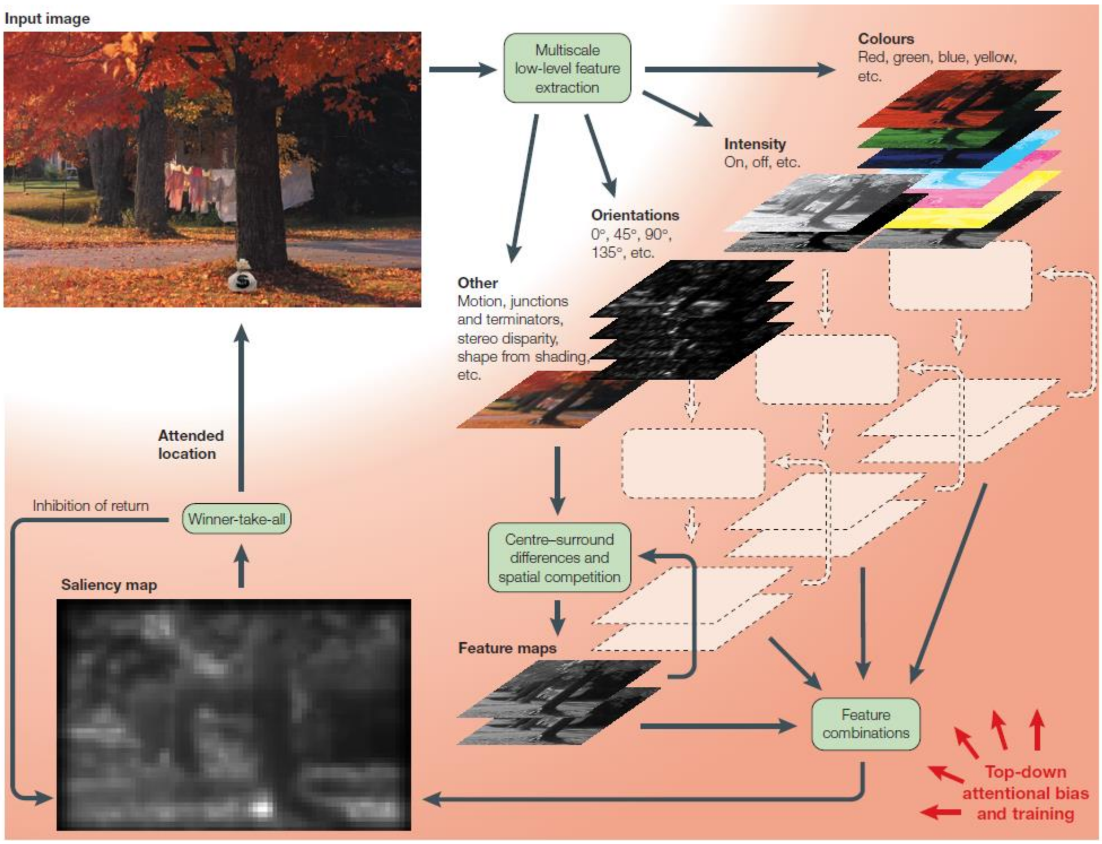


Figure 1.2 - Flow diagram of Itti and Koch's (2001) bottom-up attention model.

Cognitive Salience describes the top-down allocation of attention as a function of the viewer's prior knowledge and experience, and it relies on the construction of mental representations of spatial environments. The availability for extraction of individual objects or environmental features from these representations is taken to be a function of their Degree of Recognition (DR) and Idiosyncratic Relevance (IR). DR occurs as the degree of matching between a viewpoint-dependent observation of an object and a mental representation of that object created as a result of prior experiences. IR, on the other hand, is a measure of individual familiarity one might have with an object as a result of the object's personal, cultural, or historical significance to the observer. As such, IR increases with the number of exposures to the object and of activities related to it. For example, one's own previous place of employment or education may have particularly high Idiosyncratic Relevance, where it otherwise might have very little Perceptual or Cognitive Salience to anyone else.

Contextual Salience is a measure of the degree of attention that can be allocated to potential landmarks as a function of the type of task being carried out (Task-based Context, or TC), as well as of the mode of transportation being used and amount of resources to be allocated (Modality-based Context, or MC). During the processing of route instructions, for example, TC is defined in terms of binary relations between potential landmarks and the path selection prompted by each instruction. A saliency value is therefore assigned to each pairing of path and potential landmark within the field of view, with distance and orientation between landmark and path acting as key discriminating factors. In this model, a landmark located more proximally to a turn location will be more salient to a navigator standing within view of that decision point than a more distal landmark. Relatedly, the modality being used to navigate the environment will significantly influence the navigator's field of view and attentional allocation, so that active navigation (e.g. driving a car) will require more attentional resources than a more passive form of navigation (e.g. riding a bus). Similarly, the speed of motion (e.g. walking vs driving a motor vehicle) will contribute to the determination of a navigator's field of view.

Additionally, Caduff and Timpf's (2008) theoretical framework models the online sequence of events involved in determining landmark salience during navigation. In a first stage, sensory stimuli are stored in a Sensory Memory. Here, those stimuli undergo parallel Pre-Attentive processing whereby low-level visual properties of the stimuli are identified, individual objects discriminated, and Perceptual Representations built in Working Memory. Such representations then undergo sequential processing, implementing the top-down Cognitive Salience and Contextual Salience components, which, in turn, modulate Perceptual Salience. The objects and their respective salience profiles are then encoded or updated in Long-term Memory.

Crucially, while in its formulation this model is primarily concerned with the visual modality during active navigation, it is flexible enough to also account for the allocation of attentional resources to landmark salience determination during the processing of spatial language, and will therefore be of relevance when interpreting the results of the experiments presented in this thesis. Furthermore, as evidenced by the study by Wegman and Janzen (2011) described in this section, eye movements could potentially be extremely valuable in studying the allocation of attention to landmarks or other navigationally salient features. In Experiment 1-3 I expanded this use of eye tracking to an analysis of attention allocation to landmark words in spatial texts and to landmark regions of map-like representations. This was done in order to study how the allocation of attentional resources (measured, for example, as changes in the number and duration of fixations) may be modulated by manipulations of the reference frames implicit in the route descriptions or of the imagined spatial perspective adopted by the reader. More generally, the goal of this research was to gain some understanding into the various forms of mental representations that might mediate between the extraction of navigational information from language and its use in the process of carrying out visuospatial tasks.

In order to provide a solid theoretical foundation for the research direction outlined here, in the next few sections of this chapter I will explore the literature on mental imagery and its connections to spatial cognition and navigation. Chapter 2 will then cover key research into the processing of spatial language and the imagery processes with which it interacts. In Section 2.6, I will then explore research on how eye movements can inform our understanding of attention allocation (and of related processes) as well as mental imagery processes during active navigation, language processing, and, more generally, during spatial cognitive tasks.

1.4. The Organisation of Spatial Knowledge

In Sections 1.2 and 1.3 I introduced two key concepts for our understanding of navigation and spatial knowledge. As we familiarise ourselves with an environment, we do so by encoding the identity of salient landmarks and associating that information with an understanding of their spatial locations. These locations can be specified with respect to our own body-centred frame of reference or with respect to each other (or, indeed, both). Additionally, this knowledge must be stored and maintained in enduring representations that allow us to directly navigate an environment by, for example, following a prominent beacon-like environmental feature, but that can also support more complex navigational behaviours (e.g. mentally planning a route through an environment in which we are not currently located, or constructing linguistic descriptions of it).

The nature, format, and content of these representations have been the subject of intense research since the mid-20th century. In studying the navigational behaviour of rats, Tolman (1948) challenged the idea that spatial learning was merely due to the learning of sequences of stimulus-response associations, with the strength of these associations varying as a function of incoming sensory input. Instead, he found that the rats were able to not only learn the configuration of a maze in order to reach a reward (i.e. food or water), but that this learning also took place during non-rewarded trials. Additionally, he observed that the animals were able to plot an alternative route to a goal location (or to nearby locations) when the configuration of the maze was changed compared to their learning phase (e.g. by rotating the starting point of the maze by 180° relative to the room). He concluded that the rats could not have been relying on purely body-centred stimulus-response associations, but rather had developed a more comprehensive understanding of the spatial environment. On this basis, he hypothesised that the acquisition of spatial information is accompanied by its progressive organisation "*into a tentative cognitive-like map of the environment indicating routes and paths and environmental relationships*" (p. 192). Kuipers (1978) stated "*the cognitive map is like a map in the head. More accurately, it is like many maps in the head, loosely related, for the cognitive map certainly lacks the global consistency of a single printed map*" (p. 132). He termed a collection of loosely connected cognitive maps of varying levels of detail and at different scales a *cognitive atlas* (Kuipers, 1982), and acknowledged the phenomenological experience of cognitive maps, observing that *"some people claim to 'see' a map when they answer spatial questions"* (Kuipers, 1978, p. 132). A cognitive map was also seen as a network of streets and intersections, and a catalogue of routes, each route being *"a procedure for getting from one place to another* […]*"* (p. 132). Denis and Zimmer (1992) described cognitive maps as *"*[…] *internal representations of spatial environments, their metric properties, and the topological relationships linking their landmarks*" (p. 286).

Since then, however, the map-like nature of cognitive maps has been challenged. Tversky (1981; 1992) has presented evidence of systematic distortions and heuristics in subjects' spatial memories for locations and orientations. For example, figures within an array tend to be remembered as more closely grouped and aligned to the canonical reference axes (vertical and horizontal, or north-south-east-west) than they were in the original percept (Tversky, 1981; 1992). Additionally, curved paths are remembered as straighter than they are (Chase, 1983; Milgram & Jodelet, 1976), and landmark salience can generate asymmetries in distance judgements between salient landmarks and non-landmarks, depending on which is used as referent (McNamara & Diwadkar, 1997; Sadalla, Burroughs & Staplin, 1980). Furthermore, Holyoak and Mah (1982) observed that when participants were asked to assume a particular perspective or geographical viewpoint, they judged the distances between pairs of nearby cities (relative to the imagined viewpoint, termed *cognitive perspective*) to be larger than the distances between pairs of more distant cities. On the basis of these and more findings (for a more detailed review, see Tversky, 2000), Tversky (1993) introduced the notion of *cognitive collage* to define these error-prone representations of novel spaces resulting from the integration of multimodal information and knowledge, both spatial and non-spatial. This has more generally led to the idea that spatial cognition may rely on a multitude of different knowledge structures – ranging from more percept-like, metric and detailed (e.g.

mental images), to more abstract and topological (e.g. mental models) – computed ad-hoc from a number of different sources of information and to achieve specific goals (Mark, Freksa, Hirtle, Lloyd & Tversky, 1999). These structures will be discussed in turn in the following sections, creating a thematic bridge between spatial cognition and the broader domain of mental imagery research. This chapter will also introduce the idea of perceptual simulation as an additional form of mental imagery, potentially filling the gaps between what Tversky (2000) referred to as the Overview and View levels (corresponding to survey, or allocentric, and egocentric representations), and the Action level. However, I will begin by introducing the notion of mental imagery and offering a brief historical overview of the development of imagery as an area of research.

1.5. Mental Representations and Imagery – A Brief History

Although the scientific study of mental representations in its current incarnation was developed after the cognitive revolution of the 1950s, the Greek philosophers Aristotle and Plato were already aware of its relevance to understanding the human mind and cognition. Referring to mental images as *phantasmata*, Aristotle described them as "*a residue of the actual [sense] impression*" and considered them to be central to his theory of memory, going as far as to claim that "*It is impossible to think without an image [phantasma]*" (De Memoria 450a 1, as quoted by Thomas, 2016). Although imagery continued to play a role in the work of several philosophers, such as Descartes, Hobbes, and Locke, it wasn't until the late 19th and early 20th centuries that mental imagery began to be studied in the emerging discipline of psychology. Widely regarded as one of the founders of experimental psychology, Wilhelm Wundt championed a view of mental images that emphasised their percept-like nature, and described them as "*[…] ideas [that] do not represent things of immediate perception; briefly expressed, they originate in feeling, in emotional processes which are projected outward into the environment. This is an important and particularly characteristic group of primitive ideas. Included within it are all references to that which is not directly amenable to perception but, transcending this, is really supersensuous, even though appearing in the form of sensible ideas*" (Wundt, 1916/2013, p. 75).

The view of mental imagery as an important psychological phenomenon in early experimental psychology was short-lived. In Würzburg, Germany, Oswald Külpe, a former student of Wundt's, and his students began employing introspection and word association methods to study mental representations. Over the course of these experiments, participants frequently reported experiencing "*events of consciousness which they could quite clearly designate neither as definite images nor yet as volitions*" (Mayer & Orth, 1901, as quoted by Monson, 1993, p. 16). However, a tragic flaw of these studies (one shared by many studies at the time) was their inherent sampling bias. It was common practice at the time for experimental participants to be recruited in significant numbers from members of the experimenter's own laboratory, including the investigators themselves (Thomas, 2016). Though methodologically flawed, the results of these experiments generated what became known as the "*imageless thought*" controversy. This, coupled with the simultaneous emergence of the behaviourist approach, extinguished for decades all academic interest in mental imagery, now perceived as being of dubious cognitive importance and far too difficult to exert experimental control over.

In the 1950s, the cognitive revolution caused a paradigm shift away from behaviourism. At this time, a flurry of research focused on the hallucinogenic effect of drugs. The discovery of REM sleep (Holt, 1964) and the development of electrophysiological techniques fuelled related work on the imaginal aspects of dreaming (Dement & Kleitman, 1957). Together with studies of the vivid experiences triggered by prolonged sensory deprivation (Bexton, Heron, & Scott, 1954) and by direct temporal lobe stimulation (Penfield, 1958), these distinct lines of research raised imagery once more to the status of respectable research topic and potential tool to "*look inside the famous black box*" of cognition (Holt, 1964, p. 260). This particular conjuncture of events brought about a revival of interest both in mental imagery and, partly, in introspection.

1.6. The Role of Mental Imagery: Dual Coding versus Common Coding

Canadian psychologist Allan Paivio is perhaps more than most responsible for the significant spur during the early years of the imagery revival. Inspired by the long tradition and effectiveness of imagery-based mnemonics methods such as the method of loci, Paivio set out to explore the connection between mental imagery and memory. Lambert and Paivio (1956) observed that recall of adjective-noun pairs was facilitated when nouns preceded the adjectives, consistent with the idea that nouns, being more concrete and thus higher in imaginability, might function as conceptual supports for the encoding and recollection of the whole word pair. Paivio (1965) further explored the effect of abstractness and imagery on the paired-associate learning and recall of noun-noun pairs combining concrete (e.g. house) and abstract (e.g. freedom) nouns independently rated on the ease with which they could evoke sensory images. It was observed that recall of associates, whether concrete or abstract, was significantly higher upon presentation of a concrete stimulus. Bower (1970) expanded on these results by comparing paired-associate recall performance of concrete noun-noun pairs between three groups of participants instructed to use three distinct learning methods: rote learning, separation imagery, and interactive imagery. Separation imagery participants were instructed to visualise the two objects in each pair on opposing sides of the imagined visual field, whereas interactive imagery participants were given instructions to imagine them "*interacting in some vivid way in an integrative scene*" (p. 531). Such interactive scenes were largely described by participants as taking the form of actor-action-object relations. Results showed that interactive imagery yielded significantly higher associative recall than the other two learning methods, but no differences in stimulus recognition were observed. This was taken as indication that imagery facilitates associative learning not by making stimuli distinctive or by improving encoding, but by organising mental representations of distinct entities in coherent relational structures.

From these (and many more) findings, Paivio (1971) developed the Dual Coding Theory (DCT) of cognition. Its underlying and guiding principle is the observation that "*human cognition is unique in that it has become specialized for dealing simultaneously with language and with nonverbal objects and events. Moreover, the language system is peculiar in that it deals directly with linguistic input and output (in the form of speech or writing) while at the same time serving a symbolic function with respect to nonverbal objects, events, and behaviors. Any representational theory must accommodate this dual functionality*" (Paivio, 1986, p. 53). Accordingly, DCT assumes the existence of two main components (or codes) in human cognition, a verbal system and a non-verbal (imagery) system. These systems contain basic, modality-specific representational units called, respectively, *logogens* and *imagens*. The former are organised in associative and hierarchical networks of verbal entities, whereas the latter on the basis of part-whole relationships and similarity. Sensory systems form direct representational connections with the two codes to allow for the activation of verbal and non-verbal representations on the basis of relevant sensory input. Individual representations also form associative connections with related representations within the same code and referential connections with corresponding representations in the opposite code. As such, this model accounts for the processing advantage of concrete words by positing a higher number of referential connections between the verbal representation of a word and relevant imagery representations in the non-verbal store compared to abstract words that primarily activate verbal representations with fewer connection to representations in the non-verbal code (Holcomb, Kounios, Anderson & West, 1999).

Over four decades, Paivio's Dual Coding Theory was revisited and expanded to account for the growing experimental evidence, while having to contend with the emergence of alternative, common coding theories (CCT) of cognition. Unlike DCT, these posit the existence of a single representational format underlying all types of representations. Initially, this common code was hypothesised to be verbal in nature and to take the form of inner speech in an individual's specific native language. By the mid-1970s, however, advances in psycholinguistics and artificial intelligence had begun to spur the production of computational models in psychology. Accordingly, the common code of CCTs started to be conceptualised as akin to the abstract data structures of programming languages such as LISP (Anderson & Bower, 1973; Collins & Quillian, 1969), which came to be referred to as propositional representations, expressed not in any natural language but in a hypothetical language of thought termed *mentalese* (Thomas, 2016). Other researchers attempted to bridge the gap between DCT and CCT by formulating hybrid theories that replaced the inner-speech, natural-language verbal code of DCT with a mentalese code (Baylor, 1973; Kieras, 1978; Kosslyn, Holyoak, & Huffman, 1976), or tri-code theories that maintain a verbal, a mentalese, and an imagery code (Anderson, 1983). However, the lack of a satisfactory candidate (or neural implementation) for a hypothetical language of thought, left DCT as the model better able to account for the available evidence. This is, however, not the only point of contention surrounding imagery.

1.7. The Format of Mental Imagery: Mental Images as Analogue Representations

Although by the early 1970s Paivio's extensive work had provided convincing evidence for the relevance of mental imagery to our understanding of cognition, allowing it to once more rise to prominence in psychological research, the nature, implementation, and cognitive mechanisms underlying imagery were still unknown. As explanations started to be presented, the stage was set for the beginning of a debate the remains unresolved even today.

This early research made fruitful use of mental chronometry approaches in order to study the cognitive processes involved in visuospatial imagery. Shepard and Metzler (1971) developed the now famous mental rotation task. In this extremely elegant paradigm, participants are presented with pairs of 3D configurations of cubes. In each pair, the two figures either reflect the same configuration rotated by a certain angle, or one represents a rotated, but mirrored, image of the other that cannot be rotated to match the first configuration. When participants were tasked with judging whether the two figures represented the same, but rotated, configuration as opposed to rotated, mirror images of each other, response times were found to increase as a function of the degree of rotation. This supported the idea (confirmed via participants' anecdotal self-reports) that they were actively rotating 3D mental representations of the second figure at a steady rate to match the target one, and that larger rotations would therefore result in longer latencies. More conceptually, these results were taken as an indication that the mental representations underlying visuospatial tasks had spatial properties that made representational space isomorphic to external space, and mental objects to real ones.

Shortly after, Cooper and Shepard (1973) replicated Shepard and Metzler's original finding in a task that required participants to judge whether a letter of the alphabet, presented at various degrees of rotation on the frontal plane around the sagittal axis, was also presented in its canonical orientation or as a mirror image. Although this task did not require participants to compare two simultaneously visible figures, response times were once again found to increase with increasing angles of rotation. In following studies, Cooper (1975, 1976) once again replicated the finding of increased response latencies with increased angles of mental rotation, this time using complex irregular polygons. Experiment 2 in Cooper (1975) is particularly noteworthy, as it attempted to disentangle the time required to mentally rotate a polygon from the time required to respond to a second, test polygon. In this case, each experimental trial started with explicit instructions to mentally rotate a given polygon by a given number of degrees, clockwise or counter-clockwise, before the test polygon to be verified was presented. The participant indicated the completion of the preparation rotation via a button press, which yielded a preparation RT (RT_1) and prompted the presentation of the test polygon. Participants then had to determine whether the polygon was a normal or a mirror-image presentation of the polygon they had prepared for by executing a vocal response, and a second response time (RT_2) was recorded. RT_1 was compared to the test RT in the previous experiment, and found to follow the same linear increase as a function of angular deviation from a learnt form. RT_2, on the other hand, was not found to increase with increasing polygon rotations, indicating that, by the time participants reported having completed the preparation rotation, they had effectively rotated the polygon and were fully ready to compare their rotated mental representation to the test shape.

Around the same time, Kosslyn (1973) published one of the first studies to investigate the spatial and structural properties of visual imagery, and to explicitly test the hypothesis that the internal structure of a mental image might reflect the spatial structure of its real-world referent. In this study, participants were shown ten line drawings of common objects. They were instructed to either remember the name and appearance of each depicted object sufficiently well to construct accurate visual images of them, or to covertly generate verbal descriptions of the pictures and to be able to assign the correct description to each object name. Object encoding was followed by a response time task during which the name of one of the objects was aurally presented, followed by a second word describing a possible property of that object. Upon hearing each object's name, participants in the imagery condition ("Imagers") had to engender a mental image of the relevant object and either focus on it in its entirety or only on one prescribed end of it as if they were perceptually focusing their attention on it. Similarly, the propositional group ("Verbalisers") was instructed to begin covertly rehearsing either the whole description of the relevant object or only the part referring to a specific portion of the object. Upon hearing the property word, participants had to begin scanning their mental image or searching their verbal description for the probed property, and judge whether the object in question possessed it or not by depressing either of two buttons.

An analysis of participants' response times revealed that properties located at the point of focus for the two focus groups were verified the fastest, and that verification RTs increased as a function of distance of the properties from the focus point. This increase was more marked for the focus verbalisers, who were significantly slower than the focus imagers. Similarly, whole verbalisers were slower than whole imagers, but no significant effect of property location was observed for these groups. Globally, these results appeared to suggest an organisation of mental images as collections of perceptual (or quasi-perceptual) features in memory coherently organised in a network of spatial relations that matches the original percept. Additionally, although sections of these images may be retrieved from memory and scanned in a serial fashion, they still seemed to confer a marked advantage over the verbal representations, which did not preserve the spatial relations between components of the objects.

Follow-up experiments (Kosslyn, 1975) were carried out to further explore the nature of these mental images. By manipulating both their phenomenological size and complexity it was determined that larger (and therefore more detailed) mental images required a longer time to generate, but also that smaller mental images resulted in longer scanning time. That is, participants took longer to search and identify a probed property on a phenomenologically smaller mental image, presumably because said property was less readily “visible” on a smaller mental image. Additionally, response times increased as a function of mental image complexity (i.e. a mental image of an animal next to a wall with four digits painted on it or next to a complex, 16-cell matrix would be evaluated more slowly than a mental image of an animal accompanied by two digits or a simpler, 4-cell matrix).

The studies presented thus far have represented an intriguing way of applying mental chronometry to the study of mental imagery processes. However, providing participants with explicit imagery instructions could potentially expose a study to contamination by demand characteristics, an objection frequently raised by critics of Kosslyn’s quasi-pictorial theory of imagery (see Pylyshyn, 1973; 1981). This concern was addressed in a series of four experiments by Jolicoeur and Kosslyn (1985), providing evidence that mental image scanning time does increase with the distance scanned independently of experimenter effects or task demands.

Using the same procedure employed by Kosslyn, Ball and Reiser (1978), participants had to scan a mental representation of a map drawing whose landmark locations were learnt to criterion. Two experimenters were told to expect a U-shaped distribution of response times, a prediction justified on the basis that very close landmarks would clutter the representational space, leading to more difficult discrimination between them and therefore longer scanning times. Similarly, the largest distances between landmarks would also result in longer scanning times. The intermediate distances, on the other hand, would offer the best balance between landmark discrimination and short scanning distance, and would therefore result in the shortest scanning times. In spite of the experimenters’ expectations, however, the usual linear RT increase with increasing scanning distances was observed.

In a following experiment, using colour and black-and-white versions of the same map, four experimenters were told to expect either overall faster scanning of the colour maps compared to the black-and-white version or vice versa. Additionally, one experimenter in each group was told to expect slower scanning rates per unit of distance with the colour maps and the other experimenter in each group was instructed to expect the opposite trend. Results revealed that both mean RTs and the slopes of the scanning functions were unaffected by the expectations of the experimenters.

Experiment 3 tested for differences in experimenter effects between a perceptual scanning condition and an imagery condition. Four experimenters were provided with four different predictions concerning both overall mean scanning times and the relationship between scanning time and linear distances as a function of condition. While scanning in the imagery condition was found to take longer overall than perceptual scanning of a physical map, the same linear relationship between scanning distance and time was found across both conditions and all four experimenters.

A final experiment required participants to generate mental images of objects with a canonical orientation and to focus on either end of them. Participants were then presented with certain object properties and tasked with judging whether said properties were true of the entities currently being imagined. The object-property pairs had previously been rated as to the extent generating a mental image was necessary to determine their true/false status, and split into high-imagery and low-imagery item pairs. Although participants were never explicitly asked to "scan" their mental representations, the properties being probed could refer to features either found on the same end of the imagined object already being focused on or on the opposite end of the object. In the former case, no scanning would be necessary. In the latter case, a scanning would be necessary if the process inherently relies on imaginal processes, resulting in longer response times. Results revealed two main findings: RTs for high-imagery items were generally longer than RTs for low-imagery items; and an effect of property location was only observed for high-imagery items, with longer RTs during trials in which the property probed was located on the end of the object opposite to the one being focused on. This indicates that during those trials participants automatically scanned their representational space to move their attentional focus to the relevant end of the object in order to verify the property being probed. Taken together, the results of all four studies by Jolicoeur and Kosslyn speak to the relative imperviousness of mental imagery processes both to experimenter effects and task demands.

Globally, the results from image generation, image scanning, and property verification experiments appeared to strengthen the idea that mental images resemble their equivalent percepts, and provided important details that contributed to the development of the quasi-pictorial (or analogue) model of imagery.

1.8. The Analogue Model: Cognition and the Brain

The quasi-pictorial model was initially construed in terms of a computer graphics metaphor (Kosslyn, 1975), and fleshed out in more computational detail by Kosslyn and Shwartz (1977). This model posited the existence of two main components: a Visual Buffer, and Deep Representations. The former was conceptualised as a visual short-term memory structure ("Surface Display") onto which mental images ("Surface Images") could be generated, and that could model empirical data from image scanning studies by containing regions with different degrees of activation (i.e. a highly activated area, representing the region in focus of a mental image, and surrounding areas of decreasing resolution). The latter were construed as a long-term information storage more akin to an abstract data structure or propositional list, representing knowledge of "facts" pertaining to the objects whose visual appearance could be generated in the visual buffer. In this sense, constructing a mental image of an object was thought to involve accessing a list of properties of said object (e.g. property "has a rear tire" in the case of object "car"), in order to generate visual images of said properties that can then be compositionally assembled into a visual image of the complex object. The resulting image can then be processed by altering its apparent size (i.e. "zooming in" or "panning out"), scanning, or rotating it.

One of the main claims made by this early model (and one apparently supported by the early image scanning studies presented in Section 1.7) was that of a substantial functional overlap between cognitive and neural systems involved in perception and imagery. Evidence for a degree of structural and functional overlap between visual perception and imagery was provided by studies of electrophysiological and haemodynamic changes in visual cortical areas during imagery tasks, and by studying the parallel effects of brain damage on both perception and performance in imagery tasks (see Farah, 1988 for an early review). In a series of three positron emission tomography (PET) experiments, Kosslyn et al (1993) attempted to verify the extent to which visual mental imagery depends on the same neural substrate as visual perception. Subjects studied the appearance of upper-case letters drawn on 4x5 grids. Subsequently, in the imagery task only the grid was presented, with an X located somewhere on it, and a lower-case cue below it. Participants were instructed to mentally visualise the letter corresponding to the cue onto the grid and to determine whether it would cover the X. The same participants then took part in a perception task in which both the letter and the X were visible. In a following experiment, the perception task was modified by degrading the appearance of both the visible letter and the X in order to test the hypothesis that participants would use mental imagery to complete the noisy visual input. In a third experiment, the imagery task was modified, instructing participants to imagine either the smallest readable version or the largest, non-overflowing version of the cued letter in their mental visual field while keeping their eyes closed.

This was intended to better tease out the imagery contribution to brain activity apart from that of visual perception, but also to ascertain whether imagery activation would vary as a function of imagined visual angle subtended by the visualised object. Globally, the PET results provided evidence for a broad functional overlap between brain areas involved in visual perception and areas involved in visual imagery - especially Brodmann Areas (BA) 17 and 18, known to be topographically organised – even in the absence of any visual input.

These findings were replicated in a following PET study (Kosslyn, Thompson & Alpert, 1997) and were consistent with clinical findings of imagery impairment following occipital damage (Farah, Soso & Dasheiff, 1992). However, this cognitive model of imagery, holding the primary visual area (PVA) as a key component of the Visual Buffer shared by bottom-up perceptual processes and top-down memory and imaginal processes, did not account for all evidence. Although more recent fMRI research (Ganis, Thompson & Kosslyn, 2004) has identified occipital and temporal overlap between visual perception and imagery, more significant overlap was found in parietal and frontal areas. Additionally, a number of studies failed to report any PVA activation during imagery (see Mellet, Petit, Mazoyer, Denis & Tzourio, 1998 for a review), and evidence of a number of double dissociations between perception and imagery accumulated (see Bartolomeo, 2002 for an extensive review), as did evidence for dissociations between visual and spatial imagery.

Relatively early in the imagery debate, proponents of the analogue, depictive model were divided on whether images represented both visual and spatial information within the same cognitive, and possibly neural, structures. While Kosslyn initially advocated a strong overlap between mental imagery and the visual modality (e.g. Kosslyn, 1983), others held the position that mental images may represent spatial relations in the absence of visual information and visual phenomenology (e.g. Anderson, 1985). Over the years, neurophysiological and neuropsychological evidence was brought to bear on this question. Farah, Hammond, Levine and Calvanio (1988) described the case of a patient with bilateral temporo-occipital, right temporal, and right inferior frontal damage and displaying an uneven behavioural profile. More specifically, the participant was impaired in visual imagery tasks (e.g. colour or size comparisons) but achieved normal performance in spatial imagery tasks (e.g. 3D form rotation or mental scanning). This finding complemented previous results by Levine, Warach and Farah (1985), who studied the cases of two patients with bilateral posterior lesions, temporo-occipital for one patient and parieto-occipital for the other. The former was associated with a loss of object-colour imagery and corresponding prosopagnosia-achromatopsia in the perceptual domain, whereas the latter resulted in a loss of visuospatial imagery and visual disorientation.

This early (and later) accumulation of neurophysiological and clinical evidence meant that the Kosslynian cognitive model and its neural implementation had to be revisited in a number of ways. In subsequent formulations of the model

(Kosslyn, 1994), the topographical organisation of the Visual Buffer was reaffirmed, and the component more explicitly identified with a set of topographically organised regions of the visual cortex. These visual areas, in turn, establish both afferent and efferent connections along the ventral stream to and from a number of non-topographical areas, such as the middle and inferior temporal gyrus and the fusiform gyrus. These are implicated in visual memory and in the processing of landmark object identity (see Janzen & van Turennout, 2004 and Wegman & Janzen, 2011, discussed in Section 1.3). According to the model, these allow for both the bottom-up recognition of a stimulus and the top-down generation of images in the Visual Buffer on the basis of stored memories of object properties such as shape, colour, and texture.

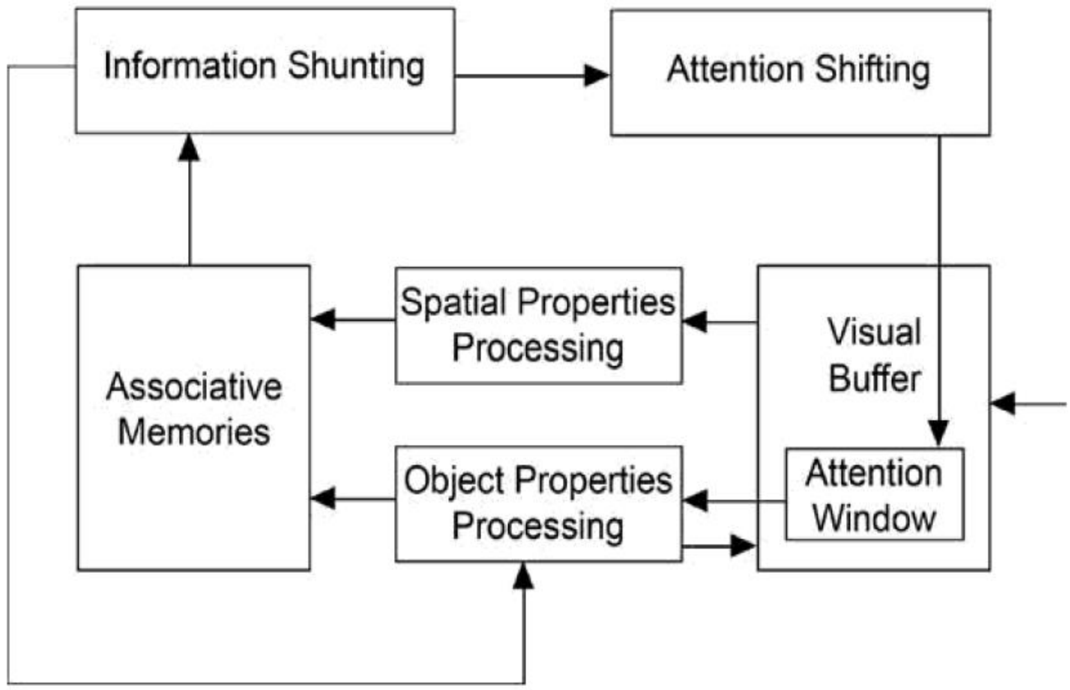


Figure 1.3 – Schematic representation of Kosslyn's model of mental imagery (Kosslyn, 2007).

Furthermore, a Spatial Properties Processing module was included in the revisited model (Kosslyn, 1994) (Figure 1.3). The Spatial images constructed by this module are postulated to be the result of activity in topographically organised parietal areas (Kosslyn, Thompson & Ganis, 2006) coding for the location, size, and orientation of the entities represented in imagery. In this sense, visual images constitute visualised portions of larger spatial mental images. The latter are construed as configurations of points in representational space, and are susceptible to the same manipulations (i.e. construction, scanning, rotation) as visual images, although these processes have not been directly explored as they apply to spatial mental images within the context of the model. Largely, research testing the various predictions of the analogue model has focused on mental representations with a strong visual component or following visual perception.

Occipito-parietal connections along the dorsal stream are involved in the processing of spatial properties of visual images, such as the relative positions of two or more objects or parts thereof. Activity associated with the processing of object and spatial properties from different modalities is then integrated in a multimodal Associative Memory component, allowing us to recall and visualise not only object identity but also its spatial context within an environment. The updated model (Figure 1.3) also construed image scanning as the result of an Attention Window component operating by selecting a region within the Visual Buffer (i.e. a certain pattern of activation within the visual areas involved) for further processing. Crucially, this selection does not solely occur in a top-down fashion within imagery itself, but also as a result of Attention Shifting within perception, as focusing on different parts or properties of observed objects brings the respective parts or properties into focus within the resulting visual image (Kosslyn, 2005).

Clearly, the analogue model and its underpinnings have grown in complexity. As a result of neurophysiological and neuropsychological evidences, imagery has come to be understood as the result of activity in a much broader network of brain regions, comprising occipito-temporal association areas, whose activity is heavily influenced by top-down parietal and frontal effects in a very task- and stimulus-dependent way (Mechelli, Price, Friston & Ishai, 2004). However, the scope of the model itself has also been significantly constrained. Although the notion of visual image was and still is central to Kosslyn's model of mental imagery, and has been at the core of the long-standing imagery debate between proponents of the analogue view and of the propositional view (Pylyshyn, 1973; 1981; 2003a,b), it became evident that it does not constitute the entire landscape of mental representational formats. Indeed, the realisation that information can be represented throughout cognition in a multitude of representational formats had been many years in the making. Imagery was no longer considered inherently visual, because visual content is not an inherently useful component when imagery is being used and manipulated during a task. The distinction between *visual* and *spatial* imagery became a central focus of research, as did the interface between mental imagery and relational reasoning in general (Knauff & Johnson-Laird, 2000). The latter has important implications for the interface between mental imagery, language, and navigation, as will be discussed in the next section.

1.9. Relational Reasoning and Mental Models

Most of the early research into mental imagery components such as mental rotation and image scanning included the presentation of perceptual input during encoding, just as most research into navigation has investigated the act of physically navigating an environment or observing a route. However, the research presented in this thesis focuses on the use of mental imagery in support of spatial cognitive processes operating in the absence of perceptual input. More specifically, it will explore the construction of spatial mental representations based on linguistic input.

In the early and mid-1970s, just as Kosslyn was beginning to explore the nature and role of visual mental images, a parallel line of research began to challenge the until then common place assumption that formal logic lay at the core of the human ability to reason and draw inferences. Prior to Johnson-Laird's (Johnson-Laird, 1975; Johnson-Laird, Legrenzi & Sonino-Legrenzi, 1972; Johnson-Laird & Wason, 1970; Wason & Johnson-Laird, 1972) seminal work, logicians and psychologists alike assumed that to draw inferences from linguistic statements, their logical form would be extracted from their syntactical form and subsequently subjected to some form of logical calculus dependent upon fundamental logical properties. For example, given the property of transitivity where R represents a given spatial relation:

For any *x*, *y*, and *z*, if *x*R*y* and *y*R*z*, then *x*R*z*.

and given the premises (Problem I):

1. **The pub is on the left of the bank.**
2. **The pet store is on the right of the bank.**

one can conclude that:

3. **The pub is on the left of the pet store.**

However, given a more complex set of premises (Problem II) such as:

1. **The pub is on the right of the bank.**
2. **The pet store is on the left of the bank.**
3. **The restaurant is opposite the pet store.**
4. **The shoe shop is opposite the pub.**

no premise explicitly states the relation between the pub and the pet store. The spatial relation between the shoe shop and the restaurant must therefore be inferred by making use of additional inferential rules and by going through additional derivations from the original premises. For example, by applying the following rule:

For any *x* and *y*, Left(x,y) ↔ Right(y,x)

to premise 2 in Problem II above, we can derive that a spatial relation between a landmark *x* and *y* and described by "on the left of," is equivalent to the spatial relation between the same landmarks in the opposite order and described by "on the right of." In other words, if "The pet store is on the left of the bank":

5. The bank is on the right of the pet store.

By conjunction of premise 1 and the new inferred relation 5, we can infer that:

6. The pub is on the right of the bank and the bank is on the right of the pet store.

Lastly, by application of the transitivity rule to step 6, we can infer that:

7. The pub is on the right of the pet store.

This finally allows inferring the spatial relation between the restaurant and the shoe shop:

8. The shoe shop is on the right of the restaurant.

The steps above are an example of the application of formal rules of logic to the kind of relational reasoning that underlies navigation in general, as well as the extraction of navigational information from linguistic propositions. As environments or routes grow in complexity, so does the amount of information that must be specified or inferred and the number of rules that must be applied in the process, in order to obtain a formal logical account of spatial descriptions. This type of logical calculus model of relational reasoning is severely limited in its ability to account for the ease with which we routinely draw logical inferences. That is because everyday relational reasoning does not (and cannot) rely merely on a syntactical analysis of the logical terms in a proposition, but on its overall meaning, on prior knowledge, the thinker's goals, and a host of contextual factors that drive the interpretation of language and sensory input. A more in-depth explanation of the theoretical and empirical reasons why inference-rule is at best a limited account of spatial reasoning lies beyond the scope of this review (but see Byrne & Johnson-Laird, 1989). Mental model theory (Johnson-Laird, 1983) provides an alternative by positing that reasoning involves the creation of "small-scale models" of reality (Craik, 1943) or "situation models" during language processing (van Dijk & Kintsch, 1983) in order to understand situations, reason on them, and anticipate possibilities. In this sense, mental models are mental simulations that are *iconic* – meaning that their structure is analogous to the structure of what is being represented. In the case of spatial descriptions, a mental model is a schematic representation of the spatial relations between the landmarks. In the case of Problem II above, the corresponding mental model might represent the environment as follows:

PET STORE ----- BANK ----- PUB

| |

RESTAURANT SHOE SHOP

Mental models can therefore make intuitively accessible spatial relations that were not explicit in the original descriptions (Byrne & Johnson-Laird, 1989; Johnson-Laird, 2010; Taylor & Tversky, 1992) in a way that is reminiscent of the visual images of Kosslyn's analogue model. However, the exact extent to which the two constructs are comparable is unclear. Johnson-Laird (1983) construed spatial models as representations of spatial relations in a symbolic 2D or 3D space, and visual images as representations of those aspects of a 3D model that can be visualised as they would appear from a specific point of view. Under Mental Model theory, Visual images are, in this sense, akin to 2½D sketches of 3D models as conceptualised in Marr's (1982) metatheory of vision, and they can support a variety of visuospatial tasks (such as conjuring a mental representation of the last known state of a room in our house to remember the location of a specific red book). However, visual images are not inherently required for the construction and use of mental models. The latter can be more abstract representations of relations whose processing does not necessitate the phenomenological experience of visual content, such as when processing the spatial relations "The mug is on the book" or "The hospital is to the west of the public library."

It is perhaps important to reiterate how the relationship between visual images and spatial images (produced by the Spatial Properties Processing component of the model; Figure 1.3), and the role of the latter in reasoning, have been left rather underspecified in Kosslyn's model. Much the same way, the role and structure of mental images in Mental Model Theory have been left relatively unexplored. These under-specifications, and the use of rather different paradigms (i.e. the study of eye movements in Mental Imagery and the use of syllogisms in Mental Model theory), make direct comparisons between the two models of imagery difficult. However, the term "image" is used vaguely enough by Kosslyn to potentially mean all forms of short-term visuospatial representations on a continuum that ranges from very sparse to highly detailed – perhaps constituting a continuum between spatial and visual. In this sense, commentators (e.g. Gottschling, 2006) have suggested that there may be situations in which Kosslyn's images effectively correspond to Johnson-Laird's visual images and spatial models, and to Marr's intermediate-level 2½D sketches.

The degree to which (and the instances in which) mental models and visual images underlie reasoning has also been the subject of considerable research in recent years. Knauff and Johnson-Laird (2000) presented participants with three- and four-term series problems, and tasked them with assessing the premises (presented one at the time) and evaluating whether or not the conclusions followed from them.

The problems entailed relations that were reported to be easy to mentally represent both visually and spatially (e.g. above-below), visually but not spatially (e.g. cleaner-dirtier), and neither visually nor spatially (e.g. better-worse). Participants' response times were recorded, and found to be slowest for the visual relations but fastest for the visuospatial relations. The authors concluded that the visual relations, which are hard to represent in a purely spatial array, require the generation of visual images, thus delaying responses until they can be constructed with sufficient detail and inspected. Visuospatial relations, on the other hand, can be constructed and evaluated without including phenomenologically visual content.

These results were subsequently replicated with the addition of relations that are easy to represent spatially but not visually (e.g. north-south) (Knauff & Johnson-Laird, 2002), and using fMRI (Knauff, Fangmeier, Ruff & Johnson-Laird, 2003). The latter study revealed that all types of problems (which were aurally presented) elicited activity in the left middle temporal gyrus, middle and inferior frontal gyri, right superior parietal cortex and bilateral precuneus, but only relations that were easy to represent visually but not spatially elicited activity in visual association cortex V2. Knauff and May (2006) further tested this visual-impedance effect by comparing the performance of sighted, sighted blindfolded, and congenitally blind participants (see also Cattaneo et al., 2008 for a review of the research in imagery and spatial processes in blindness). Unlike both sighted groups, the congenitally blind participants did not appear to be affected by the visual-impedance effect when reasoning on visual relations as opposed to visuospatial or control relations. However, their performance was consistently worse in absolute terms in all three conditions compared to the two sighted groups. This might suggest that the ability to form visual images is, in fact, important for certain types of reasoning, but that sighted individuals display a tendency for the generation of unnecessarily complex visual images when reasoning on premises that are highly imaginable. As such, the authors concluded that probing visual relations (e.g. "The dog is dirtier than the cat") triggers the spontaneous generation of visual gradations of "dirty" and impedes the generation of a mental model that can represent such gradations in a more abstract (and, crucially, more subconscious) form.

However, such an interpretation of the findings is not without problems. The idea that relational reasoning in general is impeded by visual representations rests on at least two fundamental assumptions: that transforming visual properties into non-visual (e.g. spatial) representations would necessarily be a more optimal strategy, and that solving three-term problems is an ecologically valid test of relational reasoning and navigational behaviour in the real world. Knauff and May's own findings that accuracy with visual relations was not higher in the congenitally blind group than in the sighted groups would seem to challenge the first assumption. Additionally, evidence from studies of spatial language processing would appear to challenge the second assumption. For example, Tom and Tversky (2012; see Section

2.3 for a more in-depth analysis of this study) showed that the vividness of spatial texts can facilitate the encoding and recall of plausible route descriptions.

Furthermore, Noordzij, Zuidhoek and Postma (2006) presented sighted, early- and late-blind participants with aurally presented route and survey descriptions, and compared their performance in a recognition/priming task, in a bird's-eye distance comparison task, and in a landmark naming task using a scale model of the environments described. Results revealed that, while blind participants were able to construct mental models on the basis of both types of descriptions, their performance was significantly worse following survey encoding, a pattern opposite to that observed in sighted participants. On the basis of these results, Noordzij and colleagues hypothesised that, while visual perception is not a requirement for the construction of spatial models, the ways in which we routinely interact with our surroundings (and the way we acquire and develop our navigational skills) will also influence the way we represent space on the basis of linguistic information. For example, blind participants, who, must necessarily employ egocentric navigational strategies during experiential spatial learning (i.e. learning the spatial location of objects and landmarks relative to a body-centred reference frame), might consequently be less able to process allocentric descriptions and build allocentric representations.

Similar results were found by Pasqualotto, Spiller, Jansari and Proulx (2013), who assessed the ability of congenitally blind, late-blind, and blindfolded sighted participants to perform judgements of relative direction (i.e. "Imagine you're at landmark x, facing landmark y, point to landmark z.") in a room-sized array of objects. This was explored both via egocentric locomotion (following a pre-determined route) and by studying an allocentric haptic representation (following the same sequential order, but subsequently allowing free exploration). Congenitally blind participants were found to be more accurate in judging relative directions when imagining orientations aligned with what experienced during egocentric locomotion. Sighted participants, on the other hand, were found to favour an allocentric representation of the array, thus performing more accurately when judging headings aligned with the intrinsic structure of the array. The authors concluded that visual experience facilitates the construction of allocentric representations, and that its absence during development might shape the navigational preferences of participants (Pasqualotto & Proulx, 2012).

The idea that the way we interact with our surroundings through our biological senses will shape cognition and, by extension, our ability to generate and use mental representations, had already been formulated by Johnson-Laird (1983), and is consistent with Kosslyn's suggestion that sensory and imaginal systems partly overlap, both structurally and functionally. Additionally, both mental images and mental models provide analogue representations of implicit information that can support inferences and the predictions of future possibilities based on past experiences. In recent years, this has led to the growing characterisation of mental imagery as an inherently multimodal and embodied phenomenon, allowing to bridge the gap between perception and cognition on the one hand, and action on the other. The view of imagery as an instance of embodied simulation based on perceptual symbols has important implications for visuospatial cognition, and will be discussed in the next section.

1.10. Enactive and Embodied Cognition: Mental Imagery as Perceptual Simulation

In the previous sections, I explored two of the main models of mental imagery: the quasi-pictorial, analogue model, and mental model theory. The development and testing of Kosslyn's analogue model of imagery was predominantly limited to the visual domain, but with the clear implication that different perceptual experiences will give rise to like-modality, analogue mental representations (Kosslyn, Thompson & Ganis, 2006). Similarly, although the issue of modality was not explicitly addressed by Johnson-Laird, mental models are conceived as either amodal or multi-modal (Sima, Schultheis & Barkowsky, 2013). Furthermore, both theories predict that the way we interact with our surroundings will constrain our mental representations, and that imagery (whether visual, spatial, kinaesthetic, or of other types) is a tool used to represent situations, extract implicit information from them, and predict possible outcomes. Considerable research has attempted to determine what form this interaction between perception, imagery, and action might take.

In the late 1970s and early 1980s, this new research path began to tackle the complex question of what gives rise to our conceptual representations. Much of its early impetus came from the growing field of cognitive linguistics, which saw language not as an independent construct, but as one embedded in and emergent from an individual's overall cognitive capacities and low-level sensorimotor experience of reality. In other words, the repertoire of concepts we can form and express is profoundly correlated with the conformation of our bodies and the types of interactions it allows with our environment. This was a core idea behind seminal work by Lakoff and Johnson (1980) investigating the embodied and experiential nature of metaphors (i.e. "The prices are rising" is the mapping of an experiential concept – that of objects moving in a vertical direction – to the concept of price), and built upon in a later publication (Johnson, 1987) detailing the representational medium that mediates, in this model, the relationship between sensorimotor experiences, concepts, and linguistic expression: *image schemata*.

An image schema is a schematic representation of a prototypical concept extracted from a recurrent sensorimotor experience that serves as part of the conceptual system's foundation onto which more complex concepts are built. They are unconscious and more schematic than mental images, but are also multimodal and analogue. Because of their fundamental nature, they have far-reaching consequences for models of language in general, but they are also of particular importance for theories of spatial language and cognition. Under this view, early sensorimotor experiences give origin to basic image schemata that allow us to conceptualise transferable notions of UP-DOWN, LEFT-RIGHT, FRONT-BACK, BEHIND, NEAR-FAR, CONTACT, CONTAINMENT, PATH (PATH TO and END PATH), and many other conceptual primitives that are progressively acquired as children accumulate experiences manipulating objects and locomoting, such as FORCE and TIME (Mandler & Pagán Cánovas, 2014).

Image schemata can therefore be thought of as multisensory containers for complex meanings, analogous to the notion of simulator in more recent literature on embodied cognition. One prominent model was formulated by Barsalou (1993; 1999; Goldstone & Barsalou, 1998), who proposed that cognition is a system based on perceptual symbols that are modal and analogical, rather than amodal and abstract (Figure 1.4). These cognitive units are re-enactments of original brain states in perceptual areas. They are processed unconsciously, in that they are not accompanied by phenomenological experiences, and schematic, in that they only contain a subset of the information contained in the percept that originated them. They are also multimodal, because they can arise in any sensory modality, including proprioception and introspection (Barsalou, 1999). Related perceptual symbols (e.g. a large number of multimodal experiences relating to the concept of "car") are then organised within conceptual categories called *simulators*, which allow for the generation of specific *simulations* of the concept.

Unlike perceptual symbols themselves, these simulations are conscious experiences with phenomenological properties, but they are re-activations of only part of all the multimodal perceptual information stored in memory. As such, they are relatively impoverished representations subject to distortions and simplifications, much like the distortions and heuristics found in cognitive maps by Tversky (1981; 1992). They can, however, be complex representations of whole physical scenes and situations, complete with volumetric primitives of shapes (cf. Landau & Jackendoff, 1993; Marr, 1982) and capable of assembling multiple perceptual symbols recursively and combinatorially (Figure 1.5). These processes make the perceptual symbol system productive, a crucial feature to generate imagery in the absence of perception and to communicate concepts via language (Barsalou, 1999).

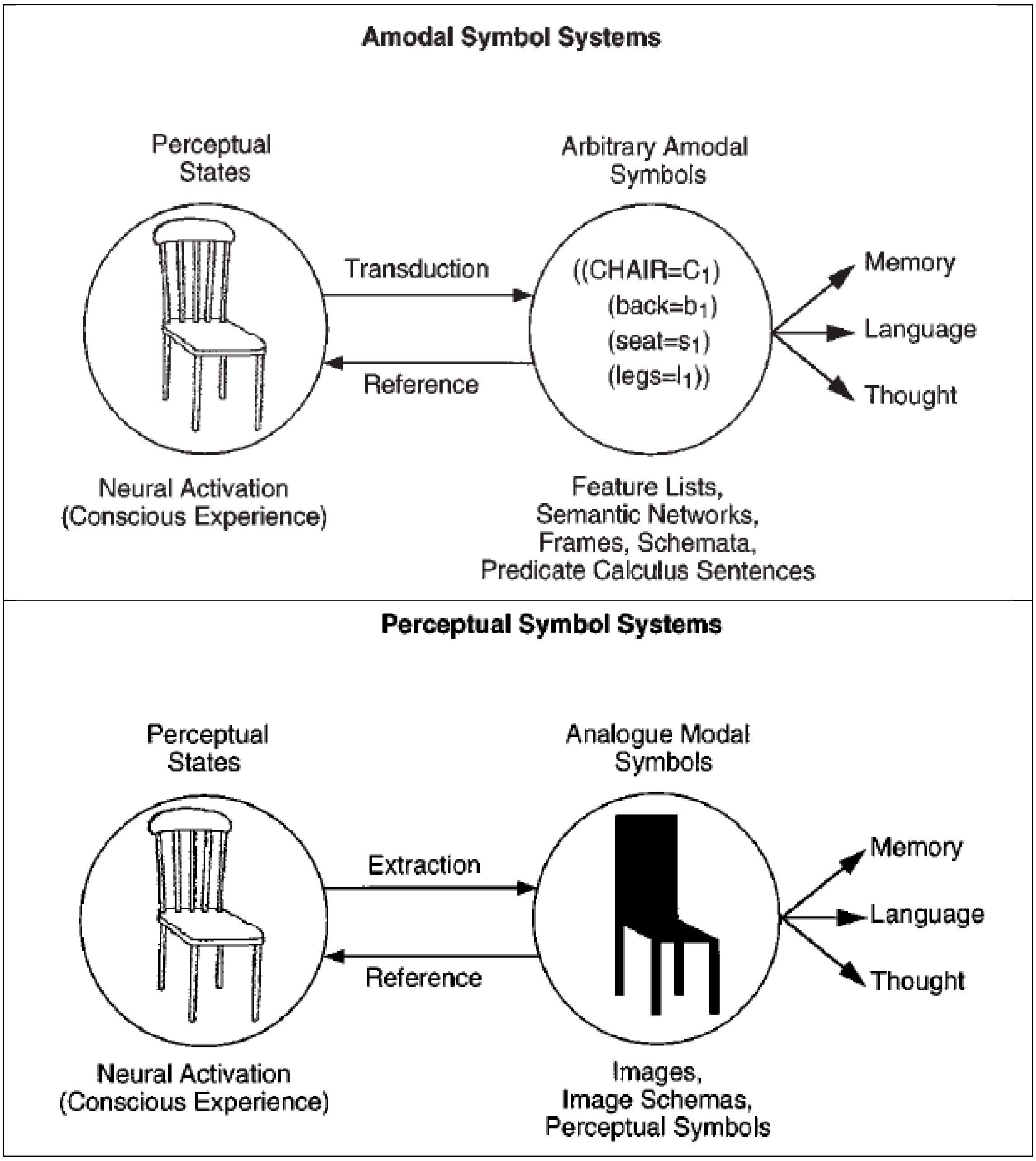


Figure 1.4 – Comparison between amodal and perceptual symbol systems (adapted from Barsalou, 1999). Perceptual experiences in amodal systems are transduced into symbolic information structures that bear no analogical resemblance to the original perceptual experiences that caused them. On the other hand, patterns of neural activation in perceptual symbol models carry salient subsets of the original information. These are then stored in long-term memory to function as symbols.

In this sense, the ability of perceptual simulations to represent entities, events, and environments beyond the current perceptual experience of the listener may be of particular relevance to the processing and production of spatial language. As we experience the world around us, linguistic symbols are acquired in parallel with the perceptual symbols for the entities to which they refer (or parts thereof). Once an association between linguistic symbols and perceptual symbols is established, future experiences with a certain linguistic input will activate the associated perceptual symbols and instantiate relevant simulations by parsing the surface syntax of the linguistic information. Given the modal, analogical, and conscious nature of perceptual simulations, they must necessarily make certain facts about their referent entities explicit. For example, if we construct perceptual simulations of entities and events taking place within a described environment, these must be grounded within a certain reference frame: we can imagine performing an action from an egocentric perspective, or we can imagine watching someone else perform the same action from an allocentric perspective. Enactive and embodied theories of mental imagery predict that the resulting mental representations will be phenomenologically and functionally different, and evidence for such a distinction has found empirical support from various lines of research.

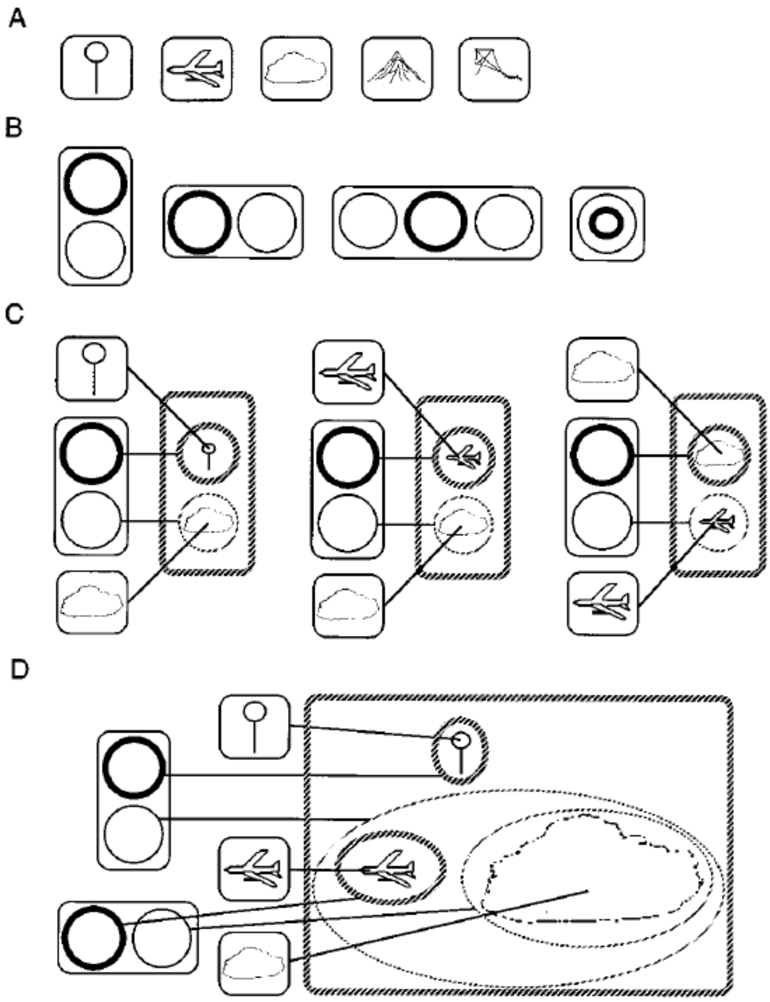


Figure 1.5 – A schematic example of how perceptual symbols for objects (A) and for spatial relations (B) – in turn extracted from repeated prior perceptual experiences with instances of “above,” “left of,” “between,” and “inside” – are used in combinatorial (C) and recursive (D) processes to generate new simulations. Boxes with thin solid lines represent simulators; boxes with dashed lines represent simulations (Barsalou, 1999).

For example, a study of spatial language processing by Brunyé, Mahoney and Taylor (2010) suggests that the simulation of perceptual and motoric events occurs during the encoding of spatial descriptions, that it can be modulated by different sounds, and that this modulation differs as a function of description reference frame.

The researchers presented participants with both route (egocentric) and survey (allocentric) written descriptions, and instructed them to read them in preparation for a memory test. Reading took place concurrently with footsteps or metronome sounds, in order to test their effects on encoding. While increasing the speed of the metronome sound increased reading speed during encoding of both description types, increasing the speed of the footsteps sound only increased the reading speed of route descriptions. This seemed to indicate that the sound of footsteps provided sensorimotor information that fostered the embodiment of spatial descriptions, but only when these expressed spatial relations within an egocentric reference frame. This increased egocentric embodiment also resulted in marginally less accurate spatial inferences, when switching to a survey representation was required to solve them. Additionally, listening to running versus walking sounds during reading resulted in increased estimations of environmental scale, indicating that the mental simulations generated during egocentric description encoding are embodied and susceptible to a variety of sensory modulations.

Further evidence in support of Barsalou's model of language comprehension as perceptual simulation was obtained via a strikingly simple behavioural paradigm: the sentence-picture verification task. Stanfield and Zwaan (2001) presented participants with sentences in which an object's orientation was tacitly implied in linguistic stimuli (e.g. "John put the pencil in the drawer" or "John put the pencil in the cup."). Barsalou's model predicts that each sentence in this pair will engender a different simulation owing to the physical properties of the object "pencil," (e.g. "thin and long"). Participants were then presented with drawings that depicted the target objects in an orientation that either matched or did not match the one implied by the sentence.

Crucially, orientation was not relevant to the task, as participants were only asked to determine, as fast as they could, whether the depicted object had been mentioned in the sentence. Filler sentences followed by sentence recall tasks were used to ensure careful reading and compliance with the task. The results showed that participants were significantly faster in identifying target objects when their depiction matched the orientation implied in the relative sentence. The findings were then extended to the visual shape of objects by Zwaan, Stanfield and Yaxley (2002), by using sentences that tacitly evoked specific shapes (e.g. "The eagle was in the sky" vs. "The eagle was in the nest.") and line drawings of the target entity's shape (e.g. an eagle with open wings vs. an eagle with wings drawn in). An even more robust effect than that observed for orientation was detected. Zwaan and Pecher (2012) later replicated these findings, further strengthening the perceptual simulation hypothesis.

Variations of the paradigm also found that the match advantage for both shape and orientation was also present when picture verification occurred with a 45-minute delay (Pecher, van Dantzig, Zwaan & Zeelenberg, 2009), and that when the pictures were presented first, increased fixations were observed during processing of mismatching sentences (Wassenburg & Zwaan, 2010). Furthermore, Coppens, Gootjes and Zwaan (2012) provided evidence of modulation of the N400 response (an electrophysiological measure of ease of semantic integration. Kutas & Hillyard, 1980; van Berkum, Brown & Hagoort, 1999) while participants read sentences mismatching the shape of objects previously presented as part of an ostensibly unrelated experiment. In yet another replication, Engelen, Bouwmeester, de Bruin and Zwaan (2011) showed the robustness of the effect of implied shape and orientation even in children between the ages of 7 and 13, and with the linguistic content provided aurally or in written form. Pelekanos and Moutoussis (2011) further showed how implied orientation in propositions appears to prime the perceptual system to detect the matching orientation at a very early stage in perception. In a similar vein, Vandeberg, Eerland and Zwaan (2012) showed that when participants observed a picture of an object or read a story about said objects, they tended to select the more transparent of two target pictures when the story they had read described the absence of the target object.

These findings seem to indicate that language processing involves (or can involve) the generation of simulations containing significant perceptual elements. As evidenced by studies adopting the picture-sentence verification task, these simulations make explicit the information that is implicit in linguistic expressions (e.g. the configuration of a bird's wings while processing sentences that describe the bird as either in flight or stationary). Additionally, the spatial perspective prompted by a spatial text or narrative numbers among the factors that can influence the embodiment of the resulting simulation. However, research into motor imagery and its links to bodily states seems to suggest that the effect of imagined perspective on embodiment is at play in imagery more generally, and not only during the construction of spatial representations. Decety, Jeannerod, Germain and Pastene (1991) compared the changes in physiological measures between both active and imagined locomotion relative to a rest baseline. Blindfolded participants were placed on a treadmill and asked to either physically locomote at 5, 8, and 12 km/h, or imagine locomoting while listening to audio recordings of the treadmill operating at the three speed levels. A control group was exposed to the audio recordings but without imagery instructions. The key finding was that both heart rate and pulmonary ventilation were found to increase during imagined locomotion as a function of imagined walking speed. Wang and Morgan (1992) used a similar paradigm to explicitly study psychophysiological changes in response to internal (i.e. imagining lifting dumbbells from a first-person perspective) or external (i.e. imagining watching oneself lift dumbbells from a third-person perspective) motor imagery. While both imagined perspective resulted in elevations of blood pressure, imagery from an egocentric perspective produced a greater increase in ventilation and greater perceived physical exertion than motor imagery from an external perspective. These findings agree with previous results by Hale (1982) and by Harris and Robinson (1986), who reported greater increases in electromyographic and oculomotor responses during egocentric simulation compared to external imagery of the same movements, and with studies of mental chronometry of imagined actions and of their neural correlates (see Decety, 1996, Grèzes & Decety, 2001, and Jeannerod, 1995 for reviews).

An additional area of research into the embodiment of visuospatial and motor representations has focused on the possible interaction of the vestibular system with the higher cognitive functions involved in mental imagery. Deutschländer et al (2009) observed increased BOLD responses in the multisensory vestibular cortex in totally blind individuals compared to sighted controls during locomotor imagery (i.e. standing, walking, and running) from an egocentric perspective. This would presumably indicate a greater reliance on vestibular input during locomotion in the absence of vision. Similarly, Péruch et al. (2011) have explored the role of the vestibular system in object-based mental transformations (i.e. rotations and translations) by testing the performance of patients with unilateral vestibular damage on three visuospatial imagery tasks, and comparing it to that of patients with bilateral damage and healthy controls. One involved the mental rotation of Shepard and Metzler's (1971) original stimuli. Participants were presented with pairs of objects and asked to determine whether they were the same (by pressing "Y") or not (by pressing "N"). During the second task, participants were presented with a map of a constructed environment containing a number of objects. They were allowed to study the map and learn the location of the objects. Subsequently they were presented with pairs of object and instructed to imagine scanning a mental image of the map following a straight line connecting the two objects. They then reported completion of the scanning by pressing the spacebar. The third task was the same as the second, but performed using a map of France (where the study was conducted) and pairs of French cities. Both clinical groups were found to be significantly impaired in all three imagery tasks, indicating a strong involvement of the vestibular system in mental rotation, in the simulation of motion, and in the estimation of metric distances. Unilateral vestibular loss was also implicated in impaired representation of external space and pointing direction in a study by Borel et al (2014). In it, vestibular patients were required to point to the spatial locations of targets that were briefly presented in peripersonal and extrapersonal space. Pointing took place in darkness, or with a visible, structured background as visual reference. Patients' representations of target configurations were found to be shifted towards the lesioned side, indicating the importance of vestibular input in the mental representation of external space.

However, the role of vestibular functioning during imagery has also been studied by using concurrent caloric vestibular stimulation (CVS), the irrigation of the external auditory canal with hot or cold air or water, normally used to study the vestibulo-ocular reflex. Mast, Merfeld and Kosslyn (2006) uncovered evidence that CVS impairs participants' performance in a high-resolution visual imagery task, consistent with findings that CVS suppresses activity in the early visual cortex (Bense, Stephan, Yousry, Brandt, & Dieterich, 2001; Deutschländer Bense, Stephan, Schwaiger, Brandt & Dieterich, 2002; Wenzel et al.,1996), and in a mental rotation task of letter stimuli. The latter has been found to rely heavily on parietal regions (Kosslyn, Digirolamo, Thompson, & Alpert, 1998; Zacks & Michelon, 2005), including areas involved in rotation perception in CVS studies (Lobel, Kleine, Bihan, Leroy-Willig, & Berthoz, 1998; Lobel et al., 1999). Interestingly, Falconer and Mast (2012) also found that CVS reduced response latencies in a body schema mental rotation task, whereas Grabherr, Cuffel, Guyot and Mast (2011) found both error rates and latencies in egocentric and object-based mental transformations increased in patients with vestibular damage compared to controls. These findings strengthen the idea of mental imagery as a quasi-perceptual experience, and make the case for more closely examining the role played by imagined spatial perspective on the degree of embodiment of, for example, route descriptions.

These data have also been complemented by brain imaging studies using fMRI and magnetoencephalography (MEG) in order to determine the extent of the neural substrate shared by real and imagined motion. Szameitat, Shen and Sterr (2007) found activation of primary sensorimotor cortices in both imagined and executed movements involving the whole body or confined to the upper extremities. Iseki, Hanakawa, Shinozaki, Nankaku and Fukuyama (2008) observed activation in the dorsal premotor cortex (PMd), supplementary motor area (SMA) and cingulate motor area (CMA) common to both the observation of gait movement from a third-person perspective and imagined locomotion from a first-person perspective. Similarly, Sharma and Baron (2013) observed activity common to both imagined and executed movements in a large network of regions involved in motor planning and control, including the contralateral primary motor cortex (PMC), cerebellum, PMd, SMA, and parietal areas. The shared activation of the contralateral primary and somatosensory cortices was confirmed by Kraeutner, Gionfriddo, Bardouille and Boe (2014) using MEG, and Taube et al (2015) found that motor imagery, and the combination of motor imagery and concomitant action observation elicited overlapping activity in the SMA, putamen, and cerebellum, areas involved in motor programming and control. Last but not least, Horner, Bisby, Zotow, Bush and Burgess (2016) recently observed grid cell involvement in the human entorhinal cortex during both navigation in a virtual environment and imagined navigation, in line with other evidence of medial temporal lobe involvement in navigational tasks in the absence of sensory cues (e.g. Marchette, Vass, Ryan & Epstein, 2014; Vass et al., 2016) and in spatial memory (e.g. Marozzi & Jeffery, 2012).

Globally, these results point strongly to the idea that generating mental imagery of actions (including those that may be involved in the mental experience of moving through an imagined spatial environment) partly relies on the same neural substrate involved in carrying out (and, to an extent, observing) those same actions. By extension, many of the areas normally involved in spatial computations in the real world (e.g. the hippocampal-entorhinal system; Buzsáki & Moser, 2013) also appear to be involved in spatial imagery and imagined locomotion. This is consistent with the idea put forward by Barsalou (1999) and with the notion of embodied and situated cognition more generally. Under this view, perception, action, and cognition are intimately related (e.g. Iachini, 2011). The environment is a source of crucial information to drive action, and the modalities through which this information is extracted will also influence the way it is processed (e.g. Wilson & Golonka, 2013). Crucially, mental imagery is a useful epistemic device to extract knowledge from this input and generate predictions that can prepare for and guide action (e.g. Moulton & Kosslyn, 2009). This process may also rely on the re-activation of patterns of brain activity and of bodily states that would be or were involved during the actual perceptual experience (e.g. Kent & Lamberts, 2008). These concepts will be explored in more detail in Chapter 6. However, this also complicates the taxonomy of forms that mental imagery can take, requiring a framework to organise the various types of mental representations and the factors that might drive the adoption of one above others.

1.11. A Theoretical Framework for Format and Reference Frame Selection

Navigating an environment and interacting with our surroundings are extremely complex tasks. This complexity is partly reflected in the variety of mental representations that may underlie visuospatial processes. In the previous sections of this chapter, I explored the literature on three different conceptualisations of mental representations. These offer different ways of representing spatial knowledge and, although they have been studied within largely distinct domains, are likely not mutually exclusive. For example, trying to build a cognitive map of the neighbourhood in which we live can yield a variety of results. We might build a schematic, allocentric representation of the relative positions of the most salient landmarks in the area. These spatial relations may reflect a broad topological organisation of the environment (e.g. Landmark A is broadly located west of Landmark B), or represent its metric properties more precisely. We may then represent more confined regions of the larger spatial model and include more perceptual aspects in our representations (e.g. a more visually accurate representation of a crossroad, with vivid representations of the surrounding shops), whether from an allocentric or an egocentric perspective. Additionally, we may simulate the experience of moving through parts of the environment, in order to plan a sequence of turns along a route in egocentric terms. Such simulations may vary in the amount of quasi-perceptual elements they contain and in their vividness.

Given the variety of forms that mental representations of space can assume and that I have described in the previous sections, I will now review some of the research into the factors that guide the selection of specific representational formats and reference frames. These were used to inform the experimental manipulations and measures used in this research.

In Section 1.2 I have already introduced the idea that, upon first interacting with an environment, allocentric representations may begin to form in parallel with egocentric representations. The findings supporting this conclusion (reviewed in the same section) run contrary to older models of spatial cognition (e.g. Siegel & White, 1975), which presupposed a hierarchical organisation of reference frames in which the construction of allocentric representations relies on the prior construction of egocentric representations (dependence assumption), to which they are functionally superior (superiority assumption). Adding to the empirical support for the parallel computation of distinct reference frames, Gramann, Müller, Eich and Schönebeck (2005) tested participants' ability to path integrate, adjusting a homing vector from the end position to the starting point of a journey presented egocentrically and in the form of simple changes in optic flow ("Tunnel task"). The key aspect of this task is that homing responses will be systematically different depending on whether an egocentric or an allocentric reference frame is adopted. The results revealed two distinct spatial strategies, whereby certain participants adapt their imagined heading in response to changes in optic flow that simulate a turn in the tunnel ("Turners") and perform the homing response from their updated egocentric orientation, whereas others registered the change in heading but responded maintaining an allocentric reference frame ("Non-Turners")(Figure 1.6). Crucially, both groups of participants received the same perceptual input and task instructions, indicating that the adoption of different strategies may have been simply due to individual preferences. Additionally, participants were able to learn to use their non-preferred strategy without significant loss of accuracy, further suggesting that both representations may be constructed in parallel but that only one may be used at any given time to complete a task (although participants may not be aware of this, as suggested by a lack of awareness during post-task interviews).

Subsequent studies (Gramann, Müller, Schönebeck & Debus, 2006; Gramann, Onton, Riccobon, Müller, Bardins & Makeig, 2010) adopted the same task to study the neural basis of egocentric and allocentric reference frame use during spatial navigation, and observed a divergence in the pattern of brain activation between Turners and Non-Turners during tunnel turns. The use of an egocentric strategy was associated with greater activity in posterior (parietal-premotor network) and frontal brain regions, whereas the computation of an allocentric reference frame was associated predominantly with activity within occipito-temporal regions, consistent with hippocampal and parahippocampal activation. Additionally, the transformation of egocentrically experienced visual flow into an allocentric representation of the route was associated with increased retrosplenial cortical activity, consistent with theories on the role of the retrosplenial cortex in translating visuospatial information between different reference frames (Ekstrom, Arnold & Iaria, 2014; Epstein, 2008; Sulpizio, Committeri, Lambrey, Berthoz & Galati, 2013; Vann, Aggleton & Maguire, 2009).

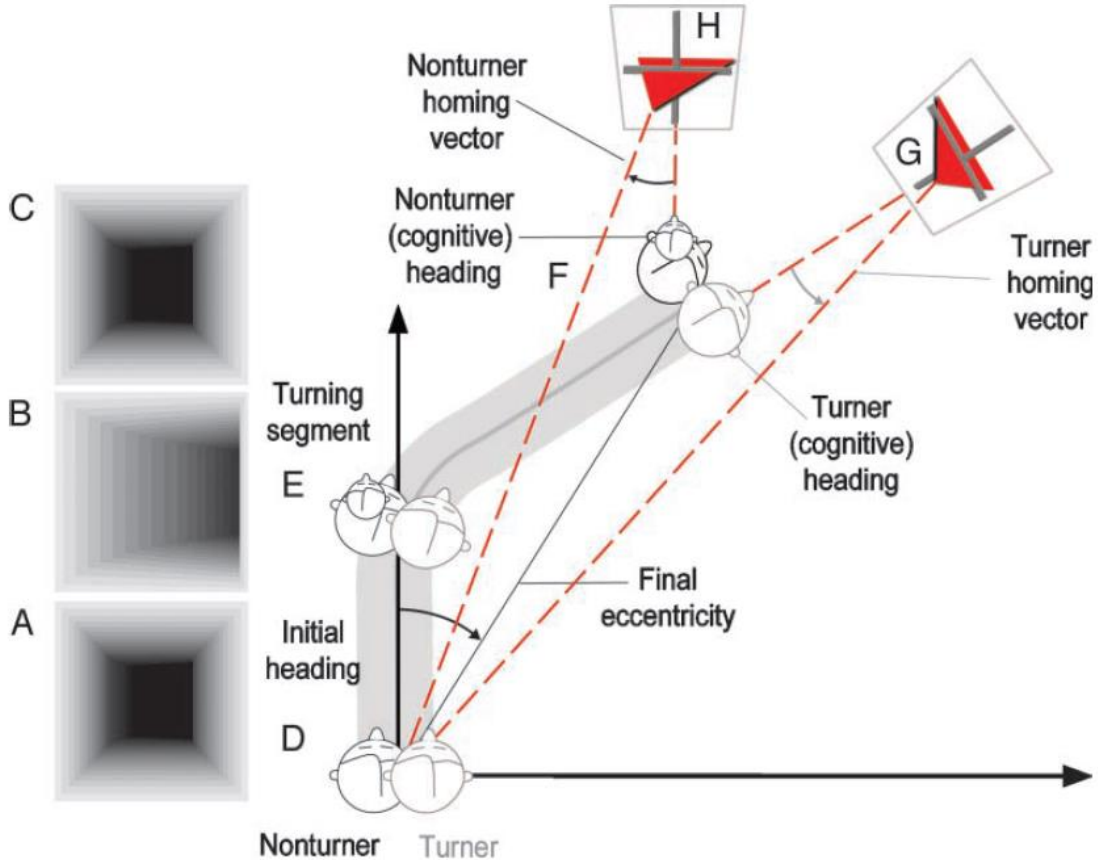


Figure 1.6 – Schematic representation of the two navigational strategies during a journey through a route comprising a starting segment (A), a turn (B), and an end segment (C). The dark grey heads show the difference between the perceptual heading information (larger head) and the cognitive heading (smaller head) of non-turners. The light grey heads represent the cognitive heading of turners, which is egocentrically updated to match the perceptual information during the turning segment (E). These different strategies result in different homing vectors (G and H) (Gramann et al., 2010).

These results provide support for the view that the neural basis of distinct spatial reference frames and representations is *soft-wired* (Gramann, 2013). This view is predicated on the idea that, while the mammalian brain is equipped with the neural substrate to compute, in a fast and efficient way, both egocentric and allocentric representations of an environment already from early exposures to it (as well as from memory), these neural structures are partly genetically determined but also plastic. Their maturation and use over the lifespan is, therefore, a complex interplay of individual biological constraints and environmental inputs – social, cultural, linguistic, and geographical. Navigators living in a certain context may be predominantly exposed to environments that foster different navigational strategies (e.g. relying more on distal landmarks as beacons in large open regions than in dense urban environments with many features extending vertically). As such, while navigators may construct both egocentric and allocentric mental representations of their surroundings, certain environmental features may preferentially lead them to select one over the other for the purpose of navigation. One crucial geographical element besides the conformation of the territory and the visibility of landmarks within it (and one that may be of relevance when exploring the construction of visuospatial mental representations on the basis of linguistic input) may be its size. In a way, visuospatial cognition can be said to involve different sets of tasks carried out at different spatial scales and depending on a variety of different factors. For example, locating items on the kitchen counter and manipulating them while making breakfast is likely to rely on different cognitive abilities, processes, and mental representations than locating and walking to our car in a small parking lot, or driving to a location on the other side of town. These three types of visuospatial tasks reflect a distinction made by Montello and colleagues (e.g. Montello, 1993) between *figural*, *vista*, and *environmental* space.

Figural space refers to personal and peri-personal space that can be apprehended from a single viewpoint, and includes both pictorial and volumetric representations of small, manipulable objects. At this scale, tasks may involve a number of processes and elements. For example, the encoding and processing of visual cues, self-to-object directions and distances; knowledge of object affordances (e.g. Klatzky, Pellegrino, McCloskey & Lederman, 1993); a sense of the physical properties that may govern interactions between objects (e.g. Battaglia, Hamrick & Tenenbaum, 2013; Hegarty, 1992; 2004; Hegarty, Kriz & Cate, 2003; Sims & Hegarty, 1997); and forms of motor imagery discussed in Section 1.10. Vista space represents the scale of environments that extend beyond the body, but can still be observed in their entirety from a single viewpoint without significant locomotion (e.g. a single room or a town square). Environmental space includes large-scale environments that require active locomotion, and the integration of sequences of viewpoints in order to build complete representations of them (e.g. an entire building, neighbourhood, or city).

Despite this theoretical distinction between different spatial scales, many of the tests used to measure navigational abilities effectively involve the encoding and manipulation of stimuli in figural space (e.g. tests of mental rotation). However, a study by Hegarty, Montello, Richardson, Ishikawa and Lovelace (2006) more closely explored the extent to which large-scale (environmental) spatial abilities can be predicted by measures of small-scale (figural) spatial abilities. During a spatial learning phase, participants actively moved through a real environment, learned the layout of a virtual environment by moving along a route through it, and passively watched a videotape of a route through a real environment. Following each learning phase, their knowledge was tested via distance and direction estimates between landmarks, and via a map-drawing task. Additionally, the following tests were administered to participants: the *Group Embedded Figures Test* (GEFT; Witkin, Oltman, Raskin & Karp, 1971), a measure of encoding and recognition of spatial patterns in embedded figures; the *Vandenberg Mental Rotations Test* (MRT; Vandenberg & Kuse, 1978), a measure of spatial visualisation and mental rotation abilities; the *Arrow Span Test* (AST; Shah & Miyake, 1996), a measure of visuospatial working memory; a test of perspective-taking ability based on materials by Huttenlocher and Presson (1973; 1979). Lastly, the *Santa Barbara Sense of Direction Scale* (SBSOD) was administered.

Exploratory factor analyses revealed that spatial learning from direct experience and learning from visual media loaded onto distinct factors, the latter including both passive learning from video and active exploration of virtual environments. Additionally, performance on small-scale spatial tests was more strongly correlated with spatial learning from media than with learning from direct experience, while the opposite pattern was observed for participants' scores on the SBSOD test (Figure 1.7). These results were taken as indication that, although measures of figural spatial abilities can be significant predictors of large-scale spatial performance, they leave a considerable amount of variance unexplained. This is consistent with a model in which the abilities involved in spatial learning at the figural and at the environmental scales are partially dissociated.

Further dissociations have also been found between performance on mental rotation tasks and on different versions of the Object Perspective Test (OPT; Kozhevnikov & Hegarty, 2001). The OPT is a pen-and-paper task in which participants are presented with a 2D, visual representation of an array of objects, and are asked to imagine standing at the location of one of the objects, facing a second, and to point to a third. In a study by Hegarty and Waller (2004), participants completed a number of tests used to assess mental rotation (including the MRT), and tests measuring perspective-taking ability (including the OPT). A model that assumed dissociable Perspective Taking and Mental Rotation abilities was found to fit the data significantly better than a model assuming a single Spatial Abilities factor. Additionally, OPT performance was found to be related to perspective-taking performance in imagined but familiar environments, i.e. the campus where the experiment was run (*Building Perspective Task*) and pairs of cities in the United States (*City Perspective Task*)(Hegarty et al., 2002). This indicates that assuming perspectives within spatial configurations is a general ability that applies to both viewed and imagined environments. Furthermore, perspective-taking performance was found to be correlated with the participants' SBSOD scores, a significant predictor of large-scale spatial cognition (Hegarty et al., 2002; 2006; Hegarty & Waller, 2004) and of survey spatial abilities (see Section 1.2).

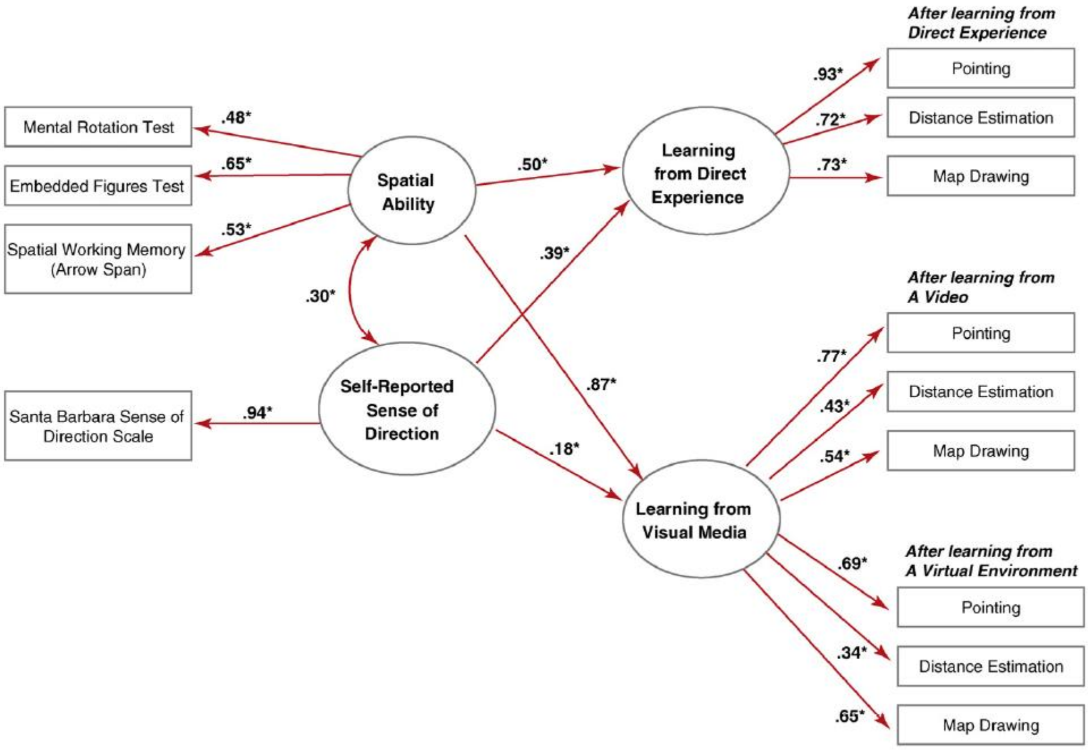


Figure 1.7 – Significant correlations between psychometric and self-report predictors, and behavioural outcomes separated by environmental learning factor (Direct Experience vs Visual Media). Wolbers & Hegarty (2010), adapted from Hegarty et al. (2006).

Similarly, Kozhevnikov, Motes, Rasch and Blajenkova (2006) used a computerised version of the OPT, providing either the canonical egocentric instructions ("Imagine you are at X. You are facing Y. Now point to Z."), or instructions that would prompt the use of an allocentric strategy involving object-based mental rotations of the array (Figure 1.8). Performance in the canonical perspective-taking task was found to be a reliable predictor of performance in navigational tasks requiring egocentric representations (i.e. finding a shortcut through a previously explored large-scale indoor environment, and pointing to non-visible targets within it), whereas performance in the array-rotation version of the OPT did not reliably predict egocentric, large-scale navigational abilities. However, both perspective-taking and mental rotation scores were significant predictors of participants' accuracy in drawing the route travelled on a floor plan of the environment, and in retracing the route after returning to the starting point. This could indicate that accurate route knowledge might rely on the generation of both egocentric and allocentric representations of an environment, consistent with the parallel computation of reference frames.

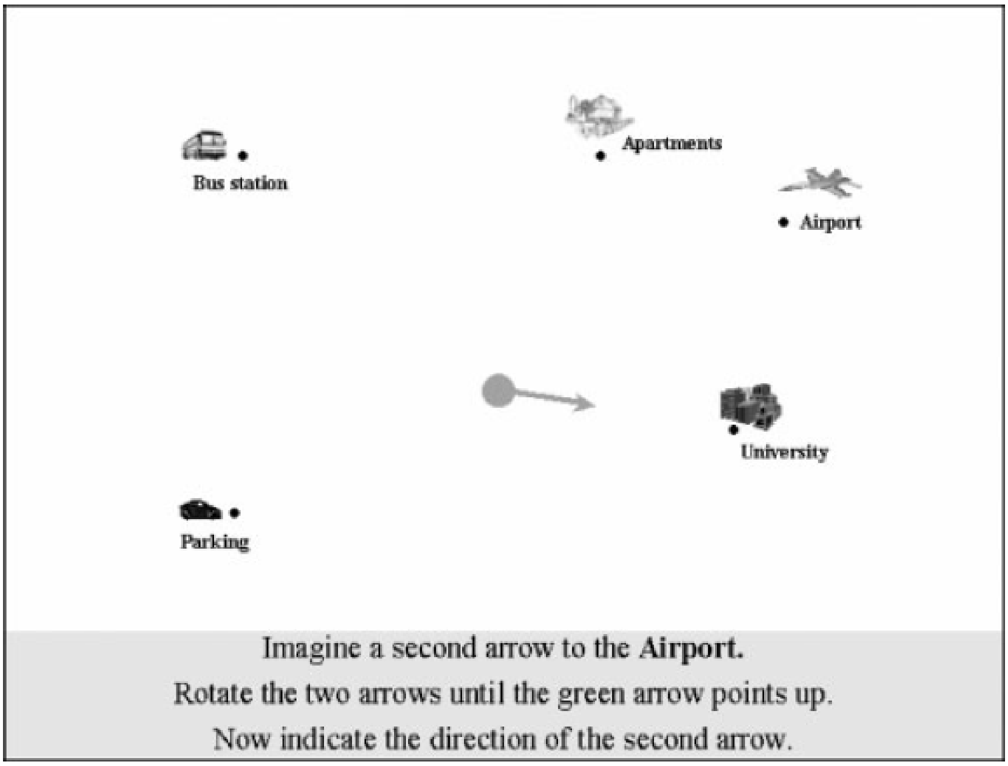


Figure 1.8 – An example of the instructions provided in conjunction with the array-rotation version of the OPT used by Kozhevnikov et al, (2006).

Ultimately, the choice of navigational and representational strategy appears to be a function of a number of factors including individual proclivities (in turn the result of innate and acquired characteristics), task demands, the type of spatial information available, and the way said information is presented (Wolbers & Hegarty, 2010). The reference frames, strategies, and underlying neural mechanisms adopted by navigators as a result of cultural influences, biological factors, and environmental features (Gramann, 2013) will be reinforced by the reference frame-specific relational terms used and, more generally, by the way spatial concepts are conveyed in the language spoken by the community (Burenhult & Levinson, 2008; Haun, Rapold, Janzen & Levinson, 2011). This makes understanding the processing of spatial language particularly important to understand the subsequent construction of mental representations and their navigational use. The studies presented in this section have largely contrasted active navigation of environments with spatial learning from visual media. However, in the experiments included in this thesis I have presented participants with linguistic descriptions of routes through plausible urban environments. One of the questions this research hoped to address was whether the construction of mental representations of plausible urban environments based on spatial linguistic information depends on figural- or environmental-scale abilities, or on both to varying degrees. For this purpose, in Experiments 1 and 2, a battery of tests (including MRT and SBSOD) was used to measure these different abilities, and to predict both performance in an allocentric representation task and eye tracking measures of reading during encoding. However, before introducing the experiments I conducted it is also important to review the literature on the individual differences and the linguistic factors that influence the processing and production of spatial language. This will be one of the goals of Chapter 2.

CHAPTER 2
External Representations and Encoding Processes

2.1. Overview

In Chapter 1 I introduced two of the basic elements that lie at the core of a functional navigational system: landmarks and reference frames. As we explore an environment, we detect the spatial location of salient landmarks, which we must then encode in terms of their position relative to us (egocentrically) or to each other (allocentrically). Evidence was presented for the parallel computation of both types of reference frames in various mental representational formats, ranging from spatial models, to visual and spatial images, to perceptual simulations. Additionally, the idea was introduced that various factors may drive the selection of specific representational formats and reference frames during visuospatial tasks were discussed, ranging from innate individual differences to environmental, socio-cultural, and linguistic influences.

Understanding these latter factors is paramount to understanding instances of navigation that are driven by the acquisition and communication of spatial information through language before directly interacting with an environment. Non-human animals are known to create neural representations of their surroundings by perceptually interacting with them (e.g. Rowland, Yanovich & Kentros, 2011; Yartsev & Ulanovsky, 2013), and all living organisms on the planet communicate in more or less complex ways, exchanging a wealth of information about themselves, each other, their immediate surroundings (e.g. Gagliano, Renton, Duvdevani, Timmins & Mancuso, 2012). Humans, however, possess the ability to exchange visuospatial information about environments beyond the immediate perceptual field (and to create internal mental representations) using complex symbolic external representations such as linguistic descriptions or sketch maps. In this sense, the present research is concerned with exploring the flow of spatial information between external and internal representation, and vice versa (Strasser, 2010). How is navigational information extracted from a certain type of external representation (e.g. a written route description)? How is it encoded within an internal mental representation (e.g. a mental model)? How is it subsequently used either to create a different type of external representation (e.g. a sketch map) or to produce a certain type of behavioural response (e.g. a judgement of relative direction)?

In order to approach these issues, this chapter will cover various lines of research on the processing and production of spatial language, with a particular focus on their relevance for theories of spatial cognition and on the individual differences that underlie them. I will also cover research into the use of sketch maps as navigational aids. Finally yet importantly, I will explore research into the role eye movements may play during the encoding of spatial information from external representations, and how they may be used as a window onto imaginal processes. This chapter will thus set the stage for Experiments 1-3, in which the encoding of written route descriptions was investigated using eye tracking.

2.2. Spatial Language and Spatial Cognition

The complexity and variety of human languages sets them firmly apart from all other forms of communication observed in the natural world. Using language we can, among other things, communicate complex philosophical ideas, describe rich visual scenes, share our favourite recipe, or provide directions to a destination. How we accomplish this has been the subject of considerable debate, one that is still not fully resolved. As the research presented in this thesis is primarily concerned with the way spatial language processing is influenced by both linguistic (e.g. the type of relational term used) and non-linguistic (e.g. imagined reference frame) factors, it is important to localise this research within the landscape of studies that have explored the interaction between language and spatial cognitive abilities.

One the one hand, certain scholars (e.g. Li & Gleitman, 2002) hold the view that the linguistic system is essentially just a collection of categories and formal structures that we use to give expression to our mental representations. These models generally involve a pre-linguistic period during which we acquire notions of objects, actions, as well as causal and spatial relations. The acquisition of this repertoire of concepts is influenced by strong biological constraints that shape our non-linguistic cognition, and is generally mediated by an abstract “language of thought” (Fodor, 1975). Subsequently, during language acquisition, children build a corresponding repertoire of linguistic forms to express the concepts they have already acquired non-linguistically. Many such models also tend to align conceptually with theories of linguistic nativism, postulating innate structural principles, such as the idea of Universal Grammar (Chomsky, 1965), that constrain language acquisition.

However, different lines of research have produced considerable evidence over the past decades to challenge the ideas of complete independence between linguistic and non-linguistic cognition, and of linguistic nativism. In the first case, evidence has emerged in support of a correspondence between linguistic and non-linguistic spatial representations. In Section 1.10 I briefly discussed the work carried out in cognitive linguistics by Lakoff and colleagues (Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 1980), and the impact it has had on the development of embodied theories of mental representations in psychology. However, that line of research was no less influential in linguistics itself, leading to the development of new theoretical approaches that radically departed from the previous, largely symbolic and computational tradition. These followed from Lakoff’s work, with the development of the cognitive grammar (Langacker, 1987; 1991) and cognitive semantics (e.g. Talmy, 2000) frameworks. The general tenets of these two cognate subfields can be summarised in a view of language in which linguistic abilities are an expression of general cognitive abilities (and of their limits) that shape the acquisition of language skills by the individual, and in which grammar and semantics are the expression of a culture and worldview shared by a community (who are likely to share many of the same sensorimotor experiences that lead to the formation of shared concepts). In this sense, the functional unit of language is seen as a form-meaning association of its semantic structure (represented as image schemata rather than propositionally. See Section 1.10.) with a lexicon of word forms and phonological labels, and grammar operates as a set of constraints on how these units can be combined to express complex meanings.

This idea that linguistic abilities rely on cognitive abilities in non-linguistic domains has considerable empirical support, as does the idea that different languages reflect different conceptualisations of the world by their respective speaking communities. In a series of experiments, Hayward and Tarr (1995) attempted to address the question of whether linguistic and non-linguistic representations share common foundational aspects in the way they encode spatial relations. They tasked participants with generating or rating the applicability of relational terms (e.g. above, below, left, right) as linguistic descriptors of the spatial relationship between object pairs in small arrays, and with recalling the position of one visually-presented object relative to another. While the use and applicability of vertical prepositions (i.e. above, below) appeared to be more graded (i.e. less categorical) than that of horizontal prepositions (i.e. left, right), results showed a primacy for locations falling directly on the extensions of a reference object's main axes (vertical and horizontal). These locations were found to be most representative of the respective linguistic labels, and the most easily remembered in non-linguistic memory tasks. This led the authors to conclude that a system of prototypical structures may underlie the ability to perceive spatial relations in vision and communicate them using language, and that this common system may code spatial relations both qualitatively and quantitatively.

The fundamental role of reference systems, as well as their parallel nature in both linguistic and non-linguistic systems, was further emphasised by studies that explored their breakdown in a variety of clinical populations and impairments. Among them, Williams syndrome (WS) has received considerable attention because of its peculiar cognitive profile and neurodevelopmental nature. Individuals with WS display profound visuospatial deficits, evident during tasks that require encoding, maintaining in working memory, and replicating a spatial configuration either by drawing it or by assembling blocks (e.g. Mervis, Robinson & Pani, 1999), but generalising also to navigational impairments (Broadbent, Farran & Tolmie, 2014; 2015). In contrast, however, they also exhibit prima facie intact linguistic abilities, with generally strong spontaneous vocabulary and linguistic fluency (Bellugi, Wang & Jernigan, 1994). This has led commentators to suggest the idea that linguistic and non-linguistic spatial cognition may be developmentally modular and independent (e.g. Bellugi, Bihrle, Neville, Doherty & Jernigan, 1992). However, more recent research has provided convincing evidence that specific deficits exist in the comprehension and use of spatial language in WS, and that these reflect non-linguistic spatial deficits.

In one such study, Landau and Zukowski (2003) tested the ability of children with WS to encode and subsequently describe visually presented motion events. These were short animations of objects (Figure objects) moving in a certain manner and along a certain path with respect to reference objects (Ground objects) (Figure 2.1). The WS group displayed preserved object naming abilities and understanding of the spatial roles of the objects, reflected in their correct syntactical encoding of subjects, objects, and prepositions. However, their specification of the paths taken by the objects and their selection of the relevant prepositions was significantly worse than that of typically developing children, with more incorrect, ambiguous or omitted terms. Furthermore, the WS group was more accurate in describing Bounded TO paths compared to Bounded FROM or VIA paths. The authors ventured that children with WS may struggle to encode and maintain accurate representations of the Ground objects when its location does not coincide with the Figure object's final location following motion (that is, in Bounded FROM and VIA paths), citing evidence for a preserved recency effect but absent primacy effect in WS (Vicari, Brizzolara, Carlesimo, Pezzini & Volterra, 1996).

Subsequently, Landau and Hoffman (2005) replicated Hayward and Tarr's (1995) design with a sample of this clinical population, in order to explore the integrity of the axial reference systems in children with WS. Those children showed a partial preservation of the typical axial system, with a pattern of results similar to that of healthy controls (e.g. a primacy for locations falling directly onto the vertical and horizontal axes) and important syndrome-specific errors. Directionality errors for the relational terms on both the vertical and the horizontal axes are normal in the acquisition of relational language. Terms referring to the vertical axis are normally acquired and mastered earlier, pointing to a particular relevance for the gravitational axis that may, speculatively, have interesting evolutionary implications. Horizontal axis directionality (left-right) errors, however, persist into adulthood in WS. This could be a sign of delayed or compromised development in aspects of spatial language in this population. More specifically, the WS group displayed a tendency to use more global terms, such as "far" or "near" to describe horizontal spatial relations. Additionally, accuracy decreased as a function of distance between reference and target objects in linguistic and non-linguistic tasks alike. This pattern of results indicates that the accuracy of the axial reference system may be subject to gradation, but also that even a noisy non-linguistic system of reference frames based on axes may be able to support the acquisition of a relational language system, albeit a less accurate one.

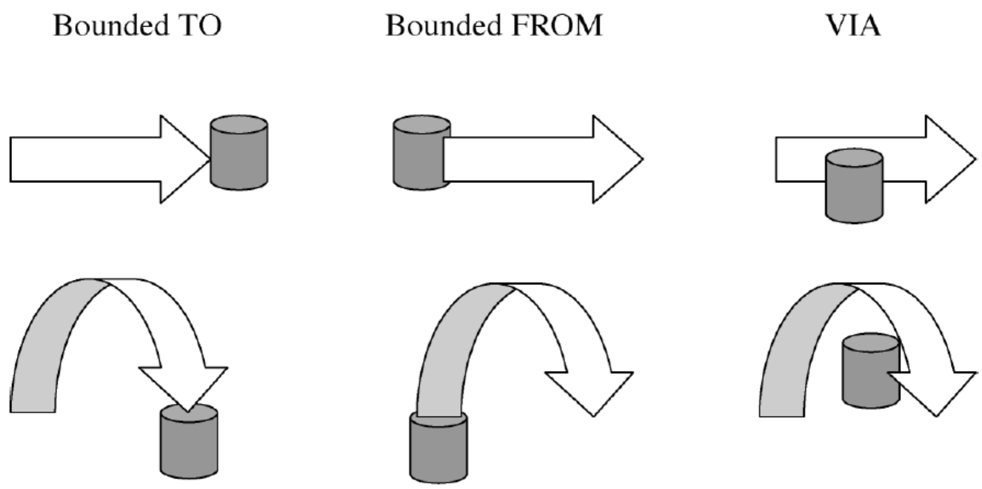


Figure 2.1 – The types of spatial relations and paths tested by Landau and Zukowski (2003) in their paradigm eliciting descriptions of motion events.

More evidence of the interaction between spatial language and broader spatial cognitive abilities in WS was produced by a number of scholars (Laing & Jarrold, 2007; Lukács, Pléh & Racsmány, 2007; Mervis & John, 2008; Phillips, Jarrold, Baddeley, Grant & Karmiloff-Smith, 2004). However, an equally important line of research has explored the interaction between linguistic and non-linguistic systems within spatial cognition not with respect to what corresponding deficits may arise in the two domains as a result of neurological or neurodevelopmental disorders, but from a cross-cultural and cross-linguistic perspective. Jackendoff (1983) and Lakoff (1987) are generally credited with reappraising Whorf's (1956) idea (or weaker versions thereof) that language can structure or influence non-linguistic cognition. This re-evaluation of the idea of linguistic relativism has engendered extensive research into the parallels between the two domains within and between groups of speakers of various languages (Lucy, 1992). The bulk of this research has focused on studying the non-linguistic use of spatial reference frames between speakers of languages that preferentially code spatial locations using different frames of reference. Employing tests such as the animals-in-a-row (Levinson & Schmitt, 1993) or the motion-maze (Pederson & Schmitt, 1993) tasks (Figure 2.2), researchers can exploit the specific logical and spatial properties that distinct reference frames display following egocentric rotation or translation. This allows to code participants' responses as unambiguously egocentric, intrinsic, or allocentric, and has produced strong correlations between linguistic and non-linguistic reference frame use (e.g. Levinson, Kita, Haun & Rasch, 2002). Namely, participants' preference for and ability in adopting specific reference frames during spatial tasks reflect the preferential use of those reference frames in their native language. Speakers of allocentric languages (e.g. Guugu Yimithirr in Australia or Tzeltal in Mexico) will preferentially encode spatial arrays in allocentric coordinates (e.g. using cardinal directions) and replicate them accordingly following rotations and movements (Majid, Bowerman, Kita, Haun

& Levinson, 2004). Such Whorfian effects have been also observed from a developmental perspective, by studying cross-linguistic variation in the acquisition of early semantic categories and spatial concepts, both in production and comprehension (Bowerman & Choi, 2001; Choi & Bowerman, 1991; Choi, McDonough, Bowerman & Mandler, 1999).

These different lines of research have shown quite convincingly that linguistic and non-linguistic domains can interact in a multitude of ways to shape our cognitive functioning. Ultimately, the exact mechanisms behind these interactions must be explained in order to successfully address the issues of how we can talk about what we see, direct our attentions to what is described, mentally visualise the content of a linguistic description, or produce linguistic content on the basis of mental representations. Although no such theory exists, a number of attested cognitive mechanisms may be involved in these processes in various capacities (Majid et al., 2004), and may be particularly relevant with respect to explaining the acquisition of spatial abilities in both domains. In so far as linguistic input is able to direct attention to specific elements of one's own surroundings, "simple" attentional processes may contribute to driving a tuning process whereby speakers of a certain language may become more or less sensitive to certain spatial categorisations depending on whether they acquire the means to express them linguistically.

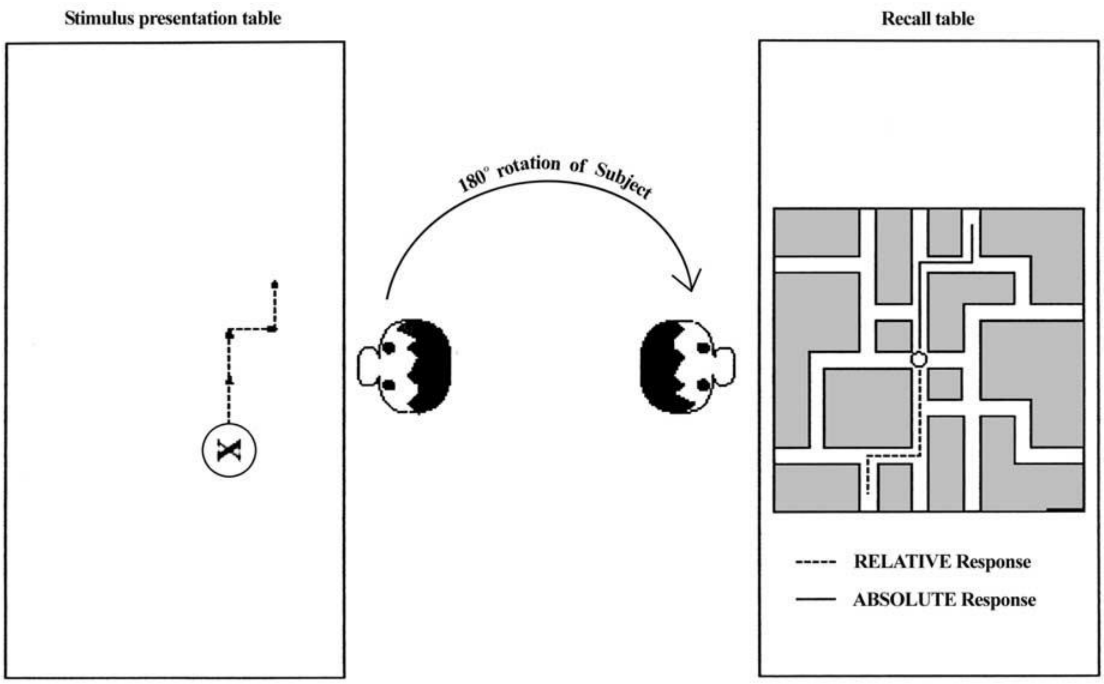


Figure 2.2 – The Motion-Maze task (Pederson & Schmitt, 1993) used to study the convergence between linguistic reference frame preference and visuospatial reference frame preference in spatial coding tasks (e.g. Levinson et al., 2002).

For example, McDonough, Choi and Mandler (2003) presented preverbal infants and adults from English- and Korean-speaking families with short clips of dynamic spatial relations using a wide range of objects. The relations probed were support (described linguistically by the English preposition "on") and containment (described linguistically by the English preposition “in”). The Korean language, however, uses the predicate “kkita” to denote “tight fit” spatial relations irrespective of containment or support, consisting therefore of subsets of both English IN and ON spatial relations (Figure 2.3), and “nohta” to denote loose-fitting support. Using a preferential-looking paradigm, participants were first familiarised with one spatial relation (tight-fitting containment or loose-fitting support) by presenting them with videos representing enactments of that particular relation simultaneously on two screens. Following the familiarisation phase, the test trials consisted of the presentation of the already familiar relation on one screen and of a novel one on the other (loose-fitting containment). Preferential looking behaviour was analysed to determine whether infants and adults would differ in their propensity to look at the novel stimulus more than the familiar one. Looking behaviour showed that while the infants of both language groups and the Korean-speaking adults saw the loose-fitting containment relations as novel compared to the tight-fitting containment relations they had been familiarised with, English-speaking adults did not.

These results were taken as indication that preverbal infants may already possess a repertoire of non-linguistic semantic categories and be sensitive to the boundaries between them, as evidenced by their ability to distinguish between loose- and tight-fitting containment. However, the preservation of this sensitivity into adulthood is conditional on the acquisition of a linguistic system that can express such nuanced differences (and resolve ambiguities when multiple perceptual features compete in the spatial categorisation process. See Choi & Hattrup, 2012.). By extension, the findings above strengthen the case for more closely exploring spatial language acquisition and processing, approaching them as potential sources of information concerning the development of spatial cognition. Perhaps one of the better-known attempts to explain the development of spatial abilities (and their interaction with language) is Gentner’s (1983) Structure Mapping Theory (SMT). The model suggests that comparison and abstraction processes are pivotal to the development of complex cognitive abilities. The underlying idea is that humans instinctively compare entities and situations, and that this process tends to build analogies that add to our conceptual understanding at increasing levels of complexity.

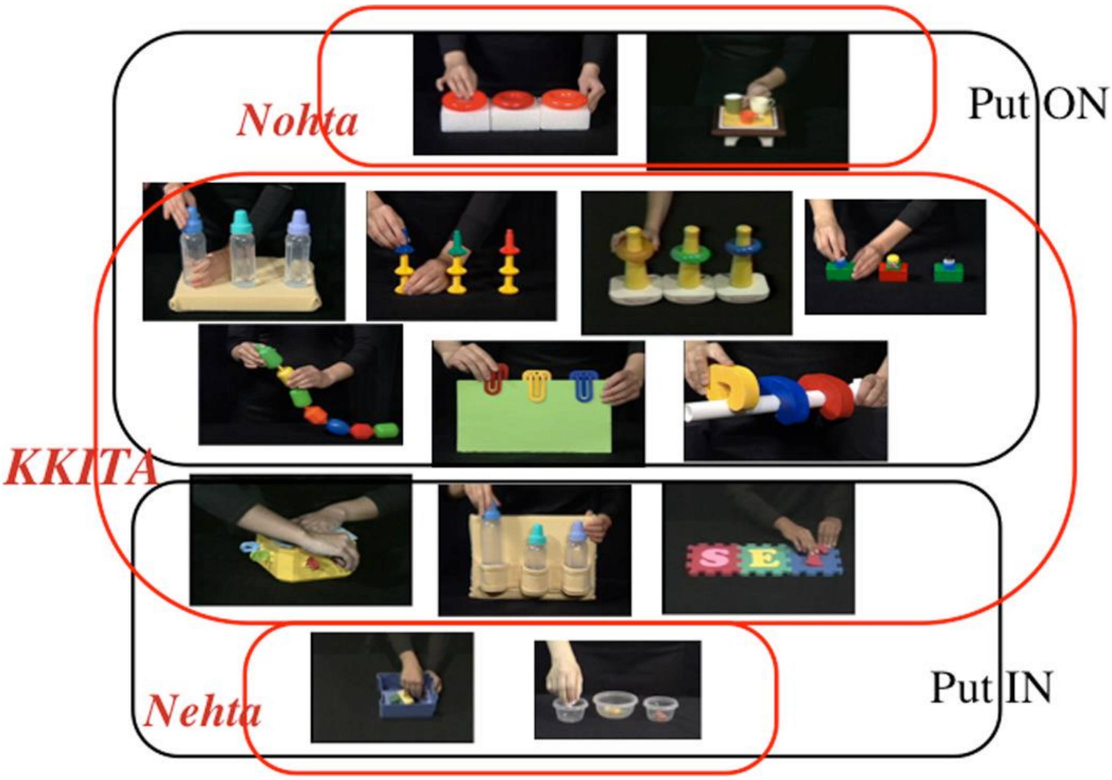


Figure 2.3 – Examples of the different grammaticalisations of spatial relations between English (black boxes) and Korean (red boxes). In the latter, both support (top) and containment (bottom) relations are differentiated with respect to the tightness of the relationship between figure objects and their reference objects (from Choi & Hattrup, 2012).

At the lowest level of complexity (and at the earliest stages of development), children tend to primarily draw comparisons between objects on the basis of their perceptual properties (i.e. the same target object can be located in two arrays in two different rooms, provided that the perceptual properties – such as shape, size, or colour – of the objects and rooms are the same). Over time, experience with a growing number of less similar objects leads to the development of strategies that rely not only on perceptual similarity between objects and situations, but also on an understanding of the spatial relations between the entities involved (DeLoache, 1987; 1995; Gentner & Loewenstein, 2001). As we experience more of the world around us, we are able to transfer our understanding of one semantic domain to instantiations of other domains. For example, a toy car on an inclined plane will roll down it in a way similar to a marble. A full-scale car without a stationary break will roll down a hill much the same way, as will a boulder down a mountain slope. At a higher level of abstraction, a situation "snowballing out of control" evokes the same basic dynamic. This is an example of a relational structure between entities being mapped from one domain to another (Gentner & Loewenstein, 2002), a process known as *structural alignment*.

A crucial turning point in the SMT model of development is the acquisition of language, particularly the ability to use relational terms. The early conceptual phase, with its reliance on close perceptual similarity between entities, is paralleled in the linguistic domain by the acquisition of object words (i.e. nouns), followed by verbs (Gentner & Christie, 2010). As children develop an understanding of structural relations, so they begin to acquire relational terms to describe those relations. The practice of assigning linguistic labels to objects and spatial relations serves multiple purposes. Giving multiple instances of an object the same linguistic label (e.g. bird), may invite children to compare those instances in order to determine what they commonalities may be despite their many perceptual differences. This process is known as *symbolic juxtaposition* (Gentner & Medina, 1998), and is also applicable to spatial relations. Different instances of “left of” may be perceptually very dissimilar, but assigning them a common linguistic label draws attention to them, allowing the underlying structural similarity between them to be encoded as an enduring representation, a process known as *reification*.

In summary, under SMT the natural tendency for analogical reasoning supports the acquisition of language, which in turn helps shape non-linguistic cognition by inviting certain specific comparisons between category exemplars. Under this mutual bootstrapping view, language operates as a *cognitive toolkit* (Gentner & Goldin-Meadow, 2003), by providing representational resources (e.g. inner speech) that coexist and operate in conjunction with other representational formats (e.g. visuo-spatial imagery) (Gentner & Christie, 2010). On a similar note, Karmiloff-Smith (1992) has proposed that a child’s semantic knowledge is not innate as such, but that its early acquisition is subject to biological constraints such as specific perceptual biases. This domain-specific knowledge then undergoes phases of progressive “re-description,” or re-encoding, into more domain-general representations over the course of the child’s interactions with the physical, social, and cultural surroundings.

More recently, Sinha and Jensen de López (2000) have also reformulated the spatial semantics acquisition process more explicitly in terms of embodiment theory, the idea that cognition is shaped by the interaction between the individual and the environment, as mediated by the individual’s biological capabilities. In doing so, they argued for the need for classical formulations of embodiment (e.g. Lakoff, 1987) to better account for the fact that the environments in which we operate and develop are not only physical, but also social and cultural, as already proposed by Vygotsky (1962). That is, just as the preservation of certain non-linguistic categories relies on the acquisition of a language system that can express them, they are also shaped by the physical interaction with the sociocultural milieu and its practices. For example, the Zapotec-speaking communities of Mexico appear to conceptualise the spatial notion of containment differently compared to English-speaking ones. While the English language differentiates canonical containment (containment within a canonically-oriented container, such as an upright cup), expressed via the preposition “in,” from containment “under” an inverted cup, Zapotec makes no such distinction. Due to the frequent use of baskets as covers (e.g. for food) in Zapotec-speaking villages, they are perceived as being just as efficient as containers in what we perceive as a non-canonical orientation. This flexibility in artefact use has permeated the Zapotec language, resulting in the use of the same body-part locative – equivalent to the English word for stomach – to describe both "in" and "under" forms of containment.

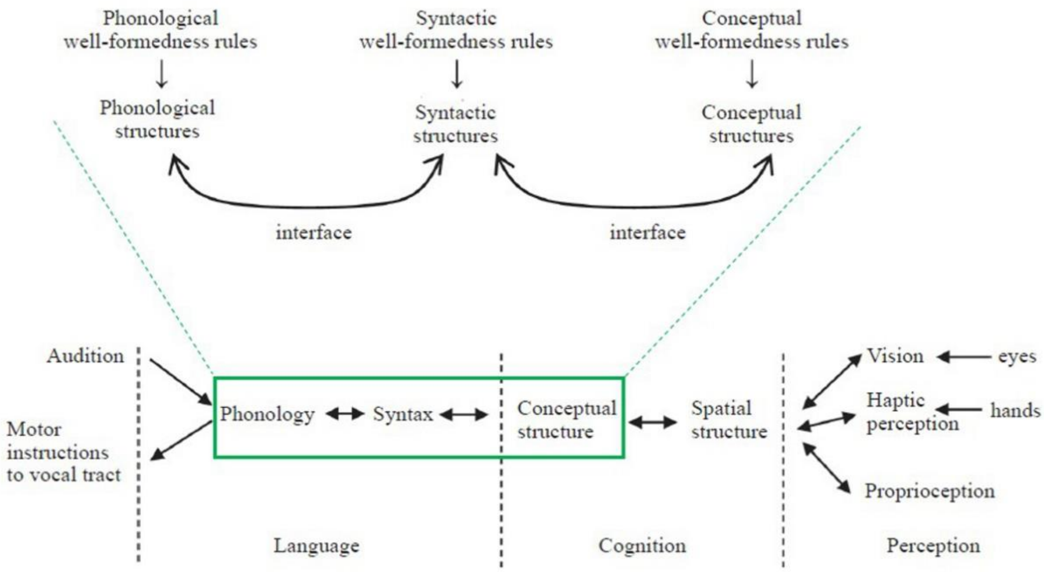


Figure 2.4 – A schematic representation of the translation of information between different formats via mental representations (adapted from Jackendoff, 2012). Note the absence of vision as a possible input source for linguistic information. This specific case was addressed in Experiments 1-3 contained in this thesis (Chapter 3). In Jackendoff's terms, this research is generally concerned with exploring how discourse-level linguistic processing leads to a creation of a conceptual structure, and how this is translated into different representations within the spatial structure depending on linguistic and imaginal factors.

Although the experiments presented in this thesis were performed on healthy, English-speaking participants, the evidence presented in this section points to the study of language as a potentially informative aspect of spatial cognitive investigations. Additionally, it justifies attempts to formulate theories that can comprehensively model the interactions between the language domain and broader cognition. It is important to understand that, with the exception of nativist approaches that appear at odds with findings of cross-linguistic diversity, the models described in this section are not mutually exclusive. Where they differ is largely with respect to the level of detail with which they describe the many different individual-environment interactions, what cognitive processes mediate them, and the many innate and acquired factors that influence them. Although the work to develop a more overarching theory that can coherently explain all of these interactions is still ongoing, Jackendoff (2012) has, I believe, convincingly explained the importance of such work and even begun to shape the core elements of such a theory. What is proposed is, in effect, a metatheory (or the need thereof, at any rate) of spatial understanding and of spatial information communication. That is, a theory that can successfully explain not only how we develop a volumetric understanding of space on the basis of sensory input, but also of "*how we talk about what we see—and how we see all the things we talk about as though they are part of the perceived world*" (Jackendoff, 2012, p 1128).

A crucial aspect of this theory involves the encoding of information within mental representations, and their translation between different formats (Figure 2.4). Perceptual information concerning the nature of objects in the environment and the relations between them is encoded into a spatial structure, and translated into a conceptual structure that can support linguistic syntax and phonology. Conversely, processing linguistic input yields a conceptual structure that can be transformed into a spatial structure and direct our perceptual attention. Jackendoff proposes that the nature of this spatial structure, e.g. for the word dog, is a *"viewpoint independent schema in spatial structure, which allows for a range of body proportions and a range of limb dispositions"* (p 1137). Additionally, he posits that a spatial structure, and the representations contained therein, must be constructed on sets of axes that can allow us to distinguish and identify the shapes of different objects based on their orientation and dimensions. Therefore, when multiple objects are part of the same mental representation, the relative spatial relations between them cannot always be efficiently encoded in precise 2D structures (akin to mental images), requiring instead more coarse but 3D representations. Similarly, movement in the world, which lies at the core of navigation, can only be instantiated if we possess an understanding of the volumetric spatial structure of the environment (including direction and distance information). Therefore, this structure cannot be encoded as two-dimensional static images such as the retinotopic representations directly derived from raw visual input. Rather, this representational system must be deeply reliant on a robust mechanism of distinct reference frames and axes (at the scale of both object-processing and of environment processing) onto which not only visual processing, but also language-derived models can be anchored, altered, and dynamically manipulated. These appear reminiscent of the coarse simulations that, Barsalou (1999; 2002; 2008) proposes, will be instantiated even for linguistic content that does not necessarily or explicitly describe spatial relations (see Section 1.10).

While the construction of 3D mental representations of objects and space on the basis of 2D visual input remains an open challenge in vision research, this aspect of Jackendoff's metatheory of spatial understanding lies beyond the purview of the research presented in this thesis. Rather, this body of work is focused on attempting to understand the feedback mechanism that exists between language comprehension and imaginal processes. Namely, how do we extract information from linguistic input to construct mental representations, and how is that extraction process influenced by concurring imaginal processes. In particular, Experiments 1-3 employed an eye-tracking paradigm to better understand vision as a source of linguistic rather than of visuospatial input, an aspect not explicitly addressed in Jackendoff's model. However, I will return to eye movements and their link to linguistic, visuospatial, and imagery processes in Sections 2.5 and 2.6. Instead, Section 2.3 and Section 2.4 will focus on the cognitive mechanisms involved in the production and processing of spatial descriptions, and the role played by individual differences in these processes.

2.3. Spatial Text Processing and Production: Mechanisms and Methodologies

So far, in this chapter, I have presented evidence that the language domain is intimately connected to other aspects of cognition, particularly the spatial domain. Individuals appear to acquire different sets of spatial concepts and to develop different ways of construing space or spatial relations depending on a number of factors during development (socio-cultural, environmental, and biological), and in a way that is reinforced by the specificities of the language in use. Additionally, we are capable of describing what we perceive, and of imagining what is described but not immediately perceptible. These abilities would seem to imply an ability to extract information from language (and other modalities) that refers to a state of affairs in the outside world, and construct a mental representation that is analogous to it – and, vice versa, to describe an imagined or remembered state of affairs. While the exact nature and format of these representations is debated, the fact that they are in some sense analogical to the situations they refer to is generally accepted. However, this was not always the case. Up until the early 1980s and prior to the diffusion of the ideas of mental models (Johnson-Laird, 1983) and situation models (van Dijk & Kintsch, 1983), many cognitive psychologists held the view that processing a spatial text involved the construction of a word-for-word representation of the surface level of the text itself. Over time, a number of subsequent studies (see Zwaan & Radvansky, 1998 for an exhaustive review) have transformed this view into the current one that sees more or less analogical mental representations as playing a much more central role in the development and use of these abilities.

Subsequent research into spatial language then developed in a variety of directions (apart from the cross-linguistic and clinical research discussed in Section 2.2). A number of studies have focused on providing evidence for the near-functional equivalence of spatial mental representations built on the basis of visual and linguistic input (Avraamides, Loomis, Klatzky & Golledge, 2004), haptic input (Giudice, Betty & Loomis, 2011), 3D sound (Loomis, Klatzky & Giudice, 2013), or actual experience (Bryant, Tversky & Lanca, 2001). For example, Avraamides et al. (2004) exposed participants to a spatial layout via visual perception and spatial language, and subsequently measured participants’ ability to perform judgements of allocentric direction between pairs of objects in the array. Results revealed that judgements made using mental representations built on the basis of linguistic input can be as accurate and as fast as those based on visual experience and visual memory. However, a frequent criticism of many such studies has been the use of exceedingly simple linguistic stimuli and spatial environments, limiting their ecological validity to an extent. Other researchers have focused on the production of ecologically valid spatial descriptions in order to isolate the cognitive processes that underlie it and their key linguistic elements.

Notably, Denis and colleagues (1996; 1997; Denis, Pazzaglia, Cornoldi, & Bertolo, 1999) built corpora of naturalistic route descriptions produced by their participants and observed operations and elements that feature consistently when spatial knowledge is externalised as linguistic content. At a basic level, the production of a route description is thought to involve the activation, in the form of visuospatial representations, of the spatial knowledge pertaining to the environment in question. This stage is then followed by a pre-verbal route-planning phase, which involves constructing a sequence of route segments within a subspace of the active mental representation that can connect the relevant origin and destination points. Route-planning is subject to a number of constraints, such as ease of navigation and ease of communicability. Following route selection, linguistic output is produced whose purpose is to convey the relevant information contained in the underlying mental representation. This stage is also subject to certain constraints, such as the limits of the addressee's processing resources, so that the resulting route description contains a limited number of statements describing only a portion of the overall information that could be extracted from a mental representation. In practice, the planned route is subdivided into paths connecting points where changes in heading occur, and the resulting linguistic output tends to rely predominantly on the identification of landmarks and on the prescription of actions to be taken at their locations. In this sense, landmarks in spatial language production serve the same functions and are subject to the same criteria for salience selection described in Section 1.3. Interestingly, although considerable between-subject variability can be observed in the number and nature of the landmarks selected, route descriptions can usually be reduced to a core structure containing the most essential aspects of a route (e.g. decision-point landmarks). Additionally, these skeletal descriptions tend to show remarkable consistency regardless of the degree of environmental knowledge, an indication that, to an extent, the perception of landmark salience might be the result of more independent metacognitive abilities (Denis et al., 1999).

However, despite the relative constancy of this core structure and in landmark selection, studies of language production have revealed considerable between- and within-subject variability with respect to the selection and maintenance of reference frames. Given the central role of reference frames in the experiments described in this thesis, it is important to consider how this variability in linguistic expression might reflect on the mental processing of reference frames and imagined perspectives. In Section 1.2 I have introduced the broad distinction between egocentric and allocentric reference frames, a central notion in much of the spatial cognitive literature and in this thesis. Subsequently, in Section 1.11, I introduced important findings suggesting that carrying out spatial tasks involved the construction and activation of different reference frames. In particular, Gramann (2013) has observed that "*it is reasonable to assume that there are more than two representations active during spatial orientation*" (p. 3) and that the egocentric-allocentric dichotomy, however useful, might be a simplification of the actual repertoire of possible mental representations of space. This sentiment has clear precedents in the linguistic literature, most notably in Levinson's (2003) extensive review of the various coordinate systems, conceptualisations, and nomenclatures adopted in different disciplines (and in different languages). While the full depth and breadth of Levinson's analysis are impossible to summarise concisely, it is nevertheless necessary to point out that the egocentric-allocentric dichotomy can broadly map onto the absolute and relative reference frames in Levinson's three-frame classification of coordinate systems (Table 2.1).

Table 2.1 – Levinson's (2003) alignment of reference frame classifications in which coordinate systems are classified on the basis of their axes' origin point (i.e. whether they are centred on the 'ego' – any body-centred sensory system – or on a ground object) and depending on their reliance on binary (i.e. between a ground object and a referent object) or ternary (i.e. between a speaker's viewpoint, a ground object, and a referent object).

INTRINSIC	**ABSOLUTE**	**RELATIVE**
Binary	Binary	Ternary
Origin = Ground	Origin = Ground	Origin = Ego
Object-centred Intrinsic-perspective 3D Model	Environment-centred	Viewer-centred Deictic-perspective 2.5D Sketch
Allocentric		Egocentric
Orientation-free	Orientation-bound	

Just as the repertoire of coordinate systems humans use under ecological conditions is probably larger than simple conceptualisations and laboratory experiments can capture, the ways in which speakers verbalise spatial relations and movement in these coordinate systems are many and varied. While much of the psychological literature has traditionally relied on the distinction between route and survey spatial descriptions (broadly mapping onto the distinction between egocentric and allocentric representations), the reality of everyday language use may be more complex. For example, Tenbrink and Salwiczek (2016) recently tested participants in a virtual reality tunnel task (Gramann et al., 2005) under different instructions (egocentric, allocentric, and neutral), and asked participants to verbalise the strategy they were using during the task in a think aloud protocol (as well as retrospectively at the end of the task). While behavioural results appeared to replicate the finding of distinct between-subject preferences for egocentric or allocentric spatial strategies, a cognitive discourse analysis (Tenbrink, 2015) of participants' verbal output revealed a partial disconnect between behavioural performance and linguistic self-reports. More specifically, the percentage of survey terms (*north, south, east, west, map, compass, above/from above*) used during verbalisations remained relatively low even when participants were provided with allocentric instructions, although participants displaying an allocentric pattern of results used them more than route terms (*left, right, front, straight, forward, back*). Additionally, less than half of the participants appeared to shift their task strategy when provided with different instructions. This pattern of results might have different explanations; for example, it is possible that the target survey terms selected were simply less likely to occur in the verbal outputs because of the nature of the task. That is, even participants imagining an allocentric view of the route and adopting an allocentric response strategy may have still described the turns in the route in egocentric terms relative to the direction of travel. Conversely, it is possible that, if participants had been given a cardinal orientation at the onset of the route (e.g. "*At the start of the tunnel, you are facing north.*"), they would have produced verbalisations with considerably more survey terms during the task.

Nevertheless, these results highlight the importance of performing clear task analyses when formulating predictions, but also that participants' self-reports can provide a wealth of information on participants' mental representations, their phenomenal experiences of their mental imagery, and their task strategy. However, they also point to the methodological challenges of analysing unconstrained linguistic production, particularly when compounded with the potential limits of participants' introspective abilities. For this reason, in Experiments 2-5 I employed a more traditional questionnaire to attempt to determine the strategy (i.e. the imagined spatial perspective) used by participants. However, before introducing Experiment 1, in the remaining sections of this chapter I will cover the issue of individual differences in spatial abilities, as well as the two key methodologies used in Experiments 1-3 – eye-tracking and map-drawing tasks.

2.4. Individual Differences in Spatial Abilities and Language

In addition to the sources of variability in linguistic production and processing discussed in the previous sections, a large number of studies appear to show that individual differences in non-linguistics visuospatial abilities play a significant role in the processing of spatial language. While consistent with the idea that processing spatial linguistic content involves the construction of mental representations that reflect certain visuospatial properties of the environments described (e.g. the relative positions of landmarks, the distances between them, their visual features, etc.), this adds a further layer of theoretical and methodological complexity to any investigation of spatial language and mental representations. Many of the studies of individual differences in spatial cognition and language have operated within the framework of Baddeley's working memory (WM) model (Baddeley, 1986; Baddeley & Hitch, 1974), which sees WM as a temporary storage and processing unit with a verbal component (verbal working memory, or VWM) and a visuospatial component (VSWM), both operating within a central executive component. The dual-task paradigm is most typically used in these studies, involving a primary task, and a concurrent task meant to selectively tax the verbal or the visuospatial components of the system. For example, De Beni, Pazzaglia, Gyselinck and Meneghetti (2005) aurally presented participants with spoken texts containing either spatial or non-spatial information, while they concurrently performed either articulatory suppression (AS), consisting of the repetition of sets of syllables, or spatial tapping (ST), which involves tapping on four buttons on a board. Sentence verification of the spatial and non-spatial content, and free recall were used as dependent measures of language processing and comprehension. The results revealed that concurrent AS reduced performance during recall of both types of texts, whereas ST interfered only with the recall of spatial descriptions, indicating the differential loading on VSWM of the two types of linguistic content.

This pattern of results was replicated and expanded upon in subsequent studies. Pazzaglia, De Beni and Meneghetti (2007) had participants perform AS and ST during either encoding or retrieval of the information, and observed that the disruptive effect of AS is found only during encoding, whereas ST interferes both with encoding and recall of the spatial texts. In yet another replication, Meneghetti, De Beni, Gyselinck and Pazzaglia (2011) presented participants with spoken route descriptions as they performed either a spatial or a verbal concurrent task. Recall tasks (a sentence verification task and a graphical representation task) were performed after three consecutive encodings of the text or after the first and third encoding. Results revealed that the verbal concurrent task interfered with sentence verification at both levels of exposure to the text. On the other hand, the interference effect of both types of concurrent tasks with the map-drawing task was noticeable only following the second encoding, likely due to a floor effect following the first encoding. Additionally, a dissociation was observed between the free recall of landmarks and their correct spatial placement on the map, in line with a result obtained in Experiment 2 presented in this thesis.

Other studies have more closely explored the role of spatial abilities in spatial text processing by measuring how these are modulated by individual differences. Meneghetti, Gyselinck, Pazzaglia and De Beni (2009) studied the susceptibility of participants to the interference of concurrent tasks as a function of their mental rotation (MRT; Vandenberg & Kuse, 1978) scores. While AS was found to disrupt the performance of both groups' on non-spatial text processing, only low-mental-rotation (LMR) participants were impaired by both types of concurrent tasks during processing of route descriptions. High-mental-rotation (HMR) participants, on the other hand, were better able to contain the interference of the ST concurrent task during spatial-text processing, presumably as a result of their greater VSWM abilities. On the other hand, Gyselinck and colleagues (Gyselinck, De Beni, Pazzaglia, Meneghetti & Mondoloni, 2007; Gyselinck, Meneghetti, Pazzaglia & De Beni, 2009) showed how participants instructed or trained to adopt imagery strategies during the processing of spatial texts obtained higher recall scores in single-task conditions, but were more susceptible to interference effects during concurrent-task conditions. This was taken as further indication of VSWM involvement during spatial language processing, particularly when mental imagery is employed.

More recently, Meneghetti, De Beni, Gyselinck and Pazzaglia (2013) explored the interaction between visuospatial abilities, the use of imagery strategies, and VSWM resources by testing LMR and HMR participants on spatial description comprehension via sentence verification and map drawing, both before and at two time points following imagery training. The training involved a theoretical introduction to the use of mental imagery, and practice with constructing mental representations of words and sentences (both concrete and abstract) as well as of practice routes and environments. During encoding of the last spatial route description of the test phase, a concurrent ST task was given in order to measure the susceptibility to VSWM interference following training. Performance scores revealed once again that HMR participants experienced no or highly reduced ST interference compared to LMR participants. The latter, however, also appeared to benefit from imagery training in the absence of a concurrent task, unlike the HMR group. This demonstrated both the natural propensity of the HMR group to adopt optimal imagery strategies (i.e. to construct a spatial representation of the described environment) already at baseline, and the potential of imagery training to improve visuospatial and navigational performance in individuals with less VSWM resources. Furthermore, HMR participants displayed a higher self-reported propensity to adopt allocentric perspectives during spatial imagery and to use cardinal directions to orient themselves. This is generally consistent with the view of mental rotation as a measure of small-scale spatial abilities, while also suggesting their partial overlap with large-scale spatial abilities, as proposed by Hegarty and colleagues (Hegarty et al., 2006; Wolbers & Hegarty, 2010).

In summary, we see that if indeed, as evidence would suggest, the processing of spatial linguistic content requires the construction of mental representations maintaining a certain degree of analogy to the situation being described, this process is complex and subject to multifactorial influences. Not only is the cultural and linguistic environment of a speaker going to interact with innate biological constraints to shape the repertoire of concepts that can be readily mentally represented, but individual circumstances and differences will also influence language acquisition and affect the way linguistic content is processed. Therefore, it seems reasonable that, to fully understand how language interacts with non-linguistic processes within spatial cognition, the online aspects of language processing must be explored much more closely. Despite this, most of the studies on spatial language have only studied the results of encoding processes indirectly by measuring differences in behavioural performance in a variety of tasks following encoding, or contextual effects on language production, and direct measures of linguistic processing are still lacking in the literature.

Table 2.2 – The first four (out of 20) statements in each description type used by Tom and Tversky (2012). These include action statements (1, 2, 6) and locative statements (3, 4, 5).

Street description	Landmark description
1. Leave the park.	1. Leave the park.
2. Follow the path that is straight ahead.	2. Follow the path that is straight ahead.
3. Go down this path which *is edged with a row of gigantic redwood trees.*	3. Go down this path *which leads to a building that was privately financed.*
4. You will then see on your left *a very bumpy and stony dirt road.*	4. You will then see on your left *an office building with small companies.*
5. On your right, there is *a road that zigzags sharply the entire way.*	5. On your right, there is *a business operated by a young couple.*
6. Turn right.	6. Turn right.

Tom and Tversky (2012), however, have attempted to relate individual imagery abilities and linguistic manipulations to the online processing of the linguistic content. Their study was a follow-up to one by Tom and Denis (2004), which had found landmark-based route instructions to be more effective (i.e. better recalled) wayfinding supports than route directions based on street names. Tom and Tversky attempted to expand on those results, by determining the extent to which the vividness of the description of the various textual elements influenced their encoding and recall. In the first of two experiments they presented participants with written descriptions of the same route through a fictitious urban environment. In a between-subject design, participants read either a street-based description or a landmark-based description. The former contained visually vivid descriptions of the salient features of each path segment, interspersed with directional changes. The latter, on the other hand, described undistinctive building-like landmarks in factual rather than perceptually salient terms. Table 2.2 contains excerpts from both description types used in the study. The descriptions, which were equated in length, were presented one statement at the time in a self-paced reading paradigm, allowing for the computation of a reading time per syllable. Participants could not go back to read previous statements, but they were allowed to read the description three times, with a map drawing task following each reading. Individual differences in spatial abilities were measured via MRT and using Money's Standardised Road-Map Test of Direction Sense (MT: Money, Alexander & Walker, 1965).

Reading times did not significantly differ between the two conditions, but reading times per syllable were longer for statements describing the locations of spatial elements than for action statements. Measures of information recall in the map-drawing task showed significantly better recall in the vivid street than in the non-vivid landmark condition, both in terms of overall recall and of spatially correct recall. Additionally, participants with high MRT scores were found to have faster reading times, and better recall in the vivid street condition, whereas MT scores positively correlated with correct spatial recall of landmarks. In a second experiment, participants read either a route text with vivid descriptions of both streets and landmarks, or a poorly vivid description. Presentation and paradigm were the same, but participants were also administered the Vividness of Visual Imagery Questionnaire (VVIQ: Marks, 1973). Participants who scored higher on any of the three imagery measures (MRT, MT, or VVIQ) were found to read either description type more quickly. Recall was generally better following encoding of the vivid description, but in this condition it was comparable between streets and landmarks. In the non-vivid description, however, participants displayed an advantage for the recall of streets over landmarks. Additionally, participants with high MRT scores obtained better recall scores following the first encoding of the non-vivid description.

These results were interpreted as suggesting a general facilitation effect of vividness on encoding and recall of information as mediated by mental imagery, regardless of the type of spatial information being encoded (i.e. landmarks or paths). The observation of a correlation between MRT scores and non-vivid recall was interpreted as evidence of dissociation between the visual and spatial aspects of mental representations, with the former facilitating the latter when present (cf. Section 1.9). In this sense, the construction of mental models of route descriptions consists in associating the information available on landmarks and streets to the relevant sections of the network of nodes and paths. The vividness of the information is thought to facilitate this association, as already observed by Paivio (1990) (see Section 1.6). However, the implication that landmarks and paths are informationally equivalent in spatial descriptions perhaps warrants some pause. It could be argued that the descriptions used by Tom and Denis (2004) represented ecologically more valid examples of the types of route directions humans normally interact with. After all, Tom and Denis (2003) had already shown how participants were not only better able to use landmark-based instructions in wayfinding tasks compared to street-based instructions, but they also tended to generate route directions that contained more landmark than route information. In fact, it is possible that describing path segments in more vivid terms may have simply had the effect of making those spatial elements behave, in a sense, like landmarks, albeit not of the point-like nature that they are normally thought to assume. Vividly-described streets might, effectively, act in a way similar to the space-defining (SD) objects described by Mullally and Maguire (2011; 2013).

In the studies by Mullally and Maguire, participants were required to mentally construct and deconstruct imagined scenes in a stepwise manner, by adding or removing one element at the time out of a set of objects they were presented with. By recording the order in which the objects were added to or removed from the scene, the authors were able to establish which categories of objects were consistently deemed necessary to maintain a cohesive "core scene". These objects can, even when imagined in isolation and apart from any spatial context, engender a sense of their local three-dimensional space without a requirement to mentally represent the contiguous surface elements (e.g. walls and floors) of the environment or a number of other objects within the same environment. For example, a grand piano was considered as more space-defining by participants than a floor light or a beanbag, because it inherently implies a floor surface to which it is anchored without having to actually represent it. Smaller, more mobile objects, on the other hand, tended to be imagined as more decontextualized, effectively "floating" within the representational space. Crucially, participants' subjective reports of differences between categories of objects were corroborated by functional brain imaging data, showing stronger activation in the parahippocampal place area (already implicated in the processing and awareness of 3D spatial layouts; Epstein, 2008) in response to SD objects. In the case of canonical route directions, landmarks might serve as SD objects during the construction of mental representations of these environments. With enough spatial information, it is possible to envision the spatial relations between them, and gain a sense of the environment's global layout, without explicitly generating vivid representations of the surface of the paths connecting them. If, on the other hand, the descriptions of path segments are intentionally made vivid (as in Tom & Tversky, 2012), the street network can take on the role of holding the representation together and of making the environment's shape and layout explicit. In such a representation, the location of non-vivid landmarks (in the canonical sense) could, in principle, be expressed in terms of its position relative to vivid path segments (e.g. that particular bench is exactly half-way down the leafy path edged on both sides with oak trees). How likely it is for this kind of mental representation to be required in most ecological circumstances is unclear, but a better understanding of the interaction between spatial structure and visual vividness in imagery might have important implications in areas such as associative learning.

More importantly, however, the study by Tom and Tversky constitutes a step in the important direction of more accurately characterising the processes underlying the encoding of spatial language. Experiments 1-3 in this thesis attempted to take a further step in this sense. In Experiments 1 and 2, eye movements were recorded as participants read route descriptions, and, in Experiment 3, also while they judged the accuracy of sketch maps following description encoding. The study of eye movements has a long history in the study of cognition, and their potential relevance to the understanding of imagery processes during language and scene processing will be explored in Section 2.6. Before that, however, Section 2.5 will review a few of the studies that have employed sketch maps to gain an understanding of their processing and production as a type of external representation complementing linguistic descriptions. An understanding of the processes involved in the construction and processing of sketch maps will be important in interpreting the results of Experiments 1-3, particularly with respect to the transformation of reference frames between linguistic encoding and visual test.

2.5. From Cognitive to Sketch Maps – Representational Congruence and Recall

As the study of the neural correlates of cognitive maps broadened following Tolman's (1948) seminal experiments, so did the experimental use of map drawing tasks meant to externalise environmental knowledge. These found application in cognitive science as in a variety of other research fields ranging from urban planning (e.g. Lynch, 1960) to geographical education (e.g. Brewster & Blades, 1989) and more. Indeed, the literature that has made use of maps (both cognitive and physical) is too vast and multidisciplinary to be adequately summarised in this section (but see Kitchin, 1994 for a review). What is perhaps more fundamental is determining whether map-drawing performance is an effective measure of the accuracy of participants' cognitive maps and environmental knowledge.

To answer this question, Blades (1990) set out to ascertain the test-retest reliability of sketch maps by tasking participants with drawing them on two sessions a week apart. Participants were asked to draw a map of the route from one local landmark to another. During the second session, one group of participants was given the same instructions, and another was asked to draw the same route but using the previous destination as origin point. Maps were scored on the number of road names and named landmarks, as well as the number of road segments drawn (including side roads drawn at intersections). A subset of the maps (20 pairs, with road and landmark names removed) was then given to two judges, who were tasked with determining which had been drawn by the same participants. Results showed that the two judges were able to correctly match a majority of the maps in both instruction conditions: 19 and 14 out of 20 in the same-instruction condition, 17 and 16 out of 20 in the different-instruction condition. Additionally, significant correlations on all three scoring measures between first- and second-session maps were found even for participants who had been given different instructions during the second session. The authors concluded that participants will consistently produce comparable sketch maps of the same route or environment, at least over short periods of time, even if task demands are altered.

Billinghurst and Weghorst (1995) further addressed the question of whether sketch maps are accurate representations of environmental knowledge when this is derived from a restricted range of perceptual cues. This was tested by having participants draw map representations of previously explored virtual environments. Participants also completed a survey probing various aspects of their subjective navigational experience (e.g. perceived realism, ease of orientation, dizziness, etc).

Maps were scored by two raters, who judged how useful they would be as navigational aids (Map Goodness), according to the number of different Object Classes included as landmarks, and to the accuracy of Relative Object Positioning. Significant correlations were found between the scoring measures Map Goodness and Object Classes, and participants' self-reported understanding of object locations and sense of orientation in the environments. This indicates that participants are, largely, aware of the quality of their own navigational performance, and of the visuospatial information they can extract from it.

Perrig and Kintsch (1985) presented participants in a between-subject design with two texts describing the spatial layout of a fictitious town, one (Route version) conveying information using egocentric relational terms and one (Survey version) using cardinal relational terms. Both texts were 24 sentences long, contained spatial and non-spatial information concerning the town, and were presented incrementally on screen one sentence every seven seconds. Within each group, half of the participants read the descriptions once, and the other half read them three times. Encoding of the descriptions was followed by a free recall task during which participants were asked to write down everything they could remember about the text, without time restrictions, paraphrasing the original texts whenever they could not remember exact propositions. This was followed by a series of true-false statements probing both non-spatial and spatial knowledge. Spatial statements probed both the locations of landmarks explicitly stated in the texts (Old Locative sentences) and spatial relations to be inferred (Inference sentences). Half of the Inference sentences were written in Route terms (i.e. "left" and "right"), half in Survey terms (i.e. cardinal directions). Finally, participants were asked to draw map of the town. Results revealed better propositional recall for the Route compared to the Survey description, and after three encodings compared to only one. Survey recall resulted in significantly more discontinuities (deviations in the order of recalled sentences compared to their occurrence in the texts) than route recall. Map drawing performance was generally poor (with no significant differences between conditions), as was performance in spatial (both locative and inference) sentence verification. A discrepancy was observed between the (relatively) good recall of text propositions (up to 47%) and the floor performance in the map drawing and spatial inference verification tasks.

In a second experiment, shorter texts and a simpler environmental layout were used. As in the first experiment, a group of participants studied a Route description and another Survey description. Reading was self-paced and up to four presentations of the texts could be requested. Additionally, a third group studied a map-like representation of the environment without time limits. All three groups performed the statement verification task following encoding. Survey and Route groups were invited to return after four days to carry out the free proposition recall and map drawing tasks. Analysis of the study time revealed that learning from the map resulted in the shortest encoding time, followed by Route and Survey text. Route description encoding resulted in better propositional recall and fewer discontinuities than Survey description encoding. This was taken to indicate that route descriptions establish more cohesive links between the different propositions contained in them, which is known to aid in encoding linguistic content into working mental models (Foos, 1980).

Map drawing performance, on the other hand, was significantly better following Survey encoding. Performance in the spatial sentence verification task showed a three-way interaction between sex, encoding text, and spatial knowledge probed. Female participants were very accurate in verifying locative sentences and inference sentences, provided the latter were written congruently with the text they had encoded. No significant difference was found between Route and Survey encoding. Male participants displayed a different pattern. Survey encoders performed much like female participants but performed worse during incongruent inferences. Route encoders, on the other hand, performed generally poorly on all Route statements, both locative and inference. Participants who had studied the map performed generally well on all locative statements. No differences in accuracy were found between survey and route inferences, but map encoders verified survey inferences significantly faster than route inferences.

Beyond demonstrating the validity of map sketches as measures of spatial knowledge, these results reveal a tendency for congruent representational formats to facilitate performance in visuospatial tasks. Namely, encoding of survey (allocentric) descriptions results in higher map drawing performance, whereas encoding of route (egocentric) descriptions results in greater propositional recall performance. This appears to be a specific case of the more general study-test congruency effect, wherein the ability to recall events or stimuli is influenced by the overlap between encoding and retrieval processes and circumstances (e.g. Morris, Bransford & Franks, 1977). Multiple studies have demonstrated this effect with a particular focus on the similarity between the physical formats used during encoding and test (e.g. Mintzer & Snodgrass, 1999; Mulligan & Osborn, 2009; Weldon & Roediger, 1987). Other studies have explored both the behavioural and neural aspects of this effect (e.g. Park & Rugg, 2008; Uncapher & Rugg, 2009). Bauch and Otten (2011) engaged participants in an incidental-learning EEG paradigm, during which they were presented with pictures and visually presented words and asked to perform size judgements on the objects represented. Following a one-hour break, their recall was probed between-subject within the same mode of presentation (picture-picture; word-word) or in the alternative mode of presentation (picture-word; word-picture) and with the inclusion of novel stimuli.

Behavioural results showed that same-presentation participants were more accurate in correctly recognising old stimuli compared to the alternate-presentation group. The nature of the stimuli (whether they were words or pictures) did not affect response accuracy, but the patterns of EEG activation during encoding of later remembered stimuli differed depending on the format. The encoding of correctly recalled words was associated with a small positive modulation over frontal electrodes compared to forgotten stimuli, regardless of the test format, a finding thought to reflect the degree of processing of a word's semantic and associative attributes (Otten, Sveen & Quayle, 2007). The authors suggested that, although participants may have generated mental images of the encoded words and attempted to retrieve that quasi-perceptual information during test in the word-picture condition, the probable lack of overlap between mental images and the visual test cues may have made the retrieval of conceptual information the most efficient strategy. This would have led to the same pattern of activation for correctly-recalled stimuli in both word-word and word-picture trials. On the other hand, the encoding of correctly remembered pictures was associated with a positive frontal modulation for the same-presentation group, but with a positive posterior modulation for the different-presentation group. In the latter case, participants presented with a word cue at test may have generated a visual mental image of it in order to match it with a stored perceptual representation of the respective picture stimulus. As such, pictures whose perceptual attributes had been more efficiently encoded as visual images during study (a process already associated with parietal activation. See Chapter 1.) would have led to better recall.

The congruency effect was further explored by Staresina, Gray and Davachi (2009), who measured the degree to which noun-colour pairs (e.g. elephant-red) were deemed "plausible" (i.e. congruous) or "implausible" (i.e. incongruous). In a second type of encoding task (a valence task), they also measured the extent to which congruence perception (and the resulting recall facilitation) may be driven not solely by a plausible semantic congruence (e.g. balloon-yellow), but also by subjective aesthetic matches. For example, following the presentation of a "cheese-green" pair, participants would have to decide whether such a semantic association (i.e. green cheese) would be "appealing" (i.e. congruous) or "unappealing" (i.e. incongruous). Recall was tested using a 3-step task consisting of old/new judgements of the nouns, a test of colour memory, and a recall test of whether a plausibility or valence task was performed on that particular item. Behavioural results revealed a significant congruency effect that facilitated not only the recall of the experimental items (noun and colour), but also memory of the task type. Furthermore, said congruency effect was present not only as a function of plausible semantic congruence, but also as a function of subjective aesthetic schema (i.e. whether a participant found a particular noun-colour match appealing).

These results are important as they suggest a more general interpretation of the congruence effect as the result of relational binding of the elements that will form an episodic memory. They are also particularly relevant with respect to the experiments presented in this thesis. Here, congruence was construed in terms of overlap between three reference frame components: implicit reference frame, imagined reference frame as an explicit encoding task demand, and test reference frame. The first was manipulated as a function of the relational terms – “left” and “right” or cardinal points – used in the route descriptions encoded by participants (Experiments 1-5). The second as a function of the spatial perspective participants were instructed to assume during encoding (Experiments 2-5). The third, as a function of the specific recall task used – allocentric map-based recall tasks in Experiments 1-3 and egocentric judgements of relative direction in Experiments 4 and 5. While Picucci, Gyselinck, Piolino, Nicolas and Bosco (2013) have previously explored the interaction of presentation format (visual virtual environment exploration vs verbal auditory route description) and verification task format (visual sketch map vs verbal sentence verification tasks) in the construction and use of spatial mental models, the influence of imagined reference frame on encoding and recall processes was not explicitly explored. Beyond assessing the effects of these factors on behavioural performance, Experiments 1-3 adopted an eye-tracking methodology to test for potential differences in spatial text encoding as a result. In the next two sections, I will review some of the relevant literature on eye movements in order to fully set up the methodological stage for the experiments that will be presented in Chapters 3 and 4.

2.6. Eye Movements in Language Processing, Scene Processing, and Mental Imagery

The idea that the eyes represent a window onto the soul is a common cultural and literary trope. Metaphysical claims are beyond the scope of this thesis, but it is undeniable that the scientific study of eye movements in psychology has a long and rich tradition. Its history goes back to Huey’s (1908) work on the pedagogy of reading (a body of work that did not fail to recognise the potential relevance of mental imagery, in the form of inner speech. Ehrich, 2006; Yaden, 1984), or perhaps even earlier (see Wade, 2010 for a review of the early history of the field). Since then, a lot has been learnt about the nature of eye movements and the many factors that influence them (see Rayner, 1998; 2009 for comprehensive reviews of the field), but several questions remain without a conclusive answer.

In the broadest sense, research in this field attempts to answer a basic question: why do we move our eyes? The simplest answer is that the anatomical limitations of the retina restrict the fovea, the region of highest visual acuity, to approximately 2° of amplitude on either side of the fixation point. However, in normal readers asymmetries may be observed on either side depending on the customary direction of reading in their language (Paterson, McGowan, White, Malik, Abedipour & Jordan, 2014; Schuett, Heywood, Kentridge & Zihl, 2008; Figure 2.5). This requires us to perform rapid movements called saccades to bring visual stimuli, whether written words or other percepts, within foveal space. During execution of these movements, perception (e.g. Matin, 1974) and processing (such as mental rotation, e.g. Irwin & Brockmole, 2000) are mostly suppressed (but see Campbell & Wurtz, 1979 for exceptions to saccadic suppression), and their amplitude and speed may depend on the nature of the task at hand. During reading, saccades have an amplitude of approximately 2° and last around 30 ms, while visual scene perception may involve, on average, larger saccades of approximately 5° and lasting around 40-50 ms (Abrams, Meyer & Kornblum, 1989). Between saccades there are periods lasting, on average, around 200-300 ms during which eye gaze is maintained on a relatively small region of the visual field, and during which visual processing can take place. Generally, eye movements are considered to be closely related to attentional processes (Hoffman, 1998) and, although covert attentional shifts are possible without overt eye movements (e.g. Posner, 1980) (see Figure 2.5), overt eye movements are always preceded by pre-saccadic, covert shifts of attention (Zhao, Gersch, Schnitzer, Dosher & Kowler, 2012). As a result, saccades and fixations are usually considered reliable proxies of attention allocation to what is being fixated, in accordance with what is known as the *mind-eye hypothesis* (Just & Carpenter, 1980), thus shifting the question from what saccades and fixations represent (i.e. shifts of attention and the allocation thereof), to what factors and cognitive mechanisms drive them. In Experiments 1-3 I have used eye tracking to record eye movements during processing of both written language (route descriptions) and non-linguistic visual stimuli (sketch maps). Accordingly, in the following paragraphs I will discuss research that has attempted to address these issues both during reading and during scene perception.

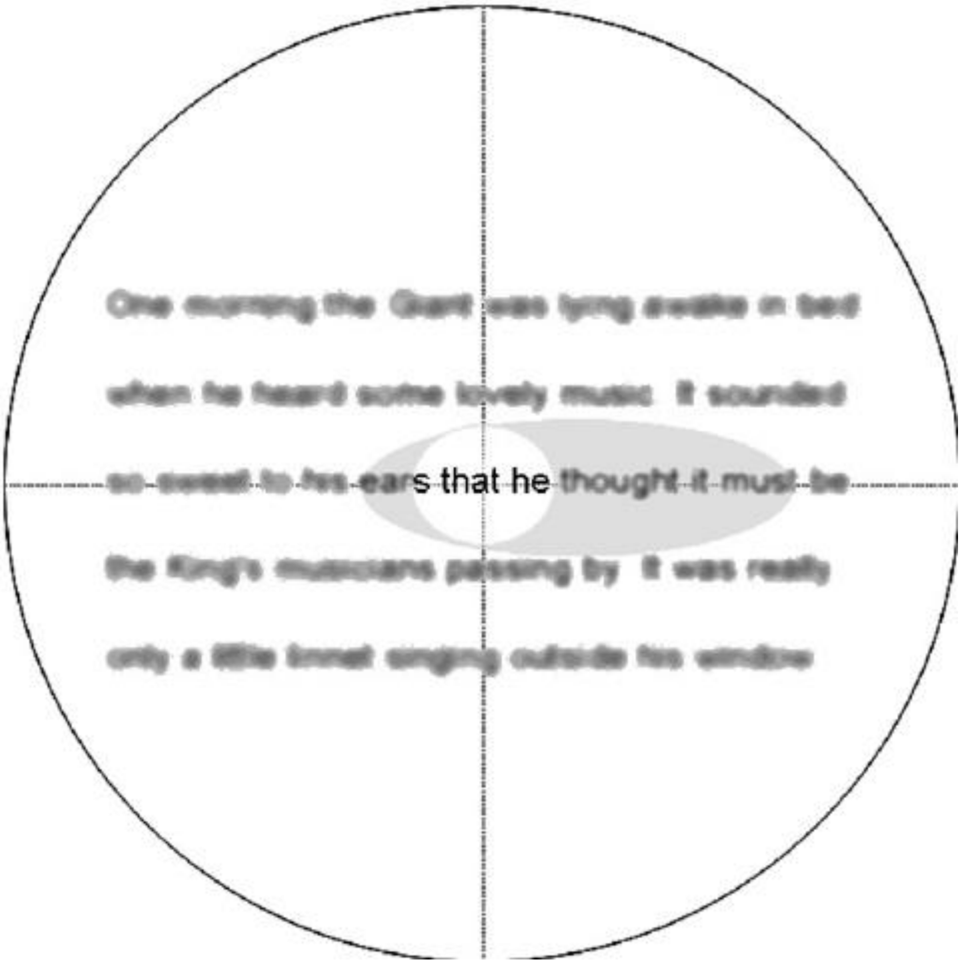


Figure 2.5 – A schematic representation (Schuett et al., 2008) of the asymmetric perceptual span of normal readers in a left-to-right language. The area at the centre of the crosshair represents the area of highest visual acuity, which decreases with distance from the fixation point (grey area). Readers may covertly attend to and begin processing words within this region to the right of the current fixation point without overtly performing saccades towards and fixating on them.

These lines of research have historically travelled along parallel paths, intersecting occasionally and displaying certain similarities. In the field of reading research, two main competing models of eye movement programming and control have received the most attention: the E-Z Reader (e.g. Reichle, Pollatsek, Fisher & Rayner, 1998) and the SWIFT (e.g. Engbert, Longtin & Kliegl, 2002) models. An in-depth analysis of the differences between the two models is not the goal of this thesis, as the eye tracking experiments presented here were not designed with the aim of testing them (however, see Rayner, 2009 for a detailed review). What is perhaps more relevant to the research presented here is the fact that both models generally rest on the assumption that word-level properties are the main drivers behind saccade programming. The location of fixations is thought to be primarily determined by low-level properties such as word length and spacing, whereas the timing of saccades and fixation duration are influenced by lexical factors such as word frequency (Rayner, 2009). Neither model adequately accounts for higher-order factors at play during online comprehension at the discourse level, despite mounting evidence challenging the purely bottom-up view of reading as "context-free decoding" (Hawelka, Schuster, Gagl & Hutzler, 2015). In this sense, the field of visual processing during scene perception has undergone a similar evolution to that of reading research. Early models of eye movements during visual scene perception and search were largely concerned with the location of fixations, and revolved around the notion of saliency maps, typically defined as the product of bottom-up perceptual factors such as contrast, colour, intensity, brightness, or spatial frequency (Itti & Koch, 2000; 2001). More recent models, however, have begun to consider also the way the effects of cognitive, behavioural, and contextual factors on eye movements (e.g. Torralba, Oliva, Castelhano & Henderson, 2006) might contribute to producing not purely perceptual saliency maps, but more comprehensive priority maps (e.g. Bisley & Goldber, 2010). These are representations of the visual field in which locations are weighted not solely on the basis of their visual salience, but also as a function of their behavioural importance and under the modulation of a number of higher-order factors.

In reading research, several studies have shown top-down effects of reading material format (e.g. single sentences in isolation vs the same sentences inserted within the context of a longer passage) or of verification task instructions (e.g. simple or complex verification questions vs multiple-choice questions) on a variety of eye tracking measures of encoding (e.g. Radach, Huestegge & Reilly, 2008). Such factors have been shown to top-down influence both early (e.g. first fixation duration) and late reading measures (e.g. total reading time), as well as measures of saccade amplitude and fixation location (assumed by established models of reading to be primarily influenced by low-level properties of words), and to also interact with word frequency. Similarly, Hawelka et al. (2015) have shown how individual differences in reading ability and speed influence participants' ability to predict upcoming words (measured by the probability of said words being skipped or refixated), with slow readers requiring more fixations to recognise a word, potentially owing to a smaller perceptual span in these participants (Figure 2.5). These results are in accordance with the predictive coding framework (Rao & Ballard, 1998), which suggests that the brain is constantly attempting to generate forward inferences on the basis of contextual cues and prior experiences in order to predict upcoming events (or words). Further evidence to this effect comes from studies of anticipatory eye movements, one of the areas of intersection between reading and scene processing research. Altmann and colleagues (Altmann, 2004; Altmann & Kamide, 1999; 2007; Kamide, 2008; Kamide, Altmann & Haywood, 2003) studied participants' eye movements during scene viewing and concurrent spoken language processing by strategically manipulating verb tenses or verb choices in their spoken linguistic stimuli. They recorded participants' eye movements as they listened to sentences containing different verb tenses (e.g. "The man will drink…" or "The man has drunk…") and viewed visual scenes containing objects contrasting on one relevant property (e.g. a full glass of beer and an empty wine glass). They observed that the future tense generated anticipatory saccades towards the full container, whereas the past tense generated anticipatory saccades towards the empty container. Similarly, during viewing of a scene depicting a boy, a cake, and other inedible objects, the onset of the verb in “The boy will eat…” prompted anticipatory eye movements towards the plausible “cake” target, whereas the verb in “The boy will move…” could not disambiguate the target object until the latter was named. These results are significant for a number of reasons. Importantly, they show that language can guide the visual search for an object (or direct attention from a reference object to a target object, such as during the processing of spatial relations), and that the interface between language processing and visual attention can be modulated by prior knowledge of objects’ affordances.

This idea was further explored in an eye tracking study by Coventry, Lynott, Cangelosi, Monrouxe, Joyce and Richardson (2010), who presented their participants with sentences describing spatial relations between top and bottom objects (e.g. “The box is above the bowl”). Each sentence was then followed by a visual stimulus depicting the two relevant objects and manipulating both the spatial and functional relations between them. Spatial manipulations saw one object vertically or horizontally displaced relative to the other one, and either near to or far from it. A third object falling from the top one (e.g. cereal falling from the box) mediated the functional relationship between the two objects. In functional scenes, the falling object was on a trajectory that would cause it to fall into the bottom object. In non-functional scenes, the falling object was on a trajectory that would cause it to miss the bottom object. Control scenes contained no falling objects. For each visual stimulus, six regions of interest were defined. One included the top object itself. Two were located just under the opening in the top object, and the falling object was depicted in either of them depending on the functional relation between top and bottom object. The latter was split into three regions: a central region where the falling object would be expected to land in functional scenes; a near-miss region where the falling object would be expected to land in non-functional scenes; and a far-miss region on the opposite end of the bottom object relative to the top one, where the falling object could not land.

After each sentence-picture pair was presented, participants performed acceptability ratings on a seven-point Likert scale. Globally, functional scenes resulted in shorter total dwell times than non-functional and control scenes. Participants spent significantly longer fixating the centre regions of the bottom objects compared to the near- and far-miss regions. Additionally, the centre regions were fixated for longer when bottom and top objects were near to each other compared to when they were more distant. In other words, participants were found to be drawn to the regions of the objects that would be involved in the end states of functional interactions between them.

These results seem to suggest that the visual inspection of static visual scenes (and language processing) can involve the generation of perceptual simulations (as discussed in Section 1.10) or, at the very least, of simpler analogue mental representations. In line with this and complementing these conclusions, the literature also provides evidence of a significant and more direct involvement of eye movements in imagery processes. A particularly fruitful paradigm in this research area involves studying participants' eye movements to empty space in response to previously presented visual stimuli or to aurally-presented linguistic stimuli. Brandt and Stark (1997) recorded the eye movements of nine participants while they viewed a simple grid-like stimulus containing a number of double black squares, and also while they visualised it after it was removed. While participants were instructed to imagine the grid they had seen, they were told that the goal of the study was to study variations in pupil size during the task. This was to control for any possible task-related demand characteristic confounds by not putting any explicit focus on the replication of eye movements during mental imagery. The results of this analysis showed that repetitive sequences of fixations (i.e. scanpaths) were observed during visual imagery, and that their patterns were closely related to those recorded during visual inspection of the respective stimuli. This was taken as further confirmation of an overlap between perception and imagery. Brandt and Stark also posited that eye movements during imagery could function as a method to scan, inspect, and link together different parts of a visual image, consistent with Kosslyn's (1988) model. However, whether the execution of eye movements is essential to the construction and processing of mental imagery or merely epiphenomenal, and whether they entail a re-enactment of the motor sequences executed during perception, remained unclear.

To address those open questions, in four different experiments Johansson, Holsanova, Dewhurst and Holmqvist (2012) studied naïve participants' eye movements during verbal recall of visual (pictures of complex scenes) or auditory (spoken descriptions of scenes) stimuli during free viewing of a blank screen, and while maintaining a central fixation either during stimulus encoding or during recall. Results showed that, after maintaining a central fixation during encoding of complex visual stimuli, eye movements corresponded to the spatial locations and directions of the objects being described during recall. Additionally, eye movements corresponded to the spatial locations and directions of objects described during recall of a spoken scene description after encoding it with a central fixation. Furthermore, maintaining a central fixation during recall of pictorial stimuli and spoken scene descriptions impaired performance, leading to poorer recall of visually presented stimuli and of their locations. Similarly, Bourlon, Oliviero, Wattiez, Pouget and Bartolomeo (2011) applied eye tracking to the study of mental imagery by asking French participants to imagine a map of France and recall the spatial locations of cities or regions. The task was found to elicit spontaneous saccades towards the locations of the probed target cities, a finding later confirmed by Fourtassi, Hajjioui, Urquizar, Rossetti, Rode and Pisella (2013).

Taken together, the studies presented in this section support the idea that eye movements serve a functional (e.g. Johansson & Johansson, 2013) and, to a degree, representational role during visuospatial mental imagery, whether generated on the basis of prior visual input, long-term knowledge, or linguistic content. Furthermore, these results point to the need for models of both reading and visual stimulus processing to consider important top-down effects. The notion of priority map – introduced earlier in this section – could be particularly important in explaining situations in which eye movements display systematic patterns that cannot be driven by bottom-up salience, such as during language-based imagery tasks. This idea will be discussed further in Chapter 6, where a model integrating language processing, mental imagery, and eye movements will be proposed in more detail and in light of the results obtained from these experiments. Before introducing the first experiment, however, the next and final section of this introductory chapter will provide a review of the few studies that have investigated eye movements during environmental exploration and wayfinding, both in laboratory and more ecologically valid settings.

2.7. Eye Movements during Map Processing and Navigation

In Section 2.6, I reviewed a number of studies that have considered eye movements as important proxies of attentional capture and allocation during language processing, scene viewing, and during mental imagery tasks. However, eye tracking has found application in a number of studies in the field of spatial cognition and wayfinding whose results might also inform this theoretical framework. A few of these studies have focused on correlating eye movements and encoding processes during map viewing. Castner and Eastman (1984; 1985) recorded participants' eye movement (via electrooculography; Castner & Lywood, 1978) as they processed a set of cartographic maps and ranked them as a function of relative visual complexity. The processing of more visually complex maps was found to require longer and more closely spaced fixations, indicating a more effortful construction of a cognitive schema on the basis of the information provided in the map. Since then, eye movements have been used as measures of processing difficulties in studies exploring issues of usability and effectiveness of map interfaces (e.g. Çöltekin, Heil, Garlandini & Fabrikant, 2009), and how these may be influenced by between-subject differences in cartographic expertise (e.g. Ooms, De Maeyer, Fack, Van Assche & Witlox, 2012). Similarly, a small study by Gunzelmann, Douglass and Khooshabeh (2008) used the number of fixations as a measure of learning across several hundred trials in a task that required participant to point to locations within allocentric representations on the basis of egocentric views of the same virtual environment. The environment was a circular space simulating a desert with a visual backdrop depicting mesas and containing random configurations of simple 3D objects. Results showed that increasing familiarity with the task required participants to shift attention between the egocentric and the allocentric view less often (resulting in fewer rounds of fixations).

Other studies of eye movements during wayfinding have focused more on how they reflect attentional allocation to landmarks and other environmental features. While these have been attempted both in real (e.g. Viaene, Ooms, Vansteenkiste, Lenoir & De Maeyer, 2014) and virtual (e.g. Peebles, Davies & Mora, 2007) environments (or in laboratory studies but using images of real environments, e.g. Emo, 2012), the former have been largely aimed at overcoming methodological difficulties or developing practical applications (e.g. pedestrian navigational aids; Kiefer, Straub & Raubal, 2012a; 2012b). On the other hand, studies involving wayfinding in virtual environments have yielded findings that are, so far, potentially more relevant to the research presented in this thesis. For example, studies by Wiener and colleagues have investigated gaze behaviour during spatial decision-making. Wiener, Hölscher, Büchner and Konieczny (2011) presented participants with screenshots of decision points in a virtual environment lacking landmarks and salient visual features in order to isolate the contribution of structural elements of the environment to wayfinding decisions. By allowing participants to freely choose which path to take to reach a target object, and by measuring their recall of path choices after exposure to a pre-set route, two main findings emerged. Participants displayed a significant tendency to choose the path leg offering the longest line of sight during free exploration, and a robust gaze bias towards the chosen path approximately 700 ms before the choice is reported. Crucially, Wiener, de Condappa and Hölscher (2011) also reported this gaze bias when participants were passively guided through a virtual environment containing landmarks on both sides of each decision point, and subsequently presented with screenshots of the various intersections (Figure 2.6). When asked to indicate the direction taken at each decision point during the learning phase, eye movements displayed a similar gaze bias towards the landmarks located in the chosen direction, a few hundred milliseconds prior to the choice being made. This is broadly consistent with findings by Janzen and colleagues (see Section 1.3) of a stronger representation in memory for decision-point landmarks during egocentric navigation. On this basis, we might expect eye movements to reflect the prioritisation of navigationally salient landmarks also during spatial language encoding or map scanning.

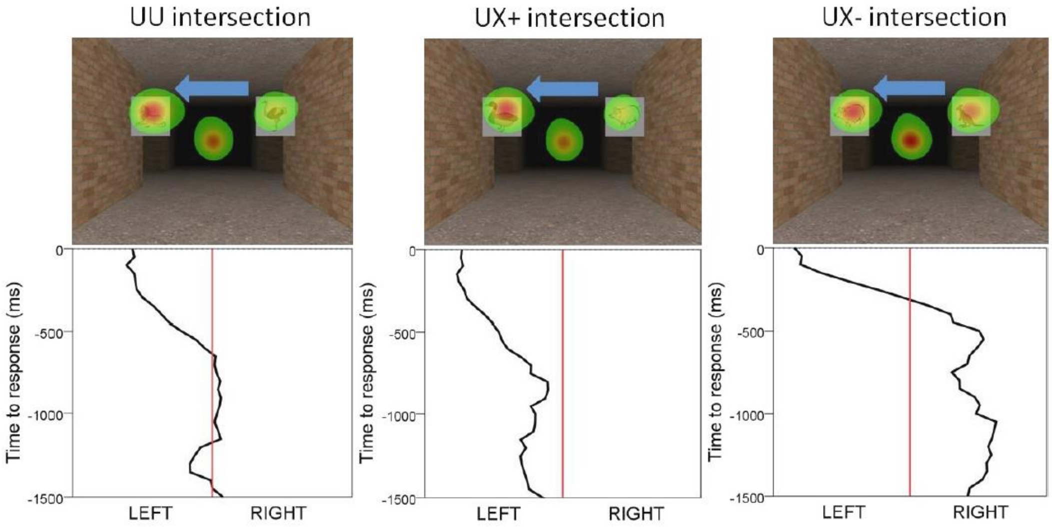


Figure 2.6 – Heatmaps and temporal plots representing the distribution of fixations with a bias towards the decision-point landmark corresponding to the correct turn direction (indicated by the blue arrows in these examples) in the study by Wiener, de Condappa and Hölscher (2011). The blue arrows were not visible to participants.

Accordingly, in Experiments 1-3 in this thesis, eye movements were recorded to ascertain whether this attentional primacy for navigationally salient landmarks would translates also to encoding measures during reading of route descriptions and during a map verification task. Additionally, the same experiments also explored whether this effect would be modulated by the reference frame of the descriptions, by the allocentric format of the maps, and by the imagined spatial perspective maintained during encoding of the spatial information. No published study has, to my knowledge, explored spatial description encoding as a function of reference frame using eye tracking. However, the expectation of top-down modulations was motivated on two grounds. Firstly, by Janzen and colleagues' finding of task demand effects on the allocation of attention to landmark categories (e.g. Wegman & Janzen, 2011). Secondly, by the observation of top-down effects on eye movement control in a number of studies of eye movements, both in reading and in scene processing (see Section 2.6).

Furthermore, Livingston-Lee et al. (2011) also attempted to investigate the interaction between reference frame and eye movements in a virtual environment. They explored the idea that navigational strategies based on egocentric and allocentric reference frames might lead to differences in eye movements during navigation in a virtual water maze. In their experiment, participants had to locate a hidden platform that was either always located in a particular cardinal quadrant of the arena with respect to distal landmarks outside of it that were visible through windows ("Place maze", their allocentric condition), or always in a different location but always marked by the same proximal landmark ("Cue maze", their egocentric condition). Because the distal landmarks in the Place maze were always above the horizon and proximal cues in the Cue maze always below the horizon, eye movements, which were recorded for one second at the beginning of each test trials, allowed to determine the average distribution of fixations during the orienting face in each of the two conditions. Starting from the second test trial, participants tended to fixate the region below the horizontal midline during Cue maze trials in order to identify the location of the cue object, and above the midline during Place trials in order to orient with respect to the distal landmarks (Figure 2.7).

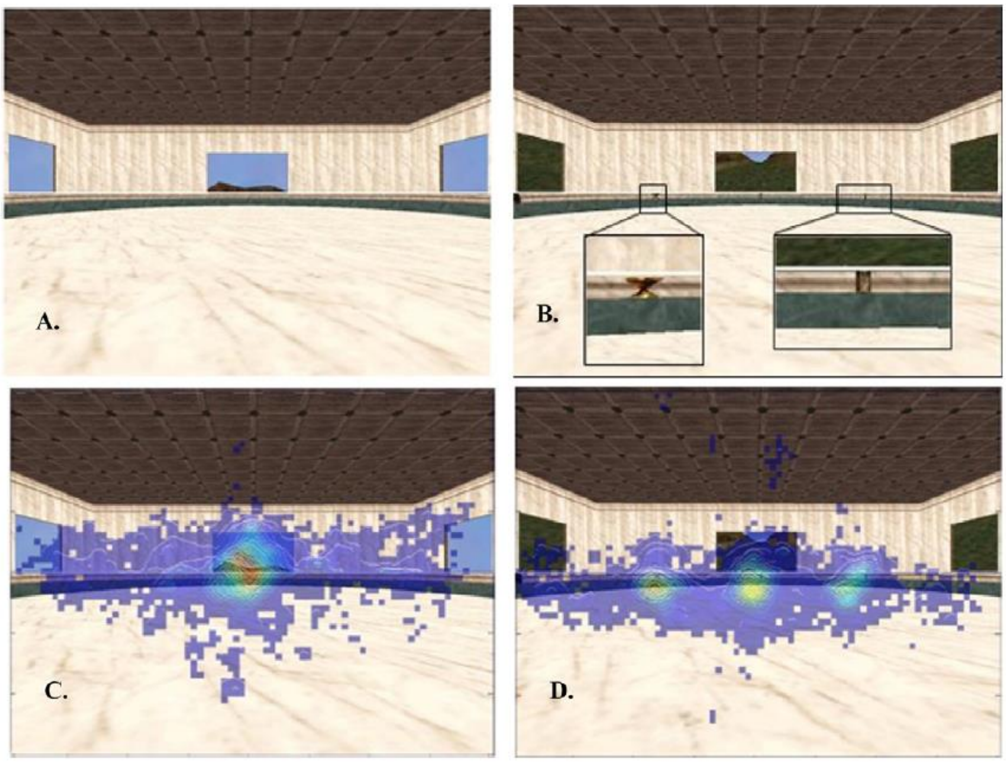


Figure 2.7 – Views of the Place (A) and Cue (B) maze used by Livingstone-Lee et al. (2011), and the mean distribution of fixations during the first second of the test trials in the same mazes (C and D). Note the allocation of attention to the distal cue in the former, in order to identify the target quadrant, and the focus on the proximal landmark cues in the latter.

Although the potential lack of ecological validity is a criticism that has often been levied against studies using virtual environments (e.g. Parsons, 2015; but see van der Ham, Faber, Venselaar, van Kreveld & Löffler, 2015, for a study comparing the validity of different forms of VR navigation as measures of human navigational abilities), these results are nevertheless evidence that different navigational strategies might correlate with different gaze patterns. This appears to also hold true when eye movements during spatial tasks are analysed as a function of participants' own spatial cognitive style (SCS) rather than of task demands. In a recent study, Piccardi, De Luca, Nori, Palermo, Iachini and Guariglia (2016) recorded participants' eye movements as they studied a route on a schematic map of an urban environment, and during a route recall phase. Participants were grouped based on their SCS test scores (Nori & Giusberti, 2006) as Route, Landmark, or Survey navigators, and various eye movement measures were recorded using target (on-route) and non-target (off-route) landmarks, as well as the white space between them, as regions of interest. Beyond obtaining greater recall accuracy scores, Survey navigators were shown to allocate less attention to target landmarks, and more attention to non-target landmarks and the white space between landmarks, than either Landmark or Route navigators. This tendency to produce more diffuse gaze patterns might be a reflection of a general property of more allocentric representations (with all the caveats concerning the limits of the egocentric-allocentric dichotomy discussed in Section 2.3), and could illuminate some of the experimental results presented in this thesis.

While the studies presented in this section are undoubtedly different from the ones conducted and presented in Chapters 3, the evidence presented so far, pointing to an overlap between visual perception, imagery, and eye movements might nonetheless suggest an intriguing possibility. Namely, that adopting a different imagined perspective during the encoding of a route description (whether top-down as a function of task demands or due to linguistic manipulations) will also lead to systematic differences in the allocation of attention to landmark words in the text, and to landmark regions of sketch maps, as reflected by eye movement measures. This hypothesis was tested explicitly in Experiments 1-3.

2.8. Summary

In Chapters 1 and 2 I have introduced a number of seemingly disparate notions drawing from different areas of cognitive investigation, and attempted to draw connections between them. These links will hopefully become clearer in light of the experimental results presented in this thesis, and Chapter 6 will attempt to tie all conceptual loose ends into a more coherent framework. However, in this section I will briefly summarise the key concepts presented so far, in order to better introduce the first experiments I conducted in Chapter 3.

At its core, this research is attempting to explore the processes through which meaning is extracted from a linguistic form (e.g. written text or spoken language) and subsequently used to support a number of tasks. The specific case under study is the processing of route directions in order to carry out spatial-navigational tasks such as drawing or evaluating sketch maps of routes (Experiments 1-3) or performing homing estimates following imagined navigation (Experiments 4-5). However, as discussed in these first two chapters, such research poses significant theoretical and methodological challenges that must be kept in mind when reviewing these experimental results. Chief among them is the fact that mental representations, given their internal nature, cannot be directly studied. Their properties (such as coarseness, vividness, spatial perspectives, topology, etc.) can only be inferred. As such, the experiments described here should be considered first and foremost as proofs of concept for ways to establish more experimental control over the study of mental representations of space, controlling for a few of the sources of variance discussed earlier. Nevertheless, in spite of its partly exploratory nature, this research also aims to actively inform the debate surrounding mental representations in spatial cognition and spatial language. More specifically, the use of eye tracking during language encoding can provide a wealth of information regarding the online processing of linguistic stimuli (and covert processes that would escape introspection) that previous psychological studies have not explored and that the unconstrained production and think-aloud protocols used in cognitive linguistics cannot, by their very nature, provide.

However, the latter tradition of research makes clear the importance of introspective reports to the elucidation of mental representations. Accordingly, Experiments 2-5 attempted to record, albeit to a limited extent, the complex phenomenal and subjective experiences which mental representations bring about. The importance of phenomenology in mental imagery research cannot be overstated: beyond its use as a sanity check to ensure key manipulations are giving rise to qualitatively different mental representations (its key application in this thesis), an in-depth consideration of subjective self-reports – whether in the form of questionnaire responses or unconstrained language production – will become, I would argue, a much more extensive and crucial part of imagery research moving forward. Without reopening the largely fruitless imagery debate, Pylyshyn (2002) was certainly correct in describing mental imagery as a field in search of a theory. To put it simply (although the issue is anything but), research in mental imagery is currently a constellation of different models conceptualising mental representations in different ways (mental images, mental models, perceptual re-enactments, embodied simulations, etc.) but without a coherent unifying framework. Formulating such a framework promises to be a daunting task and is not the goal of this thesis. However, the idea that the different conceptualisations of mental representations (described in Chapter 1) are simply snapshots of different points on a continuum of representational formats is a guiding principle of this research. This will become more obvious in Chapter 4, when moving from an exploration of allocentric representations to the testing of egocentric representations (which bring about a particular set of challenges with respect to task analysis and avoiding perilous assumptions) will make a discussion of embodiment much more salient.

While the limits of participants' ability to introspect and the possibility that self-reports might diverge from behavioural response patterns (e.g. Tenbrink & Salwiczek, 2016) should be considered, collecting self-reports should nevertheless become a norm in research practice for multiple reasons. Not only can they provide a wealth of potentially useful data about the mental processes involved in tasks, helping researchers avoid making assumptions as to the strategy (or, indeed, strategies) being used to carry them out and the underlying representations involved, but they could also be valuable correlates for both behavioural and neuroimaging data in a range of practical applications. These are also crucially important at a time when research is increasingly judged by its ability to apply to some aspect of the real world, and a field such as imagery research can ill afford to be perceived as a lofty intellectual exercise. Fortunately, the ubiquity of imagery in the daily life of most individuals makes it ideally suited not only as a way to study normal cognition and language, but also in exploring a variety of disorders that affect them (e.g. Williams syndrome, Alzheimer's disease) as well as disorders in which certain features of mental imagery might represent (however independent) trait markers of the condition (e.g. schizophrenia)(Aleman, Nieuwenstein, Böcker, & de Haan, 2000; Klein & Moritz, 2014; Oertel et al., 2009; Sack, van de Ven, Etschenberg, Schatz, & Linden, 2005). Ultimately, the possibility that mental representations might exist on a continuum of subjective experiences within normal cognition, and between it and pathological states, reinforces the need to better characterise not only the behavioural patterns produced by individuals but also their phenomenal experiences. Last, but far from least, a clearer understanding of the relationship between specific features of specific mental representations and the neural substrates onto which they rely is a crucial step in developing novel technologies that might harness them, such as human-computer and brain-computer interfaces. I will expand on these research avenues in Chapter 6. Next, however, I will provide a short overview of the first three experiments, presented in Chapter 3.

2.9. Overview of the Experiments

Following from the body of work presented in Chapters 1 and 2, the state of the art in imagery research still poses considerable methodological challenges. The key issue of determining the sorts of mental representations constructed and maintained by participants during spatial tasks has not been conclusively resolved either by examining patterns of behavioural results or by obtaining subjective self-reports. In the former case, the interpretation of results has been limited by fundamental assumptions as to the nature of the tasks involved, while in the latter self-reports have been shown to imply task strategies that did not reflect the patterns seen in the behavioural data. To overcome this impasse, the experiments presented here introduced different ways of constraining a crucial aspect of spatial representations (i.e. their imagined spatial perspectives) by manipulating all key aspects of the task: linguistic input, task instructions, and response modality. In addition, in Experiments 1-3, I complemented task performance and self-report measures with more implicit measures of attention that have recently been shown to be diagnostic of spatial cognitive style (egocentric vs allocentric) during the inspection of maps (see Section 2.7).

The use of eye-tracking measures as proxies of attention allocation during the encoding of written skeletal route descriptions provides a novel way to answer several questions concerning the acquisition of spatial information from language, its use during spatial tasks, and its structuring within egocentric and allocentric representations. In light of Sections 2.2 and 2.3, here it is assumed that, during the processing of such a linguistic form, participants will construct mental representations maintaining some degree of analogy to the state of affairs being described (i.e. the layout of an environment) rather than abstract, symbolic representations (see Section 1.4 to Section 1.10). Similarly, in light of the eye movement research presented in Sections 2.6 and 2.7, we know that factors such as the reference frame set by a task or navigational preferences can modulate the allocation of attention to features of a spatial representations (whether a map or a virtual environment) as reflected by eye gaze patterns. However, whether attention to features of a spatial descriptions (e.g. words describing landmarks) can be modulated in a similar fashion remains unclear.

On this basis, Experiments 1-3 tested a number of related ideas: whether changes in reference frames (see Section 1.2) as expressed by the relational terms in a description, and changes in imagined perspective as a function of task demands, would produce different underlying representations; whether these different representations would involve differences in the allocation of attention to different types of landmarks (see Section 1.3); and whether eye-tracking measures would be sensitive to these changes in attention during both spatial text encoding and (in Experiment 3) during map processing. To my knowledge, no study in the literature has adopted an eye-tracking methodology to explore the online encoding of spatial descriptions, particularly with respect to the possible interaction of reference frames and landmark salience (although studies of spatial language production, using predominantly qualitative analytical methods, have shown a less than straightforward mapping between relational term use and reference frame. See Tenbrink and Salwiczek, 2016). As such, Experiments 1-3 should provide important evidence concerning the nature of the relationship between eye movements and attention, and the possible modulation of this relationship by top-down and linguistic factors.

Beyond complementing the already rich literature on spatial language comprehension and production (see Section 2.3) by providing more accurate online measures of encoding, Experiments 1 and 2 also attempted to correlate both encoding and recall measures to some of the measures of individual differences already established in the literature (see Section 1.11). The recall tasks used in Experiments 1-3 involved the drawing or evaluation (during eye-tracking) of sketch maps. In this sense, these tasks provided allocentric tests of spatial knowledge to test the additional hypothesis that encoding allocentric descriptions would result in better performance during a test phase also requiring an allocentric representation (see Section 2.5). In Chapter 4, Experiments 4-5 tested for this congruency effect in the opposite direction, by testing spatial knowledge using judgements of relative directions within an egocentric reference frame, and investigating whether performance would be higher following encoding within an egocentric reference frame compared to encoding within an allocentric representation. Additionally, Chapter 5 describes three studies conducted on large groups of children in order to better describe the developmental trajectory that leads to the emergence of the ability to prioritise navigationally salient elements of an environments and to transform egocentric representations of experienced environments into allocentric, map-like representations.

More generally, these results should help elucidate the relationship between spatial language and spatial cognition, and the possible role played by various mental imagery formats in interfacing the two. In this sense, this research fits within the overarching goal set by Jackendoff (2012) of producing a theory of spatial understanding that can detail not only how we move in space on the basis of perceptual input, but also how we operate in our surroundings in the absence of these stimuli, how we talk about what we experience, and how we visualise what is conveyed through language. Chapter 6 describes steps towards the formulation of such a model with respect to the specific issues tackled in this thesis: reference frames and landmark representations.

CHAPTER 3
Testing Allocentric Representations

Experiment 1: Relational Terms

3.1. Experiment 1: Introduction

In Chapter 2, I have discussed a number of studies pointing to a complex relationship between mental imagery and eye movements during the encoding of linguistic content and the processing of visual scenes (see Section 2.3). This chapter presents three experiments aimed at elucidating this relationship further. These experiments borrow and revisit part of the paradigm used by Taylor and Tversky (1992), and Tom and Tversky (2012) but with important differences. Tom and Tversky (2012) observed how the quasi-perceptual vividness of route descriptions facilitated the recall of route elements, whether landmarks or path segments, in a map-drawing task. Additionally, they tested for the influence of vividness and individual visuospatial imagery abilities (measured via MRT, MT, and VVIQ scores) on reading time, but did so using a self-paced reading paradigm in which each statement of a description was presented in isolation and a reading time per syllable was computed for each syllable. While correlations were found between reading time measures computed on this basis, measures of imagery abilities, and recall measures, I believe a methodology allowing to measure the allocation of attention to individual words in a spatial descriptions might provide a wealth of information about the underlying mental representations being constructed.

For this reason, in Experiments 1-3 I adopted an eye-tracking methodology to better capture online attentional processes during spatial language comprehension, and to investigate whether these are subject to modulation as a result of key manipulations. More specifically, in Experiment 1 I manipulated both the navigational salience of described landmarks and the reference frame implied by the descriptions. To accomplish this, participants were presented with two types of route descriptions: egocentric and allocentric. The egocentric descriptions expressed spatial relations in terms of "left" and "right," (e.g. "Turn left at the pub."), whereas the allocentric descriptions did so in terms of cardinal relations (e.g. "At the pub, head west."). The use of these two types of relational terms and the formal distinction between egocentric (or route) and allocentric (or survey) texts is in keeping with the rest of the psychological literature in this area (but see Section 2.3 on this matter). In each description type, navigationally salient landmarks were described as being located at locations where a heading change occurred (e.g. "At the bank, head south."), whereas non-salient landmarks were located along path legs (e.g. "Walk past the clinic on your left."). In order to isolate the contribution of different reference frames, I decided to control for description vividness.

Accordingly, I presented participants with simple skeletal route descriptions (Denis et al., 1999; See Section 2.3) containing only plausible urban landmark words and generic path segments between them. The goal was to create descriptions that would generate the simplest possible spatial representations, while not sacrificing ecological validity. Crucially, both description types used here would, within the context of the spatial language literature, be classified as route instructions. That is, both egocentric and allocentric descriptions presented a sequence of motor events aimed at travelling from an origin landmark to a destination landmark. However, by manipulating the type of relational term used to describe direction changes, I tested the hypothesis that the encoding of egocentric and allocentric descriptions would result in the construction of mental representations with distinct imagined spatial perspectives. I therefore predicted that egocentric relational terms should engender phenomenologically egocentric spatial representations, while allocentric (cardinal) relational terms should engender phenomenologically allocentric spatial representations. That is, while egocentric descriptions should be encoded as the participant imagines walking through the (however schematic) environment from a first-person view, allocentric descriptions should encourage readers to adopt a more survey-like perspective, or otherwise switch between reference frames at every directional turn in order to update their egocentric perspective relative to the global environment.

Although no other study has, to my knowledge, employed eye-tracking to study the encoding of spatial descriptions and the factors that influence it, studies of eye movements in other areas of spatial cognition served to inform my predictions to an extent (see Sections 2.5 and 2.6). For example, studies of gaze bias during egocentric navigation by Wiener et al. (2011b,c) would suggest that, if a mental representation is being generated and explored egocentrically during encoding of an egocentric route description, more attention might be allocated to decision-point landmarks than to non-decision-point landmarks (see studies by Janzen and colleagues in Section 1.3). If, however, both egocentric and allocentric representations are being computed in parallel during encoding of both types of descriptions (see Section 1.2), it is possible reading patterns might not differ significantly between conditions. Accordingly, several eye-tracking measures of reading behaviour were used as dependent variables to investigate whether different linguistic reference frames would result in different mental representations, and whether eye movement patterns would vary as a result. The measures chosen (total dwell time – also known as total reading time – and regression path duration) related to the total time spent fixating regions of interest (e.g. landmark words) and the total time spent re-reading previous portions in the text. These late measures are thought to better reflect complex cognitive processes active at the level of discourse integration (e.g. Rayner, 1998). By extension, since the construction of a spatial representation from linguistic input is a complex process that requires the integration of information throughout the encoding stage, these late measures should also allow to observe the effects of a reference frame manipulation. However, to determine whether such effects are present also during early processing, dwell time was also decomposed into distinct rounds of fixations to disentangle early and later stages of discourse integration (see Section 3.3.2).

With respect to behavioural performance, Experiments 1-3 tested participants' spatial knowledge of the described environment by requiring them to draw or judge the accuracy of allocentric external representations (i.e. sketch maps). A number of different dependent measures were recorded in these tasks, ranging from participants' ability to correctly recall and place salient and non-salient landmarks (in Experiments 1 and 2), to their detection of incorrect turns (in Experiment 3). This design relies on the encoding-test congruence facilitation observed in a number of studies (see Section 2.5) to determine whether encoding route descriptions containing different relational terms does lead to the creation of mental models within distinct spatial reference frames. That is, while egocentric representations need to be transformed into map-like mental representations, allocentric representations are thought to already entail (provided encoding is successful) a map-like understanding of the environment. Therefore I hypothesised that, if allocentric encoding is achieved, performance in the recall (as measured via map drawing scores) of descriptions containing cardinal relations should be on par with, if not higher than, the recall performance of egocentrically-encoded descriptions (which should generate egocentric representations that would require a transformation during test). If, on the other hand, cardinal terms fail to automatically elicit an allocentric representation, possibly due to a natural tendency to encode sequences of spatial instructions egocentrically, task performance following allocentric (cardinal) encoding might be lower than in the egocentric condition due to the interference of competing reference frames. Lastly, if participants can successfully generate both egocentric and allocentric representations in parallel on the basis of sparse linguistic input, then performance in the map drawing task might not differ between conditions.

An additional goal of Experiments 1 and 2 was to investigate the relative contribution of different cognitive abilities to spatial language encoding and performance in map-drawing tasks. More specifically, the question to be addressed was whether building a representation of a large-scale urban environment (an instance of environmental space as defined by Montello and colleagues. Montello, 1993; Montello & Golledge, 1999) but on the basis of a route description, would depend more on small-scale spatial abilities (e.g. the mental rotation of the environmental model as a single object in figural space) or on large-scale spatial abilities (e.g. Hegarty et al., 2006) (see Section 1.11). Accordingly, participants were given a battery of psychometric measures of their spatial abilities at these different scales and of working memory. These are listed in Section 3.2.2 below.

3.2. Methods

3.2.1. Participants

A total of 24 adult English native speakers with normal or corrected-to-normal vision (7 males, 17 females, mean age 20.62 ± 2.42 years) recruited across the University of Nottingham participated for credits or in exchange for an inconvenience allowance.

3.2.2. Design and Materials

Two descriptions provided sequential instructions for how to navigate two distinct urban environments. The instructions informed the participant as to when to make a turn or keep walking forward. A total of six landmarks were distributed along the route as follows: one origin, one destination, two at salient points (changes in directions), two at non-salient points. As an added spatial dimension used to score the maps, each non-salient landmark could appear either on the left or on the right side of the road. Each route contained a total of four turns (two associated with landmarks and two at spatial locations where no landmark was present) and four path segments.

The origin and destination landmarks differed between conditions to induce, as much as possible, the generation of a different spatial model and avoid confusion. The other four landmarks were common between the egocentric and the allocentric descriptions, to allow for comparisons between conditions, particularly with respects to eye movement measures, but in different orders (see Appendix I). Furthermore, two versions of each description were prepared, to provide additional counterbalancing of the landmark presentation. Each description was presented on two pages, each containing four lines of text. See Table 3.1 below for examples of each description type.

Together with the main experimental task, we administered a set of standardised measures to assess a variety of abilities. This was motivated by the observation in the literature of significant individual differences during spatial language encoding and recall (see Section 2.3), and by the need to determine the relative contribution of spatial abilities at different scales (see Section 1.11) to the processing of spatial text and to the construction of spatial mental representations on that basis. The measures used in this study were:

- the **Mental Rotation Test A** (MRT-A) (Vanderberg & Kuse, 1978), a measure of the ability to mentally rotate abstract shapes, associated with small-scale spatial abilities (see Section 1.11). If imagined environments built on the basis of linguistic descriptions are treated as figural objects, then this measure of mental rotation abilities might predict performance in an allocentric test.

- the **Money's Standardised Road-Map Test of Direction Sense** (Money's Test or MT) (Money, Alexander, & Walker, 1965), a measure of directional sense, or left-right orientation relative to the direction of movement. The MT involves following a path traced onto an allocentric map and indicating, for each turn in the path, the egocentric direction taken (writing "L" or "R" next to it). Because of the meandering nature of the route, sections of the path follow directions misaligned with the viewer's up-forward orientation. The participant is not allowed to physically rotate the map, and the test is thought to measure the ability to perform egocentric mental rotations (e.g. Vingerhoets, Lannoo & Bauwens, 1996; Uchiyama, Mitsuishi, & Ohno, 2009). Hegarty and Waller (2004) found the MT to be a reliable measure of egocentric perspective taking abilities distinct from (albeit partly overlapping with) mental rotation abilities (e.g. MRT scores). Tom and Tversky (2012) found MT scores to positively correlate with participants' ability to correctly recall the spatial location of landmarks (see Section 2.3). However, given the nature of the test, the MT could also be considered a measure of the ability to translate between allocentric (i.e. the map-like representation) and egocentric (i.e. the imagined perspective to be assumed during turns) representations and vice versa. If participants encode an egocentric description differently from a cardinal description, and if an egocentric mental representation must be transformed into an allocentric representation during test, then participants' performance might be in part predicted by their MT scores.
- the **Santa Barbara Sense of Direction Scale** (SBSOD) (Hegarty et al., 2002), a self-report questionnaire containing questions concerning an individual's everyday wayfinding abilities, preferences and habits during navigation, ease of spatial information processing. This measure has previously been associated with spatial abilities in large-scale environmental navigation (see Section 1.11).
- the **Digit Span Test**, both forward (DFW) and backward (DBW) (Blackburn & Benton, 1957), measures of working memory storage and processing respectively.

3.2.3. Apparatus

Eye tracking was performed with an SR Research Eyelink 1000 desktop-mounted eye tracking system sampling at 1000 Hz. Viewing was binocular but only one eye was recorded. Each recording session began with a 9-point calibration and validation sequence. Stimuli were presented on a computer screen, in black font (Courier New, 18pt) on a white background, located 56 cm from the participant. Participants used a Microsoft Sidewinder USB gamepad to advance through the description pages. They were advised to look away from the text towards a post-it note attached to the frame of the screen before pressing the right trigger button. This was done so as to avoid artefactual fixations at the end of the reading phase. The appearance of each page of text was preceded by the appearance of a drift correction marker at the location of the first letter in the page. Both overlay images and eye movements were presented to the experimenter on the Host PC so that feedback could be provided.

Table 3.1 - Examples of egocentric and allocentric route descriptions as used in Experiment 1. The colour coding (shown here for explanatory purposes but not presented to participants) indicates the regions of interest investigated in the analysis of eye movements (as well as the items to be recalled by participants during the map-drawing task). Green indicates navigationally salient landmarks (i.e. turn-location landmarks) and the relational terms indicating the direction of the respective turns. Orange indicates non-salient landmarks and relational terms signalling heading changes where no landmark was present. Light blue indicates origin and destination landmarks. The double-line in the table separates the content of the first and second page of text shown to participants.

EGOCENTRIC DESCRIPTION	**ALLOCENTRIC DESCRIPTION**
Leave the house.	Leave the train station.
Turn left at the pet store.	Take the second road heading east.
Take the second right.	At the bank, head south.
Walk past the gym on your left.	Walk past the pub on your right.
Turn right at the bank.	Take the first road heading west.
Take the second left.	Walk past the gym on your left.
Walk past the pub on your right.	At the pet store, head south.
You have reached the cinema.	You have reached the town hall.

3.2.4. Procedure

Participants were seated in the experimental space and a randomised experimental sequence was selected. Each sequence contained a different order of descriptions and psychometric measures. While the two readings of each description occurred sequentially, the overall order of the experimental phases was randomised (e.g. “MRT-A, Egocentric description (x2), MT, SBSOD, DFW, Allocentric description (x2), DBW” vs “MT, SBSOD, DBW, Allocentric description (x2), Egocentric description (x2), DFW, MRT-A”). At the start of the session, participants were guided through an instruction script that explained the form of the experiment and instructed participants on how to use the gamepad to progress through the description. Calibration of the eye tracker was performed at the beginning of each script and drift correction before each page of text appeared.

During the reading sections, participants read each description twice in a row, drawing a map immediately after each reading on different pre-made templates bearing START and END markers. The templates were standard sheets of white A4 paper indicating participant number and type of description. The START and END markers were positioned at the locations of the start and end points of the original map designs from which the descriptions were derived. They were also located so as to make full use of the available drawing area. The decision to use a new template for each drawing was motivated by the need to maximise clarity and better monitor progress between drawings.

Participants drew their maps using standard felt-tip pens. A black pen was used for the first version of each map, and a red one for the second version. While participants could not erase lines directly, they were allowed to cross incorrect items off their maps during the corresponding drawing session. Drawing time was not constrained, and the experiment progressed when participants felt they were satisfied with their drawing. The first version of each map was visible for reference while participants drew the second one. Participants were informed prior to the first encoding phase that a map-drawing task would follow each reading.

For the MRT-A, one point was assigned when both target figures were correctly identified. Three minutes were given to complete the first two pages, followed by a two-minute break and by three more minutes to complete the last two pages. For the Money's Test, participants were given 45 seconds to complete as much of the path as they could, and scoring was performed by subtracting the number of incorrect or missing turns from the number of correct turns.

3.3. Results

3.3.1. Map Data

I began by coding the maps each participant had drawn after each reading on several measures:

- Number of landmarks recalled overall (out of 6)
- Number of landmarks drawn in the longest uninterrupted correct sequence (out of 6)
- Number of landmarks drawn in their correct spatial locations (out of 6)
- Number of overall correct turns (out of 4)
- Number of turns drawn in the longest uninterrupted correct sequence, i.e. without incorrect turns deviating from the described path (out of 4)
- Number of ordinal information tokens recalled (out of 2)
- Number of locations anchored to the START and END markers on the template (out of 2)

Each raw score was converted into a percentage. Raw scores were then also aggregated into two main measures as follows:

- Landmark Knowledge (LK): overall landmark recall, sequentially correct landmarks, spatially correct landmarks. (out of 18)
- Configural Knowledge (CK): overall turn recall, sequentially correct turns, ordinals recalled, correct anchoring points. (out of 12)

These were then also converted into percentages. See Figure 3.1 and Figure 3.2 for examples of the types of sketch maps drawn by participants.

Overall Landmark Recall

Overall landmark recall was analysed via a set of 2(Relational term type: allocentric vs egocentric) x 2(Landmark salience: Salient vs Non-Salient) x 2(Reading: 1st vs 2nd), Sidak-corrected within-subject ANOVAs. Significant main effects of Relational term type, $F(1,23) = 6.41$, $p = .019$, $\eta^2_p = .218$, and Reading, $F(1,23) = 47.44$, $p < .001$, $\eta^2_p =$

.674, were observed. No significant interaction between them was found, $F(1,23) = 2.04$, $p = .166$, $\eta^2_p = .082$. Landmark recall appeared to be higher in the egocentric condition, M = 74.47%, SEM = 3.40%, than in the allocentric condition, M = 64.58%, SEM = 4.85%, and generally higher after a second description encoding than after the first, M = 84.37%, SEM = 3.70% and M = 54.68%, SEM = 4.80% respectively. Additionally, a marginally significant effect of Salience was observed, $F(1,23) = 3.53$, $p = .073$, $\eta^2_p = .133$, indicating better recall of navigationally salient landmarks, M = 74.47%, SEM = 4.15%, compared to non-salient landmarks, M = 64.58%, SEM = 4.91%. However, this effect was likely limited by the low number of landmarks in the descriptions (Figure 3.3). For this reason, Landmark salience was not used as a factor in subsequent analyses of map-drawing performance. The effect of Landmark salience during test was explored in more detail in Experiment 3.

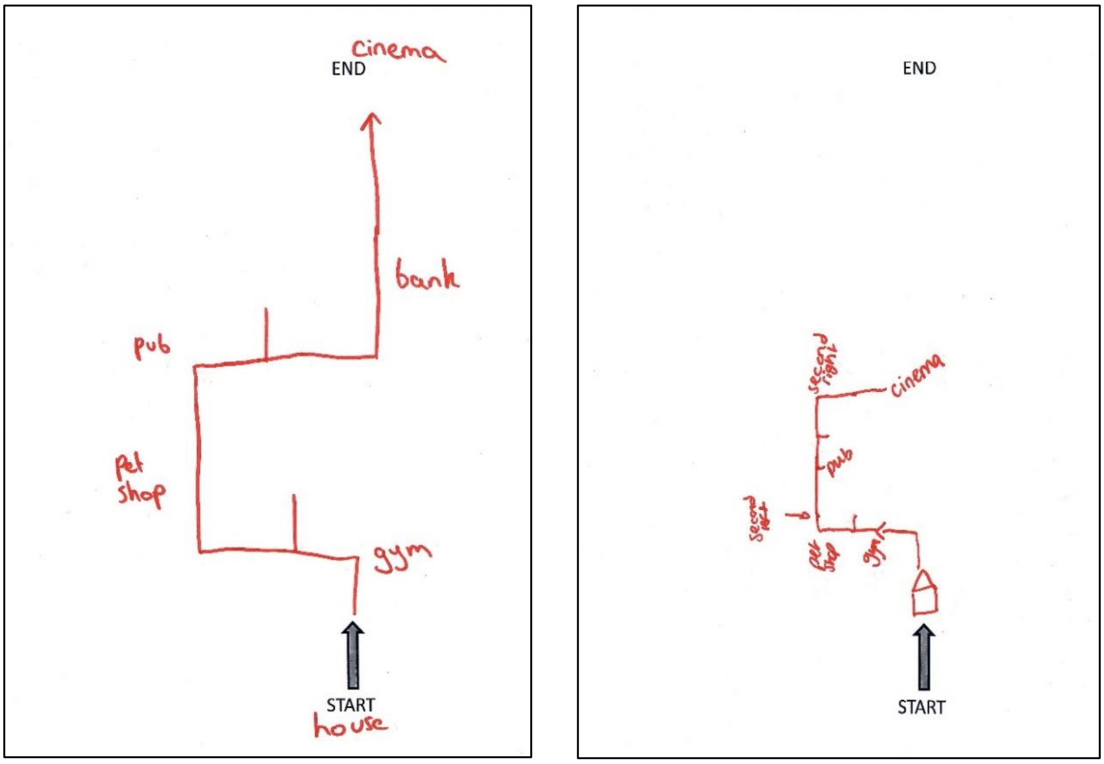


Figure 3.1 – Two examples of sketch maps drawn by participants following encoding of egocentric descriptions. Left: an accurate representation of its corresponding route description, with correctly recalled and positioned landmarks, ordinal information referring to side roads, and with a route correctly anchored to both the START and the END markers provided. Right: a sketch map of the same route, but with incorrectly located landmarks (e.g. gym), a missing landmark (i.e. bank), and with an unanchored end location.

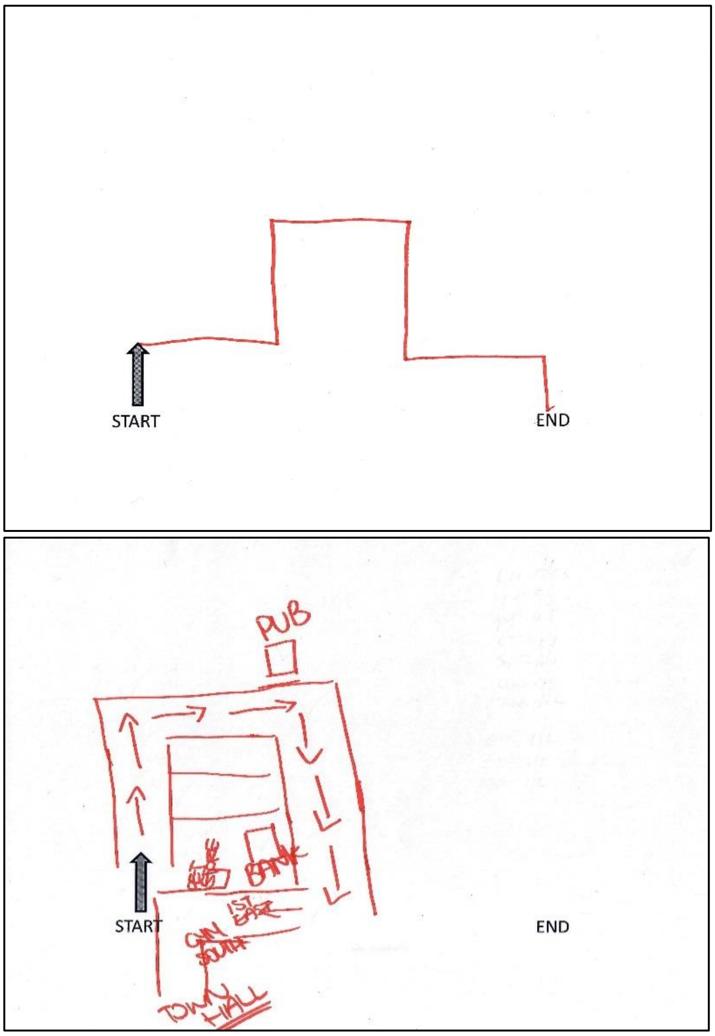


Figure 3.2 – Two examples of sketch maps drawn by participants following encoding of allocentric descriptions. Top: a poor map, with all landmarks missing and incorrect turns. Bottom: a reasonably accurate map, preserving the correct sequence of directional turns, but missing a starting landmark and failing to connect to the end marker.

Sequentially Correct Landmarks

The number of landmarks (expressed as percentage) recalled in the longest uninterrupted correct sequence (i.e. number of landmarks drawn in the order in which they are mentioned in the respective description, regardless of whether the spatial relations are preserved) was then analysed via a 2(Relational term type: allocentric vs egocentric) x 2(Reading: 1st vs 2nd) Sidak-corrected within-subject ANOVAs. A borderline significant trend for an effect of Relational term was found, $F(1,23) = 3.15$, $p = .089$, $\eta^2_p = .120$, with better recall following egocentric description encoding, M = 52.40%, SEM = 4.44%, compared to cardinal description encoding, M = 43.36%, SEM = 5.58%. A significant main effect of Reading was found, $F(1,23) = 50.28$, $p < .001$, $\eta^2_p = .686$, with better recall following the second encoding compared to the first, M = 63.16%, SEM = 5.67%, and M = 32.60%, SEM = 3.88% respectively. The interaction between the two factors was not significant, $F(1,24) = .053$, $p = .820$, $\eta^2_p = .120$. See Table 3.2 for sequential recall performance scores.

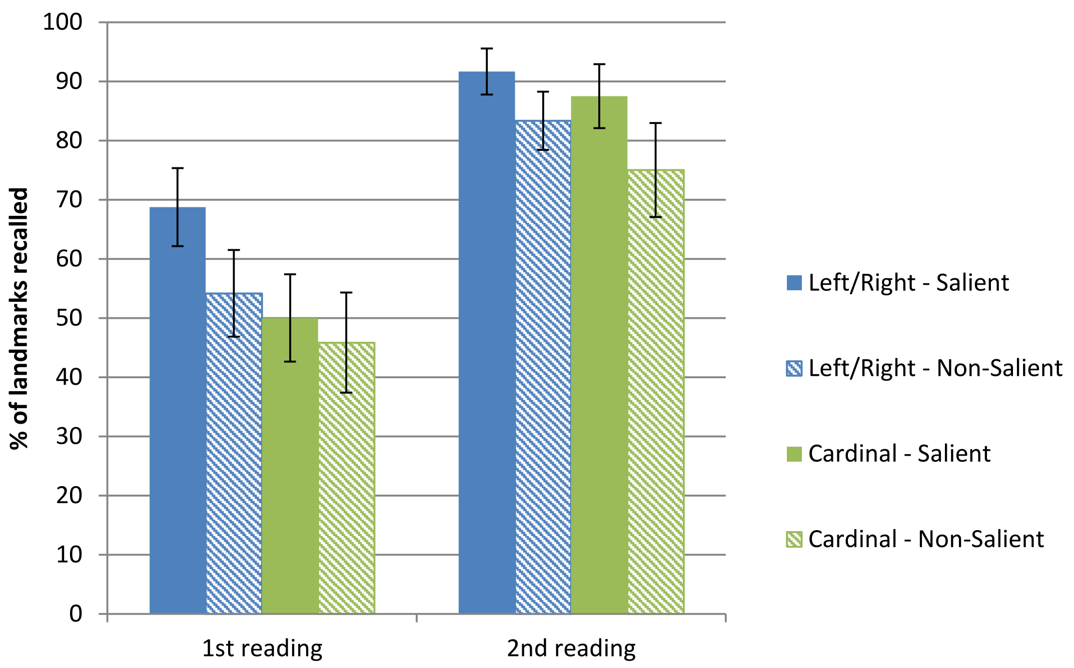


Figure 3.3 - Average landmark recall in the map-drawing task as a function of landmark salience. Error bars represent ±1 SEM.

Table 3.2 - Percentage of landmarks recalled in the correct sequence in the two conditions.

Relational Term	Reading			
	First		Second	
	M	SEM	M	SEM
Egocentric	37.46	4.28	67.33	5.89
Allocentric	27.74	5.37	58.98	6.88

Spatially Correct Landmarks

The number of landmarks (expressed as percentage) recalled in their correct spatial locations (i.e. at salient vs. non-salient locations; on the left- vs. right-hand side of the road) was also analysed via a 2(Relational term type: allocentric vs egocentric) x 2(Reading: 1st vs 2nd) within-subject ANOVA. No significant main effect of Relational term type was found, $F(1,23) = 1.80$, $p = .193$, $\eta^2_p = .073$. A significant main effect of Reading was found, $F(1,23) = 38.55$, $p < .001$, $\eta^2_p = .626$, indicating better recall performance following the second encoding, M = 62.82%, SEM = 5.28%, compared to the first, M = 38.49%, SEM = 4.75%. The interaction between Relational term type and Reading was not significant, $F(1,23) = 3.36$, $p = .080$, $\eta^2_p = .128$ (Figure 3.4).

Overall Turn Recall and Sequentially Correct Turns

I then analysed recall performance for the directional turns encountered in the descriptions. Relational term type had a borderline significant effect on the number of turns drawn in the correct direction as indicated in the descriptions, $F(1,23) = 3.59$, $p = .071$, $\eta^2_p = .135$. Directional turn recall was higher following egocentric description encoding, M = 84.37%, SEM = 4.27%, than following cardinal description encoding, M = 70.31%, SEM = 5.81%. Reading was also found to have a main effect on the percentage of correct turns recalled overall, $F(1,23) = 5.11$, $p = .033$, $\eta^2_p = .182$, with better recall following the second encoding of a description, M = 82.81%, SEM = 4.24%, compared to the first, M = 71.87%, SEM = 4.27%. The interaction between them was not statistically significant, $F(1,23) = .35$, $p = .558$, $\eta^2_p = .015$. As for the percentage of turns drawn in the longest uninterrupted correct sequence, this was found to be marginally affected by Relational term type, $F(1,23) = 3.98$, $p = .058$, $\eta^2_p = .148$ (M = 82.29%, SEM = 4.87% recall following egocentric encoding and M = 64.58%, SEM = 6.41% recall following allocentric encoding). It was also found to be marginally affected by Reading, $F(1,23) = 3.85$, $p = .062$, $\eta^2_p = .144$ (M = 68.75%, SEM = 4.12% recall following the first reading and M = 78.12%, SEM = 4.46% recall following the second reading). No interaction was observed in this instance, $F(1,23) = .05$, $p = .817$, $\eta^2_p = .002$. See Figure 3.5 for turn recall performance.

Ordinal Information

The descriptions provided included ordinals in the form of sequential turn information that was navigationally salient (e.g. "Take the **second** left" or "Take the **first** road heading south."). These denoted changes in direction where no landmarks were present, and were considered to be recalled if the participants drew a secondary path segment branching out of the main segment before or after the turn to be taken, depending on the description, or wrote the words "first" or "second" to identify turns. The ability of participants to recall this type of information was significantly modulated by Relational term type, $F(1,23) = 23.74$, $p < .001$, $\eta^2_p = .508$, with better recall when egocentric relational terms were used, M = 73.95%, SEM = 5.92%, compared to when cardinal terms were used, M = 40.62%, SEM = 5.79%. Recall was also significantly improved after the second encoding compared to the first drawing, $F(1,23) = 6.27$, $p = .020$, $\eta^2_p = .214$, M = 63.54%, SEM = 5.20% and M = 51.04%, SEM = 5.52% respectively. The interaction between the two factors was not significant, $F(1,23) = 2.379$, $p = .137$, $\eta^2_p = .094$. See Figure 3.6.

Anchoring Points

Participants drew their maps on templates bearing a START and an END mark. While these were intended to prompt participants to make use of the full A4 sheet provided so that the maps could be easily interpreted, I also included their ability to anchor the first and last landmark in the route described onto these two points in the analysis (see Gieseking, 2013). This ability was found to only be affected by Reading, $F(1,23) = 8.36$, $p = .008$, $\eta^2_p = .267$, M = 75.00%, SEM = 5.21%, and M = 83.33%, SEM = 4.16% after the first and second encoding respectively. No significant main effect of Relational term type was found, $F(1,23) = 2.24$, $p = .148$, $\eta^2_p = .089$, and no significant interaction between the two factors, $F(1,23) = 1.64$, $p = .213$, $\eta^2_p = .067$.

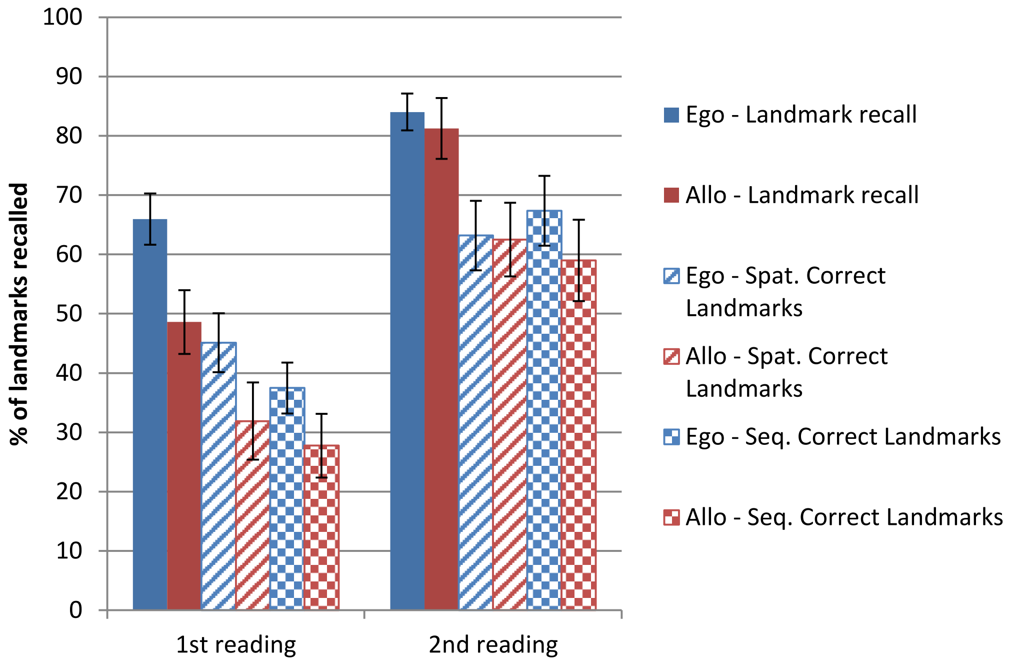


Figure 3.4 - Average performance on the various landmark recall measures in the map-drawing tasks. Error bars represent ±1 SEM.

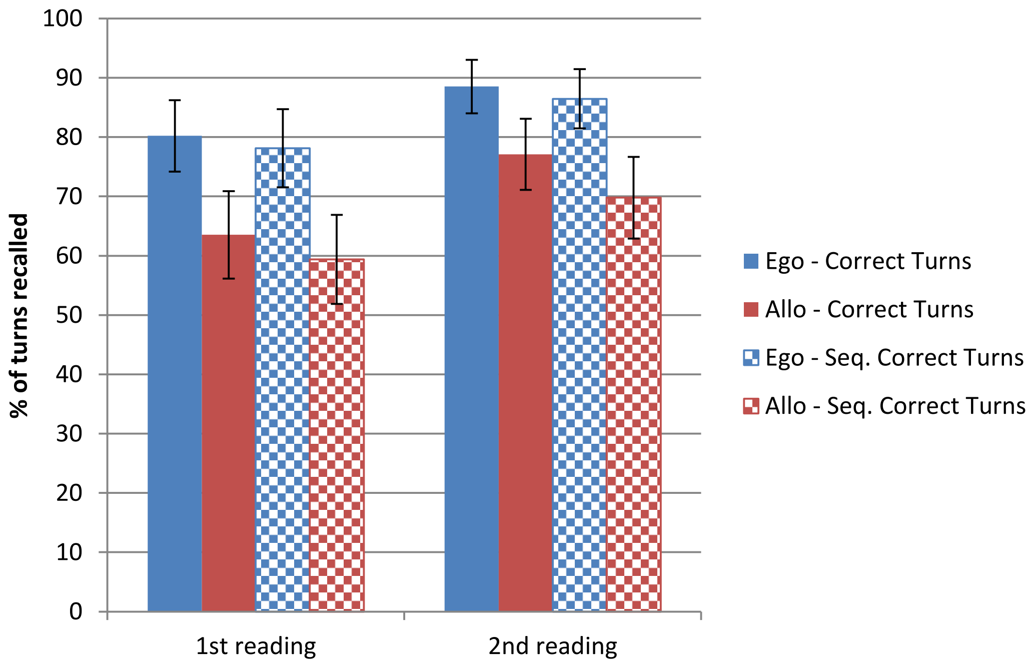


Figure 3.5 - Average performance on the two turn recall measures in the map-drawing tasks. Error bars represent ±1 SEM.

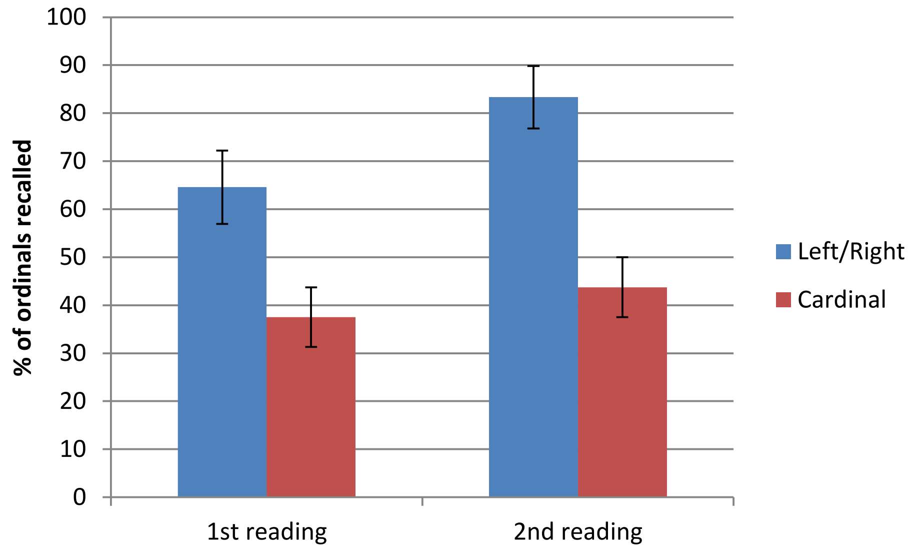


Figure 3.6 - Average performance on ordinal information recall. Error bars represent ±1 SEM.

Aggregated Map Scores

After computing aggregate measures of Landmark (LK) and Configural Knowledge (CK) as described in Section 3.3.1, we analysed them in a similar way in a 2(Relational term type: allocentric vs egocentric) x 2(Reading: 1st vs 2nd), Sidak-corrected within-subject ANOVA. Participants' LK was significantly influenced both by Relational term type, $F(1,23) = 5.57$, $p = .027$, $\eta^2_p = .195$, and by Reading, $F(1,23) = 63.47$, $p < .001$, $\eta^2_p = .734$. Overall landmark knowledge was better following egocentric description encoding, M = 60.99%, SEM = 3.77%, than following allocentric description encoding, M = 51.85%, SEM = 4.93%, and higher following the second encoding of a description compared to the first, M = 69.56%, SEM = 4.48%, and M = 43.28%, SEM = 4.06%, respectively. A marginally significant interaction between the two factors was also found, $F(1,23) = 4.14$, $p = .054$, $\eta^2_p = .153$. Participants' CK was similarly influenced by Relational term type, $F(1,23) = 8.16$, $p = .009$, $\eta^2_p = .262$, with higher scores in the egocentric condition, M = 81.59%, SEM = 3.77%, than in the allocentric condition, M = 64.23%, SEM = 4.72%. A significant main effect of Reading was also found, $F(1,23) = 8.62$, $p = .007$, $\eta^2_p = .273$, with greater performance during the second test phase compared to the first, M = 77.95%, SEM = 3.65%, and M = 67.88%, SEM = 3.26% respectively. No significant or marginally significant interaction between the two factors was found, $F(1,23) = .011$, $p = .917$, $\eta^2_p < .001$. See Figure 3.7.

Psychometric Measures

Table 3.3 includes average scores on the psychometric tests used. A round of two-tailed exploratory correlations was carried out on the psychometric tasks to check for internal consistency. Significant correlations were found between several of them, revealing broad agreement between the measures: MRT was found to correlate significantly with the SBSOD, $r(22) = .413$, $p = .045$, and even further with the MT, $r(22) = .627$, $p = .001$. Performance on the SBSOD also correlates significantly with the DBW, r(22) = .433, p = .035, which was also found to correlate significantly with MT performance, r(22) = .501, p = .013, and, unsurprisingly, DFW scores, r(22) = .414, p = .044. SBSOD did not significantly correlate with MT, r(22) = .350, p = .093.

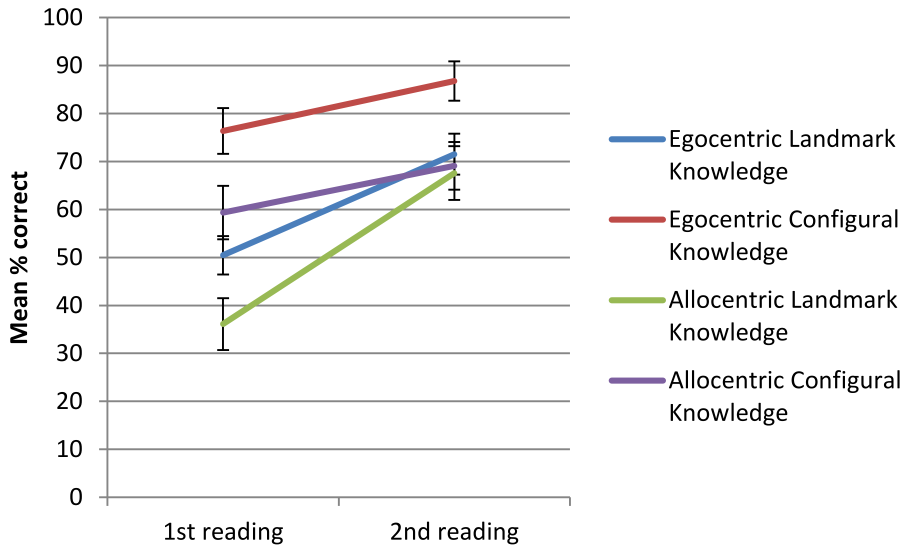


Figure 3.7 - Average Landmark and Configural Knowledge scores for all conditions and readings. Error bars represent ±1 SEM.

Table 3.3 - Average psychometric scores.

Psychometric	Average Score	SEM
MRT-A	9.58	1.09
SBSOD	61.42	3.33
Money's Test (MT)	8.75	2.69
Digit Span FW	13.21	.434
Digit Span BW	8.79	.514

Psychometric Measures and Aggregated Map Scores - Regressions

The predictor variables (i.e. psychometric scores) were then entered in a stepwise multiple regression model for each dependent variable (participants' LK and CK scores), in order to attempt to identify the best predictors of performance in the map drawing task. See Table 3.4 and Table 3.5 for a summary of the results. The primary goal of these analyses was to determine whether performance in map-drawing of imagined urban environments would be better predicted by measures of small-scale (e.g. MRT-A) or environmental-scale (e.g. SBSOD) spatial abilities (see Section 1.11). Significant predictors of each dependent variable are reported in bold. Only standardised coefficients are reported for independent variables that did not enter the model.

*Table 3.4 – Significant psychometric predictors of Landmark Knowledge for each reading in each condition. MRT-A = Mental Rotation A; SBSOD = Santa Barbara Sense of Direction Scale; MT = Money's Test; DFW = Digits Span Forward; DBW = Digits Span Backward. *p<.05; **p<.005.*

	Egocentric 1 - LK			Egocentric 2 - LK			Allocentric 1 - LK			Allocentric 2 - LK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
MRT-A	-	-	.02	-	-	-.08	-	-	-.24	-	-	.26
SBSOD	-	-	-.01	-	-	.03	-	-	.00	-	-	.11
MT	-	-	.25	**.88**	**.27**	**.56****	**.85**	**.38**	**.43***	-	-	.22
DFW	-	-	-.14	-	-	-.14	-	-	-.16	-	-	-.30
DBW	**3.44**	**1.48**	**.44***	-	-	.13	-	-	-.15	**10.44**	**2.60**	**.65****

*Table 3.5 - Significant psychometric predictors of Configural Knowledge for each reading in each condition. MRT-A = Mental Rotation A; SBSOD = Santa Barbara Sense of Direction Scale; MT = Money's Test; DFW = Digits Span Forward; DBW = Digits Span Backward. *p<.05; **p<.005. No variables were entered into the model at step 1 of the regression for CK following the first egocentric reading.*

	Egocentric 1 - CK			Egocentric 2 - CK			Allocentric 1 - CK			Allocentric 2 - CK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
MRT-A	-	-	-	-	-	.00	**2.51**	**.94**	**.49***	**1.92**	**.79**	**42***
SBSOD	-	-	-	-	-	.05	-	-	.03	-	-	-.20
MT	-	-	-	**.84**	**.26**	**.55****	-	-	.27	-	-	.14
DFW	-	-	-	-	-	.15	-	-	.17	-	-	-.18
DBW	-	-	-	-	-	.21	-	-	.28	**3.88**	**1.67**	**.40***

3.3.2. Eye movement Measures

Several eye movement measures were extracted from the raw eye tracking data during the description encoding phase, and analysed via a set of 2(Relational term type: allocentric vs egocentric) x 2(Salience: Salient vs Non-Salient) x 2(Reading: 1^{st} vs 2^{nd}) within-subject ANOVAs. Because the goal was to understand the influence of the reference frame manipulation (i.e. Relational term type) on the global understanding of the spatial descriptions, I extracted late measures of discourse-level processing referring to the entire encoding phase (see Section 3.1). These were average Total Dwell Time (DT), the sum of all fixation durations on words of interest (e.g. landmark words) across the entire reading period, and average Regression Path Duration (RPD), the total time spent re-fixating previously read words in the text before moving to the next word on the right. However, DT was also decomposed into average First-, Second-, Third-Pass Dwell Time, reflecting word processing at different stages of discourse integration. Similarly, the average number of Regressions from (Regression Out) and towards (Regression In) areas of interest was computed and analysed. The areas of interest chosen were landmark words and relational terms (see Table 3.1).

Total Dwell Time

Analysing average DT on landmark words revealed no significant main effect of Relational term type, $F(1,19) = .02$, $p = .869$, $\eta^2_p = .001$. A significant main effect of Landmark salience was found, $F(1,19) = 9.95$, $p = .005$, $\eta^2_p = .344$, indicating that navigationally salient landmarks were fixated longer than non-salient landmarks, M = 1785 ms, SEM = 188 ms, and M = 1153 ms, SEM = 119 ms, respectively. A significant main effect of Reading was also observed, $F(1,19) = 4.69$, $p = .043$, $\eta^2_p = .198$, showing an increase in total fixation time on landmark words in the second reading compared to the first one, M = 1625 ms, SEM = 161ms, and M = 1312 ms, SEM = 119 ms, respectively. A marginally significant interaction between Relational term type and Salience was also found, $F(1,19) = 4.19$, $p = .055$, $\eta^2_p = .181$. A test of simple main effects was performed to further investigate this interaction. Salient landmarks were found to generate significantly longer dwell times, M = 1944 ms, SEM = 253 ms, than non-salient landmarks, M = 961 ms, SEM = 130 ms, in the egocentric condition, $F(1,19) = 14.28$, $p = .001$, $\eta^2_p = .429$. However, this was not the case in the allocentric condition, M = 1626 ms, SEM = 244 ms, and M = 1344 ms, SEM = 161 ms, respectively, $F(1,19) = 1.11$, $p = .304$, $\eta^2_p = .055$. No other two- or three-way interaction was found to be significant, all ps > .05 (Figure 3.8).

The same analysis run on the average DT on relational terms revealed no significant main effects of Relational term type, $F(1,20) = .54$, $p = .469$, $\eta^2_p = .027$, Salience, $F(1,20) = 2.79$, $p = .110$, $\eta^2_p = .122$, or Reading, $F(1,20) = .16$, $p = .690$, $\eta^2_p = .008$. However, a significant interaction between Relational term type and their being associated or not with a landmark (i.e. Salience) was found, $F(1,20) = 16.60$, $p = .001$, $\eta^2_p = .454$ (see Figure 3.9). Testing for simple main effects revealed that egocentric relational terms associated with landmarks (e.g. "Turn **left** at the pub.") evoked a significantly longer average DT, M = 2898 ms, SEM = 464 ms, than allocentric relational terms associated with landmarks (e.g. "At the bank, head **east**."), M = 1673 ms, SEM = 209 ms, $F(1,20) = 7.66$, $p = .012$, $\eta^2_p = .277$. The opposite trend was observed for relational terms describing directional changes where no landmark was present, with those in the allocentric condition eliciting significantly longer average DT, M = 2257 ms, SEM = 280 ms, than egocentric ones, M = 1476 ms, SEM = 150 ms, $F(1,20) = 5.73$, $p = .027$, $\eta^2_p = .223$. No other two- or three-way interaction was found to be significant, all ps > .05

First-, Second-, Third-Pass Dwell Time

I further decomposed DT into distinct runs of fixations, revealing a complex pattern of effects. While no significant main effects were found on first-run DT on landmark words, (Relational term type, $p = .085$, $\eta^2_p = .148$; Salience, $p = .079$, $\eta^2_p = .154$), significant interactions were found between Relational term type and Salience, $F(1,19) = 6.08$, $p = .023$, $\eta^2_p = .243$, and between Relational term type, Salience, and Reading, $F(1,19) = 4.75$, $p = .042$, $\eta^2_p = .200$. Resolving these significant interactions revealed that the effect of Reading during the first round of fixations was predominantly localised in a difference in dwell time on non-salient landmarks in the egocentric condition, M = 225 ms, SEM = 18 ms, and M = 292 ms, SEM = 33 ms, during the first and second reading respectively, $F(1,19) = 4.87$, $p = .040$, $\eta^2_p = .204$. In a similar fashion, an analysis of the simple main effects of Relational term and Salience revealed that salient landmarks generated longer first-run DT, M = 471 ms, SEM = 100 ms, than non-salient landmarks, M = 225 ms, SEM = 18 ms, during the first reading of the egocentric descriptions, $F(1,19) = 6.19$, $p = .022$, $\eta^2_p = .246$. This difference disappeared during the second egocentric reading, M = 361 ms, SEM = 39 ms, and M = 292 ms, SEM = 33 ms, for salient and non-salient landmarks respectively, $F(1,19) = 2.75$, $p = .113$, and was entirely absent during either allocentric reading, $p = .574$ and $p = .743$ respectively.

An analysis of the average DT on landmark words during the second and third runs reveals a simpler pattern of results. In both cases we found a significant main effect of Relational term type, $F(1,13) = 5.33$, $p = .038$, $\eta^2_p = .291$, and $F(1,4) = 10.73$, $p = .031$, $\eta^2_p = .728$, respectively. Second-run DT on landmark words was generally longer during encoding of egocentric descriptions than during encoding of allocentric descriptions, M = 376 ms, SEM = 31 ms, and M = 265 ms, SEM = 25 ms, respectively. The same was true of third-run DT, M = 399 ms, SEM = 75 ms, and M = 196 ms, SEM = 14 ms, during egocentric and allocentric encoding respectively. A significant interaction was found between Relational term type and Salience for second-run DT, $F(1,13) = 11.78$, $p = .004$, $\eta^2_p = .476$, and an almost-significant interaction between Relational term and Salience for third-run DT, $F(1,4) = 7.35$, $p = .053$, $\eta^2_p = .648$. A test of simple main effects on these interactions revealed an identical pattern of results. Salient landmarks generated a significantly longer average DT than non-salient landmarks in the egocentric condition during the second run of fixations, M = 452 ms, SEM = 52 ms, and M = 301 ms, SEM = 34 ms, respectively, $F(1,13) = 5.58$, $p = .034$, $\eta^2_p = .301$. This difference did not appear to reach significance during the third run of fixations, M = 487 ms, SEM = 82 ms, and M = 312 ms, SEM = 90 ms, $F(1,4) = 4.15$, $p = .111$. However, salient landmarks still generated a significantly longer average DT in the egocentric condition compared to the allocentric condition, across readings. This was the case both during the second run, M = 452 ms, SEM = 52 ms, and M = 237 ms, SEM = 20 ms, respectively, $F(1,13) = 12.05$, $p = .004$, $\eta^2_p = .481$, and during the third run, M = 487 ms, SEM = 82 ms, and M = 195 ms, SEM = 16 ms, respectively, $F(1,4) = 17.66$, $p = .014$, $\eta^2_p = .815$.

Performing the same analysis on relational terms revealed main effects of Relational term type, $F(1,20) = 5.22$, $p = .033$, $\eta^2_p = .207$, and of Landmark presence, $F(1,20) = 5.22$, $p = .026$, $\eta^2_p = .225$, but only during the first run of fixations. In this case, cardinal relational terms were fixated for longer than egocentric relational terms, M = 415 ms, SEM = 60 ms, and M = 332 ms, SEM = 45 ms, respectively. However, relational terms denoting turns with landmarks were fixated less than other relational terms, M = 319 ms, SEM = 41 ms, and M = 427 ms, SEM = 66 ms. No other significant main effects of Relational term type or Landmark presence, or interactions between them, were found during second or third pass, all ps > .1. No significant main effects of Reading were found during any rounds of fixations, all ps > .1. See Figure 3.10 to Figure 3.13 for the results of these analyses.

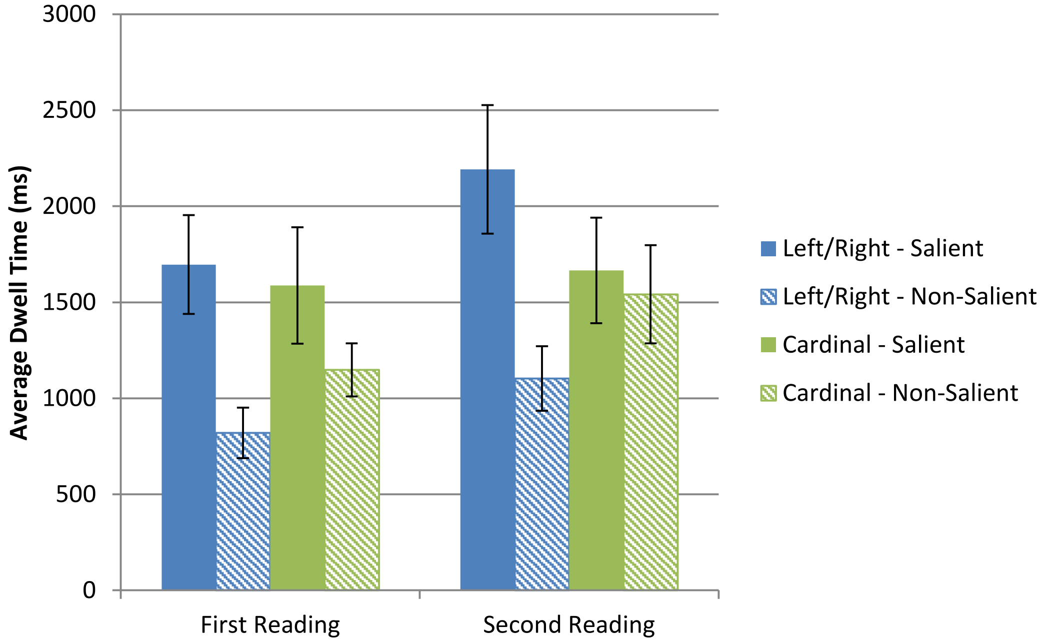


Figure 3.8 - Average dwell time on salient and non-salient landmark words during the first and second reading of each description type. Error bars represent ±1 SEM.

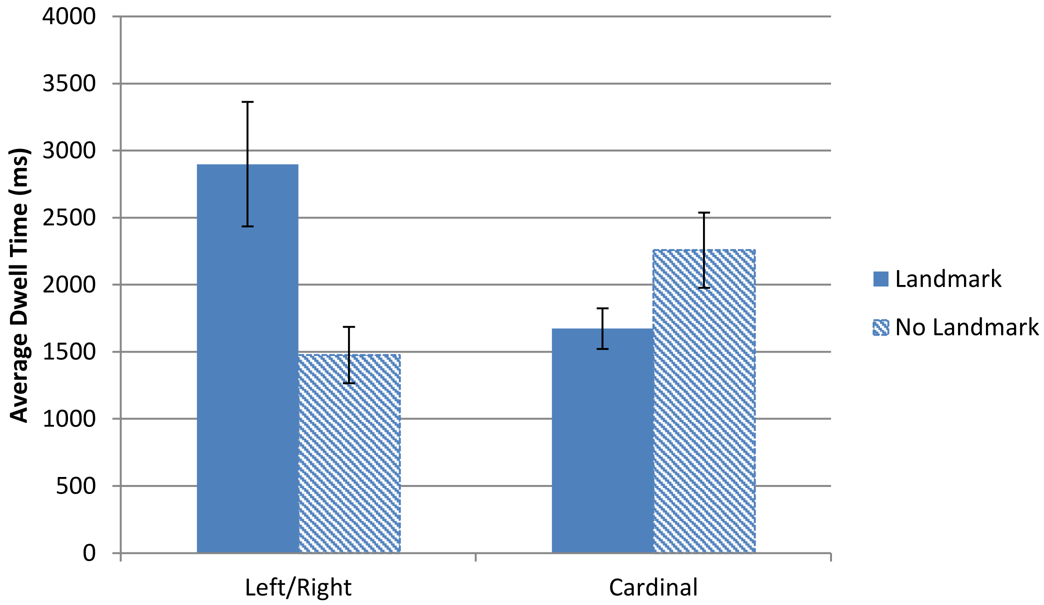
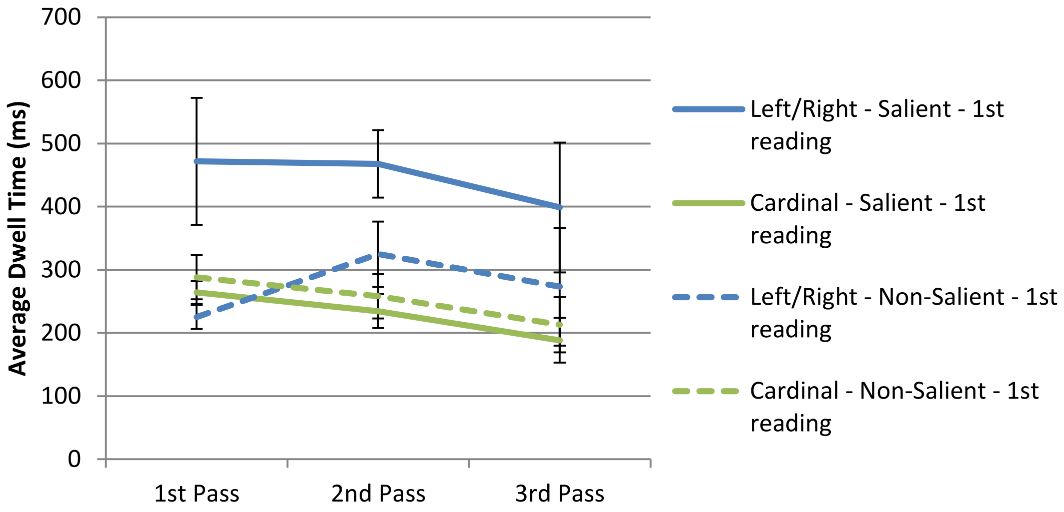


Figure 3.9 - Average dwell time on relational terms denoting turns with landmarks present and without landmarks. The dashed line represents relational terms indicating turns where no landmark was present. Error bars represent ±1 SEM.



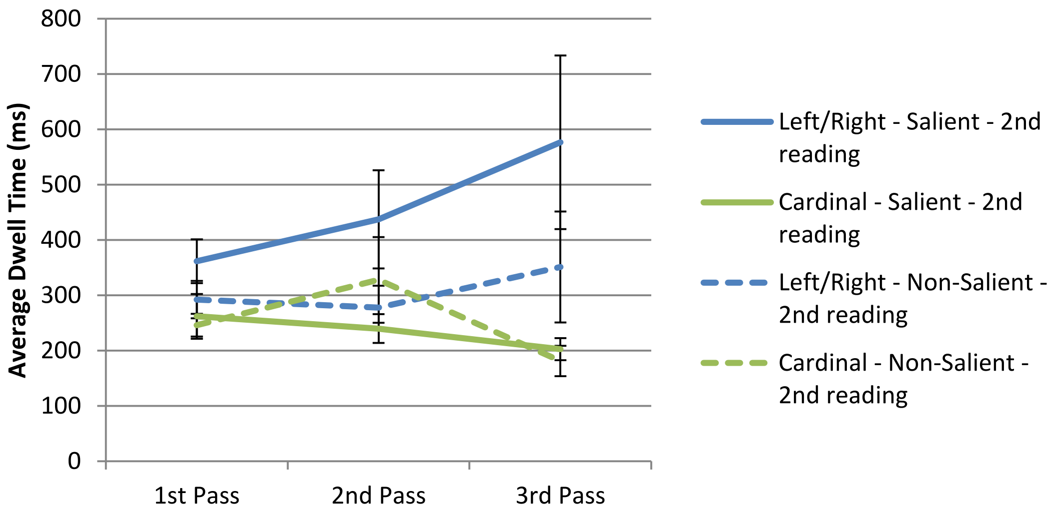
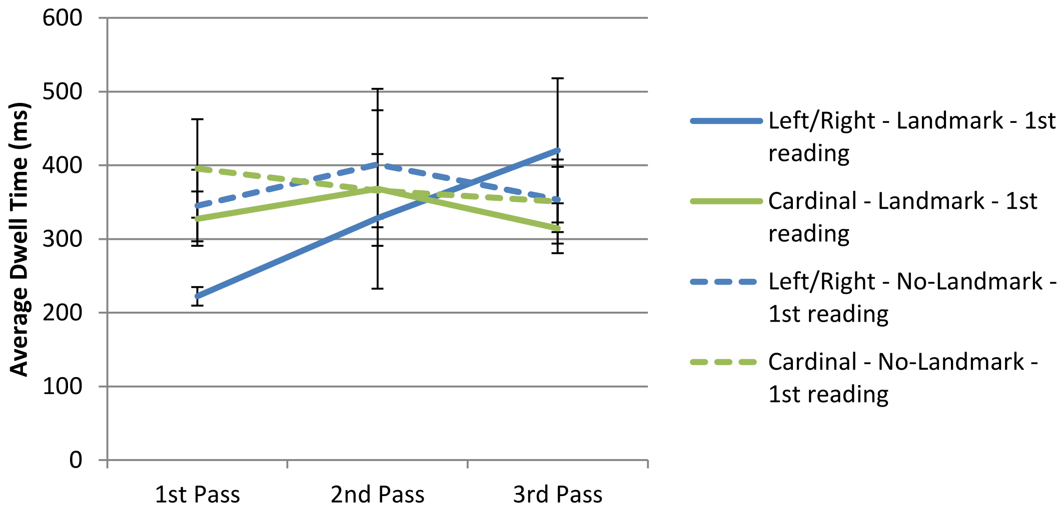


Figure 3.10 and Figure 3.11 – Dwell time on salient and non-salient landmark words during the first, second, and third (plus subsequent) rounds of fixations. Error bars represent ±1 SEM.

Regression Path Duration

An analysis of the average duration of regression paths generated by landmark words revealed no significant main effects of Relational term type, $F(1,19) = 1.45$, $p = .242$, $\eta^2_p = .071$, Reading, $F(1,19) = .47$, $p = .498$, $\eta^2_p = .025$, or Landmark salience, $(F1,19) = .24$, $p = .625$, $\eta^2_p = .013$. The interaction between Relational term type and Reading was non-significant, $F(1,19) = .015$, $p = .905$, $\eta^2_p = .001$, whereas the interaction between Relational term and Salience approached significance, $F(1,19) = 3{,}55$, $p = .075$, $\eta^2_p = .158$. A significant interaction between Reading and Salience was found, $F(1,19) = 6.89$, $p = .017$, $\eta^2_p = .266$. The three-way interaction was not significant, $F(1,19) = .13$, $p = .713$, $\eta^2_p = .007$. Tests of simple main effects showed that the difference between the RPD generated by non-salient landmarks in the two conditions approached significance, M = 856 ms, SEM = 185 ms, in the egocentric condition and M = 2089 ms, SEM = 590 ms, in the allocentric condition, $F(1,19) = 3.61$, $p = .073$, $\eta^2_p = .160$. However, the same comparison for non-salient landmarks yielded non-significant results, M = 1571 ms, SEM = 239 ms, and M = 1071 ms, SEM = 303 ms, $F(1,19) = 1.34$, $p = .261$. However, comparing the average RPD generated by salient and non-salient landmarks in the egocentric description yielded a significant difference, M = 1571 ms, SEM = 239 ms, and M = 856 ms, SEM = 185 ms, for salient and non-salient landmarks respectively, $F(1,19) = 4.79$, $p = .041$, $\eta^2_p = .202$. This was not observed during allocentric encoding, M = 1071 ms, SEM = 303 ms, and M = 2089 ms, SEM = 590 ms, $F(1,19) = 2.06$, $p = .167$. Additionally, I observed that salient landmarks generated significantly longer regression paths during first readings, M = 1609 ms, SEM = 255 ms, than during second readings, M = 1034 ms, SEM = 150 ms, $F(1,19) = 5.34$, $p = .032$, $\eta^2_p = .219$. On the other hand, non-salient landmarks generated shorter regression paths during first readings, M = 1007 ms, SEM = 258 ms, than during second readings, M = 1938 ms, SEM = 472 ms, $F(1,19) = 3.68$, $p = .070$, $\eta^2_p = .162$ (Figure 3.14 and Figure 3.15).



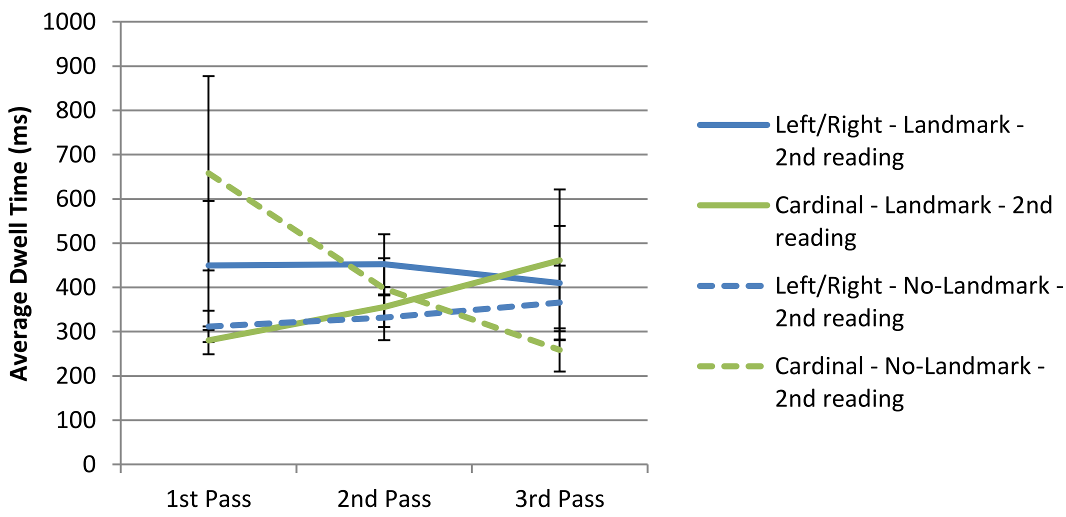


Figure 3.12 and Figure 3.13 - Dwell time on relational terms denoting turns with and without landmarks during the first, second, and third (plus subsequent) rounds of fixations. Error bars represent ±1 SEM.

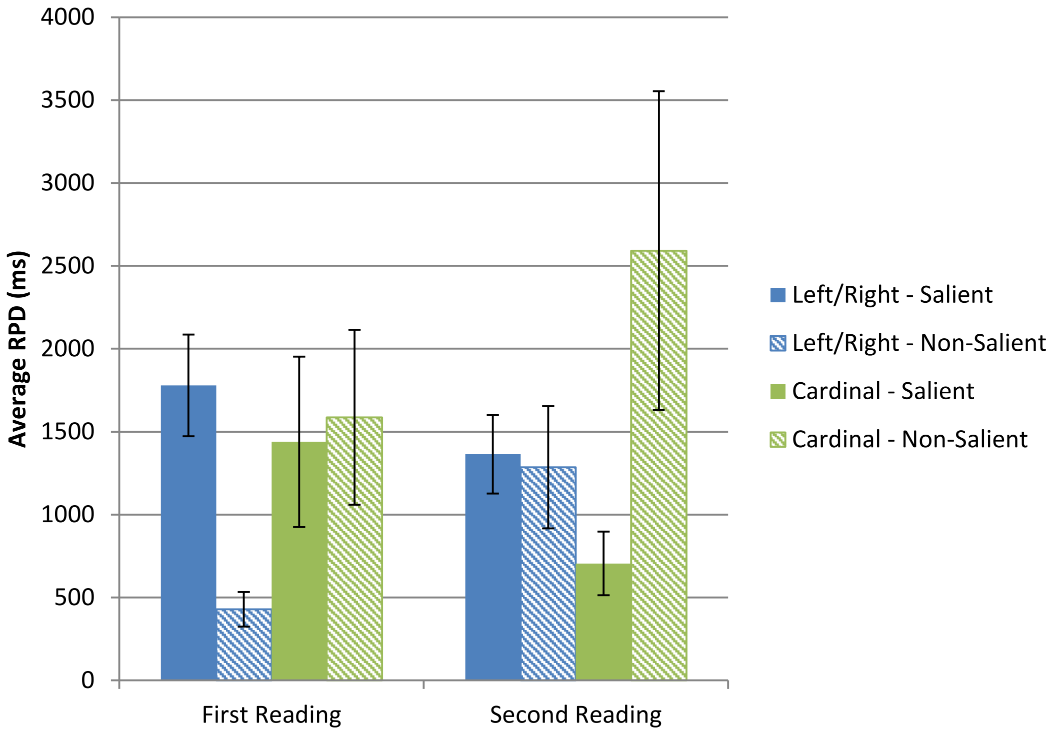


Figure 3.14 - Average regression path duration from landmark words for each reading. Error bars represent ±1 SEM.

The same analysis performed on RPD generated by relational terms revealed a significant main effect of Relational term type, $F(1,20) = 6.93$, $p = .016$, $\eta^2_p = .258$. This showed cardinal relational terms generated longer RPD than egocentric relational terms, M = 2527 ms, SEM = 363 ms, and M = 1471 ms, SEM = 206 ms, respectively. A marginally significant effect of Landmark presence was also observed, $F(1,20) = 4.11$, $p = .056$, $\eta^2_p = .170$, with relational terms denoting turns with landmarks generating shorter regression paths than other relational terms, M = 1706 ms, 249 ms, and M = 2293 ms, SEM = 271 ms, respectively. A significant interaction between Relational term type and Landmark presence was also found to affect the duration of regression paths from relational terms, $F(1,20) = 34.86$, $p = .001$, $\eta^2_p = .635$. Decomposing the interaction revealed a highly significant difference in RPD generated by salient, M = 556 ms, SEM = 112 ms, and non-salient, M = 2386 ms, SEM = 381 ms, egocentric relational terms, $p = .001$, $\eta^2_p = .536$. The same difference for the cardinal relational terms was only marginally significant, M = 2855 ms, SEM = 474 ms, and M = 2200 ms, SEM = 307 ms respectively, $p = .064$, $\eta^2_p = .162$. The duration of regression paths generated by non-salient relational terms did not differ significantly between conditions, $p = .669$, $\eta^2_p = .009$ (Figure 3.16).

Regressions In and Out

In order to clarify the above results, I analysed the number of regressions elicited by (Regression Out) and received by (Regression In) landmark words. In the case of in-going regressions, significant main effects of Relational term, $F(1,23) = 29.57$, $p < .001$, $\eta^2_p = .563$, and Salience, $F(1,23) = 9.18$, $p = .006$, $\eta^2_p = .285$, were observed.

Overall, landmark words were found to receive more regressions during encoding of cardinal descriptions, M = 1.9, SEM = .24, compared to egocentric descriptions, M = .70, SEM = .08, and salient landmarks received more regressions than non-salient landmarks, M = 1.5, SEM = .18 and M = 1.1, SEM = .12, respectively. However, a significant interaction between the two factors was also found, $F(1,23) = 18.38$, $p = .001$, $\eta^2_p = .444$. A test of simple main effects revealed that while non-salient landmarks received significantly more regressions than salient landmarks, M = .906, SEM = .14, and M = .510, SEM = .08, respectively, in the egocentric condition, $F(1,23) = 6.18$, $p = .021$, $\eta^2_p = .212$, the opposite trend was observed in the allocentric condition. In the latter, salient landmarks received significantly more regressions than non-salient landmarks, M = 2.6, SEM = .36, and M = 1.3, SEM = .18, respectively, $F(1,23) = 17.49$, $p = .001$, $\eta^2_p = .432$. No significant main effect of Reading, $F(1,23) = 1.55$, $p = .225$, $\eta^2_p = .063$, or significant interactions between Relational term and Reading, $F(1,23) = .00$, $p = .973$, $\eta^2_p < .001$, and Reading and Salience, $F(1,23) = .50$, $p = .484$, $\eta^2_p = .022$, were found. No significant three-way interaction was found, $F(1,23) = .20$, $p = .655$, $\eta^2_p = .009$.

As for the number of outgoing regressions generated by landmark words, this wasn't affected by Relational term type, $F(1,23) = .09$, $p = .758$, $\eta^2_p = .004$. However, it was significantly influenced by Reading, $F(1,23) = 4.51$, $p = .045$, $\eta^2_p = .164$, but even more so by landmark salience, $F(1,23) = 12.75$, $p = .002$, $\eta^2_p = .357$. More specifically, salient landmarks generated on average significantly more regressions than non-salient landmarks, M = 1.9, SEM = .23, and M = 1.3, SEM = .14, respectively, and more frequently during the second encoding compared to the first, M = 1.8, SEM = .21, and M = 1.4, SEM = .18, respectively. No significant interactions between Relational term type and Reading, $F(1,23) = .75$, $p = .713$, $\eta^2_p = .407$, between Relational term type and Salience, $F(1,23) = 1.22$, $p = .280$, $\eta^2_p = .050$, or between Reading and Salience, $F(1,23) = 2.05$, $p = .165$, $\eta^2_p = .082$, were found. The three-way interaction was also non-significant, $F(1,23) = 1.65$, $p = .211$, $\eta^2_p = .067$.

3.4. Discussion

The goal of this study was to explore possible variations of linguistic spatial information encoding and recall as a function of different imagined perspectives. I attempted to elicit different spatial perspectives by using different types of relational terms to describe spatial relationships between the participants and different features of the environments, at different stages of the routes described. Eye movements were recorded during description encoding to determine whether this manipulation would result in systematically different reading patterns. Participants' performance in a map drawing task was also assessed to test the hypothesis that a successful allocentric encoding should result in an equivalent, or higher, recall performance during an allocentric task relative to an egocentric encoding, whereas unsuccessful allocentric encoding could be the result of an interference effect between competing reference frames. I additionally administered participants several psychometric measures of working memory and visuospatial abilities, in an attempt to find reliable predictors of behavioural performance. I will discuss these different results in turn.

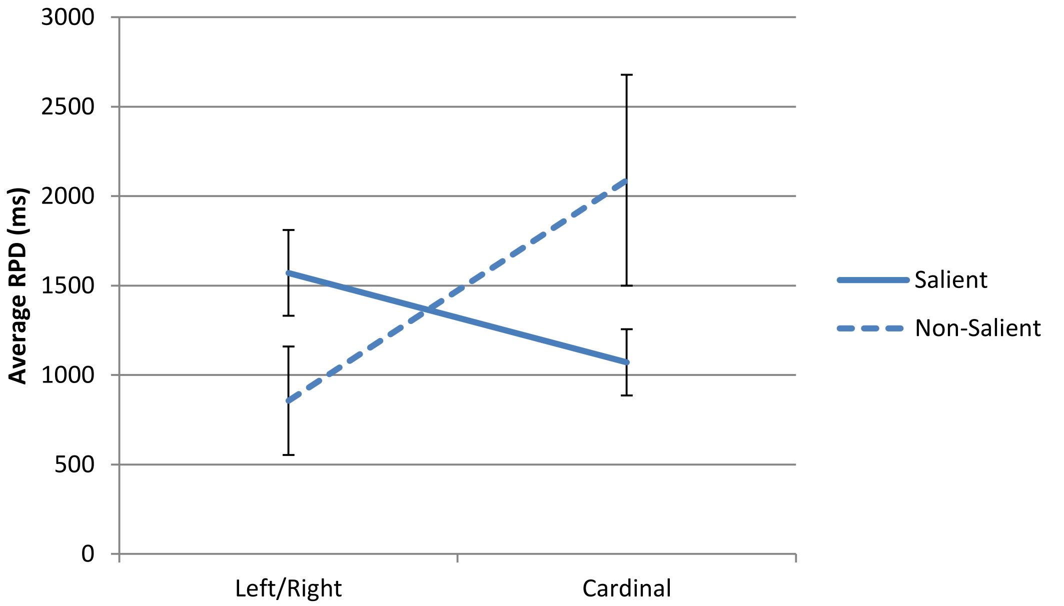


Figure 3.15 - Average regression path duration from landmark words averaged across readings. Error bars represent ±1 SEM.

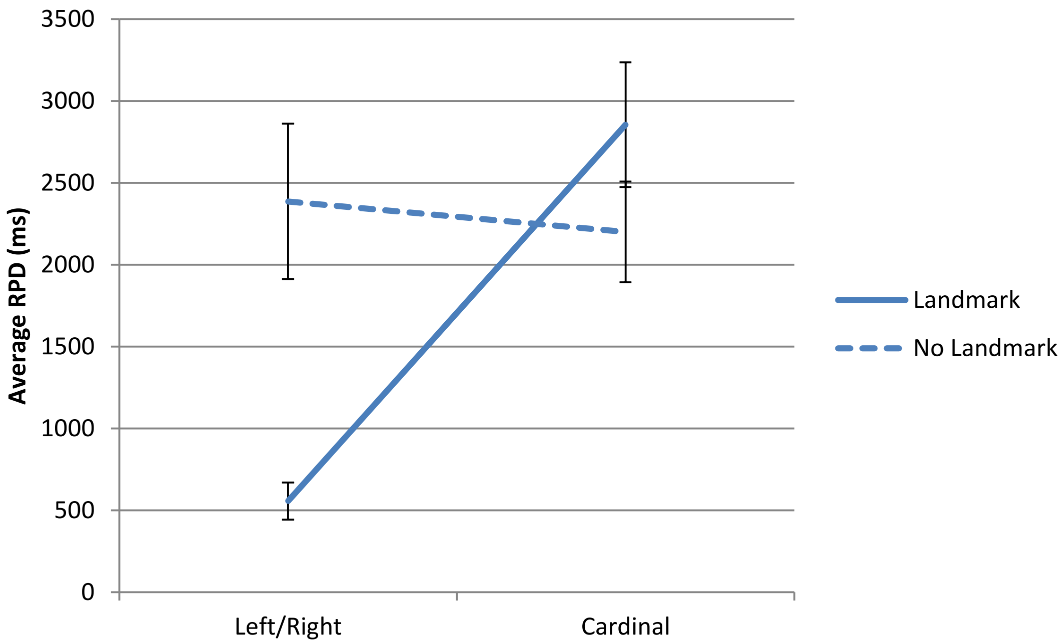


Figure 3.16 - Average regression path duration from relational terms averaged across readings. Error bars represent ±1 SEM.

Map Measures and Individual Differences

Measures of landmark recall indicate a significantly lower performance after the first reading of the allocentric descriptions compared to the egocentric condition. This difference however tended to disappear after a second reading of the descriptions. This could be indicative of an interference effect during the first encoding, whereby participants attempted to construct an egocentric representation of the sequential instructions, therefore having to remap their own position in the environment relative to the cardinal frame of reference at each turn. This could feasibly result in much higher cognitive demands and consequently lower performance. The increase in recall performance after the second reading to a level equivalent to that observed in the egocentric condition could therefore result from a change in encoding strategy and the adoption of an allocentric frame of reference. Such a change should be detectable in eye movement measures of reading performance.

Interestingly, the same pattern was not found for the measures of directional turn recall. While the effect of relational term type was only marginally significant (likely due to the low number of turns included in the descriptions), the trend points to a lower performance in the allocentric condition still after the second encoding phase. This could be due to a difficulty in incorporating configural knowledge of the environment into a spatial representation despite adopting a renewed allocentric strategy during the second reading. It is worth noting, however, that spatially and sequentially correct recall of landmarks was significantly lower than overall landmark recall in both conditions and after both readings, indicating that pure semantic recall is not necessarily a good measure of spatial understanding and should not be used as such. Overall, the aggregated map scores revealed a distinct advantage in CK acquisition for the Left/Right condition, even after a second encoding, and a deficit in LK acquisition after the first cardinal encoding. Regression analyses revealed few significant predictors of map drawing performance. Generally, mental rotation appears to be more relevant to the acquisition of CK during both cardinal encodings, whereas performance on the Money's Test appears to more strongly predict both CK and LK acquisition and recall after the second egocentric encoding. Performance on the DBW significantly predicted both CK and LK scores after the second allocentric encoding, possibly indicating higher task demands in that particular condition. This suggests that both task demands and individual differences in the availability of cognitive resources modulate the extent to which format congruence between encoding and test facilitates performance.

Eye Movement Measures

The most easily interpretable evidence for different encoding patterns as a function of different imagined perspectives comes from the analysis of average dwell time on the various areas of interest. While egocentric encoding resulted in a reliable significant difference in dwell time between salient and non-salient landmark words, this pattern was present during allocentric encoding (albeit less pronounced) only during the first reading. During the second allocentric encoding the relative difference in navigational salience of the different landmarks is not reflected in the amount of time spent fixating the corresponding words. A possible interpretation of this finding (and consistent with other findings in the literature. See Piccardi et al., 2016, and Section 2.7 for a description of their study) might suggest that, as we build progressively more survey-like representations of environments, these representations lend themselves to supporting the planning of multiple routes between the landmarks contained in them. This effectively makes the saliency profile of a set of landmarks a variable function of the particular route connecting them, and the latter a function of the way the environment is envisaged. Similarly, for the relational terms, we observe a marked reduction in the dwell time and RPD difference between salient and non-salient relational terms in the allocentric condition compared to the egocentric condition. Given the generally poorer performance in turn recall in the allocentric condition, however, this might reflect a less-than-optimal encoding as a result of conflicting reference frames. That is, participants may have been forced to continually switch between an egocentric and an allocentric imagined perspective during the encoding of cardinal descriptions, leading to an inability to attend to relational terms and correctly encode the change in direction.

The analysis of regressions generated by landmark words complicates matters further, painting a picture in which non-salient landmarks may have proven particularly difficult to integrate into a coherent spatial model in the allocentric condition, causing considerably longer regression paths in spite of a non-significant difference in the number of regressions generated by salient and non-salient landmarks. A possible explanation can be found in the fact that non-salient landmarks in the allocentric condition were still presented to the reader in egocentric terms (e.g. “Walk past the pub on your right”). For this reason, the resulting representations may not have been consistently allocentric, and the reference frame transformations may have produced greater discourse integration difficulties, contributing to generating longer regression paths. However, in that case we might expect to observe the same phenomenon – perhaps even more so – during the first allocentric reading. Additionally, salient landmarks were found to receive significantly more regressions than non-salient landmarks during both allocentric encodings, whereas the opposite trend was observed during both egocentric encodings. However, this particular pattern remains currently unexplained.

General Discussion

Contrary to the prediction of higher performance in the allocentric task after allocentric encoding, performance seemed to be generally higher following encoding of left/right descriptions, at least after the first reading, and particularly for the measures of turn recall. While it is possible that the generally lower performance in the allocentric condition might stem from participants' lack of familiarity with cardinal relational terms, the lack of an increased dwell time on cardinal terms during encoding (Figure 3.9) appears inconsistent with such a conclusion. Additionally, while changes in measures of dwell time on landmark words might be interpreted as suggesting a shift in encoding strategy between the first and the second allocentric encoding, and while the increase in landmark recall performance appears consistent with the adoption of an alternative encoding strategy as effective as the one employed during egocentric encoding, measures of regressions and performance on turn recall appear far less clear. It therefore remains unclear whether what was observed is, indeed, an instance of separable imagined spatial perspectives as a function of different relational terms or, in fact, an example of parallel computation and resulting interference between them in the allocentric condition. The possibility of an interference effect makes reading pattern also difficult to interpret. Nevertheless, the fact that significant differences can be found in the encoding patterns in the first place seems to suggest that different imagined spatial perspectives might be selectively generated during the first encoding of a route description. By extension, it is possible that the significant difference in dwell time patterns on both landmark words and relational terms between conditions (already present during the first rounds of fixations) might be indicative, at least partly, of the creation of different underlying priority maps, and further research should be carried out to investigate this possibility.

It is evident that much is still unclear as to the processes underlying the construction of spatial representations from linguistic content. The general assumption, widespread in psychological research in this field, that "survey" terms are sufficient to engender allocentric representations is not supported by these results. This is in line with the observed variability in the use of relational terms (see Section 2.3) and warrants caution when interpreting behavioural results. Much more research is needed to elucidate the cognitive and neural correlates underlying the broader processes involved in mental navigation, and the individual proclivities that determine encoding strategy, as well as the ability to translate spatial representations between different reference frames. In Experiments 2 and 3, I attempted to improve upon the current paradigm. In particular, by including explicit instructions prompting participants to imagine an egocentric or an allocentric perspective during description encoding, I intend to better characterise the pattern of interference effects between competing spatial perspectives, if present, and attempt to answer the question of whether different patterns of eye movements can be used as proxies of the construction of different mental models.

Experiment 2: Relational Terms and Explicit Imagery Instructions

3.5. Experiment 2: Introduction

In Experiment 1, I attempted to determine if different types of relational terms used to describe spatial relations in imagined environments could engender reliably different imagined spatial perspectives. I set out to test this by recording eye movements as participants read spatial route descriptions containing either egocentric or allocentric relational terms, and by analysing their performance on several measures in a map drawing task.

Based on the traditional distinction between egocentric and allocentric reference frames, and on the equally traditional assumption that route and survey spatial terms would prompt and map neatly onto the same dichotomous distinction, I predicted that an allocentric encoding would result in an equivalent, or better, performance during allocentric test than an egocentric encoding. This prediction was not confirmed. This raised the possibility that this may be the result of an interference effect due to the parallel activation of competing spatial reference frames (see Section 1.2 and Section 1.11). Participants may have attempted to maintain an egocentric representation of the environment (possibly because of a natural tendency to do so, or due to the sequential nature of a route description) even when presented with a route description containing cardinal relations. This would have forced them to update their own egocentric frame of reference relative to the cardinal frame provided by the allocentric description after every described change in direction. The cognitive load resulting from this interference may have caused the observed drop in performance.

In turn, this possibility prevented me from reaching convincing conclusions based on the eye tracking data obtained during the reading phases. While the finding that allocentric encoding resulted in a reduced dwell time difference between navigationally salient and non-salient landmarks might be seen as evidence that a different imagined spatial perspective also generates different priority maps (see Section 2.6), this must be replicated in a design that allows to control for the possible interference between different imagined spatial perspectives. That was one of the goals of Experiment 2. In a partial replication of Experiment 1, participants were presented with route descriptions containing landmarks at salient and non-salient locations, and either egocentric or allocentric relational terms. Additionally, I provided participants with explicit instructions, asking them to actively imagine either an egocentric or an allocentric perspective while reading the descriptions. Recall performance was measured via a map drawing task as in Experiment 1. From now on, experimental conditions will be referred to following the "relational term-task instruction" format, e.g. "left/right-allocentric" will denote an egocentric route description read while imagining an allocentric perspective.

The goal of this experiment was to determine whether different relational terms do engender different and separable spatial perspectives by observing the pattern of interference between them and explicit task instructions. I hypothesised that the use of egocentric or allocentric relational terms, and the active maintenance of an egocentric or allocentric imagined spatial perspective, would generate different degrees of representational interference, and result in different behavioural performance levels. In particular, I predicted that a description containing cardinal relational terms encoded while maintaining an egocentric perspective would result in a significantly lower encoding and, consequently, recall performance than observed in more compatible conditions. If the hypothesis of an interference between competing imagined perspectives in Experiment 1 is correct, then this particular combination of cardinal directions and egocentric imagined perspective should recreate the same interference pattern and cognitive costs observed in the cardinal condition in the previous study. Furthermore, the need to transform a purely egocentric representation (both in terms of relational terms and explicit task instructions) into an allocentric model during the map drawing task could result in an overall disadvantage in the left/right-egocentric condition, and in an advantage for both conditions requiring participants to actively maintain allocentric representations of the environments described. This, however, might not necessarily transpire in recall accuracy and might be more evident in tasks allowing for the recording of reaction time data.

Experiment 2 additionally attempted to examine the phenomenological experience of participants' mental representations and imagery via a self-report questionnaire administered after each reading of each description. The questionnaire probed participants' ability to imagine and maintain the spatial perspective they were instructed to, and whether this would be impacted by the type of relational term used. That is, I explored whether participants' subjective experience of imagery and task difficulty would reflect the expected cognitive load differences while mapping spatial relations described using different relational terms onto different spatial perspectives (e.g. as in the Cardinal/Egocentric condition). However, this was also a more general attempt to determine what role phenomenology might play in mental representations, an issue still unresolved in imagery research (see Section 1.9 and Section 1.10).

3.6. Methods

3.6.1. Participants

34 adult English native speakers with normal or corrected-to-normal vision (3 males, 31 females, mean age 18.38 ± .11 years) were recruited across the University of Nottingham in exchange for credits or an inconvenience allowance.

3.6.2. Design and Materials

As in Experiment 1, I created a set of spatial route descriptions containing sequential instructions for how to navigate different urban environments. The instructions informed the participant as to when to make a turn or keep driving forward. A total of six landmarks were distributed along the route as follows: one origin, one destination, two at salient points (changes in directions), two at non-salient points. Unlike in Experiment 1, non-salient landmarks were not assigned to a specific side of the road and participants were simply told they would travel past them. Each route contained a total of four turns, two associated with landmarks and two at spatial locations where no landmark was present.

Four distinct landmark sets were used and counterbalanced across conditions. Additionally, two versions of each description were created, so that the landmarks located at salient locations in one would be used as the non-salient landmarks in a different description. This produced a total of 32 different descriptions (or eight different batches), while maintaining the same map layout across all possible versions of a condition. Each participant was presented with four descriptions, one for each experimental condition. Landmark sets were created so as to maintain a similar average word length across them. As in Experiment 1, each description was presented on two pages, each containing four lines of text. See Table 3.6 below for examples of each description type, and Appendix II for a complete list of the landmark words used.

As an added factor, participants were given explicit instructions to imagine either an egocentric (described as "walking through the environments described maintaining a first-person, street-level view") or an allocentric (described as "imagining the environments described as if seen on a map") view of the environments described. These instructions were provided in the form of stylised images presented after calibration and prior to the appearance of the first page of text. A silhouette of a walking person would prompt an egocentric perspective, whereas that of a bird in flight would prompt an allocentric perspective. See Figure 3.17 for the images used.

Together with the main experimental task, a subset of the psychometric measures used in Experiment 1 (see Section 3.2.2) was administered to better understand the contribution of these abilities to task performance. These were:

- the **Money's Standardised Road-Map Test of Direction Sense** (Money's Test or MT; Money, Alexander, & Walker, 1965).
- the **Santa Barbara Sense of Direction Scale** (SBSOD) (Hegarty et al., 2002).
- the **Digit Span Test**, backward version (DBW) (Blackburn & Benton, 1957).

Table 3.6 - Examples of egocentric and allocentric route descriptions as used in Experiment 2.

EGOCENTRIC DESCRIPTION	ALLOCENTRIC DESCRIPTION
Leave the house.	Leave the park.
Turn left at the gym.	Take the first road heading north.
Take the second right.	At the pub, head west.
Walk past the clinic.	Walk past the bank.
Turn left at the school.	Take the first road heading south.
Take the first left.	Walk past the florist.
Walk past the library.	At the aquarium, head east.
You have reached the cinema.	You have reached the dentist.

Additionally, after each map drawing phase participants were presented with a five-item, self-report questionnaire in which they were asked to rate their agreement (on a five-point Likert scale) with a series of statements aimed at probing their subjective experience of information encoding. The questionnaire contained the following items:

I found it easy to imagine the perspective I was asked to imagine.

Strongly disagree 1 2 3 4 5 **Strongly agree**

I found it easy to maintain the same imagined perspective throughout the reading phase.

Strongly disagree 1 2 3 4 5 **Strongly agree**

I spent most of the reading time imagining a first-person, street-level perspective.

Strongly disagree 1 2 3 4 5 **Strongly agree**

I spent most of the reading time imagining a bird's-eye perspective.

Strongly disagree 1 2 3 4 5 **Strongly agree**

I spent most of the reading time switching between a first-person, street-level perspective and a bird's-eye perspective.

Strongly disagree 1 2 3 4 5 **Strongly agree**

Figure 3.17 - Images used to prompt explicit spatial representations. Participants were informed as to the meaning of each picture in the information sheet provided at the beginning of the session.

3.6.3. Apparatus

Eye tracking and computer setup were the same as in Experiment 1.

3.6.4. Procedure

The procedure was identical to the one used in Experiment 1.

3.7. Results

3.7.1. Map Data

Participants' map drawings were scored on the same metrics used in the previous study. See Section 3.3.1 for details.

Overall Landmark Recall

Overall landmark recall was analysed via a set of 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric) x 2(Reading: 1st vs 2nd), Sidak-corrected within-subject ANOVAs. Significant main effects of Relational term, $F(1,33) = 14.25$, $p = .001$, $\eta^2_p = .302$, and Reading, $F(1,33) = 110.54$, $p < .001$, $\eta^2_p = .770$, were observed, but no significant main effect of Imagined perspective, $F(1,33) = .02$, $p = .875$, $\eta^2_p = .001$. Landmark recall was generally better following encoding of left/right descriptions, M = 74.02%, SAM = 2.6%, than of cardinal descriptions, M = 67.27%, SAM = 2.9%, and considerably higher after a second encoding compared to the first drawing phase, M = 82.96%, SAM = 2.8%, and M = 58.33%, SEM = 2.8%, respectively. A significant interaction between Imagined perspective and Reading was found, $F(1,33) = 4.27$, $p = .047$, $\eta^2_p = .115$, but an analysis of the simple main effects of Imagined perspective revealed no significant effects during either reading, all ps > .1. Reading, on the other hand, had a significant effect on landmark recall performance in both imagery conditions, generally improving performance after the second encoding. However, this effect was stronger when participants were instructed to imagine an allocentric perspective, $F(1,33) = 117.04$, $p < .001$, $\eta^2_p = .780$, M = 56.86%, SEM = 3.3%, and M = 84.80%, SEM = 2.9%, compared to when they were instructed to engender a first-person, egocentric perspective, $F(1,33) = 48.21$, $p < .001$, $\eta^2_p = .594$, M = 59.80%, SEM = 3.2% and M = 81.12%, SEM = 3.2%. This indicates a greater rate of improvement between first and second reading for conditions requiring participants to actively build allocentric representations of the environments described, although part of this effect may be due to the generally lower performance in the cardinal-egocentric condition (Figure 3.18).

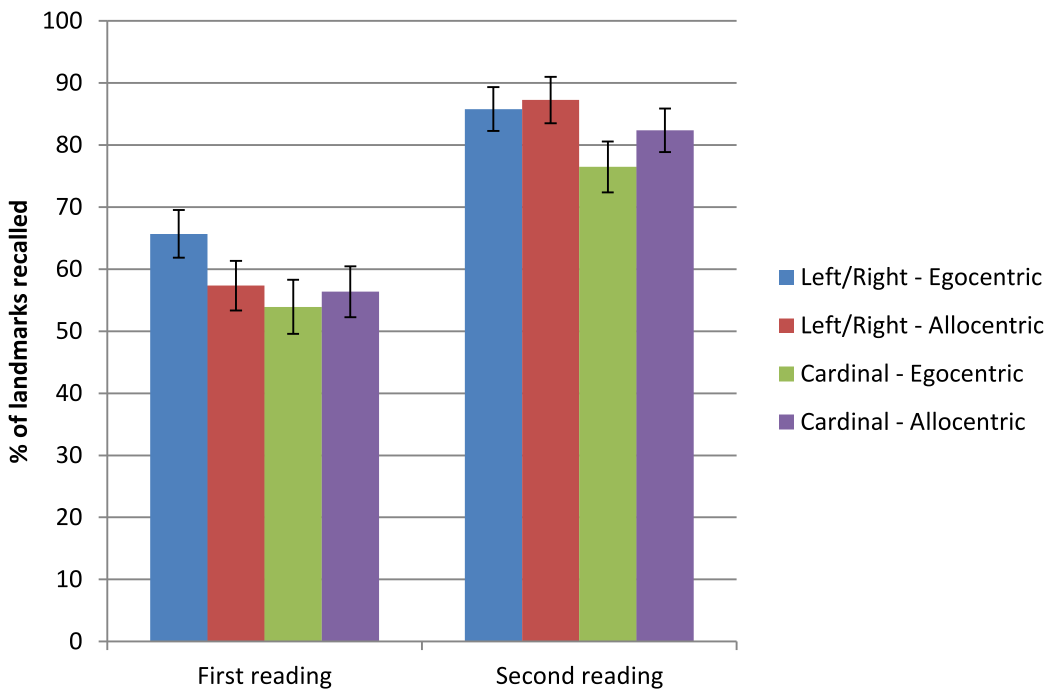


Figure 3.18 – Average overall landmark recall performance. Error bars represent ±1 SEM.

Sequentially Correct Landmarks

The number of landmarks (expressed as percentage) recalled in the longest uninterrupted correct sequence (i.e. number of landmarks drawn in the order in which they are mentioned in the respective description, regardless of whether the spatial relations are preserved) was analysed in turn. Significant main effects of Reading, $F(1,33) = 129.62$, $p < .001$, $\eta^2_p = .797$, Relational term, $F(1,33) = 7.30$, $p = .011$, $\eta^2_p = .181$, and Imagined perspective, $F(1,33) = 5.79$, $p = .022$, $\eta^2_p = .149$ were found. Sequentially correct landmark recall appeared to be higher following encoding of left/right descriptions, M = 58.57%, SEM = 3.2%, than following encoding of cardinal descriptions, M = 51.34%, SEM = 3.4%. However, performance in this measure was also better when participants were instructed to imagine an allocentric perspective during encoding, M = 58.57%, SEM = 3.3%, compared to when they were instructed to imagine an egocentric perspective, M = 51.34%, SEM = 3.5%. Performance was generally higher following the second encoding, M = 69.11%, SEM = 3.5%, than after the first, M = 40.80%, SEM = 3.1%.

Borderline significant trends for interactions between Relational term and Imagined perspective, $F(1,33) = 3.41$, $p = .074$, $\eta^2_p = .094$, and between Imagined perspective and Reading, $F(1,33) = 3.32$, $p = .077$, $\eta^2_p = .091$, were also found. The interaction between Relational term and Reading was not significant, $F(1,33) = .59$, $p = .446$, $\eta^2_p = .018$, nor was the three-way interaction, $F(1,33) = 1.634$, $p = .210$, $\eta^2_p = .047$. Visual inspection of the trend in the data seems to point to a generally reduced performance in the Cardinal-Egocentric condition (Figure 3.19).

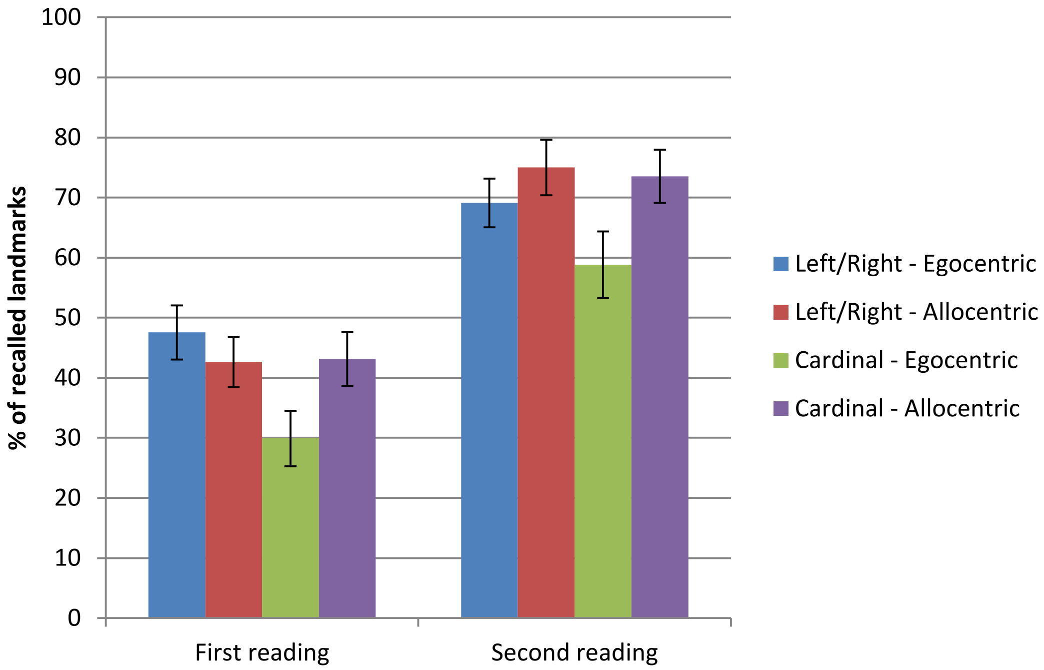


Figure 3.19 - Average sequential landmark recall performance. Error bars represent ±1 SEM.

Spatially Correct Landmarks

A significant main effect of Relational term type was found, $F(1,33) = 7.78$, $p = .009$, $\eta^2_p = .191$, as well as a significant main effect of Reading, $F(1,33) = 118.52$, $p = .001$, $\eta^2_p = .782$. These show an improvement in performance following a second encoding compared to the first, M = 69.11%, SEM = 3.6%, and M = 39.09%, SEM = 3.1%, respectively, and a generally better performance following encoding of left/right descriptions compared to cardinal descriptions, M = 57.35%, SEM = 2.9%, and M = 50.85%, SEM = 3.6%, respectively. No significant main effect of Imagined perspective was found, $F(1,33) = 2.10$, $p = .156$, $\eta^2_p = .060$, but a borderline significant interaction between Imagined perspective and Reading was also observed, $F(1,33) = 3.89$, $p = .057$, $\eta^2_p = .106$ (Figure 3.20). This shows a general trend of decreased performance in the Cardinal-Egocentric condition, which might account for much of the decrease in performance following cardinal encoding. No significant interactions between Relational term and Reading, $F(1,33) = 1.14$, $p = .292$, $\eta^2_p = .034$, and between Relational term and Imagined perspective, $F(1,33) = .53$, $p = .012$, $\eta^2_p = .393$, were found. The three-way interaction was not significant, $F(1,33) = 1.16$, $p = .288$, $\eta^2_p = .034$.

Overall Turn Recall and Sequentially Correct Turns

Only a main effect of Reading was found on both turn recall measures, $F(1,33) = 23.24$, $p = .001$, $\eta^2_p = .413$, showing an improvement in turn recall following the second encoding compared to the first, M = 90.80%, SEM = 2.2%, and M = 78.67%, SEM = 3.1%, respectively (M = 90.62%, SEM = 2.2% and M = 78.49%, SEM = 3.1% for sequential turn recall). No significant main effects of Relational term type, $F(1,33) = .41$, $p = .525$, $\eta^2_p = .012$, or Imagined perspective, $F(1,33) = 2.76$, $p = .106$, $\eta^2_p = .077$, were found. No significant or marginally significant interactions were found, all Fs < 1, all ps > 1 (Figure 3.21).

Ordinal Information

Recall of these items was found to be significantly influenced by Relational term type, $F(1,33) = 4.90$, $p = .034$, $\eta^2_p = .129$, and by Reading, $F(1,33) = 22.50$, $p = .001$, $\eta^2_p = .406$. In other words, recall was greater following encoding of left/right descriptions, M = 36.76%, SEM = 3.6%, than that of cardinal descriptions, M = 28.67%, SEM = 3.5%. No significant main effect of Imagined perspective was found, $F(1,33) = .07$, $p = .790$, $\eta^2_p = .002$. A borderline significant interaction between Imagined perspective and Reading was also found, $F(1,33) = 3.70$, $p = .063$, $\eta^2_p = .101$, but no interaction between Relational term and Imagined perspective, $F(1,33) = .87$, $p = .357$, $\eta^2_p = .026$, or between Relational term and Reading, $F(1,33) = 1.54$, $p = .222$, $\eta^2_p = .045$. The three-way interaction was not significant, $F(1,33) < .001$, $p > .999$. However, recall of these implied path segments branching out of the main route (e.g. "Take the first/second left.") was generally poor. Additionally, only two such information tokens were contained in each description (one for each turn lacking a landmark), therefore greatly reducing variance. Nevertheless, a similar pattern of performance can be found, with participants performing generally worse in the cardinal-egocentric condition (Figure 3.22).

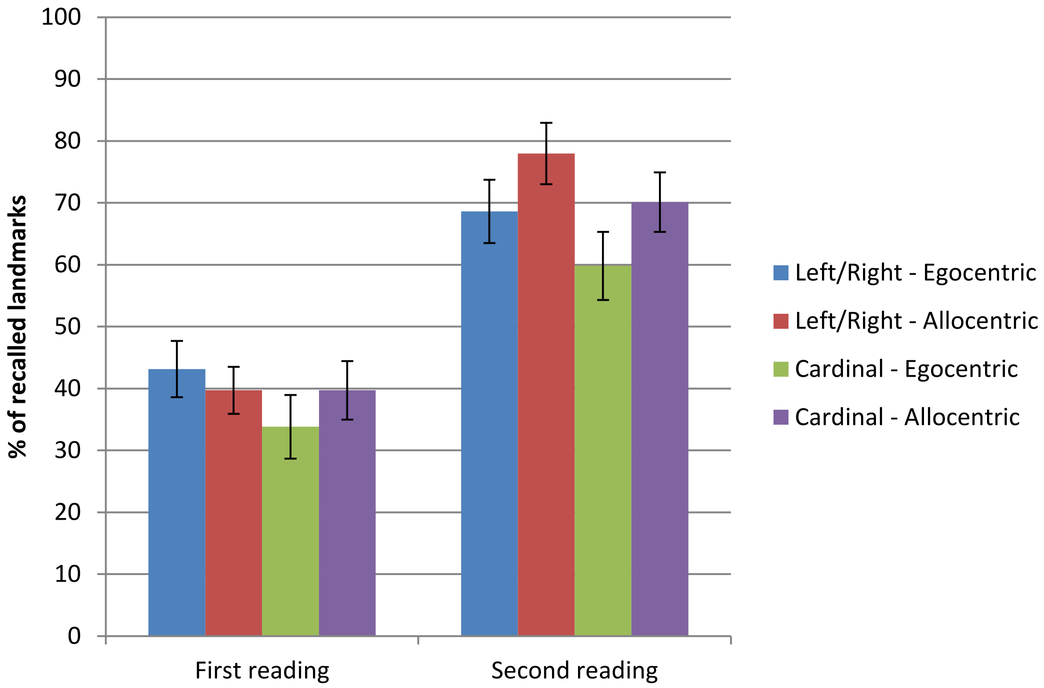


Figure 3.20 - Average spatial landmark recall performance. Error bars represent ±1 SEM.

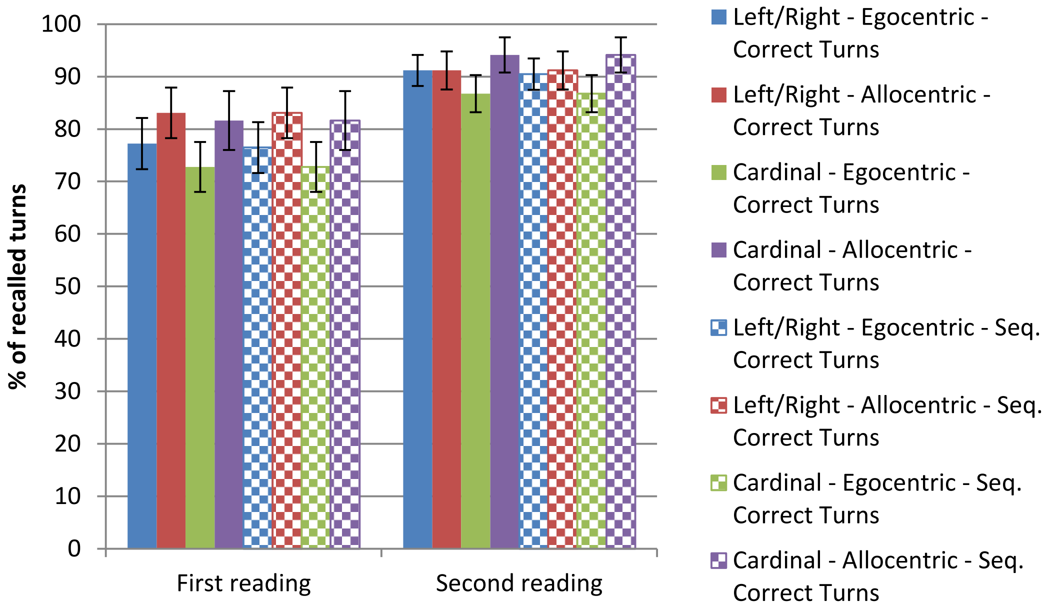


Figure 3.21 - Average performance on the two turn recall measures in the map-drawing tasks. Error bars represent ±1 SEM.

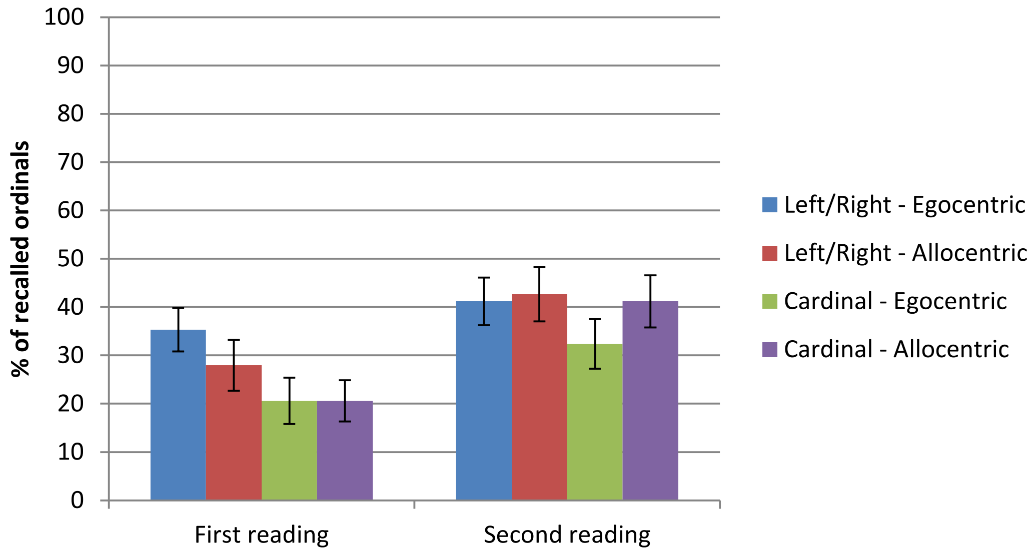


Figure 3.22 - Average performance on ordinal information recall. Error bars represent ±1 SEM.

Anchoring Points

Participants drew their maps on templates bearing a START and an END mark. While these were intended to prompt participants to make use of the full A4 sheet provided, thereby aiding map interpretation, we also included their ability to anchor the environments described onto these two points in the analysis. This was found to be affected by Reading, $F(1,33) = 11.52$, $p = .002$, $\eta^2_p = .259$, with better recall during the second drawing compared to the first, M = 79.41%, SEM = 3.1%, and M = 70.22%, SEM = 2.9%, respectively. Performance in this measure was also found to vary as a function of Imagined perspective, $F(1,33) = 26.50$, $p = .001$, $\eta^2_p = .445$. Participants were more likely to be able to map the environments onto the template when asked to generate an allocentric representation of the environments, M = 82.35%, SEM = 3.0%, than when imagining walking through them, M = 67.27%, SEM = 3.1% (Figure 3.23). No significant main effect of Relational term type was found, $F(1,33) = .014$, $p = .908$, $\eta^2_p < .001$. All interactions were non-significant, all ps > .1. Generally, it appears that participants were better able to rescale and project their mental representations onto the templates provided when given instructions to construct allocentric representations.

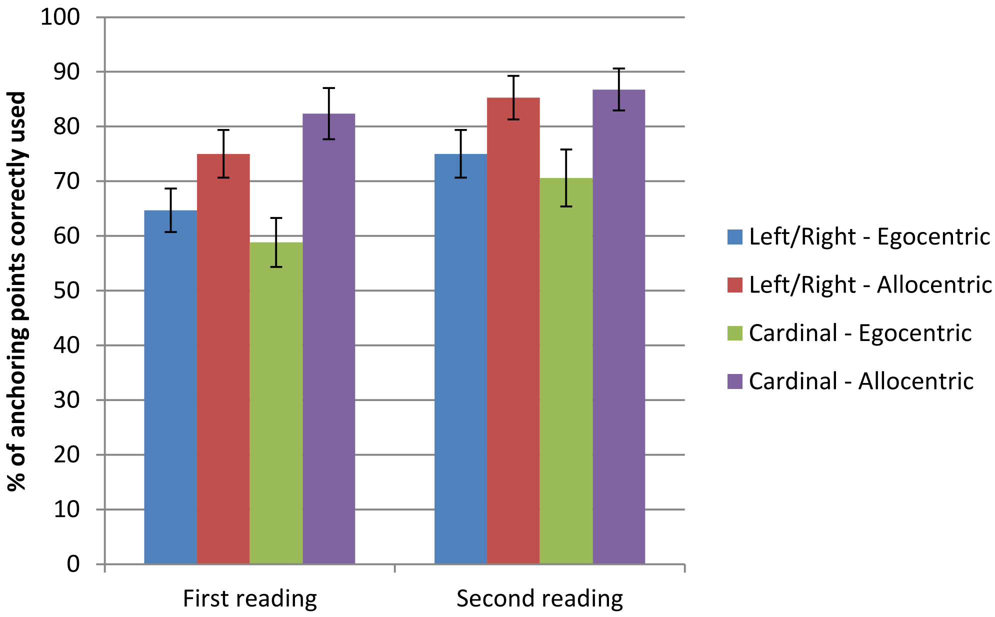


Figure 3.23 - Ability to anchor origin and destination points onto the START and END points provided. Error bars represent ±1 SEM.

Aggregated Map Scores

After computing measures of Landmark (LK) and Configural Knowledge (CK) as in Experiment 1, these were analysed via 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: Egocentric vs Allocentric) x 2(Reading: 1^{st} vs 2^{nd}), Sidak-corrected within-subject ANOVAs.

Significant main effects of Relational term type, $F(1,33) = 12.72$, $p = .001$, $\eta^2_p = .278$ and of Reading, $F(1,33) = 141.91$, $p = .001$, $\eta^2_p = .811$, were found to significantly affect landmark knowledge performance. Aggregated recall of landmark information was better following encoding of left/right descriptions, M = 63.31%, SEM = 2.7%, than following encoding of cardinal descriptions, M = 56.49%, SEM = 3.2%. LK recall also increased between the first and second drawing, M = 46.07%, SEM = 2.8%, and M = 73.73%, SEM = 3.2%. No significant main effect of Imagined perspective was observed, $F(1,33) = 2.65$, $p = .113$, $\eta^2_p = .074$. A significant interaction between Imagined perspective and Reading was also observed, $F(1,33) = 5.32$, $p = .027$, $\eta^2_p = .139$, and analysed further. Performance was found to improve following a second encoding both under egocentric imagery instructions, M = 45.67%, SEM = 3.4% and M = 69.77%, SEM = 3.7%, $F(1,33) = 63.43$, $p < .001$, $\eta^2_p = .658$, and under allocentric imagery instructions, M = 46.48%, SEM = 3.2% and M = 77.69%, SEM = 3.2%, $F(1,33) = 152.93$, $p < .001$, $\eta^2_p = .823$. However, performance following a second encoding was significantly higher under explicit instructions to generate an allocentric perspective compared to egocentric instructions, M = 77.69%, SEM = 3.2%, and M = 69.77%, SEM = 3.7%, respectively, $F(1,33) = 8.93$, $p = .005$, $\eta^2_p = .213$. No such effect was observed following the first reading, M = 46.48%, SEM = 3.2%, and M = 45.67%, SEM = 3.4%, respectively, $F(1,33) = .05$, $p = .816$, $\eta^2_p = .002$.

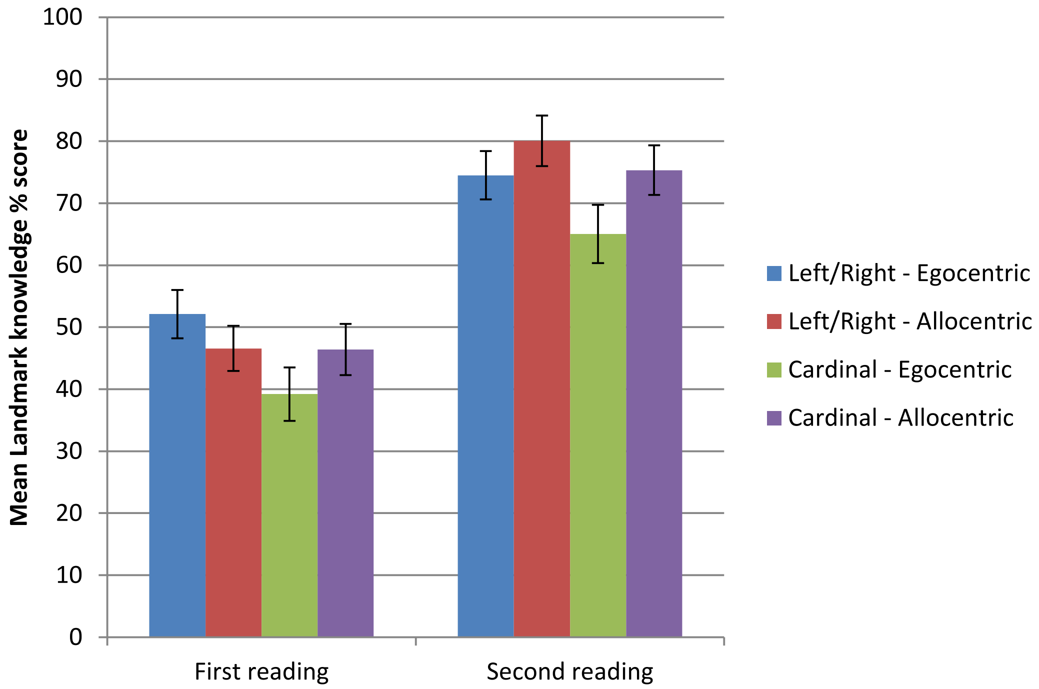


Figure 3.24 - Aggregated Landmark Knowledge performance scores. Error bars represent ±1 SEM.

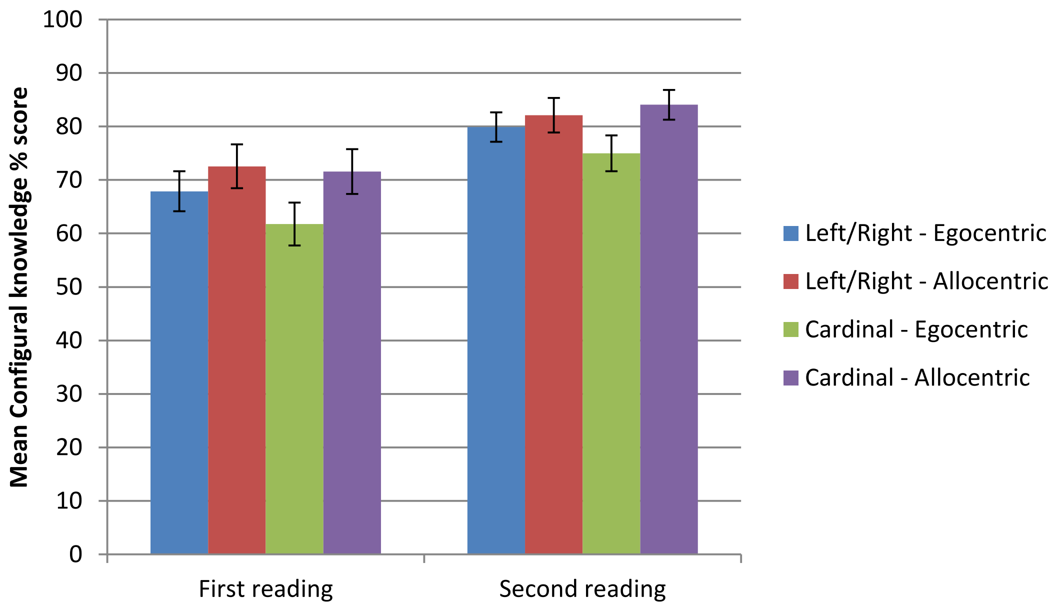


Figure 3.25 - Aggregated Configural Knowledge performance scores. Error bars represent ±1 SEM.

Configural knowledge, on the other hand, was found to be only affected by Reading, $F(1,33) = 31.897$, $p = .001$, $\eta^2_p = .492$, and Imagined perspective, $F(1,33) = 5.560$, $p = .024$, $\eta^2_p = .144$. Participants' recall of these measures was better following encoding while imagining an allocentric perspective, M = 77.57%, SEM = 2.2%, than while imagining an egocentric perspective, M = 71.14%, SEM = 2.6%. No effect of Relational term type or interactions were found, all ps > 1 (see Figure 3.24 and Figure 3.25). Overall, these scores reflect the general pattern observed in the individual behavioural measures, painting a picture of reduced performance in the Cardinal-Egocentric condition.

3.7.2. Questionnaire Results

We then proceeded to analyse participants' responses to the self-report questionnaire items. These were intended to verify that manipulating relational terms and task instructions was having an effect, and to explore and quantify, as much as possible, the subjective, phenomenological experience of spatial imagery construction and manipulation during encoding. However, the questionnaires were administered after each drawing phase rather than immediately after each reading phases, so as not to interfere with encoding and recall of information. An effect of Reading was found only for the first questionnaire item (Ease of perspective generation), indicating that participants found it easier to imagine the perspective they were instructed to generate during the second description encoding, $F(1,24) = 8.15$, $p = .009$, $\eta^2_p = .254$. Responses were therefore averaged across readings, and analysed using 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric), Sidak-corrected within-subject ANOVAs. I must note that the questionnaire was introduced into the experimental design starting with the 10th participant, so responses for only 24 participants were recorded.

Ease of Spatial Perspective Generation

Item 1 on the questionnaire asked participants to rate their agreement with the statement that they had found it easy to generate the spatial perspective they had been explicitly asked to adopt. This ease of imagery was found to be affected only by Imagined perspective, $F(1,24) = 17.52$, $p = .001$, $\eta^2_p = .422$. Relational term type had no significant effect, $F(1,24) = 1.16$, $p = .292$, $\eta^2_p = .046$, and the interaction between the two factors was similarly non-significant, $F(1,24) = .50$, $p = .486$, $\eta^2_p = .020$. Participants generally found it easier to generate allocentric representations of environments, M = 3.6, SEM = .20, rather than egocentric ones, M = 2.6, SEM = .16 (Figure 3.26).

Ease of Perspective Maintenance

Item 2 asked participants to rate their agreement with the statement that they had found it easy to maintain the same spatial perspective throughout the reading phase. Average ratings for this item were also found to be affected by Imagined perspective, $F(1,24) = 7.40$, $p = .012$, $\eta^2_p = .236$. No significant main effect of Relational term type, $F(1,24) = 1.121$, $p = .300$, $\eta^2_p = .045$, or interaction between the two factors, $F(1,24) = .03$, $p = .853$, $\eta^2_p = .001$, were found. As for the previous item, participants reported finding it easier to maintain an allocentric imagined perspective, M = 3.5, SEM = .22, than an egocentric one, M = 2.7, SEM = .21 (Figure 3.27).

Perspective Time Estimates

Items 3 and 4 asked participants to rate their agreement with statements that they had spent most of the encoding time imagining an egocentric or allocentric perspective respectively. With respect to Item 3, only a main effect of Imagined perspective was found, $F(1,24) = 29.44$, $p = .001$, $\eta^2_p = .551$. Relational term had no significant effect, $F(1,24) = 2.05$, $p = .164$, $\eta^2_p = .079$, and no interaction between them was found, $F(1,24) = 3.80$, $p = .063$, $\eta^2_p = .137$. Participants reported spending most of the encoding time imagining an egocentric perspective when they had been instructed to do so, M = 3.2, SEM = .16, but not when they had been instructed otherwise, M = 1.7, SEM = .18.

A significant main effect of Imagined perspective was found to influence responses to Item 4, $F(1,24) = 46.18$, $p < .001$, $\eta^2_p = .658$, indicating that participants reported spending most of the encoding time imagining an allocentric perspective when instructed to do so, M = 4.1, SEM = .17, compared to when they were instructed to do otherwise, M = 2.5, SEM = .17. A significant interaction between Relational term and Imagined perspective was also found to influence responses to Item 4, $F(1,24) = 4.96$, $p = .036$, $\eta^2_p = .171$. Resolving the interaction reinforced the finding, showing that, when instructed to imagine an allocentric perspective, participants did so regardless of whether the description they were reading contained egocentric, M = 4.2, SEM .22, $F(1,24) = 53.52$, $p < .001$, $\eta^2_p = .690$, or cardinal, M = 4.1, SEM = .20, $F(1,24) = 15.36$, $p = .001$, $\eta^2_p = .390$, relational terms . However, when given instructions to imagine an egocentric perspective, participants were more likely to imagine an allocentric one while reading a cardinal description, M = 2.8, SEM = .21, than when reading a left/right one, M = 2.2, SEM = .24, $F(1,24) = 4.89$, $p = .37$, $\eta^2_p = .169$. In summary, it appears that encoding of route descriptions in the Cardina-Egocentric condition made participants unable to maintain a stable mental representation, whether egocentric or allocentric. Figure 3.28 presents a summary of participants' agreement ratings for both questionnaire items.

Perspective Switching

The last questionnaire item asked participants to rate their agreement with the statement that they spent most of the encoding time switching between an egocentric and an allocentric perspective. Only a main effect of Imagined perspective was found for this measure, $F(1,24) = 5.76$, $p = .041$, $\eta^2_p = .163$, indicating increased reported perspective switching under instructions to imagine and maintain an egocentric perspective, M = 2.4, SEM = .19, compared to when allocentric imagery instructions were given, M = 1.9, SEM = .19. However, no effect of Relational term type, $F(1,24) = .267$, $p = .610$, $\eta^2_p = .011$, or interaction between the two factors,

$F(1,24) = .147$, $p = .705$, $\eta^2_p = .006$, was found. While, surprisingly, the average rate of reported perspective switching was generally low and within the “disagree” range for all conditions, it was nevertheless higher in conditions requiring participants to generate egocentric representations of the described routes, but especially in the Cardinal-Egocentric condition (Figure 3.29).

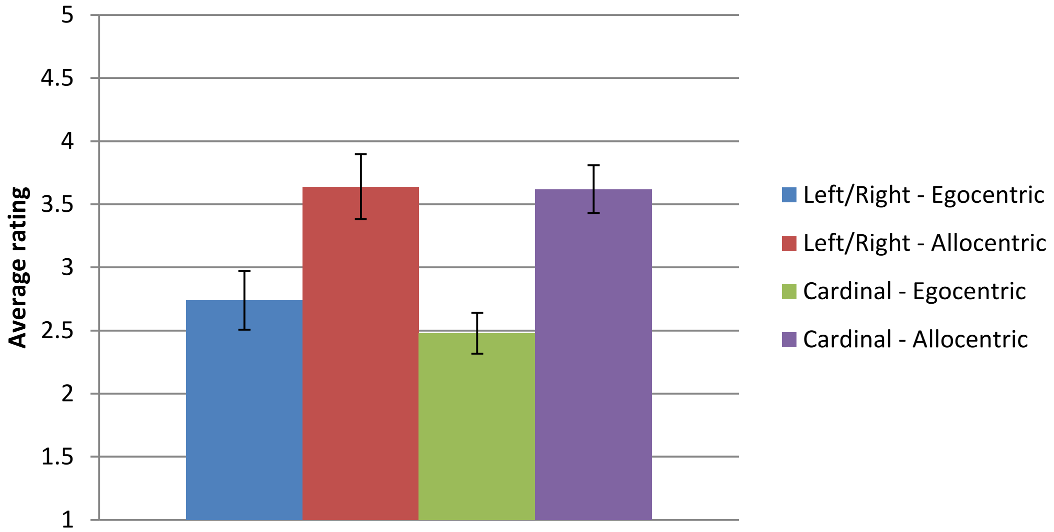


Figure 3.26 - Average rating for the "ease of perspective generation" item. Error bars represent ±1 SEM.

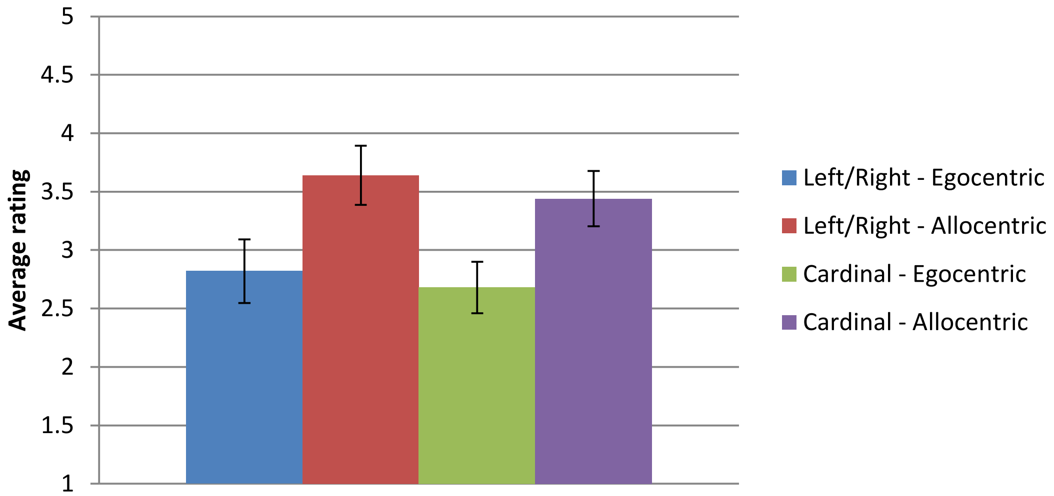


Figure 3.27 - Average rating for the "ease of perspective maintenance" item. Error bars represent ±1 SEM.

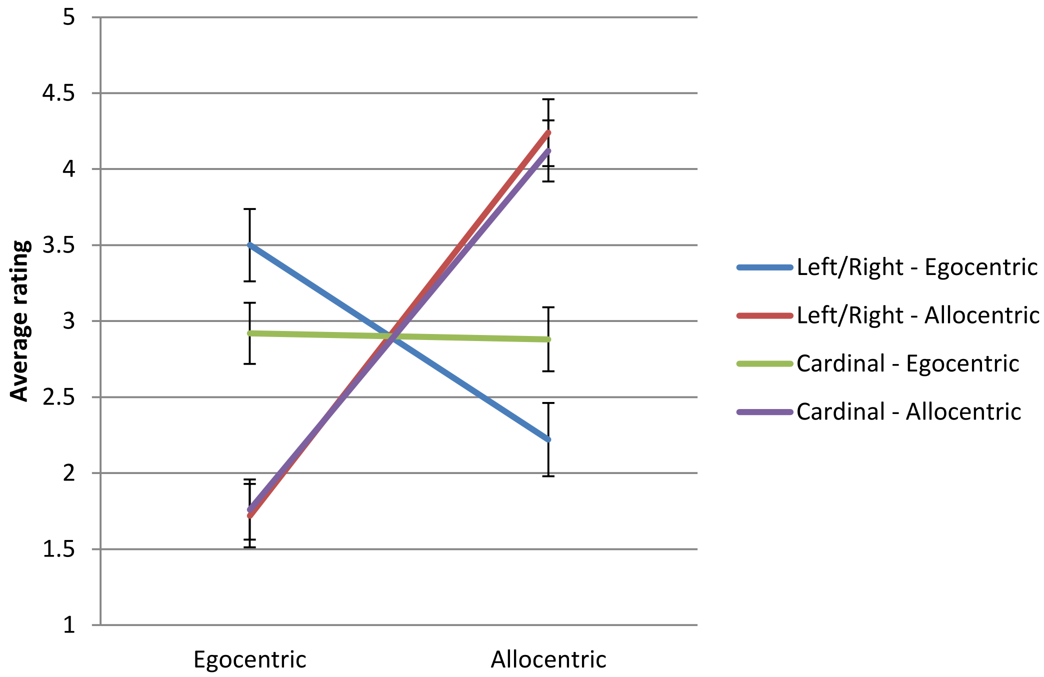


Figure 3.28 - Summary of participants self-report estimates (Y axis) of the perspective they spent most of the encoding time maintaining (X axis) in the different conditions. Error bars represent ±1 SEM.

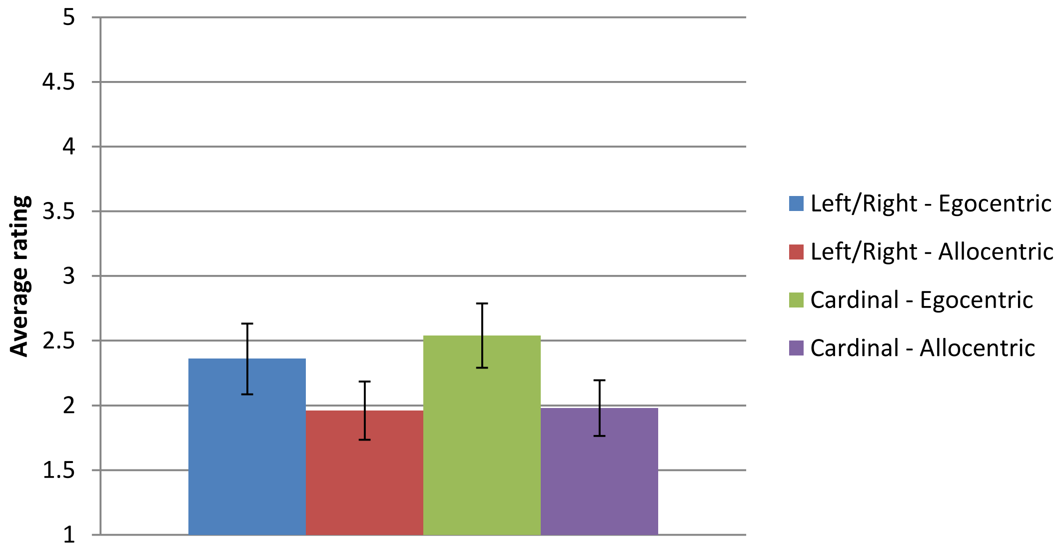


Figure 3.29 – Average rating for the "perspective switching" item. Error bars represent ±1 SEM.

3.7.3. Psychometric Data

A round of one-tailed correlations was performed between the psychometric measures administered and the aggregated map performance scores in order to test the hypothesis that higher performance in the map drawing task would correlate with higher measures of working memory, sense of direction, and navigational skills. See Table 3.7 for average psychometric performance and Table 3.8 for significant correlations.

Table 3.7 – Average psychometric scores.

Psychometric	Average Score	SEM
SBSOD	52.12	1.94
Money's Test (MT)	6.76	2.12
Digit Span BW (DBW)	8.62	.358

*Table 3.8 – One-tailed correlations between aggregated map drawing performance scores and performance on psychometric measures. * p < .05; ** p < .01.*

Condition	Map Knowledge	Reading	SBSOD	DBW	MT
L/R – Ego.	Landmark	First	**-.325***	**-.365***	**-.495****
		Second	-.178	-.095	-.118
	Configural	First	.033	.205	**.366***
		Second	.050	.058	-.042
L/R – Allo.	Landmark	First	-.132	-.019	.157
		Second	-.227	.093	.227
	Configural	First	-.256	.097	**.329***
		Second	-.070	.179	.241
Card. – Ego.	Landmark	First	-.030	.117	-.105
		Second	-.041	**.330***	.171
	Configural	First	.143	**.327***	050
		Second	.131	**.364***	.275
Card. – Allo.	Landmark	First	-.192	**-.297***	-.018
		Second	-.227	-.113	-.098
	Configural	First	.083	.281	.052
		Second	-.137	.217	**.382***

3.7.4. Psychometric Measures and Aggregated Map Scores - Regressions

As in Experiment 1, participants' psychometric scores were entered stepwise into regression models as potential predictors of each dependent variable (LK and CK scores). See Tables 3.9 and 3.10 for a summary of the significant results. Significant predictors of each dependent variable are reported in bold. Only standardised coefficients are reported for independent variables that did not enter the model. Empty cells represent instances in which a particular predictor did not enter the model during step 1.

*Table 3.9 – Significant psychometric predictors of Landmark Knowledge for each reading in each condition. SBSOD = Santa Barbara Sense of Direction Scale; MT = Money's Test; DBW = Digits Span Backward. *p<.05; **p<.005.*

	L/R-Ego. 1 - LK			L/R-Ego. 2 - LK			L/R-Allo. 1 - LK			L/R-Allo. 2 - LK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
SBSOD	-	-	-.26	-	-	-	-	-	-	-	-	-
MT	**-.90**	**.28**	**-.49****	-	-	-	-	-	-	-	-	-
DBW	-	-	-.22	-	-	-	-	-	-	-	-	-
	Card.-Ego. 1 - LK			Card.-Ego. 2 - LK			Card.-Allo. 1 - LK			Card.-Allo. 2 - LK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
SBSOD	-	-	-	-	-	-	-	-	-	-	-	-
MT	-	-	-	-	-	-	-	-	-	-	-	-
DBW	-	-	-	-	-	-	-	-	-	-	-	-

*Table 3.10 - Significant psychometric predictors of Configural Knowledge for each reading in each condition. SBSOD = Santa Barbara Sense of Direction Scale; MT = Money's Test; DBW = Digits Span Backward. *p<.05; **p<.005.*

	L/R-Ego. 1 - CK			L/R-Ego. 2 - CK			L/R-Allo. 1 - CK			L/R-Allo. 2 - CK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
SBSOD	-	-	-.01	-	-	-	-	-	-	-	-	-
MT	**.64**	**.28**	**.36***	-	-	-	-	-	-	-	-	-
DBW	-	-	.08	-	-	-	-	-	-	-	-	-
	Card.-Ego. 1 - CK			Card.-Ego. 2 - CK			Card.-Allo. 1 - CK			Card.-Allo. 2 - CK		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
SBSOD	-	-	-	-	-	.05	-	-	-	-	-	-.18
MT	-	-	-	-	-	.16	-	-	-	-	-	.09
DBW	-	-	-	**3.41**	**1.54**	**.36***	-	-	-	**.49**	**.21**	**.38***

3.7.5. Preliminary Discussion

The analyses performed on the behavioural performance scores in the map-drawing task are broadly in agreement with the hypothesis that the use of allocentric relational terms (cardinal points) can result in an interference effect if an egocentric spatial perspective is being maintained, replicating the observation of generally reduced performance following cardinal description encoding in Experiment 1. This interference is not present when participants are explicitly asked to generate a survey-like representation of the environments described, regardless of the relational terms used in the descriptions. This is particularly evident in measures of sequential and spatial landmark recall, as well as in the measure of participants' ability to anchor origin and destination landmarks on the START and END markers provided. Despite the reduced variance due to there being only two potential anchoring points per map, the difference is striking. If confirmed, this might indicate that, following egocentric encodings, participants retrieve information sequentially as they experienced it, starting from the START marker and paying little attention to the END marker until the end of the route. Conversely, allocentric encodings, and the resulting survey-like representations, prompt participants to be more mindful of the global configuration of the environment, enabling them to scale their spatial representation to match the template provided.

This pattern of behavioural results is also found in the agreement ratings participants gave to the questionnaire items. When given explicit instructions to build a map-like, survey mental representation of the environments described, they reported doing so and maintaining such representations significantly more easily than when they were instructed to imagine walking through the environments described. Participants found it especially difficult to maintain any given spatial perspectives while encoding cardinal-egocentric descriptions, a sign of considerable interference between competing reference frames. However, they were no more likely to report a considerable rate of perspective switching in this particular condition compared to the other three, a possible indication of introspective limitations that should be taken into account. In summary, an allocentric encoding does appear to be easier and more stable, allowing for higher performance during map drawing tasks. A purely egocentric encoding can be a successful but more cognitively taxing strategy, likely due to the need to transform a first-person representation of the environment into a map-like network of spatial relations. As predicted, this was especially difficult when spatial relations were provided allocentrically (i.e. using cardinal terms), requiring participants to remap their mental cardinal compass after each egocentric turn.

Despite a number of significant correlations, few reliable psychometric predictors of map drawing performance were identified in the stepwise regression models. The most reliable pattern reveals negative correlations between landmark knowledge performance after the first reading of the purely egocentric condition (Left/Right-Egocentric) and all three psychometric measures used. This could be a possible indication that reliance on sequential recall of information (if, indeed, it is more prominent when information is encoded egocentrically) might be a preferential strategy for individuals with a lower ability to manipulate complex spatial representations in working memory. However, it is also likely that the pattern of correlations might change if participants were presented with even more complex descriptions with a higher number of informational items to encode and process. This makes interpreting these correlations difficult. Performance on the DBW significantly predicted CK acquisition or use after the second encoding of cardinal-egocentric descriptions, possibly reflecting higher task demands and working memory load in that condition (consistent with results in the previous study).

Performance on the Money's Test, on the other hand, negatively correlated with (and was a significant predictor of) performance on measures of landmark knowledge after the first encoding of Left/Right-Egocentric descriptions. This is in contrast to findings by Tom and Tversky (2012), who found MT scores to positively correlate with correct recall of landmarks' spatial locations (see Section 2.3). Performance on both SBSOD and DBW followed the same pattern of negative correlations with this particular condition. However, performance on the Money's Test positively correlated with (and was a significant predictor of) CK measures after the first Left/Right-Egocentric encoding. This seems to suggest that increasing ability on the Money's Test corresponds to an increased ability to derive an environment's global configuration from spatial route descriptions, but to the detriment of landmark knowledge when a first-person perspective is maintained. This is consistent with the somewhat lower performance in sequential and spatial landmark recall measures in the Left/Right-Egocentric condition.

3.7.6. Eye Movement Measures

As in Experiment 1, several eye movement measures were extracted from the eye tracking data. These were: average Total Dwell Time (DT); average First-, Second-, Third-Pass Dwell Time; average Regression Path Duration (RPD). They were analysed via a set of 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: Egocentric vs Allocentric) x 2(Salience: Salient vs Non-Salient) within-subject ANOVAs.

Total Dwell Time (DT)

Analysis of the total DT on landmark words revealed no significant main effects of Relational term type, $F(1,33) = 3.09$, $p = .088$, $\eta^2_p = .086$, or Imagined perspective, $F(1,33) = .15$, $p = .698$, $\eta^2_p = .005$. It did, however, reveal a significant main effect of Landmark salience, $F(1,33) = 45.13$, $p < .001$, $\eta^2_p = .578$, indicating that, overall, navigationally salient landmark words were fixated significantly longer than non-navigationally salient ones, M = 1842 ms, SEM = 143 ms, and M = 1074 ms, SEM = 76 ms, respectively. Additionally, a significant interaction between Relational term type and Salience was found, $F(1,33) = 13.59$, $p = .001$, $\eta^2_p = .292$. A test of simple main effects showed a significant main effect of landmark salience on both levels of Relational term type, with salient landmark words being fixated longer than non-salient landmark words. However, this effect was stronger in conditions employing egocentric relational terms, $F(1,33) = 32.10$, $p = .001$, $\eta^2_p = .493$, than in conditions employing cardinal terms, $F(1,33) = 11.11$, $p = .002$, $\eta^2_p = .252$. Salient landmarks were fixated considerably longer than non-salient landmarks during the encoding of descriptions using "left" and "right" to define spatial relations, M = 2186 ms, SEM = 221 ms, and M = 974 ms, SEM = 90 ms, respectively. However, this difference was markedly reduced during the encoding of cardinal descriptions, M = 1497 ms, SEM = 130 ms, and M = 1174 ms, SEM = 111 ms, for salient and non-salient landmark words respectively. Additionally, while salient landmark words were fixated significantly longer than non-salient landmark words when egocentric relational terms were used compared to when cardinal terms were used, $F(1,33) = 9.45$, $p = .004$, $\eta^2_p = .223$, no such difference was found for non-salient landmarks, $F(1,33) = 2.29$, $p = .140$, $\eta^2_p = .065$ (Figure 3.30).

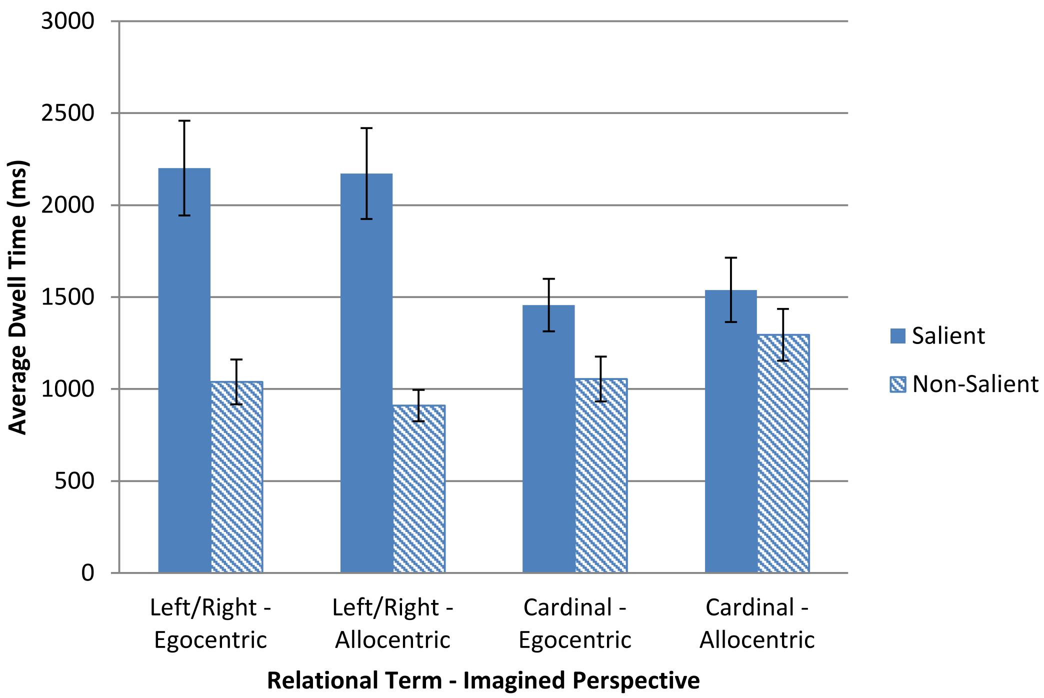


Figure 3.30 - Average dwell time on landmark words. Error bars represent ±1 SEM.

A similar analysis repeated on the average DT on relational terms revealed a significant main effect of Relational term type, $F(1,33) = 23.59$, $p < .001$, $\eta^2_p = .417$, and a significant main effect of Salience (Landmark presence), $F(1,33) = 13.31$, $p = .001$, $\eta^2_p = .287$, but no main effect of Imagined perspective, $F(1,33) = .50$, $p = .483$, $\eta^2_p = .015$. Overall, egocentric relational terms were fixated longer than cardinal relational terms, M = 1951 ms, SEM = 135 ms, and M = 1252 ms, SEM = 125 ms, respectively, and relational terms denoting a turn with landmark longer than other relational terms, M = 1859 ms, SEM = 134 ms, and M = 1344 ms, SEM = 125 ms, respectively.

The analysis also revealed two significant interactions: one between Relational term type and Salience (Landmark presence), $F(1,33) = 27.08$, $p = .001$, $\eta^2_p = .451$, and one between Imagined perspective and Salience (Landmark presence), $F(1,33) = 7.79$, $p = .009$, $\eta^2_p = .191$. Resolving these interactions revealed a main effect of Relational term type on average DT on salient relational terms (i.e. relational terms describing turns where landmarks were present), $F(1,33) = 37.92$, $p = .001$, $\eta^2_p = .535$, but no such effect on average DT on non-salient relational terms, $F(1,33) = .06$, $p = .806$, $\eta^2_p = .002$. In other words, during encoding of cardinal descriptions relational terms denoting turns where landmarks were present were fixated significantly less than during descriptions containing egocentric relational terms, M = 1140 ms, SEM = 109 ms, $F(1,33) = 29.31$, $p < .001$, $\eta^2_p = .470$, and M =

2578 ms, SEM = 227 ms, $F(1,33) = 1.89$, $p > .1$, $\eta^2_p = .054$, respectively. Additionally, explicit instructions to imagine an allocentric perspective during encoding appeared to increase the average DT on relational terms unaccompanied by landmarks compared to when egocentric instructions were given, $F(1,33) = 6.97$, $p = .013$, $\eta^2_p = .174$, M = 1467 ms, SEM = 143 ms, and M = 1220 ms, SEM = 123 ms, respectively. However, no such difference was found for relational terms describing turns with landmarks, $F(1,33) = 1.82$, $p = .186$, $\eta^2_p = .052$, M = 1787 ms, SEM = 146 ms, and M = 1931 ms, SEM = 143 ms (Figure 3.31). Ultimately, the resulting pattern is one of reduced DT difference between salient and non-salient relational terms when these are cardinal terms rather than left/right.

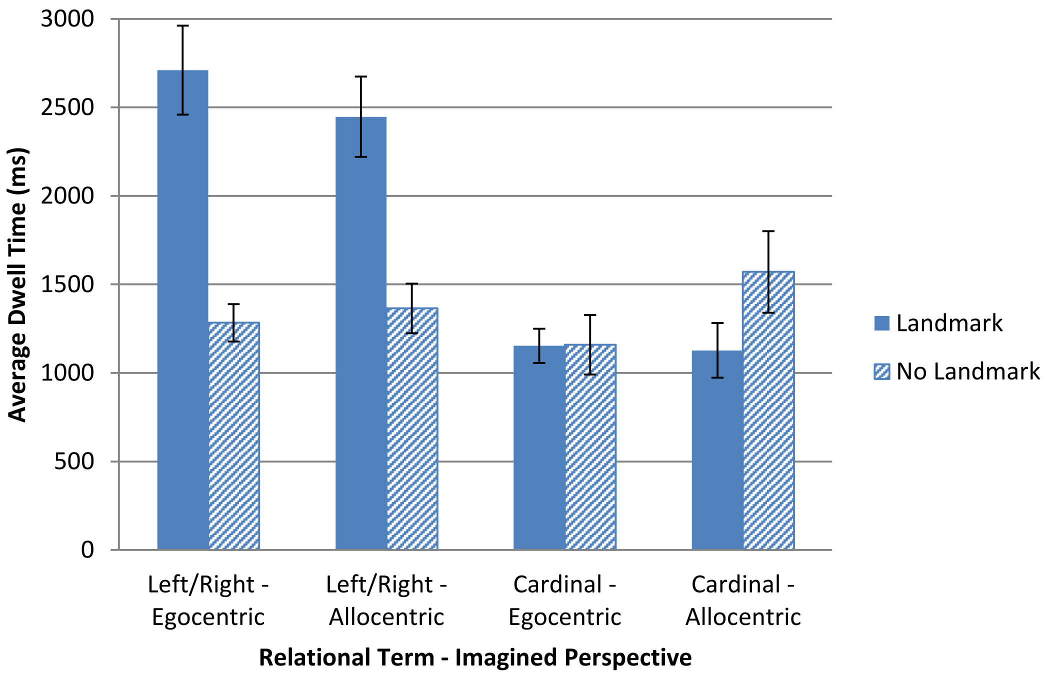


Figure 3.31 - Average dwell time on relational terms. Error bars represent ±1 SEM.

First-, Second-, Third-Pass Dwell Time

As in Experiment 1, average DT was split into different rounds of fixations, providing a measure of salience for the areas of interest at different stages of information integration. For DT on landmark words, a significant main effect of Relational term type was found during all three rounds of fixations, $F(1,32) = 18.89$, $p < .001$, $\eta^2_p = .371$, $F(1,30) = 14.60$, $p = .001$, $\eta^2_p = .327$, and $F(1,21) = 14.09$, $p = .001$, $\eta^2_p = .40$. This showed that egocentric relational terms were fixated longer than cardinal relational terms during first, M = 446 ms, SEM = 37 ms, and M = 296 ms, SEM = 15 ms, second, M = 520 ms, SEM = 51 ms, and M = 362 ms, SEM = 25 ms, and third pass, M = 578 ms, SEM = 73 ms, and M = 310 ms, SEM = 18 ms, respectively. No main effect of Imagined perspective or Landmark salience was found during first or second pass, both ps > .1. However, a main effect of Landmark salience on DT was observed during third pass, $F(1,21) = 6.95$, $p = .015$, $\eta^2_p = .249$, indicating longer DT on salient landmark words, M = 525 ms, SEM = 61 ms, compared to non-salient landmark words, M = 363 ms, SEM = 37 ms.

Additionally, an interaction between Relational term type and Landmark salience was found on all three measures. This was strongest during first pass, $F(1,32) = 18.77$, $p = .001$, $\eta^2_p = .370$, but still present during second, $F(1,30) = 8.26$, $p = .007$, $\eta^2_p = .216$, and third pass, $F(1,21) = 7.00$, $p = .015$, $\eta^2_p = .250$. During first pass, salient landmarks were fixated significantly longer when egocentric relational terms were used, M = 525 ms, SEM = 55 ms, than when allocentric cardinal terms were, M = 261 ms, SEM = 19 ms, $F(1,32) = 25.28$, $p = .001$, $\eta^2_p = .441$. However, no such difference was found for non-salient landmarks, $F(1,32) = 1.40$, $p = .245$, $\eta^2_p = .042$, M = 368 ms, SEM = 31 ms, and M = 330 ms, SEM = 20 ms. Similarly, salient landmark words were fixated significantly longer than non-salient landmark words during first-pass encoding of left/right descriptions, $F(1,32) = 10.33$, $p = .003$, $\eta^2_p = .244$, for a difference of M = 157 ms. During first-pass encoding of cardinal descriptions, non-salient landmark words were fixated longer than salient landmark words, $F(1,32) = 8.97$, $p = .005$, $\eta^2_p = .219$, for a difference of 68 ms.

During second pass, only a main effect of Relational term type, $F(1,30) = 14.60$, $p = .001$, $\eta^2_p = .327$, and a significant interaction between Relational term type and Landmark salience were found, $F(1,30) = 8.26$, $p = .007$, $\eta^2_p = .216$, all other $ps > .1$. The main effect indicated that landmark words in general were fixated longer during second-pass encoding of left/right descriptions, M = 520 ms, SEM = 51 ms, compared to cardinal descriptions, M = 362, SEM = 25 ms. Resolving the interaction revealed that salient landmarks elicited longer average DTs during encoding of egocentric descriptions, M = 585 ms, SEM = 56 ms, relative to allocentric descriptions, M = 335 ms, SEM = 26 ms, $F(1,30) = 21.31$, $p = .001$, $\eta^2_p = .415$. No such difference was found non-salient landmark words, M = 454 ms, SEM = 56 ms, and M = 388 ms, SEM = 35 ms, $F(1,30) = 1.73$, $p = .198$, $\eta^2_p = .055$. Similarly, salient landmark words were fixated significantly longer than non-salient landmark words during second-pass encoding of left/right descriptions, $F(1, 30) = 7.94$, $p = .008$, $\eta^2_p = .209$, but no such difference was found during second-pass encoding of cardinal descriptions, $F(1,30) = 2.12$, $p = .155$, $\eta^2_p = .066$.

Still during third pass, significant main effects of Relational term type, $F(1,21) = 14.09$, $p = .001$, $\eta^2_p = .402$, and of Landmark salience, $F(1,21) = 6.95$, $p = .015$, $\eta^2_p = .249$, were found. These indicated longer overall dwell times on landmark words during third-pass encoding of left/right descriptions, M = 578 ms, SEM = 73 ms, compared to cardinal descriptions, M = 310 ms, SEM = 18 ms, and longer overall dwell times on salient landmark words than on non-salient landmark words, M = 525 ms, SEM = 61 ms, and M = 363 ms, SEM = 37 ms, respectively. Additionally, an interaction between Relational term type and Landmark salience was found, $F(1,21) = 7.00$, $p = .015$, $\eta^2_p = .250$. A test of simple main effects on the interaction revealed that salient landmark words elicited significantly longer DTs than non-salient landmark words when egocentric relational terms were used, M = 743 ms, SEM = 124 ms, and M = 413 ms, SEM = 48 ms, respectively, $F(1,21) = 7.98$, $p = .010$, η^2_p = .275. However, no such difference was found during third-pass encoding of cardinal descriptions, $F(1,21) = .02$, $p = .868$, $\eta^2_p = .001$, M = 306 ms, SEM = 24 ms, and M = 314 ms, SEM = 32 ms, respectively (Figure 3.32).

Analysing different rounds of fixations on relational terms, however, yielded more varied results. While no main effects of Relational term type or Imagined perspective were found, both ps > .1, a main effect of Landmark presence (i.e. salience) was found on average DT during the first round of fixations, $F(1,31) = 5.28$, $p = .028$, $\eta^2_p = .146$. Relational terms describing turns not associated with landmarks were fixated significantly longer than their landmarked counterparts, M = 437 ms, SEM = 43 ms, and M = 363 ms, SEM = 28 ms, respectively. No interactions were found between any of the factors, all ps > .1.

During second pass, no significant main effect of Relational term type, Imagined perspective or Landmark presence were found, all ps > .1, but a significant interaction between Relational term type and Landmark presence was observed, $F(1,26) = 4.72$, $p = .039$, $\eta^2_p = .154$. Decomposing this interaction revealed that, when “left” and “right” were used to describe spatial relations, these were fixated significantly longer when they described turns accompanied by landmarks, M = 508 ms, SEM = 55 ms, than when they described turns without landmarks, M = 390 ms, SEM = 26 ms, $F(1,26) = 6.23$, $p = .019$, $\eta^2_p = .193$. No such difference was observed for cardinal relational terms, M = 379 ms, SEM = 40 ms, and M = 462 ms, SEM = 62 ms, $F(1,26) = 1.34$, $p = .257$, $\eta^2_p = .049$.

During third pass, no significant main effect of Relational term type, Imagined perspective or Landmark presence were found, all ps > .1, but a significant interaction was found between Imagined perspective and Landmark presence, $F(1,19) = 4.72$, $p = .043$, $\eta^2_p = .199$. Simple main effects revealed a significant difference in average DT between relational terms describing turns with or without a landmark, M = 493 ms, SEM = 51 ms, and M = 362 ms, SEM = 35 ms, respectively, $F(1,19) = 4.93$, $p = .039$, $\eta^2_p = .206$, but only when participants were explicitly instructed to generate and maintain an egocentric, street-level perspective. When instructed to generate and maintain an allocentric, map-like perspective, no such difference was found, M = 426 ms, SEM = 57 ms, and M = 474 ms, SEM = 74 ms, $F(1,19) = .38$, $p = .544$, $\eta^2_p = .020$ (Figure 3.33).

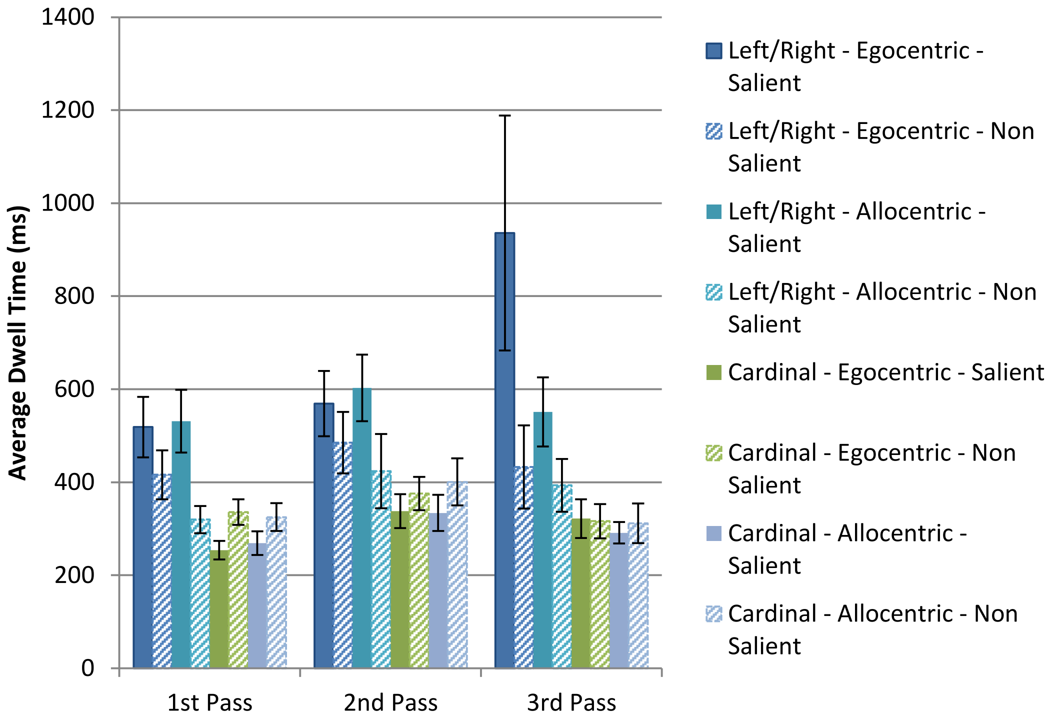


Figure 3.32 - Dwell time on salient and non-salient landmark words during the first, second, and third (plus subsequent) rounds of fixations. Error bars represent ±1 SEM.

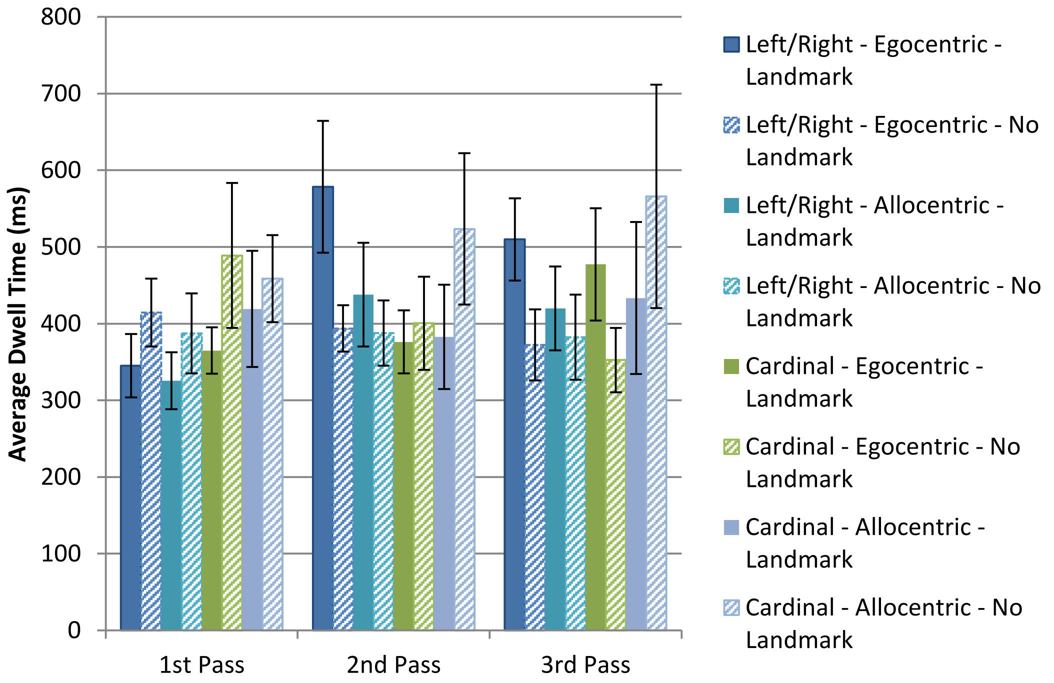


Figure 3.33 - Dwell time on relational terms denoting turns with and without landmarks during the first, second, and third (plus subsequent) rounds of fixations. Error bars represent ±1 SEM.

Regression Path Duration (RPD)

Analysing the duration of regression paths generated by landmark words I observed significant main effects of Relational term type, $F(1,32) = 8.47$, $p = .007$, $\eta^2_p = .209$, and Salience, $F(1,32) = 64.05$, $p = .001$, $\eta^2_p = .667$, as well as a borderline significant main effect of Imagined perspective, $F(1,32) = 3.89$, $p = .057$, $\eta^2_p = .108$. These resulted in longer RPD being generated by landmark words during encoding of left/right descriptions, M = 3526 ms, SEM = 284 ms, compared to cardinal descriptions, M = 2777 ms, SEM = 256 ms, but also in marginally longer RPD while imagining an allocentric spatial perspective, M = 3395 ms, SEM = 255 ms, as opposed to a first-person perspective, M = 2908 ms, SEM = 280 ms. Non-salient landmark words were also found to generate longer RPD than navigationally salient ones, M = 4606 ms, SEM = 397 ms, and M = 1697 ms, SEM = 147 ms, respectively.

A significant interaction between Relational term type and Salience, $F(1,32) = 6.04$, $p = .020$, $\eta^2_p = .159$, and an almost significant interaction between Imagined perspective and Salience, $F(1,32) = 3.49$, $p = .071$, $\eta^2_p = .098$, were also detected. Decomposing the significant interaction revealed a simple main effect of Relational term type on RPD generated by salient landmarks, with these generating significantly longer regression paths when egocentric relational terms are used, M = 2424 ms, SEM = 247 ms, than when cardinal relational terms are used, M = 971 ms, SEM = 133 ms, $F(1,32) = 29.95$, $p = .001$, $\eta^2_p = .483$. No such effect was found for non-salient landmarks, M = 4628 ms, SEM = 447 ms, and M = 4584 ms, SEM = 477 ms, $F(1,32) = .009$, $p = .926$, $\eta^2_p = .001$ (Figure 3.34).

Main effects of Imagined perspective, $F(1,31) = 19.15$, $p < .001$, $\eta^2_p = .382$, and Landmark presence, $F(1,31) = 28.28$, $p < .001$, $\eta^2_p = .477$, were also found to influence RPD generated by relational terms, but no main effect of Relational term type, $F(1,31) = .02$, $p = .872$, $\eta^2_p = .001$. In this sense, relational terms generated longer RPD while participants actively imagined an allocentric perspective, M = 2334 ms, SEM = 207 ms, compared to when they were imagining an egocentric perspective, M = 1400 ms, SEM = 100 ms. Furthermore, relational terms denoting turns at landmark locations generated longer RPD than other relational terms, M = 2417 ms, SEM = 213 ms, and M = 1317 ms, SEM = 76 ms, respectively.

In addition, significant interactions between Relational term type and Imagined perspective, $F(1,31) = 8.75$, $p = .006$, $\eta^2_p = .220$, and between Imagined perspective and Landmark presence, $F(1,31) = 8.62$, $p = .006$, $\eta^2_p = .218$, were found. Relational terms describing turns associated with landmarks were found to elicit significantly longer regression paths, M = 2208 ms, SEM = 212 ms, than other relational terms, M = 591 ms, SEM = 61 ms, when participants were instructed to imagine egocentric perspectives, $F(1,31) = 45.55$, $p = .001$, $\eta^2_p = .595$. This difference only approached significance when participants were asked to imagine an allocentric view of the environment described, $F(1,31) = 3.75$, $p = .062$, $\eta^2_p = .108$, M = 2625 ms, SEM = 330 ms, and M = 2043 ms, SEM = 146 ms, respectively. However, relational terms describing turns not associated with landmarks were also found to elicit significantly longer regression paths when allocentric instructions were provided, M = 2043 ms, SEM = 146 ms, compared to when egocentric instructions were, M = 591 ms, SEM = 61 ms, $F(1,31) = 78.99$, $p = .001$, $\eta^2_p = .718$ (Figure 3.35).

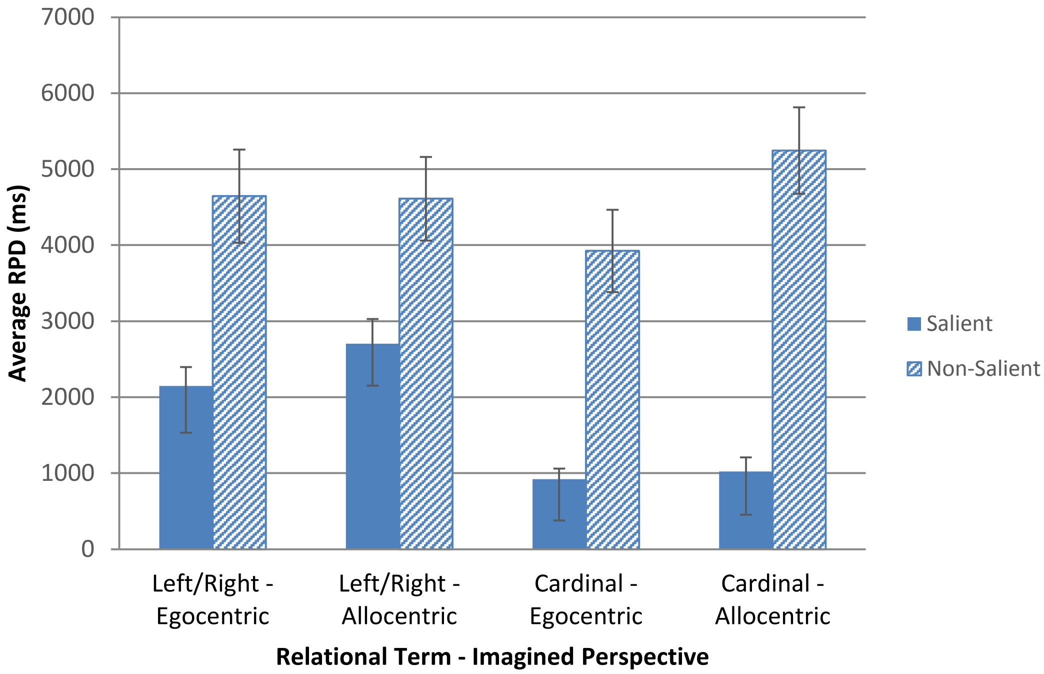


Figure 3.34 - Mean duration of regression paths (RPD) generated by salient and non-salient landmark words. Error bars represent ±1 SEM.

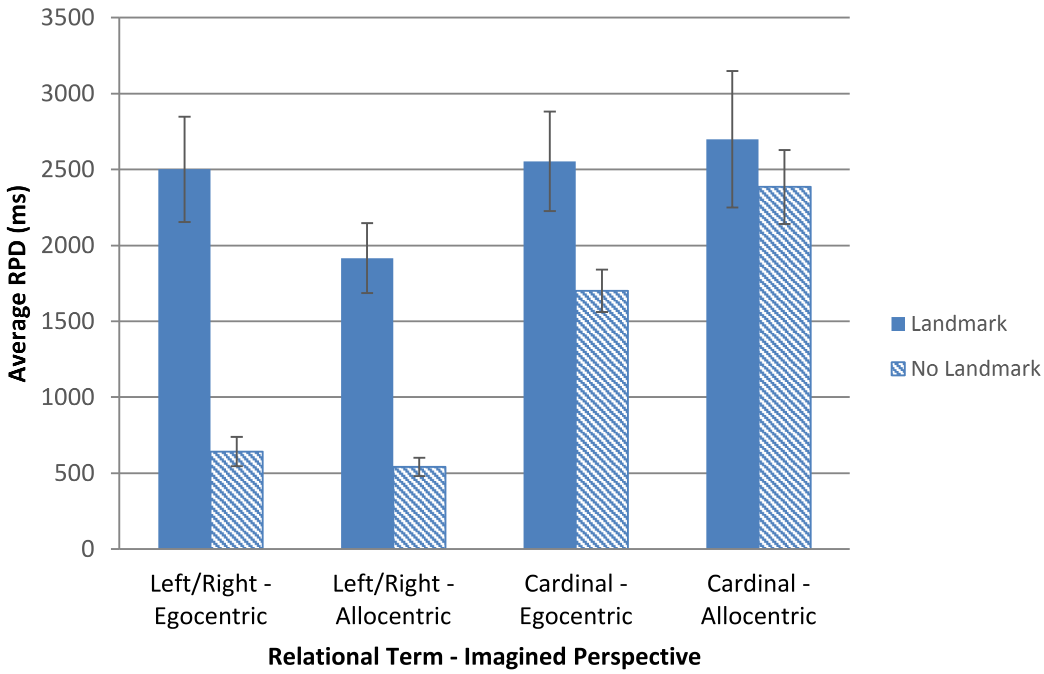


Figure 3.35 - Mean duration of regression paths (RPD) generated by relational terms. Error bars represent ±1 SEM.

3.8. General Discussion

The aim of this experiment was to clarify previous results and to attempt to disentangle different imagined spatial perspectives during encoding and retrieval of spatial route descriptions. By measuring behavioural performance, encoding measures, and self-reported phenomenology, I intended to investigate the possible interference pattern that could be expected to arise when participants are provided with explicit representational instructions (in essence, explicit strategies to observe during task performance) that may or may not interfere with the representations engendered by different types of relational terms.

In particular, I predicted that attempting to encode spatial relations described in allocentric, cardinal terms while maintaining an egocentric view of the environment should generate the highest degree of reference frame interference (due to the need to remap environment-centred cardinal directions after each turn), and possibly mirror the performance observed in the allocentric condition in the previous experiment. This was indeed the case, both in behavioural and phenomenological terms, with average recall performance in the map-drawing task being lowest in this condition, and participants reporting no significant stable preference for either imagined perspective while encoding this type of description. Besides confirming the hypothesis, the reported interference also works as a good confirmation of the critical manipulation and of participant compliance.

However, the reported rate of perspective switching was not significantly higher in the cardinal-egocentric condition. This could possibly indicate limitations intrinsic in the ability of participants to introspect and quantify certain aspects of their phenomenological experience under considerable interference, or the use of a different, unexpected strategy in order to attempt to accomplish the task. Nevertheless, by the second recall phase, performance in the cardinal-egocentric condition was manifestly lower in most map scoring measures. Additionally, while not always significant, we identified a general trend of increased performance in recalling left/right-allocentric descriptions relative to purely egocentric descriptions, which could support the hypothesis formulated in Experiment 1 concerning the cognitive cost of transforming a purely egocentric representation into a map-like representation during map drawing phases. Furthermore, as in Experiment 1, performance in simple landmark recall, and finer-grained recall of sequentially and spatially correct landmarks appeared to be dissociated. This finding should caution against using measures of simple semantic recall as a gauge of spatial understanding.

A particularly interesting finding involved the ability of participants to connect both the beginning and end of their route drawings to the START and END points provided on the template. While the use of only two anchoring points leads to rather low variance, perhaps cautioning against reading too much in this particular measure, the pattern is nonetheless quite interesting. Participants appeared far more likely to correctly anchor their representations to both markers provided when an allocentric representation was prompted, with 44% of the variance in this measure explained by the type of perspective being actively imagined during encoding. My tentative interpretation of this result would see explicit egocentric instructions as generating a strictly sequential encoding and recall, and a reduced attention to the global layout of the environment. This could feasibly affect the scalability of the resulting spatial representations and participants' ability to map them onto the templates provided. This particular finding appears to be corroborated by participants' reported ease in generating and maintaining allocentric representations, when prompted to do so, regardless of the relational terms used.

With regards to the eye movement measures, I was able to replicate the finding of a reduced difference in dwell time between navigationally salient and non-salient landmarks during cardinal description encoding. This was particularly evident when participants were instructed to imagine an allocentric perspective, whereas some degree of difference persisted when an incompatible egocentric perspective was prompted. This latter pattern appears to more closely match that found in the allocentric (cardinal) condition in Experiment 1 (Figure 3.8), supporting the conclusion that in that case participants tended to maintain an egocentric view of the environment, resulting in interference. Unfortunately, this study cannot speak as to whether that tendency was a general preference, partly supporting traditional hierarchical theories of spatial model constructions, or simply the result of the sequential nature of the route descriptions provided. However, when cardinal relational terms were used, regardless of the spatial perspective being actively imagined, the time spent fixating navigationally salient landmarks dropped significantly compared to when egocentric relational terms are used (again, regardless of the instructions provided).

Cardinal relational terms also appeared to significantly reduce the duration of regression paths generated by salient landmarks (Figure 3.34). Interestingly, when egocentric relational terms were paired with instructions to maintain an allocentric perspective, the average duration of regression paths generated by said relational terms was shorter than during purely egocentric encoding (particularly for relational terms describing turns not associated with landmarks). This could indicate that an allocentric, map-like representation (which participants reported finding easier on average) paired with the more familiar "left" and "right" might make overall encoding easier, leading to fewer regressions being required to integrate turns into the global environment. However, the generally observed pattern of regression path durations in this experiment was rather different compared to that found in Experiment 1. Here, non-salient landmarks generated considerably longer regression paths than salient landmarks in all conditions, whereas the opposite was true for the egocentric condition in the previous experiment. A similar reversal between experiments was found for the average duration of regression paths generated by relational terms. These discrepancies in regression path cannot easily be accounted for at present.

Globally, this experiment has yielded interesting insights into different aspects of spatial cognition as it applies to language. On the one hand, the available data speak to the possibility of, at least partly, separating different imagined spatial perspectives for the purpose of analysing their different behavioural, phenomenal, and neural correlates. The current paradigm appears to be a solid factorial design that could allow for the analysis of the relative, but simultaneous, contributions of different relational terms and explicit imagery instructions to spatial mental model construction. However, empirical questions remain as to how to improve the task used to study spatial information recall. In particular, the relatively low number of information tokens in our descriptions reduces the overall variance available in the map measures and may hide what could otherwise be stronger effects observable in our behavioural scores. Although the current number was motivated by a necessity for task feasibility, the case could be made that more complex descriptions could aid in the study of encoding difficulties and in determining the threshold at which progressive spatial disorientation begins to set in. Even more crucially, a higher number of landmarks might allow testing for the effect of navigational salience also on measures of landmark recall. On the other hand, the fact that different reading patterns and condition-specific strengths and weaknesses can arise already during the first encoding of route descriptions supports the idea that participants can be made to generate and commit to specific representational formats during the encoding of language-mediated spatial information. Although the separation of different imagined perspectives may run against the ecological tendency towards parallel processing (Gramann, 2013), the possibility of experimentally distinguishing, at least in part, between egocentric and allocentric perspectives could be extremely valuable to the study of spatial language and cognition in general.

In Experiment 3 I maintained the same full-factorial design used in Experiment 2, but replaced the map-drawing task with a map-verification task during eye-tracking. This modification was intended to allow better comparison between dependent variables during encoding and during test. More specifically, I investigated whether the patterns of attention allocation (measured as total DT) to landmark words would translate to similar patterns of attention allocation to landmark regions of sketch maps.

Experiment 3: Eye Movements in Map Verification

3.9. Experiment 3: Introduction

In Experiments 1 and 2, the high temporal and spatial resolution afforded by eye-tracking was exploited to gain new insights into the processes that underlie spatial route description encoding under a variety of linguistic and cognitive conditions. I strategically modulated the reference frame implied by the relational terms used in the descriptions used, as well as the spatial perspective actively imagined by participants during encoding by providing them with explicit encoding strategies. The goal of this combined manipulation was to experimentally control for the ecological tendency of participants to construct multiple reference frames in parallel (Burgess, 2006; Gramann, 2013) when acquiring spatial information, in order to better tease out the behavioural and eye movement correlates of distinct reference frames during spatial language processing.

The key finding was a modulation of landmark salience as a function of the relational terms used in the description. More specifically, the use of cardinal terms (as opposed to the egocentric “left” and “right”) reduced the difference in dwell time between navigationally salient and non-salient landmarks (i.e. between those located at turn points vs those located along path legs). However, the lack of eye-tracking measures during the test phase in Experiments 1 and 2 prevented the acquisition of an equally fine understanding of the time-course of spatial information recall, and of goal-oriented manipulation of mental representations. More specifically, it did not allow to test whether the same modulation of landmark salience observed during spatial language encoding would also affect the recall and use of landmark information during a spatial task.

Experiment 3 attempted to address such limitations by replacing the map drawing task with a computer-based map verification task. This presents a significant advantage. The recording of eye movements across both phases of the experimental paradigm allows comparing related measures between encoding and test. In particular, Experiment 3 focused on Dwell Time (DT) as a measure of attention allocation, and compared DT on landmark words during description encoding to DT on landmark regions of a map to be verified. This adaptation allows for the investigation of modulations of landmark salience during map verification as a result of the linguistic and imagery manipulations during the encoding stage. More specifically, Experiment 3 investigated whether strategic biases towards more navigationally salient landmarks are present during map verification whenever they are present during description encoding, or whether these are eliminated during map verification when absent during encoding (e.g. following encoding of cardinal descriptions). This change in paradigm also allows for the acquisition of parametric measures that could not be obtained from the map-drawing task used in Experiments 1 and 2, such as accuracy rates and response times. Yet another advantage is the possibility to present participants with a considerably higher number of trials than could be accomplished in Experiments 1 and 2.

Two key modifications were implemented in this study with respect to stimulus presentation. The reading time was capped at 60 seconds per description in order to further increase the number of trials that could be included in a session. Additionally, the descriptions were presented on a single page of text rather than two. This choice was motivated by existing research showing that text continuity in route descriptions can affect spatial information recall, particularly when forming an egocentric representation. Sugimoto and Kusumi (2014) presented their participants with texts containing three sentences that described the spatial relations between four landmarks in different environments, and then asked them to draw a map of each environment after reading the relevant description once. In their study, reading was self-paced and texts were presented one sentence at a time. These texts could either describe a route through the environments or provide the global layout of the environments from a survey perspective as defined by Taylor and Tversky (1992).

Sugimoto and Kusumi also defined text continuity as the extent to which each sentence in a description depends on the sentences immediately preceding it (see also Perrig and Kintsch, 1985), and manipulated this factor by transposing the order of the sentences in the texts, thereby changing the sequence in which landmarks were introduced. They could either describe the four landmarks (A, B, C and D) in the same order as they would appear on a physical path connecting them (from A to D), or describe the spatial relations between pairs of landmarks in different orders (i.e. B-C A-B C-D or A-B C-D B-C). Spatial information recall was measured as the number of pairs of landmarks whose names and spatial relations were correctly recalled. They tested the hypotheses that: 1) text discontinuity would decrease spatial information recall by forcing participants to maintain two segments of the route in working memory until the end of the description, when they would acquire a third and connecting one; and 2) that text discontinuity would increase reading time. Both predictions were broadly confirmed in their study, and egocentric route learning (i.e. the learning of spatial descriptions analogous to the ones used in the experiments presented here) appeared to be particularly affected by text discontinuity compared to survey learning. As such, presenting route descriptions on a single page of text constituted an attempt to minimise the possible confounding effect of text discontinuity on information encoding, mental representation construction, and information recall.

One of the goals of this study was to replicate the findings from Experiment 2. On the basis of previous observations of main effects of relational term type on description reading measures (as in Experiments 1 and 2), I predicted that during encoding of left/right descriptions a significant difference in total dwell time between salient and non-salient landmark words would be observed, with navigationally salient landmark words fixated significantly longer than non-salient landmark words. Additionally, I posited that during cardinal description encoding the difference in total dwell time between salient and non-salient landmark words would be reduced.

Furthermore, the recording of eye movements during the test phase allows testing for changes in the landmark salience profile between description encoding and map inspection. Two alternative hypotheses can be formulated in this respect. On the one hand, a similar pattern of fixations could be observed while participants study the maps, with salient landmark regions being fixated for longer than non-salient landmark regions following encoding of left/right descriptions, and no significant difference between salient and non-salient landmark regions during map inspection following cardinal description encoding. In this case we would conclude that the spatial priority maps (see Chapter 3) being constructed during encoding of spatial descriptions is preserved during test. However, the observation of significant main effects of (or interactions involving) imagined perspective on map drawing performance scores in Experiment 2 might support a prediction of changes in eye movement patterns between encoding and test. More specifically, Dwell Time measures during the map verification phase could be found to differ between the Left/Right-Egocentric and the Left/Right-Allocentric conditions, and between the Cardinal-Egocentric and the Cardinal-Allocentric conditions, indicating significant effects of imagined perspective.

For example, in Experiment 2, DT measures in the Left/Right-Allocentric condition were found to resemble those found in the Left/Right-Egocentric condition, with navigationally salient landmark words being fixated significantly longer than non-salient ones. However, both behavioural scores and self-report measures indicated differences between the two conditions, arguably due to the fact that, in the former, participants were generating a bird’s-eye view of the environment being described. Phenomenologically, this may have appeared similar to a map-like representation onto which participants were tracing the described route and “pinning” landmarks. The use of egocentric relational terms may have stressed the motor-sequential nature of the route, resulting in more attention being paid to landmarks located at those locations of the visualised map where the route turned. However, the result would still have been an allocentric representation, albeit with certain landmarks more available in memory than others. It is possible that the primacy of turn landmarks could be suppressed in such a representation by the time encoding is complete. In a map verification task during which eye movements are recorded, this might therefore result into a more equal distribution of attentional resources to all landmark regions following Left/Right-Allocentric encoding.

With regard to measures of accuracy and response time, I predicted that having encoded a route description while maintaining an allocentric perspective would result in a cognitive advantage both in terms of accuracy and response time.

More specifically, more accurate and faster map verifications should be recorded in the Left/Right-Allocentric and Cardinal-Allocentric conditions, and slower and less accurate map verifications in the Left/Right-Egocentric and Cardinal-Egocentric conditions, due to the need to transform a first-person, egocentric representation into a map-like, allocentric one. On the basis of findings, in Experiment 2, of significant reference frame interference in the Cardinal-Egocentric condition, I further predicted this condition would produce the least accurate and the slowest responses.

3.10. Methods

3.10.1. Participants

30 adult English native speakers with normal or corrected-to-normal vision (13 males, 17 females, mean age 20.70 ± .62 years) were recruited across the University of Nottingham in exchange for credits or an inconvenience allowance.

3.10.2. Design and Materials

The study followed a 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric) design. A total of 48 different route descriptions (12 per condition) were written. The descriptions were of the same format as those used in Experiments 1 and 2, and described routes through plausible urban environments containing six different landmarks: one origin, one destination, two at salient points (changes in directions), and two at non-salient points. Six different sets of 12 landmarks were created, for a total of 72, and their spatial location was counterbalanced across each stage of the route. Each route contained a total of four turns, two associated with landmarks and two at spatial locations where no landmark was present. See Appendix III for examples of the descriptions and for a full list of the landmark words used.

A total of 48 maps were also drawn using GIMP (GNU Image Manipulation Program), one for each route description. These were simple schematic representations of the routes, inspired by the maps drawn by participants themselves in Experiments 1 and 2. This was an attempt to increase the ecological validity of the eye tracking data, ensuring the collection of similar scanpaths to those that would be generated as participants inspected sketch maps hand-drawn by someone else. The maps were all between 500x700 and 500x800 px in size, in both portrait and landscape orientation to accommodate routes developing in different map-centred cardinal directions. They represented landmarks using landmark words written in Sans font, 18 px size. These were surrounded by rectangular boxes, which also represented the boundaries of the regions of interest used by the eye tracking software to output RoI reports. Path legs and boxes were created using the paths tool, and a line width of 5 px. Black was used for both text and lines. Ordinal information in the descriptions (e.g. “Take the first left.”) was represented as paths diverging from the main route as described. Half of the maps were correct representations of their respective routes, while the other half contained errors in the form of incorrect turns. Each map contained only one error, at the second, third, or fourth turn. The location of the landmarks was always correct in terms of their being at a turn location (correct or otherwise) or along a path leg. This was to avoid confounding our measures of dwell time on our landmark regions of interest. See Table 3.11 for examples of egocentric and allocentric descriptions, and Figure 3.36 and Figure 3.37 for examples of the maps used.

3.10.3. Apparatus

Eye tracking and computer setup were the same as in Experiments 1 and 2. Participants used a Microsoft Sidewinder USB gamepad to categorise each map as a correct or incorrect representation of the route description they had just read. Each description was preceded by a drift correction marker at the location of the first letter of the first word. Both overlay images and eye movements were presented to the experimenter on the Host PC so that feedback could be provided. No drift correction marker was presented before the presentation on screen of the maps in order to avoid restricting or influencing participants' visual scanning patterns. Runtime randomisation of trials was used.

3.10.4. Procedure

Participants were seated by the eye tracker, and provided with an explanation of the task. They were allowed to assume a comfortable position and to familiarise themselves with the gamepad. They were instructed to hold it at all times, and to use the trigger buttons to perform the task (pressing the right trigger if they judged each map to be correct, and the left trigger if they judged it to be incorrect). They were also informed that, although reading time was capped at 60 seconds, they were allowed to skip to the map verification phase by pressing a button on the gamepad if they felt confident they had understood the description. Calibration of the eye tracker was performed at the beginning of the script and drift correction before each description appeared.

As in Experiment 2, each description was preceded by either of the two images used to explicitly prompt a given imagined perspective (Figure 3.17), and participants were informed as to the meaning of each image both in the information sheet and verbally prior to starting the experiment. Unlike in Experiments 1 and 2, however, the route descriptions were presented on a single page of text, and participants were given a maximum of 60 seconds to read them. After 60 seconds, a map would appear on the screen for participants to inspect and assess. The map was not preceded by a fixation point, in order to avoid biasing the location of the first fixation on the map (and, potentially, of the resulting scanpath). The map inspection phase had no time limit, but participants were instructed to answer as quickly as they could while maintaining accuracy. The test stimuli remained onscreen until participants made a response. They were then extinguished and a new drift correction marker would appear prior to the presentation of the following route description.

Table 3.11 – Examples of egocentric and allocentric route descriptions as used in this study.

EGOCENTRIC DESCRIPTION	ALLOCENTRIC DESCRIPTION
Leave the house.	Leave the aquarium.
Turn left at the veterinary.	At the sushi bar head north.
Take the first right.	Take the second road heading east.
Walk past the university.	Walk past the Indian restaurant.
Turn left at the newsagent.	Take the first road heading north.
Take the second right.	Walk past the bank.
Walk past the library.	At the park head east.
You have reached the cinema.	You have reached the estate agent.

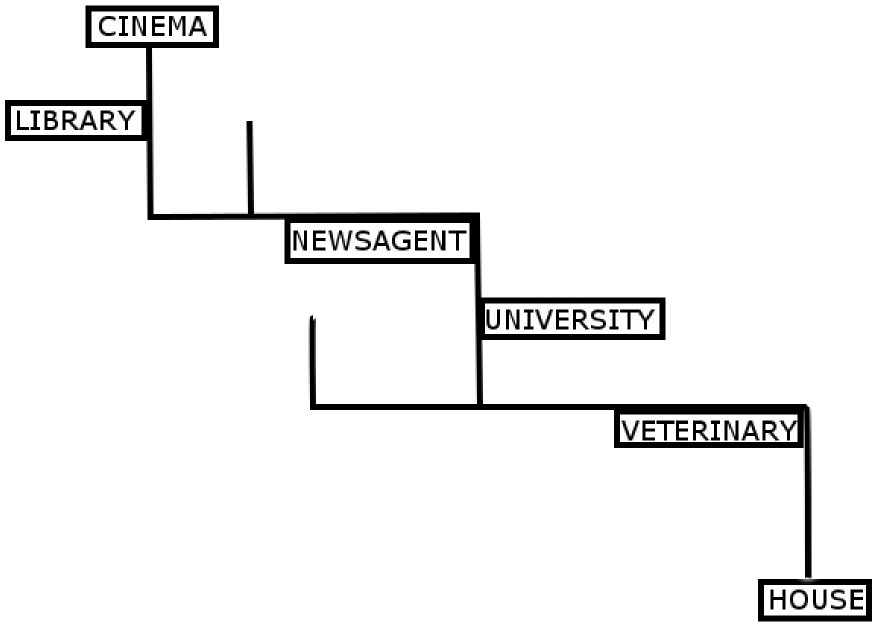


Figure 3.36 – An example of a correct map as presented to participants (in this case correctly representing the egocentric description in Table 3.12).

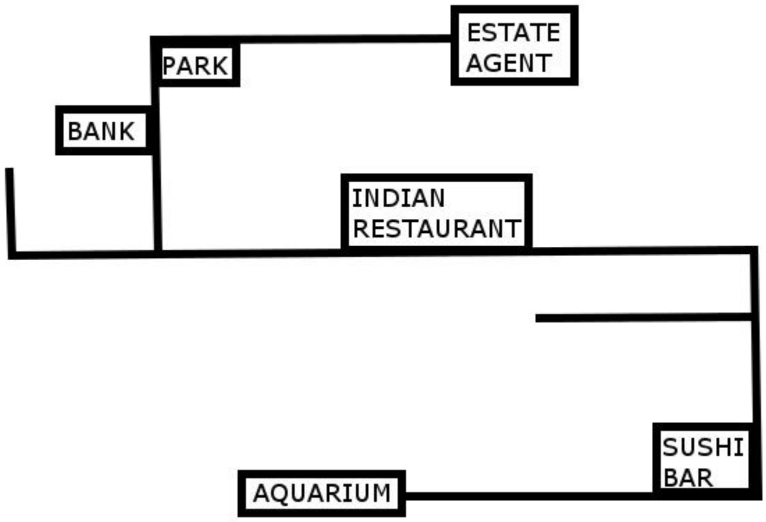


Figure 3.37 – An example of an incorrect map as presented to participants (following the allocentric description in Table 3.12). In this instance, the second (with "aquarium" as origin point) turn is in the wrong direction (west rather than the correct east).

3.11. Results

Accuracy

Accuracy was computed as a percentage of correct responses (i.e. whenever a participant correctly judged a map to be correct or to contain errors) for each condition per participant and then analysed via 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric), Sidak-corrected within-subject ANOVAs. However, no significant main effect of Relational term type or Imagined perspective were found, $F(1,29) = .719$, $p = .403$, $\eta^2_p = .024$, and $F(1,29) = .659$, $p = .423$, $\eta^2_p = .022$, respectively. No statistically significant interaction between the two factors was present, $F(1,29) = .028$, $p = .868$, $\eta^2_p = .001$. Accuracy rates were generally high (see Table 3.12).

Table 3.12 – Descriptive statistics for response accuracy rates in percentages for all conditions.

Left/Right-Egocentric		Left/Right-Allocentric		Cardinal-Egocentric		Cardinal-Allocentric	
M	SEM	M	SEM	M	SEM	M	SEM
81.94	2.56	80.27	3.00	79.72	2.70	78.61	2.75

Response Time

RTs for incorrect responses were excluded from analysis. This led to an overall loss of 18.05% of cases for the Left/Right-Egocentric condition, 19.72% for the Left/Right-Allocentric condition, 19.44% for the Cardinal-Egocentric condition, and 20.55% for the Cardinal-Allocentric condition, leaving 1160 individual data points for analysis.

On this basis, mean response times for each condition per participant were computed and analysed. These data are presented in Figure 3.38. A Shapiro-Wilk test of normality provided evidence of significant skew in the RT distribution in the Cardinal-Egocentric and in the Cardinal-Allocentric conditions, S-W = .774, $p < .001$ and S-W = .882, $p = .003$ respectively. Accordingly, response times were log-transformed, but the pattern and magnitude of results did not differ from those obtained from the untransformed data. In the following sections, results are reported for the untransformed data and in the original units for ease of interpretation.

No significant main effect of Relational term type was found, $F(1,29) = 1.88$, $p = .181$, $\eta^2_p = .061$. However, a significant main effect of Imagined perspective was observed, $F(1,29) = 4.84$, $p = .036$, $\eta^2_p = .143$. The direction of the effect revealed significantly faster response times for allocentric perspectives, M = 11979 ms, SEM = 614 ms, compared to egocentric ones, M = 13101 ms, SEM = 774 ms (Figure 3.20), a mean difference of 1122 ms, SEM = 509 ms. The interaction between the two factors was not significant, $F(1,29) = .40$, $p = .530$, $\eta^2_p = .014$.

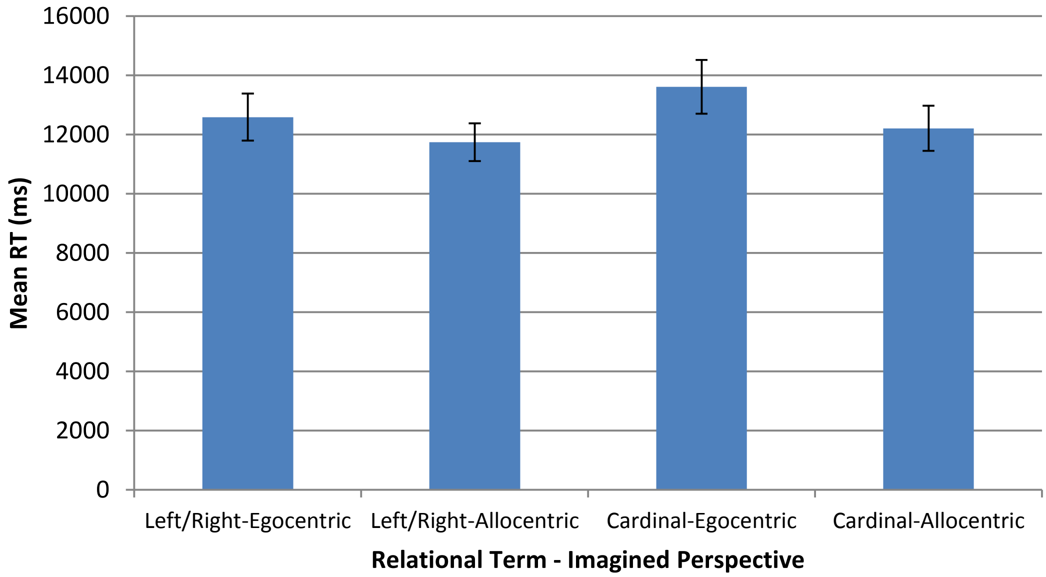


Figure 3.38 - Average response time for each condition. Error bars represent ±1 SEM.

Reading Time and Description Skipping

I subsequently analysed the time it took participants to read the descriptions, and explored the rate at which they opted to advance to the map verification phase before the 60-second cut-off. I computed average reading times per condition for each participant, and then analysed them via 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric), Sidak-corrected within-subject ANOVAs.

The analysis yielded a significant main effect of Relational term type, $F(1,29) = 4.93$, $p = .034$, $\eta^2_p = .145$, revealing a reading advantage of 1158 ms for left/right descriptions compared to cardinal ones, M = 45285 ms, SEM = 2129 ms, and M = 46443 ms, SEM = 2088 ms, respectively. No significant main effect of Imagined perspective was found, $F(1,29) = .60$, $p = .443$, $\eta^2_p = .020$. The interaction between the two factors was not significant, $F(1,29) = .23$, $p = .631$, $\eta^2_p = .008$. See Table 3.13 for average reading times for all conditions.

Neither Relational term type, $F(1,29) = .88$, $p = .355$, $\eta^2_p = .030$, nor Imagined perspective, $F(1,29) = .23$, $p = .633$, $\eta^2_p = .008$ appeared to influence participants' tendency to skip to the map verification stage. No significant interaction between the two factors was found, $F(1,29) = .027$, $p = .871$, $\eta^2_p = .001$. Averaged across levels of Imagined perspective, the mean skipping rate was 58.20% during encoding of left/right descriptions, SEM = 6.8%, and 56.25% during encoding of cardinal descriptions, SEM = 6.9%.

Table 3.13 – Descriptive statistics for description reading times for all conditions in milliseconds.

Left/Right-Egocentric		Left/Right-Allocentric		Cardinal-Egocentric		Cardinal-Allocentric	
M	SEM	M	SEM	M	SEM	M	SEM
45506	2238	45064	2132	46898	2107	45988	2186

Total Dwell Time – Description Encoding

Total Dwell Time (DT) was computed for each Region of Interest (RoI; the two salient and the two non-salient landmark words) in each description for each participant, and then averaged across descriptions and across RoIs in each condition. The result was an average DT value for salient landmark words and one for non-salient landmark words in each condition for each participant. These were then analysed via 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric) x 2(Salience: Salient vs Non-Salient) within-subject ANOVAs.

Results revealed significant main effects of both Relational term type, $F(1,29) = 17.043$, $p = .001$, $\eta^2_p = .370$, and Salience, $F(1,29) = 4.522$, $p = .042$, $\eta^2_p = .135$, as well as a significant interaction between the two, $F(1,29) = 5.197$, $p = .030$, $\eta^2_p = .152$. No significant main effect of Imagined perspective was found, $F(1,29) = .189$, $p = .667$, $\eta^2_p = .006$. The interaction between Imagined perspective and Relational term was non-significant, $F(1,29) = 1.181$, $p = .286$, $\eta^2_p = .039$, as was the interaction between Imagined perspective and Landmark salience, $F(1,29) = 2.608$, $p = .117$, $\eta^2_p = .083$. The three-way interaction between the factors was also non-significant, $F(1,29) = .012$, $p = .914$, $\eta^2_p < .001$. A test of simple main effects on the interaction between Relational term type and Landmark salience revealed that salient landmark words were fixated for longer than non-salient landmark words during encoding of left/right descriptions, $F(1,29) = 10.261$, $p = .003$, $\eta^2_p = .261$, M = 1956 ms, SEM = 182 ms, and M = 1464 ms, SEM = 112 ms, respectively. However, this difference was not present during encoding of cardinal descriptions, $F(1,29) = .292$, $p = .593$, $\eta^2_p = .010$, M = 2172 ms, SEM = 151 ms, and M = 2079 ms, SEM = 134 ms, respectively. To sum up, navigationally salient landmark words were fixated longer than non-salient ones during reading of descriptions containing egocentric relational terms (Left and Right), but no such difference was observed during reading of cardinal descriptions (Figure 3.39). These results replicate what was found during encoding in Experiment 2.

Total Dwell Time – Map Verification

As with the encoding phase data, an average DT value for salient landmark RoIs and one for non-salient landmark RoIs in each condition for each participant was computed also for the stimuli in the test phase (both correct and incorrect maps). These were then analysed the same way, via 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: egocentric vs allocentric) x 2(Salience: Salient vs Non-Salient) within-subject ANOVAs.

No significant effects of Relational term type, $F(1,29) = .50$, $p = .484$, $\eta^2_p = .017$, or Imagined perspective, $F(1,29) = .04$, $p = .827$, $\eta^2_p = .002$, were found. However, a significant main effect of Landmark salience was observed, $F(1,29) = 17.82$, $p < .001$, $\eta^2_p = .381$, showing longer DT on navigationally salient landmark regions, M = 737 ms, SEM = 41 ms, compared to non-salient ones, M = 563 ms, SEM = 39 ms. No significant interaction between Relational term type and Imagined perspective, $F(1,29) = .70$, $p = .408$, $\eta^2_p = .024$, or between Relational term type and Landmark salience, $F(1,29) = .70$, $p = .409$, $\eta^2_p = .024$, was found. However, a significant interaction between Imagined perspective and Salience was found, $F(1,29) = 13.34$, $p = .001$, $\eta^2_p = .315$. An analysis of simple main effects revealed that salient landmark regions were fixated significantly longer than non-salient landmark regions following explicit maintenance of an egocentric perspective during encoding, $F(1,29) = 25.42$, $p < .001$, $\eta^2_p = .467$, M = 757 ms, SEM = 49 ms, and M = 536 ms, SEM = 36 ms, respectively. However, this difference was reduced following explicit maintenance of an allocentric perspective during encoding, $F(1,29) = 6.47$, $p = .017$, $\eta^2_p = .182$, M = 689 ms, SEM = 41 ms, and M = 590 ms, SEM = 47 ms, for salient and non-salient landmark regions respectively. Therefore, it appears that during map verification the DT difference between salient and non-salient landmark regions was reduced when participants were instructed to maintain an allocentric, bird's-eye view perspective (Figure 3.40).

Predicting RT from DT measures

By analysing DT measures, I observed that these appeared to change between description encoding and map verification in two of our conditions (Left/Right-Allocentric and Cardinal-Egocentric), but remained consistent between the two experimental phases in the Left/Right-Egocentric and Cardinal-Allocentric conditions. Namely, in the Left/Right-Egocentric condition, salient landmarks were fixated longer than non-salient ones both during description encoding and during map verification. Conversely, in the Cardinal-Allocentric condition, salient landmark words were not fixated longer than non-salient landmark words during description encoding, and the difference was reduced during map verification compared to the Left/Right-Egocentric condition.

This raised the possibility that a general tendency of allocentric representations of environments may be towards the “equiavailability” (Ruggiero, Iachini, Ruotolo & Senese, 2010) of all landmarks. Relatedly, it begs the question of whether eye movement patterns could be used to determine the degree of allocentricity (or egocentricity) of a participant’s spatial representations. If that is the case, then they might also be predictive of behavioural performance in tasks that require specific imagined perspectives. To investigate this possibility, I computed a mean difference in DT (Salient vs Non-Salient landmark regions) for each participant across both correct and incorrect map stimuli (Figure 3.41). DT difference values closer to 0 ms mean that salient and non-salient landmark regions were fixated equally during map verification. Larger positive values indicate that salient landmark regions were fixated for longer during map verification.

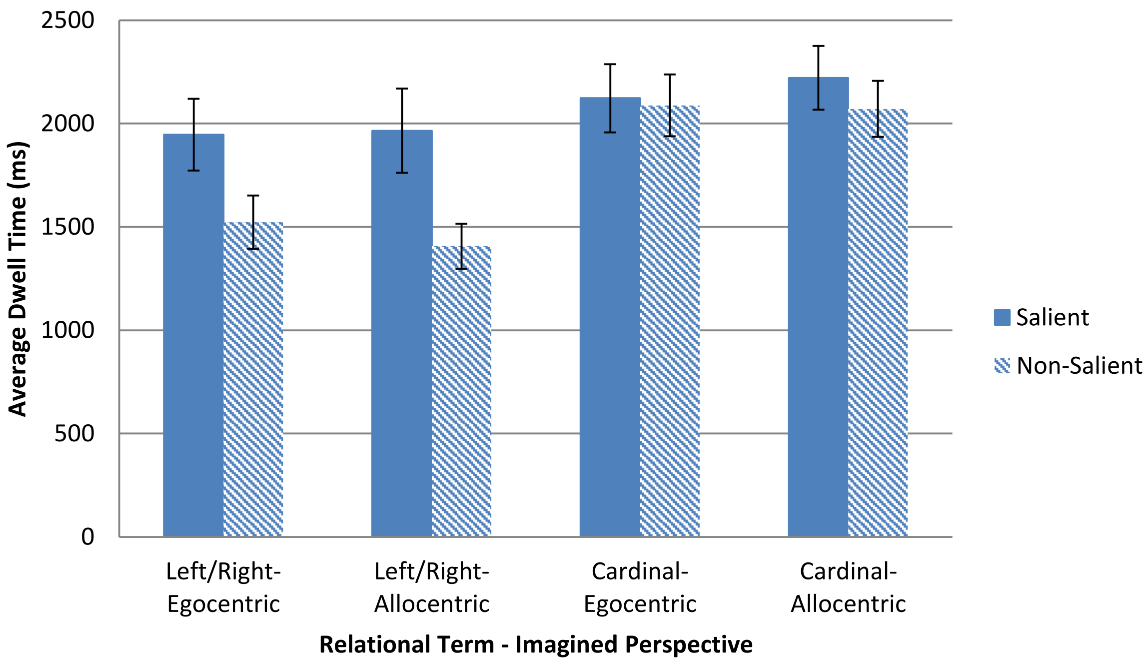


Figure 3.39 - Average dwell time on landmark words. Error bars represent ±1 SEM.

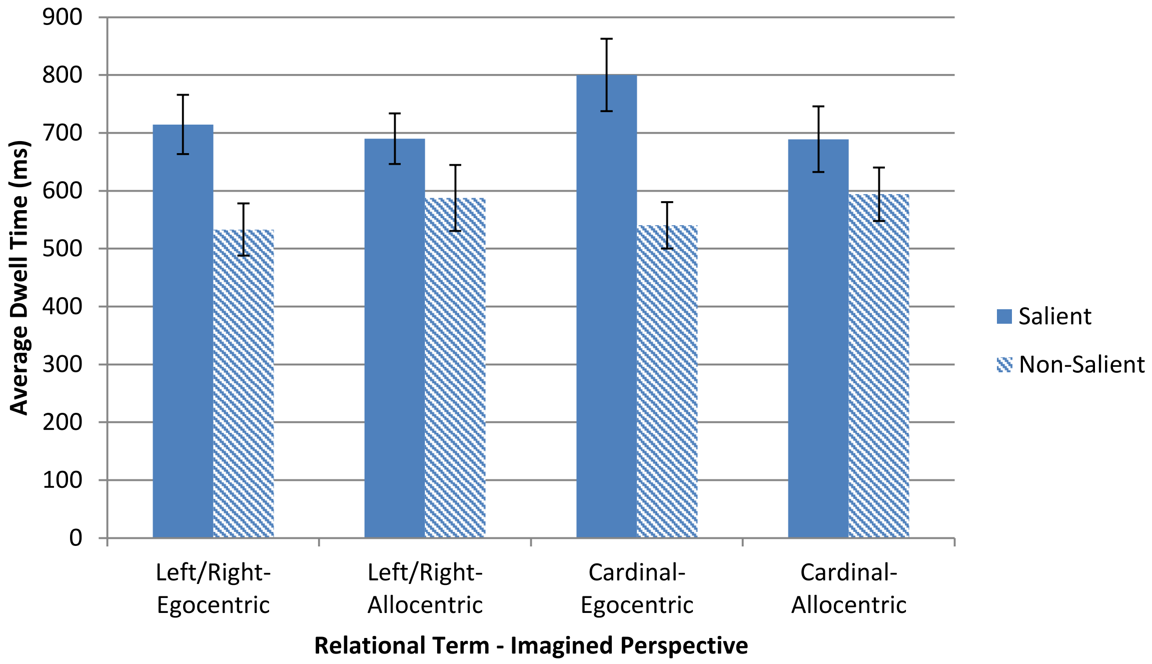


Figure 3.40 - Average dwell time on landmark regions of the maps. Error bars represent ±1 SEM.

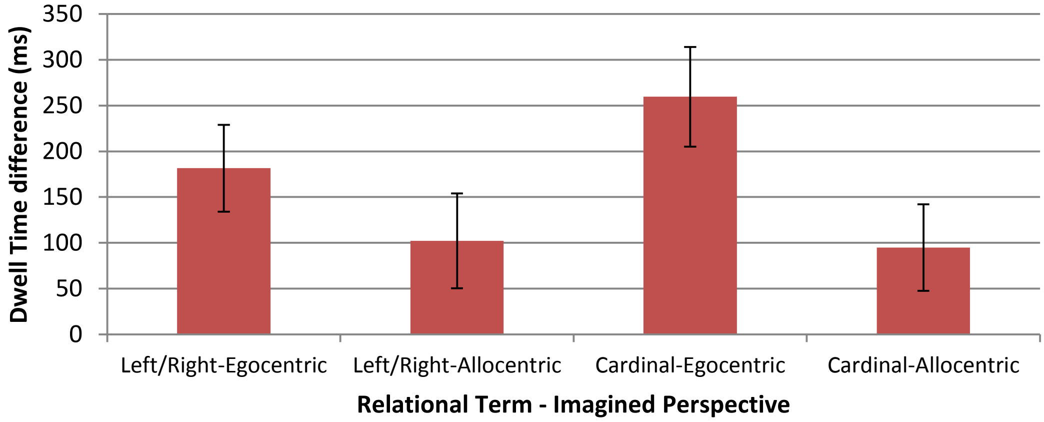


Figure 3.41 – Average difference in dwell time between salient and non-salient landmark regions during map verification.

This DT difference for the map phase was then regressed against verification RT in each condition to determine how allocentric-like patterns of fixations on the landmark regions might predict performance (Figure 3.20). In three conditions, map DT difference was found to be a significant predictor of RT performance (see Table 3.14 for regression coefficients and statistical significance). DT difference explained a significant proportion of variance in verification RT in the Left/Right-Egocentric condition, $R^2 = .19$, $F(1,29) = 6.84$, $p = .014$, in the Left/Right-Allocentric condition, $R^2 = .20$, $F(1,29) = 7.01$, $p = .013$, and in the Cardinal-Egocentric condition, $R^2 = .37$, $F(1,29) = 16.68$, $p < .001$. The regression for the Cardinal-Allocentric condition was marginally significant, $R^2 = .12$, $F(1,29) = 3.99$, $p = .055$. Therefore, it appears that a smaller DT difference during map verification (i.e. less primacy for salient over non-salient landmarks) can be a measure of allocentricity, and an indicator of better performance in an allocentric task (Figure 3.42).

Although plotting the data revealed a potential extreme score in the Cardinal-Egocentric condition, the Mahalanobis distance for that data point computed during the regression was not significant at the .001 alpha level recommended in the literature (Tabachnick & Fidell, 2012). Therefore, the decision was made to preserve the entire dataset (except RTs for incorrect responses) for analysis.

*Table 3.14 – Results of the regression between DT difference (DT-) during map verification and map verification RT. *p<.05; **p<.001; - = NS.*

	L/R-Ego. RT			L/R-Allo. RT			Card.-Ego. RT			Card.-Allo. RT		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
DT-	7.42	2.83	.443*	5.53	2.08	.448*	10.21	2.50	.611**	5.67	2.84	.353

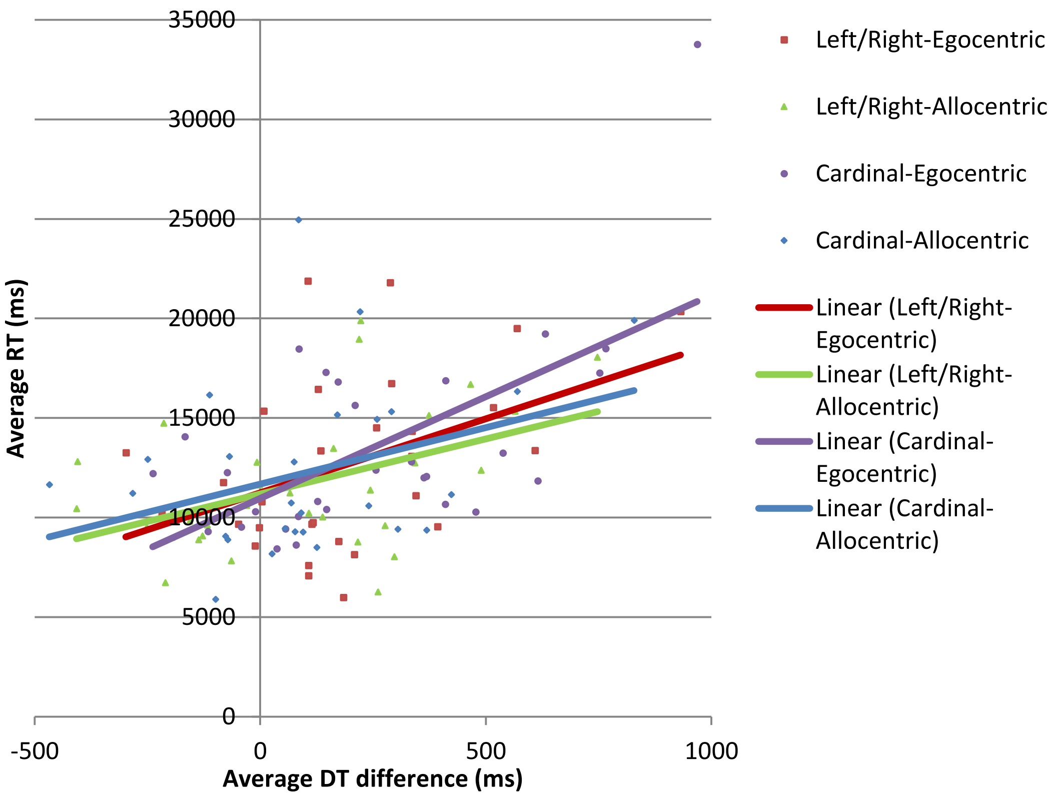


Figure 3.42 – Regression plot between mean dwell time difference between salient and non-salient landmark regions during map vision (X axis) and mean response time (Y axis) for all conditions (differently coloured markers and trend lines).

3.12. Discussion

This experiment tested participants' ability to encode route descriptions and to use the resulting spatial representations to judge whether sketch maps accurately depicted them. Although no significant differences in accuracy rates were found between conditions, asking participants to generate an allocentric representation of the environments during encoding appeared to reduce the amount of time required for them to correctly judge whether the respective maps were correct or not. This seems to suggest that imagined spatial perspective might be an important component of spatial computations, and that transforming a mental representation from one spatial format into another can incur cognitive costs.

If it were possible to compare accuracy measures across experiments, I might be better able to determine if decreasing text discontinuity had a positive effect on performance, in accordance with findings by Sugimoto and Kusumi (2014). Unfortunately, the measure of accuracy in this study bears little resemblance to the measures of accuracy acquired from participants' map drawings in Experiments 1 and 2, precluding attempts to determine whether presenting the descriptions in continuous texts increased overall performance. However, the lack of noticeable decrease in accuracy for the Cardinal-Egocentric condition (found to elicit reference frame interference and result in lower map drawing performance in Experiment 2) could, to an extent, be related to this difference in stimulus presentation. Presenting route descriptions in their entirety on a single screen (unlike in Experiments 1 and 2) may have made it easier for participants to integrate the information contained in Cardinal-Egocentric descriptions, despite the conflict between the reference frame implied by the cardinal relational terms and the first-person, egocentric perspective participants were asked to imagine.

This study was also partly intended to replicate findings from Experiments 1 and 2 concerning the allocation of attention to landmark words of different salience levels during encoding of route descriptions. On the basis of those experiments, I predicted that a decrease in the dwell time difference between salient and non-salient landmark words would be observed during encoding of cardinal descriptions (regardless of imagined perspective). This was indeed the case (Figure 3.39), although the pattern of results displayed a fundamental difference from that observed in the previous experiment. In Experiment 2 cardinal descriptions resulted in a markedly reduced average dwell time on salient landmark words but no difference in dwell time on non-salient landmark words compared to left/right descriptions. In this study no change in dwell time on salient landmark words was observed, but rather an increased dwell time on non-salient landmark words (in a manner more similar to the cardinal condition of Experiment 1). See Figure 3.43 for a comparison of encoding DT measures between Experiments 2 and 3. Although this ultimately resulted in the predicted lack of DT difference between salient and non-salient landmark words during cardinal encoding, the cause of this variability between experiments warrants further investigation.

It is worth noting how the average DT on non-salient landmark words during Left/Right description encoding was also higher in this study than in Experiment 2, although still significantly lower than the average DT on salient landmark words. It is possible that the changes to the way descriptions were presented in this study (on a single page of text rather than two) may have caused participants to change their reading patterns. Additionally, capping the available reading time at 60 seconds may have prompted participants to also fixate navigationally non-salient information for longer so as to maximise information intake and better ground their mental representations, thereby explaining the increased DT on non-salient landmark words compared to our previous studies.

Lastly, I investigated whether measures of DT on the landmark regions of the maps would follow a pattern similar to that found during the encoding of the descriptions, or whether these would differ. I observed that while DT patterns were comparable between description encoding and map recognition phases in the Left/Right-Egocentric and Cardinal-Allocentric conditions, they were reversed for the two remaining conditions. In other words, in conditions in which relational terms and actively imagined perspective could have induced potentially conflicting reference frames during description encoding, DT patterns appear to change between the two experimental phases. During encoding they appear to be influenced by the type of relational term used, and during map verification by the type of imagined perspective prompted explicitly.

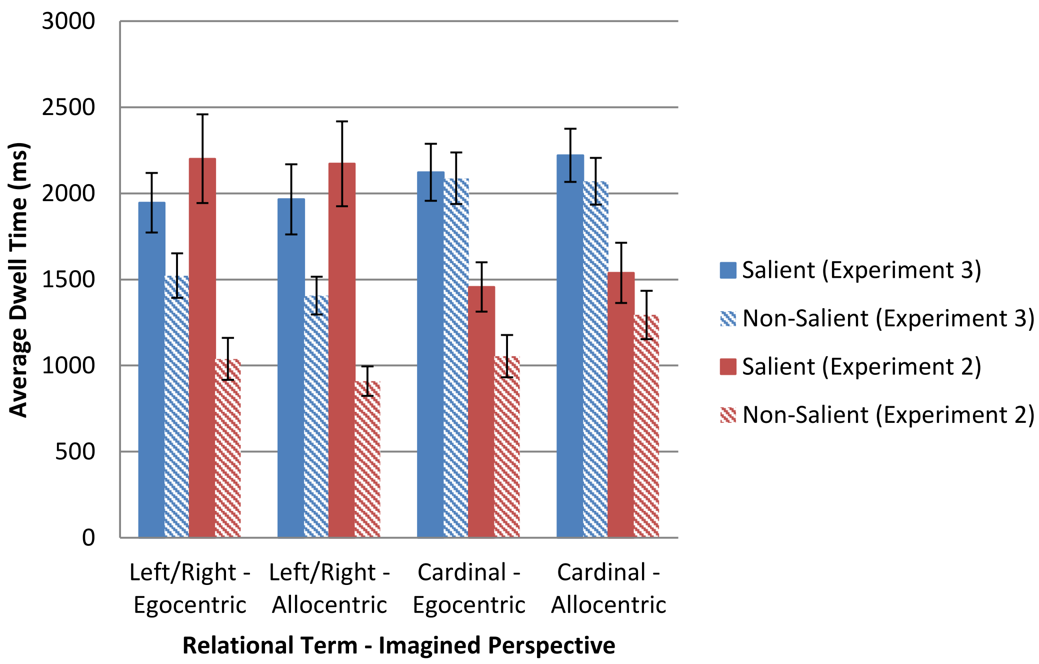


Figure 3.43 – Comparison of average DT measures on landmark words during description encoding between Experiment 2 and Experiment 3. Error bars represent ±1 SEM.

More specifically, while salient landmark words were fixated for longer during the encoding of Left/Right-Allocentric descriptions, that difference was reduced while participants studied the maps. Conversely, while no difference in DT was found between salient and non-salient landmark words during Cardinal-Egocentric description encoding, such a difference was present during the Map phase, and was in fact more pronounced than in the Left/Right-Egocentric condition. Effectively, DT measures during Left/Right-Allocentric encoding resembled those found in the purely egocentric encoding, but became more similar to those found in the purely allocentric condition during test. On the other hand, DT measures during Cardinal-Egocentric encoding display a similar pattern to that found during purely allocentric encoding, but appeared more similar to those found in the purely egocentric condition during test. This effect appears to be less prominent in the Left/Right-Allocentric condition, which one would expect (both theoretically and on the basis of Experiment 2) to be more compatible and lead to less interference compared to the Cardinal-Egocentric condition.

The finding that a lower DT difference between navigationally salient and non-salient landmark regions during map verifications can significantly predict lower response times can inform our understanding of language derived mental representations on several grounds. It appears that the process of encoding purely allocentric representations manipulates the salience profile of landmarks so that equal attention is given to each of them regardless of the navigational salience implied by the description itself. This is consistent with the idea that a survey representation of an environment is necessary to afford flexible navigation (i.e. comparing alternative routes, planning shortcuts, etc.: Chrastil & Warren, 2012). This, in turn, should require the ability to flexibly prioritise different landmarks depending upon the configuration of a specific route. Rendering all potential landmarks in an environment "equiavailable" (Ruggiero, Iachini, Ruotolo & Senese, 2010) may be part of this process of flexible attentional allocation, and Picardi et al. (2016) have also provided results consistent with this view (see Section 2.7).

As such, these data also suggest that the balance of attentional allocation to salient and non-salient landmark regions during map inspection can be diagnostic of how allocentrically participants are thinking of that particular environment. Interestingly, map verification DT difference did not significantly predict verification RT in the Cardinal-Allocentric condition (although the regression approached significance). This might confirm that the equiavailability of landmark words during description encoding already leads to the construction of allocentric spatial representations, and that any further improvement in behavioural performance is not significantly predicted by the same measure of equiavailability (DT difference) during map verification, but rather by other factors not considered in this experiment. Crucially, the continuous nature of this relationship agrees well with the idea of a continuum of representational possibilities between purely egocentric and purely allocentric.

The finding that participants chose to spend, on average, 1158 ms longer on the description screen in cardinal conditions compared to left/right conditions warrants further investigation. On the one hand, while much of the effect appears to be driven by longer reading times of Cardinal-Egocentric descriptions (see Table 3.14), it is possible that cardinal descriptions in general may have taken longer to process due to a general lack of familiarity with cardinal relational terms. In that case, this pattern would represent at least a partial trade-off between reading time and response time in the Cardinal-Allocentric condition. Alternatively, increased reading times during cardinal encoding may be simply due to cardinal descriptions containing slightly longer turn statements (i.e. "Take the second road heading east" vs "Take the first right").

On the other hand, participants were not explicitly instructed to read the descriptions and skip to the map verification phase as quickly as possible and before the 60 seconds of maximum reading time were up. It is therefore possible that many participants may have simply not chosen to take advantage of the possibility even when they felt confident that they understood the description, or that they may have done so more frequently later in the experiment as fatigue set in. In this case, measures of reading time should be approached with caution, as they may not fully reflect encoding difficulty.

3.13. Conclusions

In these three experiments, I set out to explore the processes via which we build spatial mental representations on the basis of linguistic input, and I attempted to answer a number of questions in this regard. In Experiment 1 I placed particular focus on the way the reference frame of a spatial description, as conveyed by the relational terms contained in it, might influence the imagined spatial perspective readers will assume during encoding. This was motivated by the need to determine whether the ecological tendency to compute both egocentric and allocentric reference frames in parallel, as observed in a number of navigation studies (see Chapter 1), would also be observed during the construction of spatial mental representations from language, and what factors might modulate it.

To investigate this, I attempted to ascertain whether differences in the resulting mental representations might be accompanied by differences in reading patterns as measured via eye-tracking, and whether these, if found, would influence behavioural performance during an allocentric spatial task. Additionally, I sought to determine whether performance would also be modulated by individual differences on small- and large-scale spatial abilities. In Experiment 2, a further manipulation was included in the form of explicit instructions to imagine an egocentric or allocentric spatial perspective, in order to better dissociate the effects of distinct imagined perspectives. These instructions were accompanied by a self-report questionnaire, in an attempt to determine whether exploring participants' phenomenological experience during imagery could help elucidate these processes further. Effects of encoding reference frame and imagined perspective on spatial task performance were explored by assessing the production of sketch maps following description encoding. In Experiment 3, I replaced the map-drawing task used in Experiments 1 and 2 with a map-verification task carried out during eye-tracking, in order to better relate measures of encoding and test. These tasks were ultimately tests of reference frame and spatial perspective specificity of spatial representations. It was hypothesised that if different encoding strategies are adopted during the course of reading route descriptions with different relational terms, these should be reflected in eye-tracking measures of reading. It was also hypothesised that, if the resulting mental representations are constrained by the descriptions' reference frames (or explicit imagery instructions), recall processes and performance during map drawing and map verification should also vary between conditions. In particular, performance in these allocentric tasks should be facilitated in conditions that we might expect to favour the generation of allocentric mental representations.

These predictions were partly verified. The analysis of eye movements yielded a number of interesting measures of attention allocation during spatial text reading (particularly with regard to the online processing of landmark salience), with differences between egocentric and allocentric processing. The differential allocation of attentional resources to landmark words in a description or to landmark regions on a map as a function of implied reference frame and of imagined perspective points to landmark salience as an interesting factor in the investigation of mental imagery and spatial mental models. While previous research has investigated the functional activity in brain regions involved in salience processing following navigation accompanied by visual input (see Section 1.3), more effort should go into determining the neural contributions of these areas to the establishment of a landmark salience profile in egocentric and allocentric mental representations during processing of spatial linguistic information. These potential areas of future research will be discussed in Chapter 6 of this thesis.

Behaviourally, several dependent measures point to a performance advantage when a test of spatial knowledge probes a spatial perspective congruent with the one generated during encoding. For example, Experiment 2 revealed a significant advantage in correctly recalling the spatial locations of landmarks (Figure 3.20) or in scaling an allocentric spatial mental model to an external template during map drawing (Figure 3.23) following encoding within an imagined allocentric perspective. Additionally, participants reported finding it easier to generate and maintain allocentric mental representations of the described environments when instructed to do so and regardless of the relational terms used in the description. Furthermore, Experiment 3 provided evidence of significantly faster response times in a map-verification task after encoding route descriptions within an imagined allocentric perspective, albeit potentially offset by differences in total encoding time that warrant further investigation. This generally adds support to the idea that allocentric representations can be computed just as easily as (and possibly in parallel with) egocentric representations (see Section 1.2), but with an important qualification. It appears that the reference frame implied by the relational terms used and the imagined perspective imposed top-down onto the emerging representation can be integrated and operate synergistically when congruent, but interfere with the encoding of information when incongruent. As a result, a number of dependent measures have shown that an egocentric representation will be easier to construct when mapping onto it spatial relations expressed in terms of “left” and “right”, as opposed to using cardinal terms. Similarly, there are hints that increased familiarity with egocentric relational terms might facilitate performance slightly in the left/right-allocentric condition compared to the, also congruent, cardinal-allocentric condition, but this trend did not generally reach statistical significance.

Globally, these results appear to suggest that linguistic reference frame and cognitive spatial perspective are partly overlapping but distinct elements that operate to construct a mental representation of space. Accordingly, it remains unclear whether the simple manipulation of relational terms in Experiment 1 was sufficient to elicit allocentric representations. The generally poorer performance in the allocentric (cardinal) condition, particularly following the first encoding of the descriptions, contradicts to the idea that a successful allocentric encoding should facilitate performance in an allocentric task such as map drawing. While this could simply have been due to participants’ lack of familiarity with cardinal relational terms (although these were not fixated longer and did not generate longer regression paths), or an insufficient number of trials for a practice effect to become evident, it is also possible that participants may have tended to imagine an egocentric perspective. This may have been incompatible with the cardinal directions provided, requiring mental translations between different reference frames in order to process and integrate the information provided. The tendency to adopt an egocentric perspective may have been due to the sequential nature of the descriptions used (typical of egocentric navigation and of route instructions). Additionally, the specification of non-salient landmarks as being on the left or right side of the road (e.g. “Walk past the school on your left”) may have led participants to envision the route egocentrically.

As such, even though differences in encoding were observed in Experiment 1 (e.g. a decrease in attention allocated to salient landmark words relative to non-salient ones during cardinal encoding), these cannot conclusively be identified as proxies of allocentric mental imagery processes. However, their replication in both Experiments 2 and 3 suggests that these are real effects deserving of further study. Additionally, the fact that different psychometric measures were found to significantly predict different aspects of behavioural performance in the various conditions would suggest that, to some extent, the manipulation was successful and partly distinct cognitive processes may have been involved in executing the task under different task demands. In particular, in Experiment 1, mental rotation performance was found to be a significant predictor of configural knowledge (awareness of the turns of the route and overall structure of the environment) following encoding of cardinal descriptions, but not of left/right descriptions.

Relatedly, participants' SBSOD scores (measures of navigational abilities in large-scale environments) did not significantly predict any measure of behavioural performance. This would therefore appear to suggest that constructing a spatial mental model from a description, and scaling it to fit the templates provided in Experiments 1 and 2, may have more in common with the mental manipulation of single objects in figural space than with the processes involved in the active exploration of larger environments. Similarly, performance on the Money's Test significantly predicted both landmark and configural knowledge following the second encoding of left/right descriptions, and landmark knowledge following the first encoding of cardinal descriptions. This makes sense, in light of the fact that the MT is a test involving a high degree of reference frame translation.

In Experiment 2, however, MT performance correlated negatively with landmark knowledge and positively with configural knowledge in the Left-Right/Egocentric condition, significantly predicting both. It did not predict performance in either area in the Cardinal/Egocentric condition, which I expected to more closely resemble the Allocentric condition of Experiment 1. Instead, performance in this condition was more significantly predicted by a more general working memory (processing) component, measured via the backward version of the digit span task. This might partly indicate that a lower ability to translate reference frames (necessary to quickly indicate the direction of each turn in the MT), might lead to the adoption of a more landmark-based navigational strategy, whereas a greater ability in this domain might allow to more easily reconstruct the global configuration of an environment even from a sequential route description. Globally, more research is needed to tease out the individual difference factors that might predict mental model construction from linguistic descriptions and their use during behavioural tasks.

Experiment 3 was especially interesting, as it allowed to directly relate eye-tracking measures during both encoding and test, and to relate these to a more parametric measure of behavioural performance (i.e. RT). Beyond replicating once more the finding of salience modulation of both landmark words and regions, this experiment established a reduced salience for turn-location landmarks as a possible marker of allocentric representation construction. This obviously warrants further replications under a variety of different conditions, starting with the adoption of more complex and varied route descriptions. Additionally, replications with neuroimaging components will be necessary to relate these results to those of other studies that have focused on landmark processing (see Section 1.3). Nevertheless, this particular design appears to be a valid means of studying the processing of spatial language, the relevant imaginal processes involved, and the contribution of both linguistic and top-down manipulations.

In conclusion, the results from Experiments 1, 2, and 3 so far generally support the idea that mental imagery plays a functional role in spatial cognitive tasks. They also show that, while mental imagery can, to an extent, be influenced by the linguistic content of a spatial text, the construction of a spatial model (and the processing of the text itself) is also amenable to top-down modulation. The eye-tracking data suggest that mental imagery can drive and influence eye movement patterns during both language and image processing, and that our phenomenological experiences of different imagined spatial perspectives are one of the loci of this function rather than merely epiphenomenal. However, the experiments presented in this chapter only tested participants' acquired spatial knowledge within the context of an allocentric task. In the next chapter, I present two experiments aimed at exploring whether the same pattern of results is observed when participants are being tested in an egocentric reference frame and imagining an egocentric perspective. The different nature of egocentric representations poses certain empirical and methodological challenges, requiring a change in both stimulus presentation and spatial test task. Despite this, Experiments 4 and 5 maintained the basic approach of creating different degrees of interference between linguistic frames of reference and imagined spatial perspectives during spatial description encoding.

CHAPTER 4
Testing Egocentric Representations

Experiment 4: Bearing Estimation

4.1. Experiment 4: Introduction

4.1.1. Egocentric Representations: Primitives and Computations

In Chapter 3, I explored the possible advantage in representing spatial knowledge within an allocentric reference frame when this knowledge is subsequently tested within a congruent reference frame (i.e. in a map-drawing or map-verification task). Allocentric representations have received considerable attention in spatial cognition, likely owing to a series of factors. These include the influence of Tolman's (1948) work and the ensuing popularity of the notion of cognitive map; initial hierarchical models of spatial knowledge that saw allocentric representations as the end result of experience with an environment (e.g. Siegel & White, 1975); and the growing identification of hippocampal and parahippocampal structures with the cognitive map following O'Keefe and Nadel's (1978) work. (e.g. McNamara & Shelton, 2003). Additionally, considerable work has attempted to explore the spatial localisation of (and, consequently, the hippocampal involvement in) episodic memory (e.g. Burgess, Maguire & O'Keefe, 2002).

In recent years, however, arguments have been put forward for a better characterisation of the nature and role of egocentric representations as integral parts of a spatial cognitive and navigational system, and not merely as the initial, transient input to enduring, allocentric cognitive representations. Perhaps most significantly, episodic memory has been proposed to be inextricable from the subjective egocentric experience of self, referred to as autonoetic consciousness (Tulving, 2002; Vogeley & Fink, 2003; Vogeley et al., 2004). This phenomenal experience is thought to be crucial not only to reflect on past events and behaviours, but also to plan for the future. In this vein, Byrne, Becker and Burgess (2007) have proposed a neural network integrating long- and short-term memory with mental imagery, wherein long-term memory is stored in the form of medial-temporal allocentric representation but inspected under directed attention in the form of egocentric parietal short-term representations built bottom-up from perceptual input or top-down from imaginal processes. The translation from one format to the other is mediated by proprioceptive information and motor efference signals, both real and simulated to enable mental navigation in time and in space. Although Byrne et al.'s (2007) proposal has not been extensively tested so far, Gomez, Rousset and Baciu (2009) showed increased incidental learning (i.e. better conscious recall in a Remember-Know-Guess paradigm; Gardiner, 2001) of non-spatial information obtained from passive egocentric exploration of an environment (i.e. viewing a video of a route) compared to allocentric study of the same environment.

However, the research presented in this thesis is primarily concerned with the acquisition of spatial knowledge. More specifically, this chapter is concerned with how spatial knowledge is extracted from language and encoded within egocentric representations. In the spatial domain, a number of different accounts have been proposed with regard to reference frame dominance in spatial computations and navigation. While most accounts, as already mentioned, have historically tended to focus on allocentric representations as the primary locus of computation (e.g. Gallistel, 1990; O'Keefe & Burgess, 1996), other accounts see egocentric representations and their update during self-motion as the central component of the navigational system (e.g. Wang & Spelke, 2000). Indeed, Filimon (2015) has recently argued, on the basis of evidence from behavioural, neuropsychological, and neuroimaging investigations, that all spatial representations can be construed as egocentric representations. This claim appears to be motivated by several findings, such as evidence of viewpoint- or direction-specificity during allocentric processing (such as when we attempt to align our body-centred reference frame to a cartographic north: e.g. Frankenstein, Mohler, Bülthoff & Meilinger, 2012). Other findings invoked to justify the centrality of egocentric reference frames include the reliance on gaze-centred reference frame during reaching (e.g. Selen & Medendorp, 2011), and the partial overlap in neural activation between regions thought to subserve egocentric and allocentric processes (e.g. Galati et al., 2010). Under Filimon's account, established evidence of separability between egocentric and allocentric reference frames appears to be largely dismissed as task- or strategy-specific effects, whereby egocentric representations are activated to varying degrees together with non-spatial object recognition processes.

It is unclear what advantages doing away with allocentric representations altogether might confer, and much of this proposal appears to rest on two foundations: a lack of unequivocal definitions of "egocentric" and "allocentric", and the possibility of conceiving of both egocentric and allocentric strategies to perform many of the spatial tasks adopted in the literature. Although I find purely egocentric accounts unconvincing, these two points are nevertheless important and warrant addressing. This is true even if we tend to accept more moderate accounts, suggesting that both egocentric and allocentric representations can coexist in parallel (e.g. Gramann et al., 2013) and that the dominance of one type over the other is the result of multiple factors such as individual differences and/or task demands (e.g. Gramann et al., 2005; see Section 1.11). This is because no account has thus far conclusively prevailed over the others, and a review of the literature reveals a wealth of varied and often conflicting results. These would appear to be the result of a number of factors, but I will focus on two that strike me as particularly important. These are a failure to establish control over what information is encoded and how this occurs, and a failure to determine what the task used is actually probing (or a tendency to assume that it probes a certain aspect of spatial knowledge without carrying out an appropriate task analysis).

In light of these ambiguities, a discussion of Experiments 4 and 5 presented in this chapter should probably begin with at least a brief a theoretical introduction to delineate what egocentric representations can, in principle, represent, and how the term is operationalised in this context. Klatzky (1998) carefully dissected the notions of allocentric and egocentric reference frames, highlighting aspects that may be considered primitive in the respective representations ("primitive parameters") and aspects that are likely computed on the basis of representational primitives ("derived parameters"). In this framework, an egocentric representation defines the location of a point in terms of its distance from the *ego* position (the navigator's location), and its egocentric bearing as the angle between the egocentric distance vector and the navigator's intrinsic axis of orientation. Since the navigator's heading can only be computed with respect to an external reference direction (i.e. allocentrically), one's heading in a purely egocentric representation is always 0°. As the navigator moves, its egocentric distance and bearing from objects in the environment changes and must therefore be re-computed. The processes that allow these parameters to be updated following perceptual exposure to and travel through (predominantly small-scale) environments have been the subject of considerable research (e.g. Loomis, Da Silva, Fujita & Fukisima, 1992; Rieser, 1989). A few studies have also compared spatial updating abilities following real and imagined movement (e.g. Presson & Montello, 1994). These can be broadly divided into two groups, depending on the specific task used.

A first group of studies explored spatial updating following movement within an array of objects surrounding the participant. In a study by Farrell and Robertson (1998), blindfolded participants were asked to point to the location of previously seen objects following a body rotation. The pointing task was run under four different conditions. In the *Updating* condition, participants rotated to face another direction and had to then point back to a named target object. In the *Imagination* condition, participants had to imagine facing a different direction and point to the target object as if they were. In the *Ignoring* condition, participants physically rotated towards a new direction but were instructed to ignore the rotation and point to the target as if still in the starting orientation. In the *Control* condition, participants rotated in one direction and then back in the opposite direction by the same amount, so that they ended facing the original direction. For all initial physical rotations, participants were touched on either shoulder to indicate direction, began rotating, and were told to stop when they had reached the new orientation. They were not told what object they were now facing. They then pointed to the named target object.

Results showed that mean angular error did not differ significantly between the three experimental conditions, but responses in the *Updating* condition were significantly faster than in the *Imagination* and *Ignoring* conditions. In the latter two conditions, a significant relationship was found between rotation magnitude, and both bearing estimation error and response latency. This suggests that in the *Imagination* and *Ignoring* conditions, participants performed mental rotations to align actual and imagined reference frames, and that the time required to do so increased as a function of distance. However, being unconstrained by a physical direction of motion, participants could imagine rotating in either direction so as to cover the shortest imaginary distance to the target orientation. This led to a curvilinear, rather than linear, relationship in the *Imagining* and *Ignoring* conditions. Additionally, the increased response latencies in the *Ignoring* condition indicate that participants had automatically updated their heading during motion using proprioceptive cues, and more time was necessary to mentally "undo" the physical rotation. Generally, the finding that bearing estimation error was not significantly higher in the imagery condition compared to the physical updating condition might suggest that spatial updating in imagery and during physical motion could rely on shared cognitive and neural mechanisms, just as imagery and perception appear to (see Section 1.7 and 1.8).

A second group of studies employed a triangle completion task to investigate the spatial updating of self-position and orientation under a variety of conditions, including imagined locomotion along routes. The task involves being exposed to a two-legged route and performing bearing estimates from the terminus location to the origin point. In one such study by Klatzky, Loomis, Beall, Chance and Golledge (1998), participants were split into five route exposure conditions. In the *describe* condition, blindfolded participants listened to verbal descriptions of routes with two legs and a turn (Turn 1) between them. The descriptions were of the type "Go forward 3 m, turn clockwise 90°, then go forward 3 m." Participants did not move during encoding, but at the end of the presentation had to turn in place (Turn 2) by the same amount they would have to in order to face their origin point if they had actually been standing at the destination. In the *watch* condition, participants watched the experimenter walk the two legs of the path, then closed their eyes and performed the bearing estimation as if they had. In the *walk* condition, blindfolded participants were physically led along the route, at the end of which they executed their bearing response. Two simulated conditions were also included. In these, participants sat on a rotating stool and wore head-mounted displays (HMD) showing the routes as optic flow changes consistent with motion through a virtual environment. In the *real-turn* condition, changes in optic flow associated with Turn 1 in the route were accompanied by physical motion, as the stool was rotated by the experimenter. In the *visual-turn* condition, no physical motion accompanied the optic flow changes.

Their analysis of bearing errors suggested that, in all conditions in which Turn 1 was not physically executed (i.e. when proprioceptive input was not available), participants failed to update their own perceived heading. Therefore, even though they were able to construct an allocentric representation of the path layout, their imagined heading at the end of the second leg was the same as in the first leg. This led them to systematically overestimate Turn 2 by the value of Turn 1 (Figure 4.1). On the basis of such results, one might conclude that proprioception is necessary to align one's own perceived egocentric reference frame with one's own actual position within the environment. This idea appeared to be broadly confirmed by a study by Chance, Gaunet, Beall and Loomis (1998), in which participants performed bearing estimations to target objects from a terminus location of longer routes through virtual environments. Over three sessions (at least one week apart), participants experienced different virtual environments through a HUD by physically walking, by turning in place to steer the constant forward motion in the optic flow, or by steering it with a joystick. An important difference emerged, however. Although navigation with physical walking led to lower absolute error in the bearing estimates, this was only the case during the second and third session, potentially indicating that vestibular and proprioceptive input, although important contributors to egocentric spatial updating, may not be automatically integrated during every instance of locomotion through an environment.

Gramann and colleagues (see Section 1.11) further explored the possibility that egocentric updating may occur in the absence of proprioception and vestibular input. They did so by using the tunnel task, a variant of the triangle completion task in which a two-legged route is presented using sparse visual input in the form of changes in optic flow, giving the impression of movement through a tunnel. These experiments suggested that egocentric updating in the absence of vestibular and proprioceptive input is possible, that the automatic adoption of an egocentric or allocentric strategy may be influenced by individual differences, and that participants are able to switch to their non-preferred strategy without considerable difficulty. These findings point to individual differences as one of the aforementioned sources of the lack of experimental control in these spatial tasks. However, they also suggest that practical measures may be taken to constrain participants' spatial reference frame during encoding, mental representational format, and imagined perspective during testing. Accordingly, in Experiments 4 and 5 I attempted to constrain route description encoding the same way I did in Experiments 2 and 3. Participants were presented with descriptions of two-legged routes (as in the triangle completion task used by Klatzky et al., 1998) in which the heading change was presented in egocentric or cardinal relational terms. Concurrently, participants were required to imagine and maintain an egocentric or an allocentric imagined perspective. It was hoped that establishing better control over language encoding processes might allow these experiments to partly clarify the discrepancies between studies that observed egocentric spatial updating following imagined motion in described environments (e.g. Wraga, 2003) and studies that failed to (e.g. Avraamides, 2003).

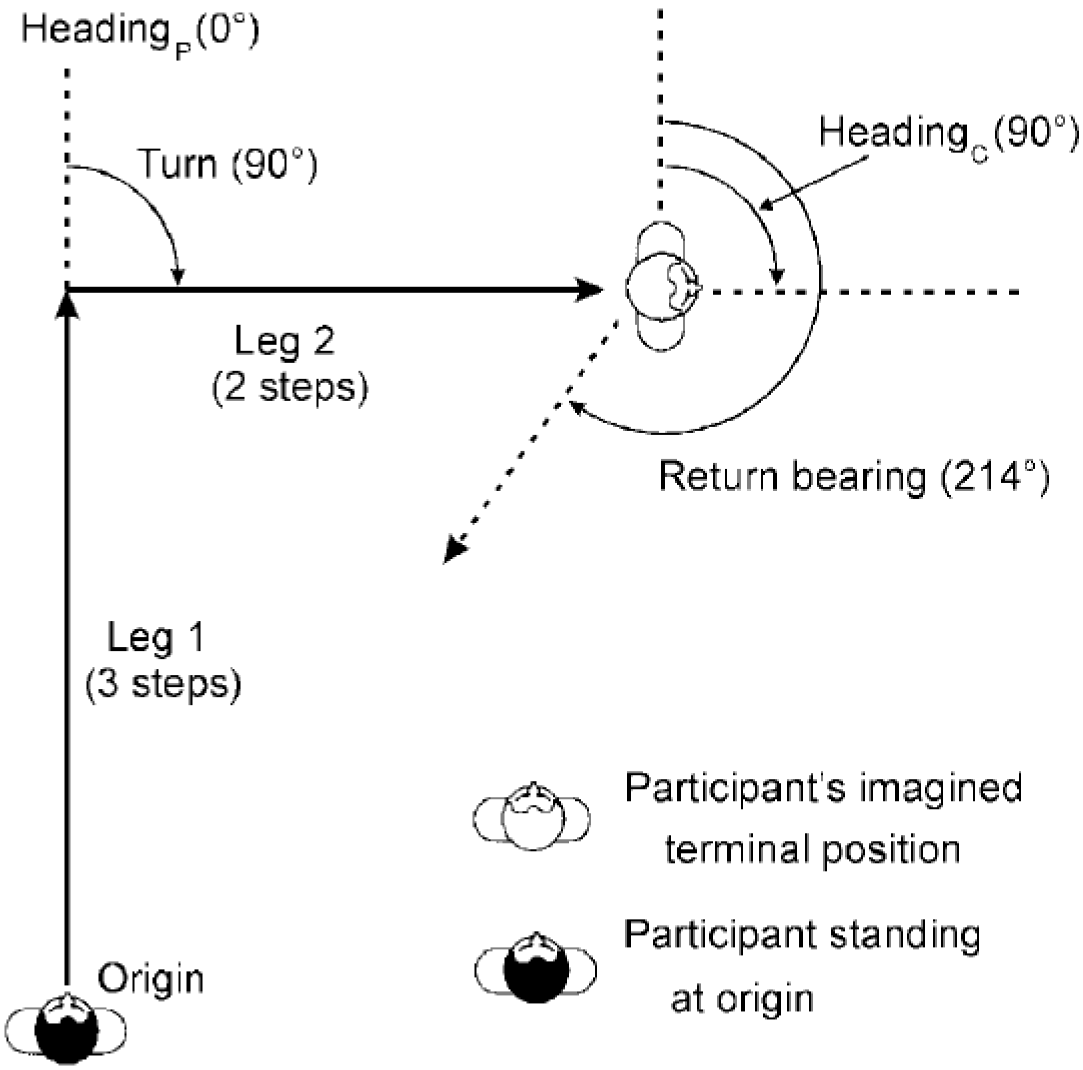


Figure 4.1 – Schematic representation of a triangle completion task (Avraamides et al., 2004), in which a failure to update the egocentric bearing following a turn leads to a bearing estimation using the 0° heading. The resulting bearing (214°) is the correct response (124°) overshot by the same amount as the turn (90°).

4.1.2. Encoding-Test Congruence: Egocentric Task Selection

Crucially, Avraamides, Klatzky, Loomis and Golledge (2004) have suggested that response modality may be what determines updating success in these studies. In their study they contrasted verbal and body turn homing responses following exposure to two-legged paths in a *Describe* and *Watch* condition. Unlike in Klatzky and colleagues' (1998) work, however, their path legs were described in terms of the number of steps to walk rather than in metric distances. Turns were described in degrees of angle, and verbal responses were prompted in the form of "relational term, number of degrees." Additionally, participants received training in understanding and expressing angular measures prior to the experimental trials. Verbal responses were found to be extremely accurate (with angular errors as low as 1°), whereas body turn responses were subject to angular overestimation by the same value as the turn, as found by Klatzky and colleagues (1998). Similarly, Wraga (2003) had participants locate objects in a circular array following real and imagined rotations using a traditional pointer aligned with their egocentric axis of orientation, a verbal response, and six directional arrow keys (front, back, front-left, front-right, back-left, back-right), but spatially scrambled so as to be offset from the participants egocentric reference frame. As a result, a spatial updating advantage was found for physical movement when a centred pointer was used, but for imagined movements when the offset pointer and verbal responses were used. The theoretical justification for this was that manual responses, being grounded on the physical reference frame of the participant, might create a conflict with the egocentric reference frame to be imagined. As a result, spatial updating during imagined self-rotation might rely on a process *"in which the role of the physical body is minimised"* (Wraga, 2003, p. 1004).

Although response modality is likely to have a systematic effect on performance, this particular conclusion seems counterintuitive in light of the abundant evidence for embodiment during language processing and imagery (see Section 1.10). On the contrary, one could argue that egocentric spatial updating following imagined motion will be facilitated by a manual response that favours embodiment at test, provided that embodiment is also successful during encoding. In other words, egocentric spatial updating will be successful provided the spatial information is encoded egocentrically, and provided that participants adopt an egocentric strategy to carry out the task. Considerable assumptions are made about these two aspects of spatial tasks in the literature, particularly with respect to linguistic stimuli. Accordingly, Experiments 4 and 5 in this chapter adopted the same encoding procedure used in Experiments 2 and 3. I presented participants with descriptions of two-legged routes conveyed using egocentric or cardinal terms, and with explicit instructions to imagine and maintain a specific spatial perspective during encoding. With regard to the test phase, both experiments were ultimately triangle completion tasks, and participants were instructed to imagine standing at the end of the route and to imagine pointing towards the direction of the origin landmark. However, the way participants performed the response differed between Experiment 4 and Experiment 5, and response modalities were chosen to modulate the degree of egocentric embodiment at the moment of testing. In Experiment 4, participants expressed their bearing from the terminus of the route to the origin point by marking the location where the vector would intersect the circle presented on the template in Figure 4.3. This response modality was inspired by the Object Perspective Test (OPT) developed by Kozhevnikov and Hegarty (2001) to dissociate mental rotation (e.g. the ability to mentally rotate an array of objects) and perspective-taking abilities (e.g. the ability to reorient the imagined self) (see Section 1.11). Although the canonical version of the OPT was found to be associated with egocentric navigational abilities, and although the instructions provided in Experiment 4 were essentially the same, albeit applied to a triangle completion task (i.e. “Imagine you are at the final location. Point back to your origin location.”), the template used in Experiment 4 was different in a simple but significant way. Where the centre of the original OPT template was a point without an intrinsic axis of orientation, the template used here represented the navigator from an allocentric perspective and with an allocentric heading with respect to the template coordinates. That is, the navigator’s nose and forward-facing hands (i.e. the navigator’s egocentric heading of 0°. See Klatzky, 1998) were aligned with the canonical cartographic coordinates (i.e. forward was north was top of the sheet). This provided the means to, at least partly, determine whether participants had updated their imagined egocentric bearing to the origin point during imagined locomotion and/or whether they had constructed and egocentric or allocentric representation by the time the homing estimate was to be performed. The argument and hypothesis in this respect are as follows.

Given that allocentric heading is not a representational primitive in an egocentric representation (Klatzky, 1998), the origin point of a two-leg route will always be located somewhere behind the participant’s terminus location in a purely egocentric representation. If, on the other hand, the route is represented allocentrically, the final heading of the participant (what Klatzky referred to as response heading) will be determined by the direction (and angle, although not relevant in Experiments 4 and 5) of the turn in the route (Figure 4.2). If participants are maintaining an allocentric representation of the route at the time of testing and responding with an allocentric strategy (akin to that used by Non-Turners in Gramann’s tunnel task), the angular error of their response should be expected to increase with the difference between the response heading and the upward/northward heading implied by the template. In these cases, the participant would be performing a mental rotation of the global configuration of the route so that their imagined final heading matches the heading of the centre figure on the template.

Ultimately, the following predictions were formulated for Experiment 4. In line with Experiments 2 and 3, I hypothesised that the combination of egocentric relational terms and of explicit instructions to imagine a first-person spatial perspective would result in the most egocentric-like spatial representation during and following encoding. This, in turn, should result in better performance in an egocentric spatial task, because of the need to transform an allocentric or mixed-perspective representation into a fully egocentric representation in the other conditions. However, said advantage should not be expected to manifest if the task is undertaken following an allocentric strategy. Such a strategy will see participants mentally rotate the configuration of the route so that the final heading matches the template orientation, just as participants might mentally rotate an array of objects to perform judgements of relative direction (JRD) from a perspective (imagined or otherwise) different from the study one. In this case, testing for an effect of final heading on angular errors will yield a significant effect.

4.2. Methods

4.2.1. Participants

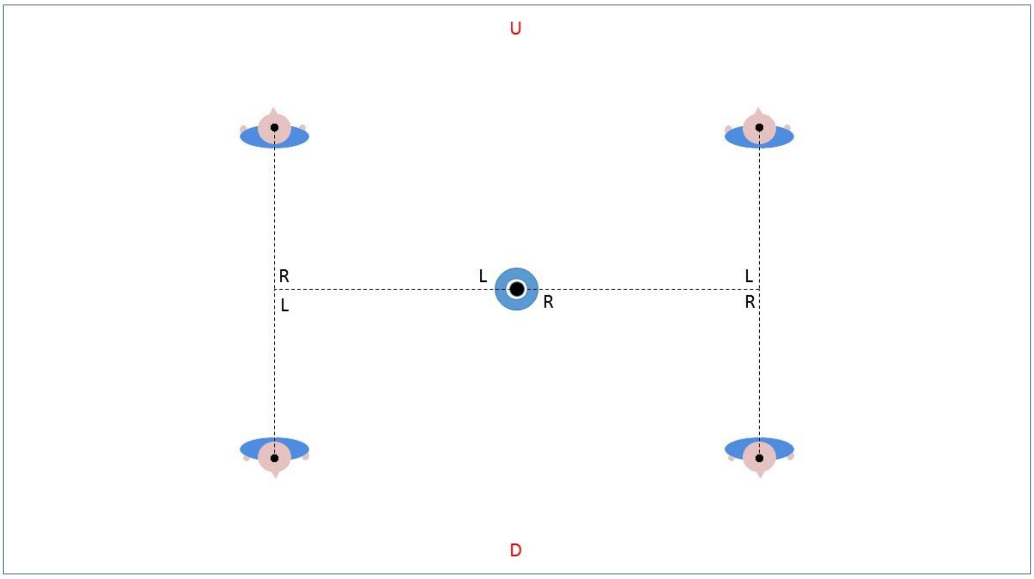
20 adult English native speakers (6 males, 14 females, mean age 20.65 ± 1.15 years), with normal hearing and no history of neurological or psychiatric disorders, were recruited across the University of Nottingham in exchange for credits or an inconvenience allowance.

4.2.2. Design and Materials

As in the previous experiments, a 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: Egocentric vs Allocentric) within-subject design was used. 64 audio recordings were recorded by a male speaker and equated in runtime at approximately 15 seconds. Each recorded route description detailed the directions from a location A to a location C, via a location B where a directional change occurred. The routes were divided into four blocks of 16 stimuli, with each block representing a different permutation of relational terms and actively imagined perspectives. Each route description contained two relational terms (both either egocentric or allocentric) describing the initial heading and a turn at location B, and two metric distances (50, 100, 150, or 200 metres). The latter were required to determine the correct angular measurements to compare against participants' responses, and were designed to be ecologically valid urban walking distances, in keeping with the descriptions of plausible urban environments used in Experiments 1-3. 48 different landmarks were used, and presented in a different counterbalanced order in each block to prevent any similarities between routes. See Table 4.1 for examples of an egocentric and an allocentric route description, and Appendix IV for a complete list of the landmark words used.

Table 4.1 – Examples of egocentric and allocentric route descriptions as used in Experiment 4.

EGOCENTRIC DESCRIPTION
Leave the botanical garden and turn left. Walk for 200 metres. Turn left at the archaeological museum and walk for 200 metres. You have reached the playground.
ALLOCENTRIC DESCRIPTION
Leave the pub and head north. Walk for 50 metres. At the butcher head east and walk for 50 metres. You have reached the vintage shop.



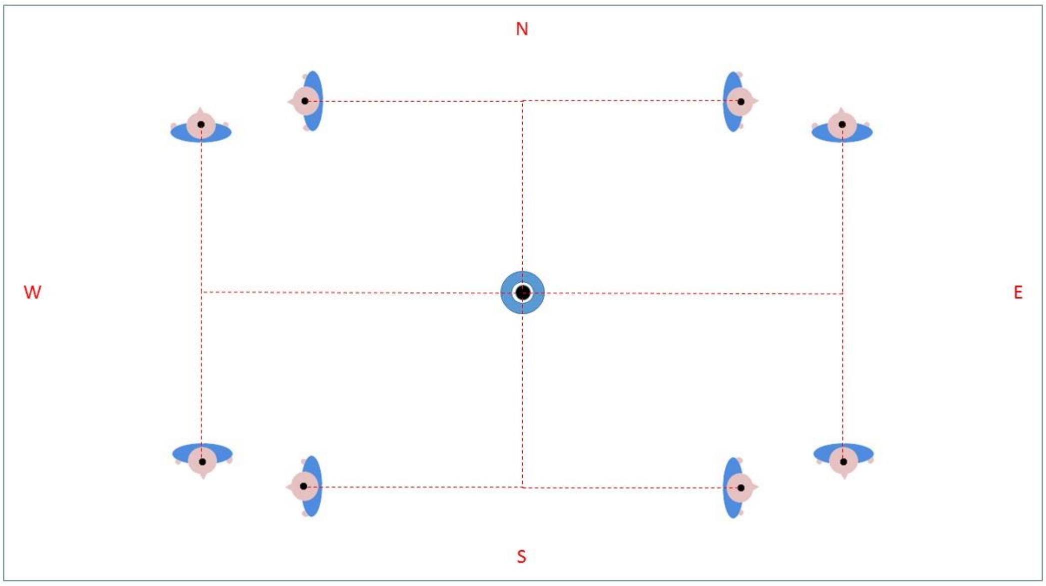


Figure 4.2 – All possible final headings resulting from allocentric encoding of different combinations of left/right (Top) and cardinal (Bottom) relational terms. Different sequences of turns away from the origin point (black circle) will produce final headings matching [(U)P, (N)ORTH] or mismatching [(D)OWN, (E)AST, (W)EST, (S)OUTH] the implied heading of the figure in the test template (Figure 4.3).

Each description encoding was followed by the presentation of a template (Figure 4.3) printed onto standard A4 sheets (landscape orientation) onto which participants marked their bearing response. At the end of each block of route descriptions, participants were also administered the same self-report questionnaire used in Experiment 2, with the addition of the following two items:

I found it easy to point back to my starting location.

Strongly disagree	1	2	3	4	5	**Strongly agree**

I found it easy to picture different walking distances in my head.

Strongly disagree	1	2	3	4	5	**Strongly agree**

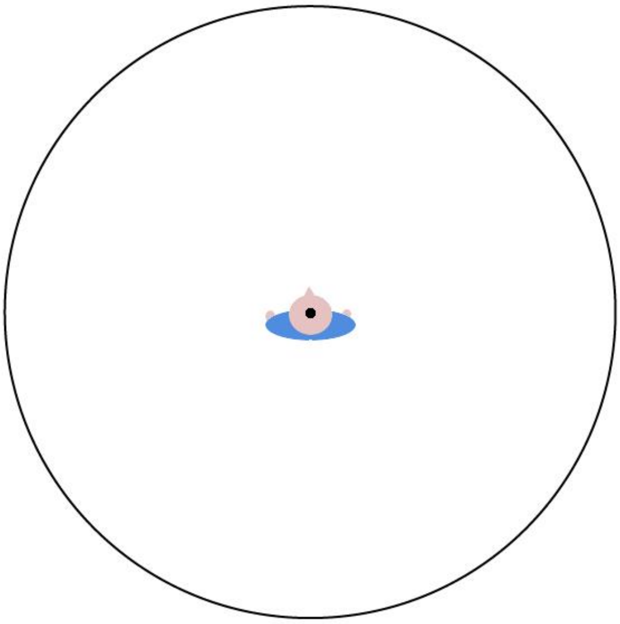


Figure 4.3 – Template used by participants to determine the relative direction of each route's origin location. The stylised figure in the centre is meant to represent the participant at the final location.

4.2.3. Apparatus

Participants were seated at a comfortable distance from the laptop used to play the recordings, and listened to them through a headset at a comfortable volume. Throughout each block, the relevant image used to explicitly prompt a specific spatial perspective (the same two images were used as in Experiments 2 and 3. See Figure 3.15.) remained visible on screen. For each trial, a new A4 template was placed between the laptop and the participant.

4.2.4. Procedure

The order of the four condition blocks was counterbalanced so that each participant was exposed to a different permutation. The relevant image used to explicitly prompt the required imagined perspective appeared on screen at the beginning of each block and was kept visible until a block with different explicit instructions started. Participants listened to each recording once and with their eyes closed, and were then prompted to mark the direction of the origin point relative to the terminus. This was done by marking an X on the template at the location where the circle surrounding the stylised figure intersected the imagined homing vector. After each block of 16 trials, the self-report questionnaire was administered, and a new block of descriptions was presented.

4.3. Results

4.3.1. Bearing Estimation

A bearing vector was drawn between the centre point of each circle and the marking made by participants on its circumference, and between the circle's centre point and its 0° circumference point. The unsigned difference in degrees between the resulting angle and that of the correct response was computed for each participant to derive an average error rate per item and per condition. As a result, in a trial whose correct bearing response was 180° (i.e. exactly behind the participant), a participant's angular error was 10° whether their response was 170° or 190°. Angular errors were then averaged across all items in a block, effectively averaging across levels of metric distance, because the goal of the experiment was not to determine whether angular error would increase differently over different imagined metric distances. However, participants' ability to imagine path segments of different lengths was probed using the self-report questionnaire in order to ascertain whether this would be easier while maintaining an egocentric or an allocentric imagined perspective (see Section 4.3.2).

Mean angular errors were then entered into a 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: Egocentric vs Allocentric) repeated-measures ANOVA. Comparing average bearing estimate error between conditions revealed a marginally significant effect of Relational term type, $F(1,19) = 3.78$, $p = .067$, $\eta^2_p = .166$, and significant effect of Imagined perspective, $F(1,19) = 6.40$, $p = .020$, $\eta^2_p = .252$, but no significant interaction between them, $F(1,19) = 1.36$, $p = .258$, $\eta^2_p = .067$. Overall, average bearing estimation error was found to be lower following encoding of left/right descriptions, M = 33.37°, SEM = 4.5°, than following encoding of cardinal descriptions, M = 40.29°, SEM = 3.1°. Bearing error was also significantly lower when participants encoded the route descriptions while maintaining an imagined allocentric perspective, M = 33.70°, SEM = 3.6°, compared to when an egocentric perspective was imagined, M = 39.96°, SEM = 3.7° (Figure 4.4).

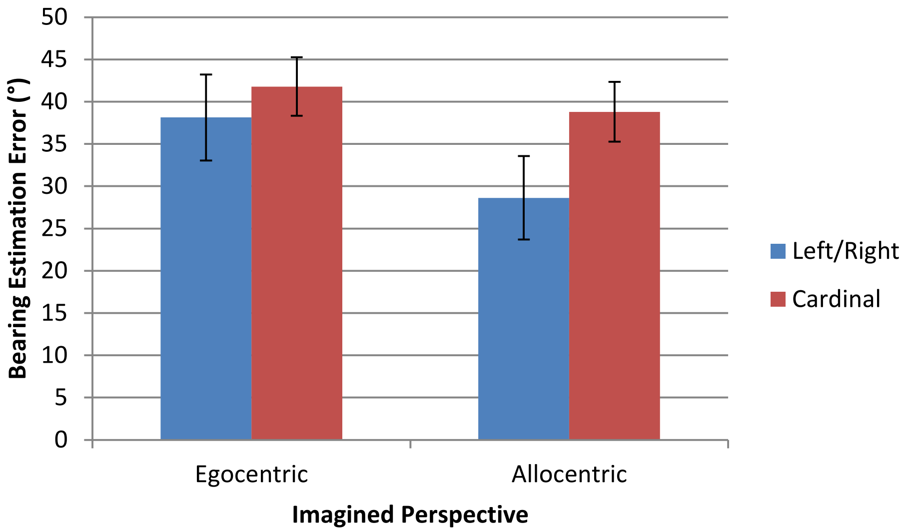


Figure 4.4 – Average bearing estimation error per condition in degrees. Error bars represent ±1 SEM.

I subsequently tested for an effect of final heading on bearing performance. Final heading (i.e. "response heading" in Klatzky's terms) was recorded as a function of what direction (in template/cartographic coordinates: Figure 4.2) the navigator was facing at the end of each route. For cardinal descriptions, the final heading of a route was "north" if, at the terminus, the navigator would be drawn as facing the upper edge of the template (corresponding to a canonical cartographic north). For left/right descriptions, the final heading was either "up" (equivalent to the cardinal north) or "down" (equivalent to the cardinal south). Due to the different number of levels of the Final heading factor for left/right and cardinal descriptions, angular errors were entered in separate models depending on the relational terms used in the descriptions. Bearing estimate errors for responses following encoding of left/right descriptions (i.e. Left/Right-Egocentric and Left/Right-Allocentric) were entered in a 2(Final heading: Up vs Down) x 2(Imagined perspective: Egocentric vs Allocentric) repeated-measures ANOVA. In this analysis, bearing estimates were found to be affected by both Final heading at the terminus location, $F(1,19) = 9.18$, $p = .007$, $\eta^2_p = .326$, and by Imagined perspective, $F(1,19) = 5.08$, $p = .036$, $\eta^2_p = .211$. The mean error rate was significantly lower when the final heading was "upward" relative to the template coordinates (or a hypothetical, allocentric mental map), M = 24.77°, SEM = 4.7°, compared to when it was "downward," M = 41.98°, SEM = 6.2° (Figure 4.5). Additionally, error rates were significantly lower if description encoding was performed within an imagined allocentric perspective, M = 28.62°, SEM = 4.9°, compared to an imagined egocentric perspective, M = 38.13°, SEM = 5.0°. No significant interaction between the two factors was found, $F(1,19) = .007$, $p = .934$, $\eta^2_p < .001$.

The model for the Cardinal-Egocentric and Cardinal-Allocentric conditions was a 4(Final heading: North vs South vs East vs West) x 2(Imagined perspective: Egocentric vs Allocentric) repeated-measure ANOVA. In this case, only Final heading had a significant effect, $F(3,57) = 14.81$, $p = .001$, $\eta^2_p = .438$, with a final "northward" heading yielding significantly lower errors, M = 19.12°, SEM = 2.9°, than a southward, M = 43.65°, SEM = 4.7°, eastward, M = 50.13°, SEM = 5.1°, or westward, M = 48.26°, SEM = 4.7°, final heading (Figure 4.6). No significant main effect of Imagined perspective, $F(1,19) = .88$, $p = .358$, $\eta^2_p = .045$, or interaction between Imagined perspective and Final heading was found, $F(1,19) = .22$, $p = .879$, $\eta^2_p = .012$.

In other words, if an allocentric plotting of a route meant that the heading at the final location was north (or up, in an imagined cartographic map), participants' homing estimates were significantly more accurate than if the final heading was different in allocentric-template coordinates.

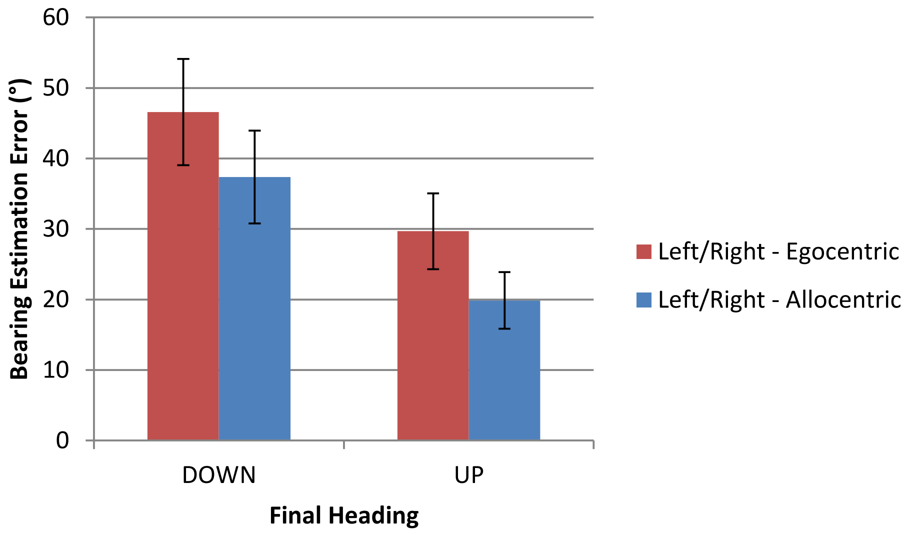


Figure 4.5 – Average bearing estimation error as a function of final heading in left/right conditions. Error bars represent ±1 SEM.

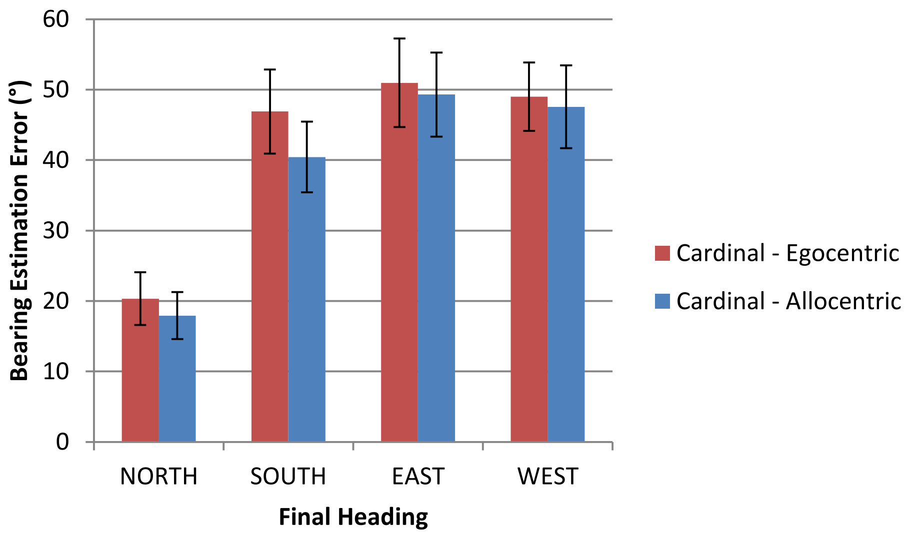


Figure 4.6 – Average bearing estimation error as a function of final heading in cardinal conditions. Error bars represent ±1 SEM.

4.3.2. Questionnaire Results

Table 4.2 includes the descriptive statistics for all questionnaire items for all conditions. Mean responses were computed and analysed using a 2(Relational term type: Left/Right vs Cardinal) x 2(Imagined perspective: Egocentric vs Allocentric) repeated-measure ANOVA.

Table 4.2 – Self-report questionnaire average scores. Item 1: Ease of perspective generation; Item 2: Ease of orientation; Item 3: Ease of perspective maintenance; Item 4: Estimate of time spent imagining an egocentric perspective; Item 5: Estimate of time spent imagining an allocentric perspective; Item 6: Average rate of perspective switching; Item 7: Ease of distance visualisation.

Condition	Item 1		Item 2		Item 3		Item 4		Item 5	
	M	SEM	M	SEM	M	SEM	M	SEM	M	SEM
L/R – Ego.	3.62	0.30	2.87	0.15	3.43	0.27	3.81	0.20	2.06	0.21
L/R – Allo.	3.37	0.23	2.62	0.20	3.75	0.19	1.87	0.25	4.06	0.21
Card. – Ego.	2.62	0.31	2.75	0.25	2.87	0.27	3.68	0.19	2.18	0.24
Card. – Allo.	3.56	0.25	2.87	0.22	3.56	0.24	2.06	0.28	3.87	0.28

Condition	Item 6		Item 7	
	M	SEM	M	SEM
L/R – Ego.	2.62	0.27	3.12	0.22
L/R – Allo.	2.00	0.24	2.81	0.31
Card. – Ego.	3.25	0.30	2.87	0.34
Card. – Allo.	2.25	0.34	2.81	0.22

Ease of Spatial Perspective Generation

Item 1 on the questionnaire asked participants to rate their agreement with the statement that they had found it easy to generate the spatial perspective they had been explicitly asked to adopt. The ANOVA revealed no significant main effect of Relational term type, $F(1,19) = 2.72$, $p = .115$, $\eta^2_p = .125$, but a marginally significant main effect of Imagined perspective, $F(1,19) = 4.25$, $p = .074$, $\eta^2_p = .159$. This indicated greater ease of spatial perspective generation when participants were asked to imagine an allocentric perspective, M = 3.4, SEM = .16, compared to when they were asked to imagine an egocentric perspective, M = 3.0, SEM = .20. A marginally significant interaction between the Relational term type and Imagined perspective was also found, $F(1,19) = 4.25$, $p = .053$, $\eta^2_p = .183$ (Figure 4.7). A test of simple main effects revealed that participants found it easier to imagine an egocentric perspective when presented with left/right descriptions, M = 3.4, SEM = .26, than with cardinal descriptions, M = 2.6, $F(1,19) = 5.9$, $p = .025$, $\eta^2_p = .237$. Conversely, they found it equally easy to construct an allocentric representation when presented with left/right, M = 3.4, SEM = .21, and cardinal, M = 3.4, SEM = .22, description, $F(1,19) = < 1$, $p = .1$, $\eta^2_p < .001$. Relatedly, the presentation of cardinal descriptions made it easier for participants to imagine an allocentric perspective of the route, M = 3.4, SEM = .22, than an egocentric one, M = 2.6, SEM = .26, $F(1,19) = 13.36$, $p = .002$, $\eta^2_p = .413$. On the other hand, the presentation of left/right descriptions afforded a comparable ability to map the route onto an egocentric, M = 3.4, SEM = .26, or an allocentric, M = 3.4, SEM = .22, perspective, $F(1,19) = .01$, $p > 1$, $\eta^2_p = .001$. Participants' anecdotal reports at the end of the study generally confirmed this trend, indicating that the generation of an egocentric perspective of an allocentric description was perceived to be more difficult compared to the other, more compatible conditions.

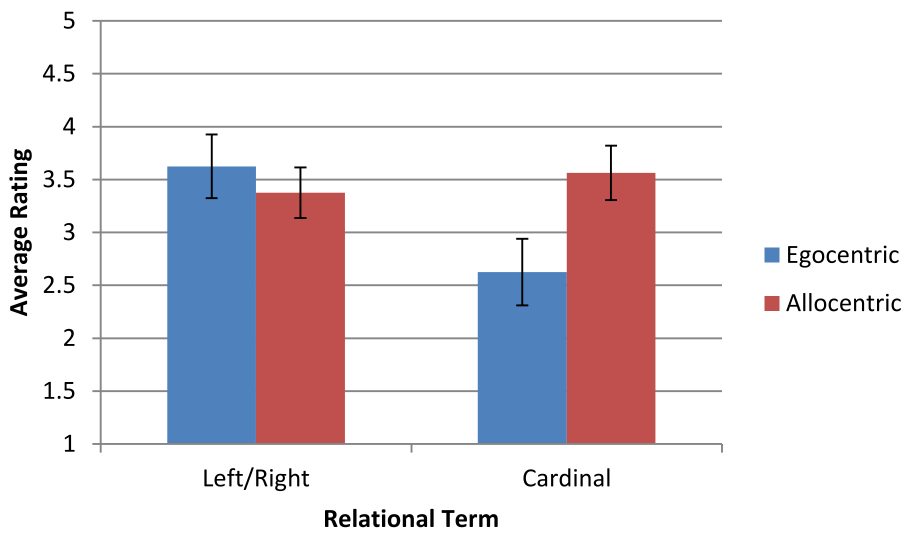


Figure 4.7 – Average rating for the "ease of perspective generation" item. Error bars represent ±1 SEM.

Ease of Orientation

Item 2 asked participants to rate their agreement with the statement that they had found it easy to point back to their starting location. No significant main effects of Relational term type, $F(1,19) < .001$, $p > .999$, $\eta^2_p < .001$, or Imagined perspective, $F(1,19) = .58$, $p = .453$, $\eta^2_p = .030$, were found. The interaction between them was also not significant, $F(1,19) = .74$, $p = .400$, $\eta^2_p = .037$.

Ease of Perspective Maintenance

Item 3 asked participants to rate their agreement with the statement that they had found it easy to maintain the same spatial perspective throughout the reading phase. Marginally significant main effects of Imagined perspective, $F(1,19) = 4.22$, $p = .054$, $\eta^2_p = .182$, and of Relational term type, $F(1,19) = 3.86$, $p = .064$, $\eta^2_p = .169$ were observed, but no significant interaction between them, $F(1,19) = .04$, $p = .832$, $\eta^2_p = .002$ (Figure 4.8). Generally, participants found it marginally easier to maintain the assigned perspective during encoding of left/right descriptions compared to cardinal ones, M = 3.5, SEM = .16, and M = 3.1, SEM = .17, respectively, but they generally found it easier to maintain a stable representation when instructed to generate an allocentric one, M = 3.6, SEM = .16, compared to an egocentric one, M = 3.1, SEM = .21.

Perspective Time Estimates

Items 4 and 5 asked participants to rate their agreement with statements that they had spent most of the encoding time imagining an egocentric or allocentric perspective respectively. In both instances, only a main effect of Imagined perspective was found, $F(1,19) = 81.27$, $p < .001$, $\eta^2_p = .811$, and $F(1,19) = 57.15$, $p < .001$, $\eta^2_p = .751$. No main effects of Relational term type were found, $F(1,19) = .44$, $p = .514$, $\eta^2_p = .023$, and $F(1,19) = .15$, $p = .697$, $\eta^2_p = .008$, and no significant interactions between the two factors, $F(1,19) = 3.11$, $p = .094$, $\eta^2_p = .141$, and $F(1,19) = 1.22$, $p = .283$, $\eta^2_p = .060$. Figure 4.9 presents a summary of participants' agreement ratings for both questionnaire items. Globally, participants reported spending most of the encoding time imagining an egocentric perspective when instructed to do so, M = 3.7, SEM = .14, compared to when they had received the opposite instructions, M = 1.8, SEM = .19. A similar pattern was observed for encoding time spent imagining an allocentric perspective, M = 4.1, SEM = .15, and M = 2.2, SEM = .19, for same and opposite instructions respectively.

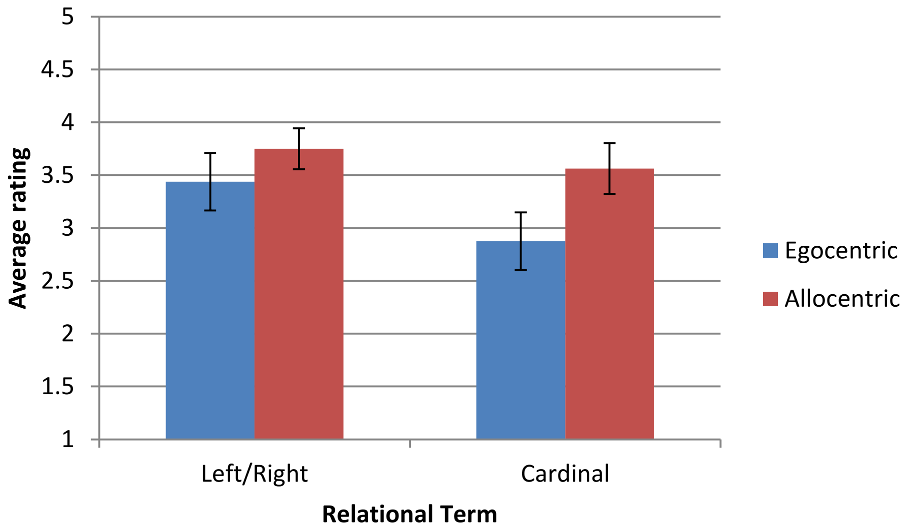


Figure 4.8 – Average rating for the "ease of perspective maintenance" item. Error bars represent ±1 SEM.

Perspective Switching

Item 6 asked participants to rate their agreement with the statement that they spent most of the encoding time switching between an egocentric and an allocentric perspective. Both a main effect of Relational term type, $F(1,19) = 7.122$, $p = .015$, $\eta^2_p = .273$, and of Imagined perspective, $F(1,19) = 6.498$, $p = .020$, $\eta^2_p = .255$, were found for this measure. The interaction between the two factors was not significant, $F(1,19) = .921$, $p = .349$, $\eta^2_p = .046$ (Figure 4.10).

Globally, the ratings showed increased reported perspective switching during encoding of cardinal descriptions compared to the encoding of left/right descriptions, M = 2.5, SEM = .22, and M = 2.1, SEM = .16, respectively. Additionally, participants reported switching imagined perspective more frequently during encoding when instructed to imagine an egocentric perspective, M = 2.7, SEM = .23, compared to when an allocentric perspective was required, M = 1.9, SEM = .22. A comparison between the Left/Right-Egocentric and the Cardinal/Egocentric conditions revealed a significant main effect of Relational term type, with the latter condition resulting in higher reported perspective switching, M = 3.0, SEM = .28, than the former, M = 2.4, SEM = .24, $F(1,19) = 5.516$, $p = .030$, $\eta^2_p = .225$.

Ease of Distance Visualisation

Item 7 asked participants to rate their agreement with the statement that they found it easy to imagine different walking distances. No significant main effects of Relational term type, $F(1,19) = .51$, $p = .483$, $\eta^2_p = .033$, or Imagined perspective, $F(1,19) = .94$, $p = .347$, $\eta^2_p = .059$, were found. The interaction between them was not significant, $F(1,19) = .45$, $p = .510$, $\eta^2_p = .029$. The average scores point to a general difficulty in consciously, mentally visualising path segments of varying lengths across all conditions (Figure 4.11).

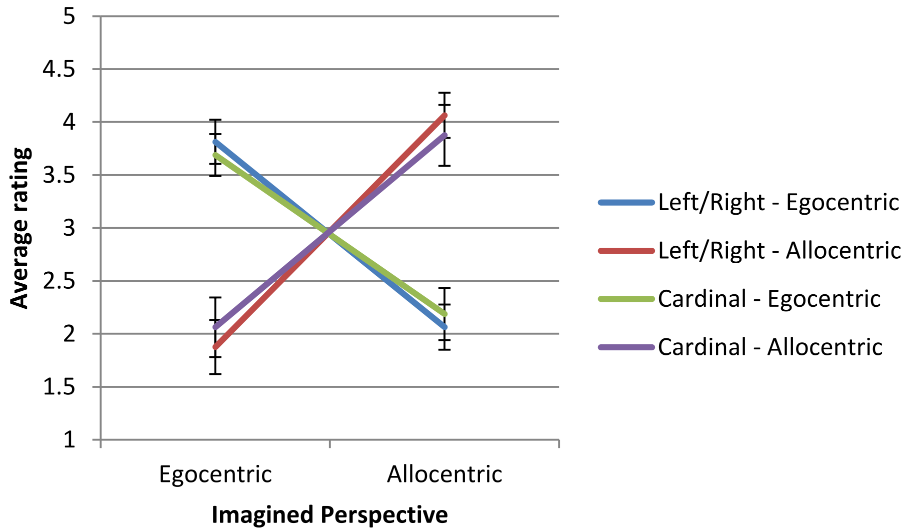


Figure 4.9 – Summary of participants self-report estimates (Y axis) of the perspective they spent most of the encoding time maintaining (X axis) in the different conditions. Error bars represent ±1 SEM.

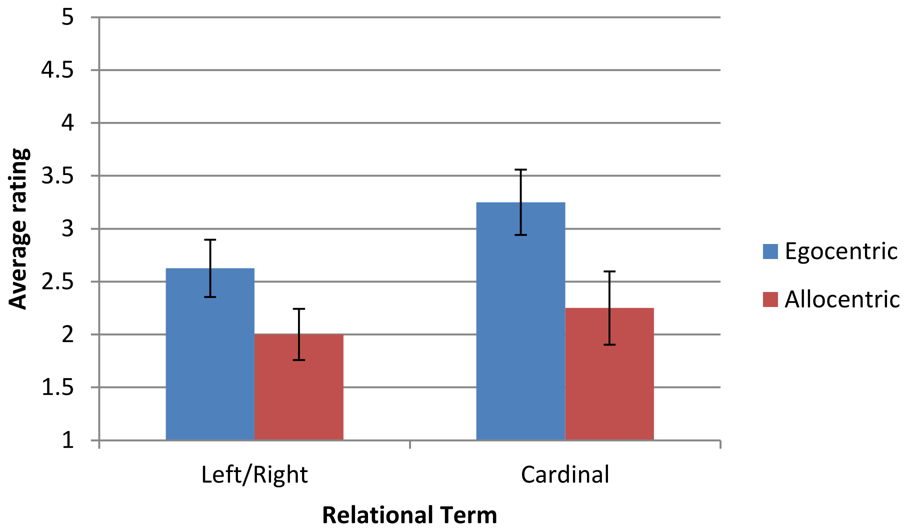


Figure 4.10 – Average rating for the "perspective switching" item. Error bars represent ±1 SEM.

4.4. Discussion

While the overarching goal of the experiments in this chapter was to test the other face of the encoding-test congruence effect by probing spatial knowledge within an egocentric reference frame, this aim poses a number of methodological issues. Chief among them is the issue of determining whether the spatial tasks being used actually probe the type of representations we assume. Although it is probably safe to assume that map-drawing and map-verification tasks do require allocentric representations to be active and dominant at test (an assumption seemingly supported by Experiments 1-3 in Chapter 3), finding a spatial task that is reliably egocentric has proven problematic in the literature. In light of evidence of parallel activation of egocentric and allocentric representations, and of significant individual difference effects in determining the adoption of either spatial strategy, attempting to establish some degree of control over participants' mental representations seems paramount in order to study them efficiently.

The approach taken in Experiment 4 was to manipulate language encoding strategies as done in Experiments 2 and 3, so as to modulate the degree of interference between an imagined spatial perspective and the reference frame in which the route information is provided. The rationale behind the encoding manipulation was the same as in Experiments 2 and 3: successful encoding of spatial information within a compatible imagined perspective (e.g. Left/Right-Egocentric) should either facilitate or hinder performance depending on the perspective from which spatial knowledge is probed during test. In Experiment 4, the task was intended to be an egocentric bearing estimation task. However, as discussed in Section 4.2, response modality might influence the strategy (and, therefore, the perspective) adopted at the time of testing. As such, Experiment 4 was also a test of whether bearing estimations produced using the template in Figure 4.3 would rely on allocentric or egocentric representations.

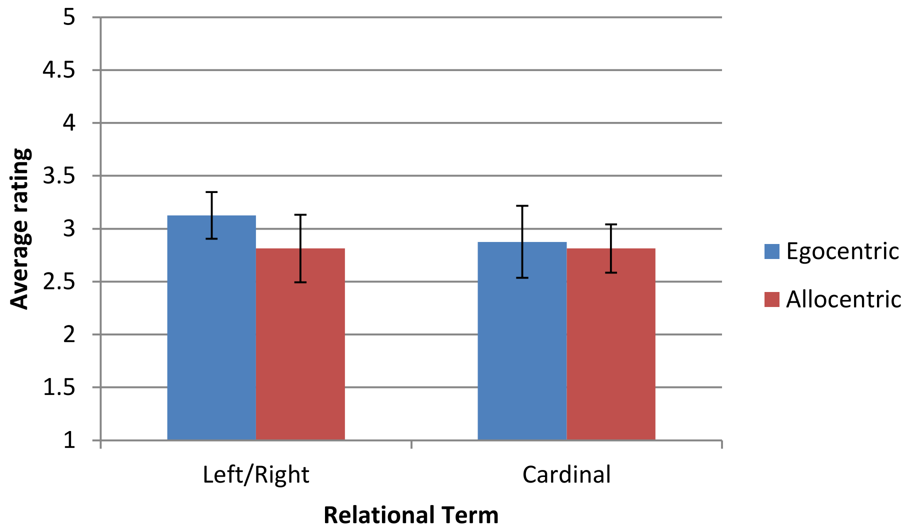


Figure 4.11 – Average rating for the "distance visualisation" item. Error bars represent ±1 SEM.

In line with Experiments 2 and 3, I predicted that a successful encoding of route descriptions in the Left/Right-Egocentric condition should produce more accurate bearing estimates if the task was performed while maintaining an egocentric imagined perspective (i.e. if participants imagined standing at the terminus), because responses would not require a reference frame transformation between encoding and test. This prediction was not confirmed, and mean bearing error appeared to be lowest in the Left/Right-Allocentric condition. On the other hand, this pattern of results was consistent with the possibility that participants were assuming an allocentric perspective during test. As a further confirmation of this, the response heading at the terminus location, which is defined in allocentric terms (whether in a cardinal reference frame or using template-space coordinates. See Figure 4.2.), significantly modulated the degree of error in bearing estimates. Namely, whenever response heading deviated from a north/up orientation, bearing error increased significantly, indicating that participants may have been mentally rotating an allocentric mental representation of the route configuration until the navigator's heading at the final location matched the heading of the navigator on the response template. As a result, any advantage conferred by an efficient egocentric encoding was likely negated by participants' tendency to approach the test phase allocentrically.

To monitor the success of the encoding manipulation, the self-report questionnaire used in Experiment 2 was administered to participants in Experiment 4. The evidence is encouraging. Interestingly, while a few of the self-report questionnaire ratings continue to point towards a degree of reference frame interference in the Cardinal-Egocentric condition (including an increased reported rate of perspective switching not observed in Experiment 2 – Cf. Figure 4.10 and Figure 3.28), this interference is not apparent in participants' responses to Items 4 and 5 (Figure 4.9). In this experiment participants reported spending most of the encoding time maintaining an egocentric representation in the Cardinal-Egocentric condition, as per explicit task instructions, but reported finding this significantly harder than in the other conditions (Figure 4.8). Whether this discrepancy results from a general difficulty with introspection and self-reporting, or from task-related effects is currently unclear. However, one explanation is that the route described in Experiment 4 contained only one directional turn, unlike the four used in Experiments 1-3. As a result, while switching perspective during the turn may have increased the difficulty of integrating the spatial information in a coherent egocentric representation, the actual proportion of trial time spent switching perspective may have been perceived as limited. Future studies will need to explore the interaction between route length and reference frame-perspective interference.

Generally, the pattern of results obtained from the bearing estimation task suggests that, as predicted, the degree of embodiment experienced at the moment of testing is an additional aspect that will determine whether any advantage afforded by a successful egocentric embodiment during encoding is brought to fruition during test. Experiment 5 aimed to provide a test of this idea by altering the way participants produced their bearing estimates.

Experiment 5: Bearing Estimation with Pointing Device

4.5. Experiment 5: Introduction

Experiment 4 highlighted the difficulty in maintaining experimental control over participants' mental representations during the test phase. More specifically, the pattern of results observed appears to suggest that participants, at the moment of estimating their homing vector, were maintaining an allocentric imagined perspective of the routes presented. We suspected this was due to the seemingly bird's-eye view perspective represented on the templates used by participants to perform their estimates (Figure 4.3). To verify this, Experiment 5 was conducted under the same experimental conditions but with the crucial difference of using a compass-like pointing device (Figure 4.12), placed in front of the participant, with the goal of prompting a purely egocentric response (or as close to one as empirically possible). I hypothesised that this would more effectively prompt participants to assume an egocentric, first-person view at the moment of pointing, possibly increasing the sense of embodiment of the egocentric reference frame (or helping to maintain what was established during information encoding), and leading to a reduction in error rates in the Left/Right-Egocentric condition compared to the previous experiment. The same predictions were formulated as in Experiment 4.

4.6. Methods

4.6.1. Participants

20 adult English native speakers with normal or corrected-to-normal vision (7 males, 13 females, mean age 24.50 ± .99 years) were recruited across the University of Nottingham in exchange for credits or an inconvenience allowance.

4.6.2. Design and Materials

The design and materials used were the same as in Experiment 4 (see Appendix IV). In this case, however, a compass-like pointing device was used in place of the paper templates employed in the previous study. The device in question was a sheet of hard, grey plastic 50 cm in width and 22.5 cm in height. A pivoting needle 11.5 cm long rotated about the intersection point of the rectangle's axes. A 360° protractor was used to acquire participants' responses, using the upper vertical axis of symmetry as the 0° origin point.

4.6.3. Apparatus

As in Experiment 4, participants were seated at a comfortable distance from the laptop used to play the same recordings used in Experiment 4, and listened to them through a headset. Throughout each block, the relevant image used (also in the previous experiment) to explicitly prompt a specific spatial perspective remained visible on screen. The pointing device was placed between the participant and the laptop.

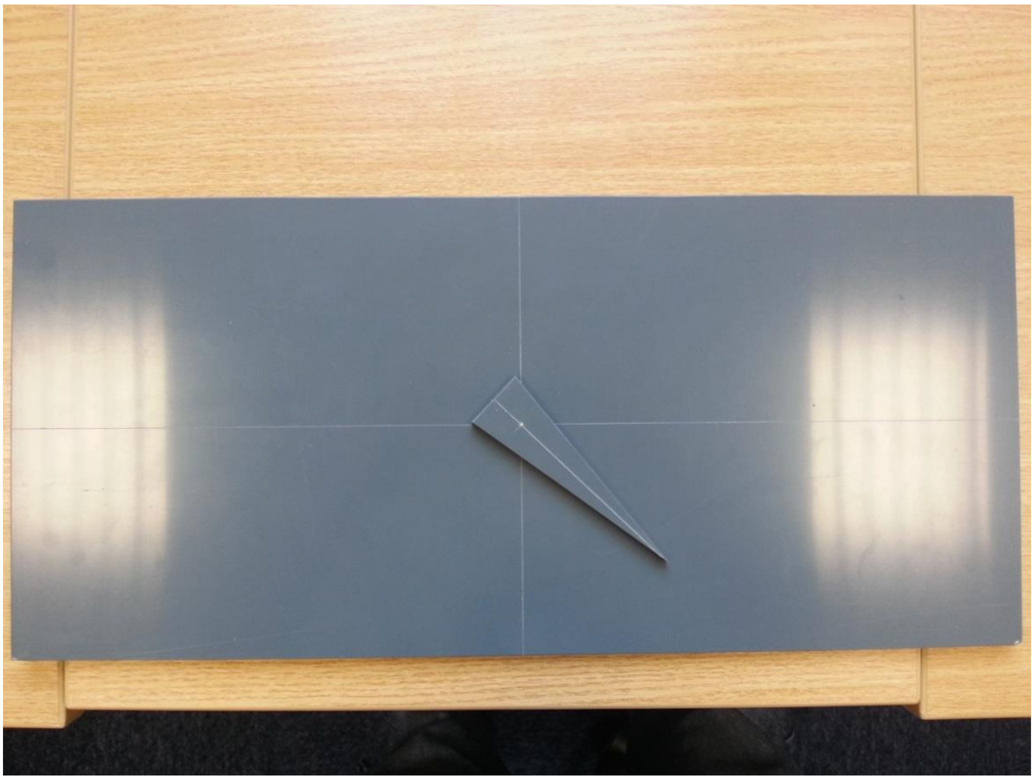


Figure 4.12 – Pointing device used to record participants' egocentric relative position judgements.

4.6.4. Procedure

Participants were each assigned a different permutation of the four condition blocks. The relevant image used to explicitly prompt the required imagined perspective appeared on screen. They then listened to each recording once, with their eyes closed, and were then prompted to mark the direction of the origin point relative to the last location visited. Unlike in Experiment 4, this was done by positioning the compass-like needle on the imaginary vector leading back to the origin point. Once the participant had positioned the needle, the position was recorded using a 360° protractor and the pointer reset to the 0° (upward) position. After each block of 16 trials, the self-report questionnaire was administered, and a new block started.

4.7. Results

4.7.1. Bearing Estimation

Bearing error was computed as in Experiment 4. Comparing average bearing estimation error between conditions revealed a significant effect of Relational term type, $F(1,19) = 15.61$, $p = .001$, $\eta^2_p = .451$, but no significant main effect of Imagined perspective, $F(1,19) = .883$, $p = .359$, $\eta^2_p = .044$, or interaction between them, $F(1,19) = .019$, $p = .892$, $\eta^2_p = .001$. Overall, the average bearing estimation error was found to be significantly lower after listening to descriptions containing egocentric relational terms, M = 27.31°, SEM = 4.6°, than following encoding of cardinal descriptions, M = 43.12°, SEM = 3.7°, regardless of the perspective imagined during encoding. As in Experiment 4, I further investigated the effect of final allocentric heading on pointing accuracy, and found it to be non-significant in left/right conditions, F(1,19) = 2.07, p = .166, η^2_p = .098 (M = 21.61°, SEM = 6.4°, for a final "up" heading, and M = 33.01°, SEM = 5.7°, for a final "down" heading). It was however still significant in cardinal conditions, F(1,19) = 5.32, p = .008, η^2_p = .219, albeit with a smaller effect size relative to Experiment 4. Additionally, unlike in the previous study, a decrease in average error was also observed for a final south heading. Estimate errors in cardinal conditions were M = 31.85°, SEM = 7.3°, for a northward heading, M = 31.73°, SEM = 4.9°, for a southward heading, M = 55.95°, SEM = 5.6°, for an eastward heading, and M = 52.95°, SEM = 6.6°, for a westward heading.

As a follow-up analysis, data from Experiments 4 and 5 were pooled, and angular error analyses were then repeated including Response type as a between-subject factor (Template vs Device) in the original design. The goal was to assess the impact a manual pointing response may have had on egocentric perspective embodiment (and, consequently, on performance) compared to a 2D template. Analysing overall bearing estimate error in such a way revealed a significant main effect of Relational term type, F(1,38) = 18.01, p < .001, η^2_p = .322, indicating lower bearing error following encoding of left/right descriptions, M = 30.34°, SEM = 3.2°, compared to encoding of cardinal descriptions, M = 41.71°, SEM = 2.4°. No significant main effect of Imagined perspective was found, F(1,38) = 1.01, p = .321, η^2_p = .026, nor a significant main effect of the between-subject factor Response type, F(1,38) = .10, p = .75, η^2_p = .003. However, an interaction between Imagined perspective and Response type was found, F(1,38) = 5.73, p = .022, η^2_p = .131.

An analysis of simple main effects revealed that, using the template in Experiment 4, participants were globally more accurate in their bearing estimates when instructed to maintain an allocentric reference frame during encoding, M = 33.71°, SEM = 3.8°, than when instructed to imagine an egocentric perspective, M = 39.96°, SEM = 3.8°, F(1,38) = 5.78, p = .021, η^2_p = .132. Use of the pointing device in Experiment 5, on the other hand, reduced the bearing error under egocentric imagery conditions, M = 33.94°, SEM = 3.8°, so that it did not differ significantly from the error rate under allocentric imagery conditions, M = 36.49°, SEM = 3.8°, F(1,38) = .96, p = .332, η^2_p = .025. Lastly, I ran a simple mixed-effects ANOVA with Condition as a four-level within-subject factor, and Response type as a two-level between-subject factor, in order to determine what condition had seen the most significant decrease in bearing estimate error with the use of the manual pointing device. With a mean difference of 12.25°, SEM = 6.8°, the Left/Right-Egocentric condition was the only condition in which bearing estimate errors had decreased in Experiment 5 compared to Experiment 4, M = 25.87°, SEM = 4.8°, and M = 38.13°, SEM = 4.8°, respectively, although the difference only approached significance, $F(1,38) = 3.23$, $p = .080$, $\eta^2_p = .079$. The three other conditions remained unchanged, all Fs < 1. Figure 4.13 compares average bearing estimation error in the different conditions in Experiments 4 and 5. See Figure 4.14 and Figure 4.15 for comparisons of the final heading effects between the two experiments.

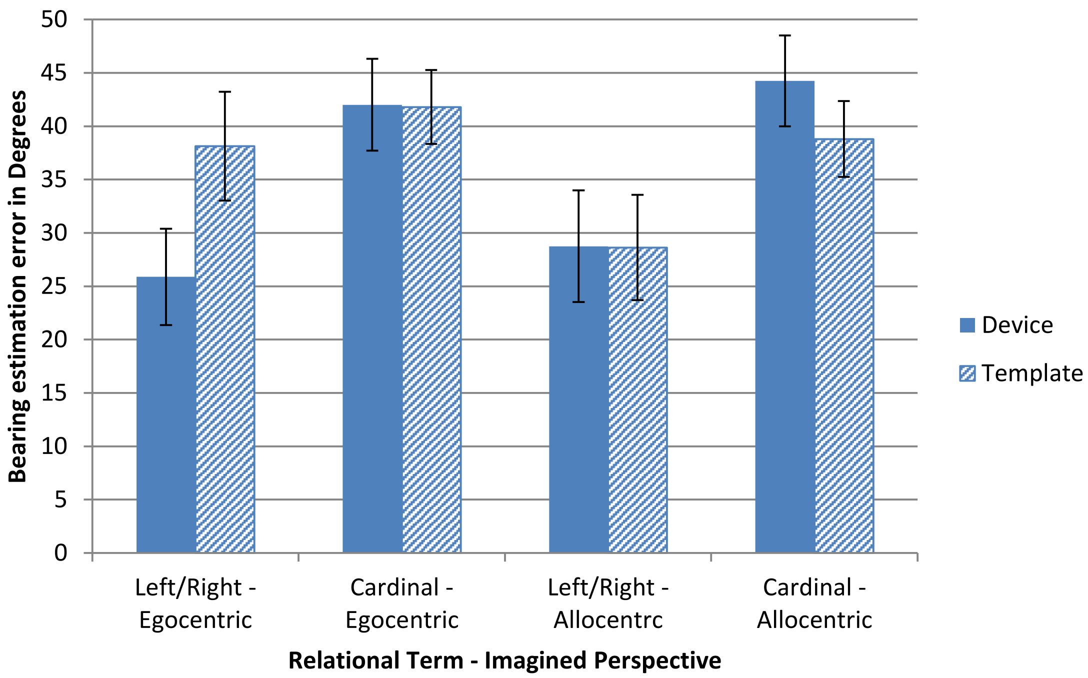


Figure 4.13 – Average bearing estimation error in the four conditions as a function of test mode. Error bars represent ±1 SEM.

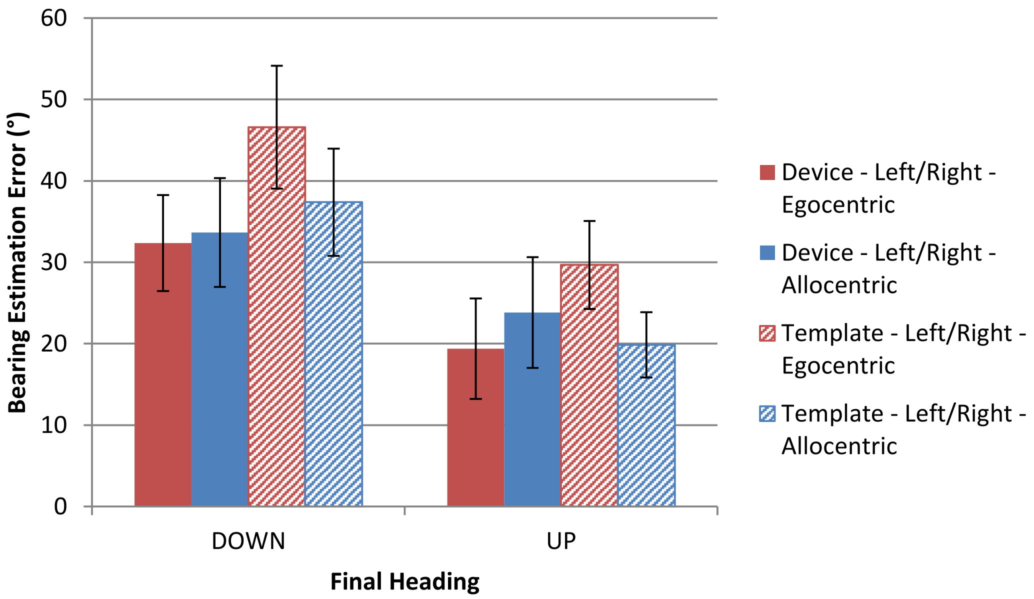


Figure 4.14 - Bearing estimation error in left/right conditions as a function of final heading and test mode. Error bars represent ±1 SEM.

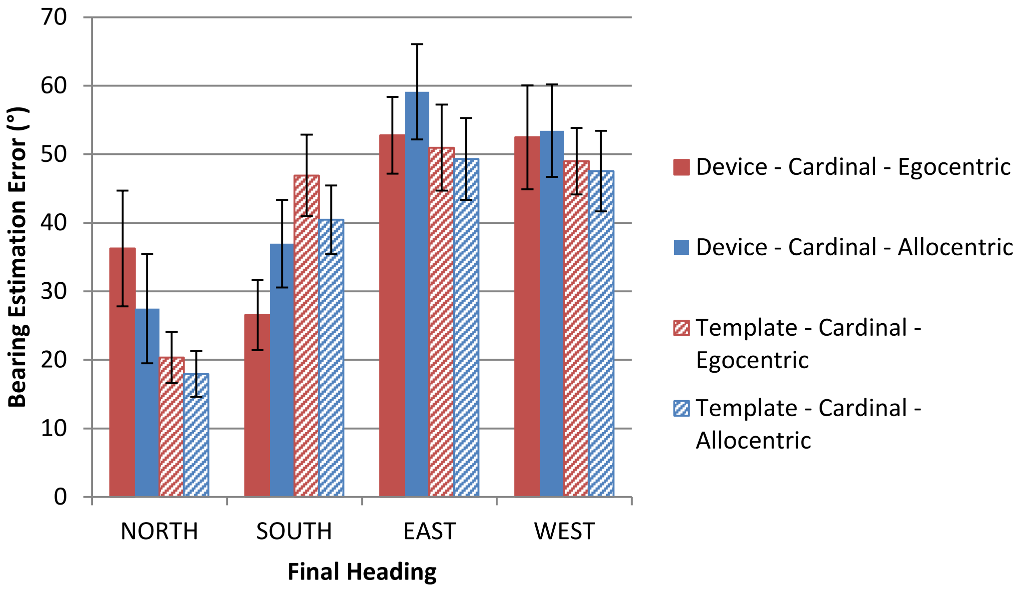


Figure 4.15 - Bearing estimation error in cardinal conditions as a function of final heading and test mode. Error bars represent ±1 SEM.

4.7.2. Questionnaire Results

Table 4.3 includes the descriptive statistics for all questionnaire items for all conditions. These are discussed individually in the following sections.

Table 4.3 – Self-report questionnaire average scores. Item 1: Ease of perspective generation; Item 2: Ease of orientation; Item 3: Ease of perspective maintenance; Item 4: Estimate of time spent imagining an egocentric perspective; Item 5: Estimate of time spent imagining an allocentric perspective; Item 6: Average rate of perspective switching; Item 7: Ease of distance visualisation.

Condition	Item 1		Item 2		Item 3		Item 4		Item 5	
	M	SEM	M	SEM	M	SEM	M	SEM	M	SEM
L/R – Ego.	3.10	0.27	2.70	0.23	3.20	0.27	3.75	0.19	1.85	0.24
L/R – Allo.	3.30	0.26	2.75	0.21	3.40	0.24	2.05	0.18	3.85	0.23
Card. – Ego.	2.75	0.23	2.50	0.21	3.15	0.23	3.60	0.23	2.20	0.23
Card. – Allo.	2.95	0.29	2.50	0.23	3.55	0.28	2.15	0.22	4.05	0.18

Condition	Item 6		Item 7	
	M	SEM	M	SEM
L/R – Ego.	2.20	0.24	2.45	0.25
L/R – Allo.	2.35	0.23	2.90	0.24
Card. – Ego.	2.20	0.21	2.50	0.25
Card. – Allo.	2.10	0.20	2.55	0.25

Ease of Spatial Perspective Generation

Item 1 on the questionnaire asked participants to rate their agreement with the statement that they had found it easy to generate the spatial perspective they had been explicitly asked to adopt. No significant main effects of Relational term type, $F(1,19) = 2.99$, $p = .100$, $\eta^2_p = .136$, or Imagined perspective, $F(1,19) = .62$, $p = .438$,

η^2_p = .032, were found. The interaction between them was also not significant, F(1,19) < .001, p > .999, η^2_p < .001.

Ease of Orientation

Item 2 asked participants to rate their agreement with the statement that they had found it easy to point back to their starting location. No significant main effects of Relational term type, F(1,19) = 1.08, p = .311, η^2_p = .054, or Imagined perspective, F(1,19) = .01, p = .897, η^2_p = .001, were found. The interaction between them was also not significant, F(1,19) = .04, p = .841, η^2_p = .002.

Ease of Perspective Maintenance

Item 3 asked participants to rate their agreement with the statement that they had found it easy to maintain the same spatial perspective throughout the reading phase. No significant main effects of Relational term type, F(1,19) = .10, p = .748, η^2_p = .006, or Imagined perspective, F(1,19) = .91, p = .350, η^2_p = .046, were found. The interaction between them was also not significant, F(1,19) = .25, p = .618, η^2_p = .013.

Perspective Time Estimates

Items 4 and 5 asked participants to rate their agreement with statements that they had spent most of the encoding time imagining an egocentric or allocentric perspective respectively. In both instances, only a main effect of Imagined perspective was found, F(1,19) = 27.61, p = .001, η^2_p = .592, and F(1,19) = 38.43, p = .001, η^2_p = .669. This indicated that participants spent most of the encoding time imagining an egocentric perspective when instructed to do so, M = 3.6, SEM = .18, but not when instructed otherwise, M = 2.1, SEM = .18. Similarly, they had spent most of the encoding phase imagining an allocentric perspective when instructed to do so, M = 3.9, SEM = .17, but not when instructed otherwise, M = 2.0, SEM = .21. No main effects of Relational term type were found, F(1,19) = .02, p = .871, η^2_p = .001, and F(1,19) = 2.34, p = .142, η^2_p = .110, and no significant interactions between the two factors, F(1,19) = .922, p = .349, η^2_p = .046, and F(1,19) = .280, p = .603, η^2_p = .015. Figure 4.16 presents a summary of participants' agreement ratings for both questionnaire items. In general, participants seemed to comply with explicit task instructions, generating and maintaining the required spatial perspectives without any sign of reference frame interference.

Perspective Switching

Item 6 asked participants to rate their agreement with the statement that they spent most of the encoding time switching between an egocentric and an allocentric perspective. No significant main effects of Relational term type, F(1,19) = .74, p = .398, η^2_p = .038, or Imagined perspective, F(1,19) = .01, p = .905, η^2_p = .001, were found. The interaction between them was also not significant, F(1,19) = .85, p = .367, η^2_p = .043.

Ease of Distance Visualisation

Item 7 asked participants to rate their agreement with the statement that they found it easy to imagine different walking distances. Relational term type had no significant effect on the ability of participants to visualise path legs of different lengths, $F(1,19) = .60$, $p = .445$, $\eta^2_p = .031$. A significant main effect of Imagined perspective was found, $F(1,19) = 5.58$, $p = .029$, $\eta^2_p = .227$, yielding a higher average rating for allocentric conditions, M = 2.7, SEM = .21, than for egocentric conditions, M = 2.4., SEM = .22, but no significant interaction between the two factors was found, $F(1,19) = 1.74$, $p = .202$, $\eta^2_p = .084$. By contrasting only the two left/right conditions a significant main effect of Imagined perspective was observed, $F(1,19) = 7.02$, $p = .016$, $\eta^2_p = .270$, with a reported advantage in the ease of distance visualisation when imagining an allocentric perspective, M = 2.9, SEM = .24, compared to an egocentric perspective, M = 2.4, SEM = .25. However, ratings for this measure were generally quite low (Figure 4.17).

4.8. Discussion

Experiments 4 and 5 were an attempt to measure distinct patterns of JRD performance in a triangle completion task following encoding of described routes. By using the same manipulation during the language encoding phase, coupled with the same task instructions, Experiment 5 specifically tried to isolate the contribution of an alternative response format to differences in performance in order to address specific hypotheses. Namely, that egocentric encoding of a route description can support spatial updating processes in the absence of proprioceptive and vestibular input (contra Klatzky et al., 1998). Subsequently, that successful egocentric updating during imaginal locomotion should afford an advantage during a subsequent egocentric task. Furthermore, I hypothesised that said advantage will manifest provided the response format preserves the encoding embodiment. Relatedly, I proposed that body-referred responses (e.g. using a physical pointer) can be used to achieve such preservation of embodiment (contra Avraamides et al., 2004, and Wraga, 2003). As a corollary, the diverging patterns of results between Experiment 4 and Experiment 5 were used to differentiate the processes likely involved in the two types of responses.

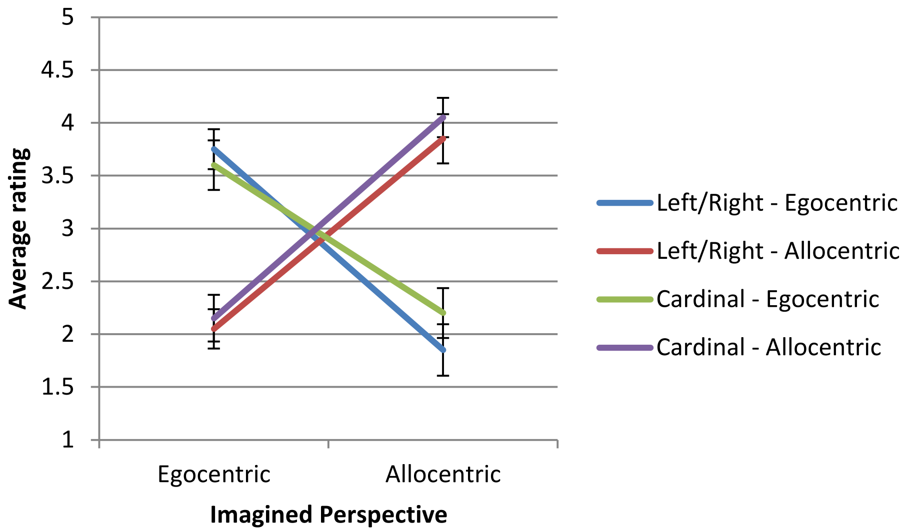


Figure 4.16 - Summary of participants self-report estimates (Y axis) of the perspective they spent most of the encoding time maintaining (X axis) in the different conditions. Error bars represent ±1 SEM.

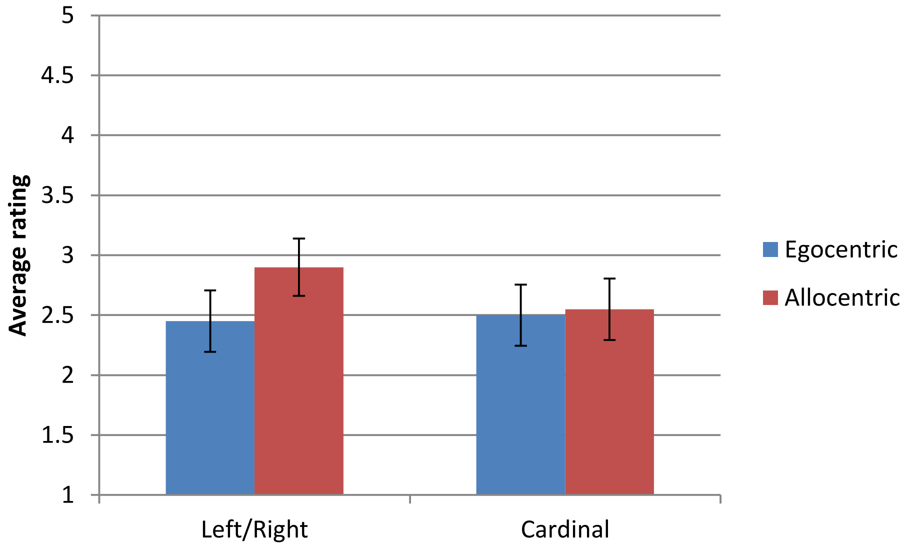


Figure 4.17 - Average rating for the "distance visualisation" item. Error bars represent ±1 SEM.

Results revealed that estimating the bearing to the origin from a terminus location using a physical pointer following imagined locomotion appeared to maintain the embodiment generated by Left/Right-Egocentric, leading to a 12.25° reduction in bearing error in that condition in Experiment 5 compared to Experiment 4. While the difference only approached significance (possibly due to the relatively small sample size), the possible practical implications warrant further attention. Namely, would a 12.25° reduction in bearing error make the difference between correctly returning to one's destination (e.g. via a different route) and getting lost under ecological conditions?

Additionally, the use of physical responses appeared to reduce the effect of response heading, used in Experiment 4 as a hallmark of allocentric processing. However, this warrants further investigation. The effect was still present at the group level in Experiment 5, indicating that the issue may be better explored via an individual-difference approach, similar to that taken by Gramann and colleagues (2005;2006;2010; Gramann, 2013). In light of these results, the self-report questionnaire used in Experiments 2, 4, and 5, will require further validation. While the consistency on several items between Experiments 2 and 4 hints at its potential usefulness in exploring the phenomenology of spatial imagery, ratings in Experiment 5 appeared less consistent. Beyond a sample size issue, the five-point scales used may have been far too coarse to capture some of the nuances between different types of representations.

It is important to note how response accuracy was comparable between the left/right-egocentric and the left/right-allocentric conditions. This would seem to run contrary to the idea that different encoding processes in the two conditions will result in different mental representations. However, this pattern of results might be explained in two ways. The use of the more familiar egocentric relational terms from an allocentric imagined perspective might have allowed participants to approach the task in two ways. On the one hand, if they performed the task allocentrically, it might have made it easier for participants to mentally rotate the path configuration to match the perspective required at test. This could have made their responses more accurate than in the cardinal-allocentric condition. Alternatively, it might simply have made it easier for them to “zoom in” to the terminus location and assume an imagined egocentric perspective at the time of testing compared to the cardinal-allocentric condition and on par with the left/right-egocentric condition. The current data do not allow to disambiguate between these two possibilities, but a paradigm allowing to record response times between description offset and pointing might. That is, if participants are encoding the route egocentrically (and updating their bearing during encoding) in the left/right-egocentric condition, then we might expect this particular combination of relational terms and imagined perspective to yield the fastest response times, as no reference frame translation would be required. Similarly, if participants in the left/right-allocentric and cardinal-allocentric conditions are encoding the route allocentrically, the reference frame translation process should delay their responses compared to the left/right-egocentric condition. Lastly, if the construction of an allocentric model and the “zooming in” to an egocentric imagined perspective is made easier by the use of more familiar egocentric relational terms, then responses should be faster in the left/right-allocentric condition compared to the cardinal-allocentric condition. Such a paradigm could also attempt to incorporate a brain-imaging element, to further establish whether different processes are involved during encoding and response in the different conditions.

In spite of these limitations, Experiment 5 was globally successful in testing the hypotheses formulated on the basis of the literature and of previous experiments described in this thesis. Namely, it showed that congruence between encoding and test yields better performance (i.e. bearing accuracy), and that response modality plays a role in determining whether congruence between linguistic reference frame and imagined perspective will produce an advantage. More specifically, it revealed that maintaining reference frame embodiment during test might be an important factor during egocentric spatial perspective taking. Additionally, it reinforced the need, already presented with Experiments 1-3, to constrain mental representations of space at all stages of investigation in order to interpret results in any meaningful way. Relatedly, it stressed the importance of task analysis (e.g. Marr, 1982) in understanding the processes involved in performing a task and, more generally, in the cognitive system under investigation. Most significantly, Experiment 5 demonstrated the ability of participants to estimate egocentric bearings following imagined locomotion with angular errors in line with (sometimes lower than) other results in the literature. These include conditions involving physical motion (e.g. 20-40° in the physical walk condition in Chance et al., 1998), and all conditions (except Walk) used by Klatzky et al. (1998) for a Turn 1 value of 90°.

As such, these results could be an important part of the debate surrounding the embodied nature of egocentric mental imagery (both during language processing and in general) and the possible internal simulation of proprioceptive and vestibular input (Byrne et al., 2007), and will require further investigation. Future research possibilities will be discussed in Chapter 6, following a summary of the main results presented in this thesis. Before that, however, Chapter 5 will present three studies attempting to better elucidate developmental aspects of two key navigational elements that Experiments 1-5 have already explored in typical adults: landmark salience and reference frame transformation.

CHAPTER 5
Landmark Salience and Reference Frames during Development

5.1. Landmark Salience during Development

Experiments 1-5 in this thesis investigated the encoding and use of spatial language in healthy adults. In particular, we focused on two cognitive mechanisms that appear central to visuospatial behaviour and navigation: the ability to identify salient landmarks in an environment (or in a description of it), and the ability to fluidly translate imagined spatial perspectives in order to perform different types of spatial tasks. Interestingly, analyses of eye movement patterns during both spatial text encoding and map verification suggest that a host of linguistic and cognitive factors can actively modulate landmark salience. More specifically, these experiments have explored the effects of an individual's imagined spatial perspective and of the reference frame implied by the text itself in modulating attentional allocation to landmark words and landmark map regions. The most consistent result thus far appears to be a reduction in the allocation of attention to navigationally salient landmarks during processing of cardinal route descriptions and when constructing an allocentric representation of a described environment. This effect hints at a complex interaction between landmark knowledge, the perception of landmark salience, and the mental manipulation of spatial perspectives.

Landmark salience and the role of landmarks in place learning were explored in Section 1.3, with reference to a number of studies (e.g. Janzen & van Turennout, 2004) that investigated how objects along a route are encoded differently depending on their navigational salience (e.g. their being located at a turn location or midway along a straight section of a route) and their cognitive salience (e.g. due to task instructions prioritising a certain semantic category of objects). The basic paradigm used in those studies involved the presentation of an egocentric route (either active or passive) through a virtual environment containing landmark objects at various locations (i.e. decision- and non-decision points), followed by an object recognition task aimed at testing differential landmark activation in memory. Behavioural and brain imaging results generally revealed the ease and rapidity with which healthy adults can perceive certain features in the environment as being more navigationally useful than others, but also the flexibility with which this landmark selection process is amenable to top-down effects. However, the goal of this chapter is to explore these processes within the context of the acquisition of spatial knowledge during childhood, and of the development of our abilities to make use of it. I will begin by reviewing some of the relevant developmental literature, before introducing three developmental studies examining landmark use and reference frame transformation in children.

The development of landmark memory and of children's more general wayfinding abilities has been an active area of research for the past few decades. Siegel and White (1975) formalised one of the first models of spatial knowledge acquisition. It postulated a sequential development of spatial knowledge as a function of direct experience with an environment: during a first stage of experience, children acquire knowledge of and develop memory for the perceptual features of landmarks in an environment (landmark knowledge). Subsequent exposures to the same environment then allow them to develop a memory for the sequence of direction changes in a path, and to establish relationships between turns in a path and the landmarks located in their proximity (route knowledge). This finally allows them to combine their egocentric perceptual experiences into a cohesive, allocentric cognitive maps that can support a variety of spatial tasks (variously referred to as survey knowledge or configural knowledge).

In support for this model, Herman and Siegel (1977) tested the ability of 20 children between the ages of 5 and 11 to recall 19 landmarks and their spatial locations in a classroom-sized model town after walking through it repeatedly. Each landmark was labelled and described upon encountering it (providing also non-spatial context as to its function within the city, e.g. "*This is the schoolhouse of the town. All of the children of this town go here to study very hard.*"). After each of three consecutive exposures to the model, or only after the third one, children were tested on their ability to recall whether a set of landmarks had appeared in the town, as well as on their ability to place them in their correct locations in the model. Children's topological accuracy (the ability to place landmarks in the correct quadrant of the town) and Euclidean accuracy were measured, and both were found to improve with age and with repeated exposures to an environment, with the younger children requiring more familiarisation with the environment to achieve Euclidean accuracy comparable to that of the older children.

Subsequently, Allen, Kirasic, Siegel and Herman (1979) compared children between the ages of 6 and 10 (second and fifth grade) to university students on their ability to select landmarks at navigationally salient locations along a route. Participants were presented with a series of 52 photographs depicting a first-person view of a route through a commercial neighbourhood. The route contained ten critical areas, defined as locations with an actual or potential change in heading (i.e. junctions). After viewing all 52 photographs in a slideshow format, participants were presented with all of them on a rectangular display, and asked to select those that would most help them identify their location along the route.

The number of photographs depicting the critical areas of the route and chosen by the participants was compared between the three age groups. Globally, this number was found to increase as a function of grade level, and the children (in both age groups) were found to often select perceptually salient but spatially ambiguous scenes. In a follow-up experiment, a different sample of second- and fifth-grade children were again shown the route walkthrough, and presented with the scenes selected as navigationally salient in the previous experiment, either by their peers or by the adult group. In each case, they were iteratively presented with a target scene, and asked to select the one closest to it in the route, until all scenes had been used as point of origin. Results revealed that fifth graders were significantly more accurate in their distance judgements when they were tested using adult-selected scenes, compared to when they were tested using peer-selected scenes. This appeared to indicate a dissociation between the ability to select salient landmarks within the context of an environment, and the ability to use that information to inform and support wayfinding, with the latter remarkably preceding the former.

Rowen and Hardwick (1983) later demonstrated that an ability to use available environmental features as landmarks, and to take advantage of them to recall spatial locations, can already be observed in kindergarten children (mean age 5.6 years). In their study, children were escorted through a network of interconnected hallways in search of a target object. The location of the target object could either be unmarked, marked by a low-salience landmark (e.g. a chair), or by a high-salience landmark (e.g. a bright orange highway cone). Once found, they were instructed to carry it back to the location where their search started by continuing to walk in the same direction. Once at the starting point, they were instructed to carry the object back to the location where they had found it, either by walking in the same direction, or by reversing their direction of travel. Replacement error (the distance between the target's original location and the replacement one) was used as measure of performance.

While direction of travel had no significant effect on performance, the presence of a landmark – and its salience – was found to significantly modulate performance. This primacy for perceptual salience in the selection of environmental features as landmarks, seemed therefore to offer an explanation for the sometimes conflicting findings concerning the facilitation effect of landmarks in children's spatial task performance (see Feldman & Acredolo, 1979). However, studies that tested children's spatial abilities during direct experience with large-scale environments were still lacking. Cornell, Heth, and Broda (1989) attempted to do this by escorting children (a group of six year olds and a group of 12 year olds) on a tour across an unfamiliar university campus and then asking them to retrace their steps. Prior to the test phase, children were either not informed that they would need to lead the way on the return journey, generally instructed to pay attention to the route because they would have to, or explicitly directed to pay attention to either distal or junction landmarks. Children who had been prompted to pay attention to landmarks generally performed better than other groups, but whereas the 12 year olds were able to make use of distal landmarks, the younger children could only take advantage of landmarks when these where located at junctions.

These findings complemented those of Allen et al (1979), suggesting that salient landmark selection is not automatic in younger children, and that when prompted to encode landmarks, distance is an important factor for them.

The developmental changes in landmark use and landmark-based navigation in large-scale environments were further explored by Heth, Cornell and Alberts (1997). Eight- and 12-year-old children were guided along a route through a university campus, and their attention was directed to landmarks located at T and Y intersections. After experimenters had changed the location or orientation of certain landmarks, the children were guided on a return journey to the starting point. The return journey included four detours off the original path, with four test locations off route and four on route. At these locations the children were asked whether they were on or off the original path, and to point in the direction to take to return to the original path (or to the starting point, if already on the original path). Additionally, they were asked to explain what motivated their choices, and, at the end of the route, they were shown line drawings of the designated landmarks and asked to determine whether anything about them had changed.

Younger children were generally more likely to identify on-route junctions where a landmark had been manipulated as being off the original path, and they were also less adept at finding their way back to the original path after a detour. Additionally, while the two age groups did not differ in their tendency to make use of the designated landmarks that had been pointed out to them during the original route (or in their tendency to report doing so), the older children were significantly more likely to identify additional landmarks in the environment. In doing so, they were also significantly more likely to select stable landmarks (as opposed to movable objects) and distal landmarks.

More recently, Clearfield (2004) investigated the development of landmark use in 8-, 11-, and 14-month-old infants. The study focused on the interaction between mode of locomotion (crawling vs walking) and locomotive experience in the development of cue learning (the association of a spatial location to a visible landmark) and place learning (the use of two or more distal landmarks to identify a spatial location). In an adaptation of the Morris water maze task, the infants were placed inside an octagonal arena and tested on their ability to move towards their parent (who was visible outside the arena during baseline trials, but remained hidden behind the arena walls during test trials).

During the test of place learning abilities, infants had to rely on distal landmarks located outside the arena, whereas during the test of cue learning, the parent’s location was marked with a visible landmark. Results revealed that cue learning generally increased memory for spatial locations, but also that infants’ locomotive experience in their respective modes of locomotion was a key factor in their ability to complete the tasks: experienced crawlers performed better than novice crawlers and novice walkers, and experienced walkers performed better than novice walkers. This suggests that spatial understanding is intimately connected to the way we experience and interact with space, and that a few of the spatial abilities acquired during crawling do not persist with the acquisition of a new mode of locomotion. Such findings point to embodiment as a crucial aspect in the acquisition and use of spatial abilities, but could be of relevance to the study of mental representations of space built in the absence of the proprioceptive cues generated by most experiential interactions with our surroundings.

In this vein, Jansen-Osman and Fuchs (2006) studied the role of different types of landmarks, and of their associations to spatial locations, on human wayfinding behaviour and spatial knowledge acquisition during exploration of a virtual environment. 60 children from two grade levels (second and sixth) were compared to adults on their ability to: 1) find, and learn to criterion, the shortest route through a maze to a target landmark; 2) estimate the direction back to the starting point; 3) find the shortest alternative route back to the starting point, followed by the shortest alternative route to the target landmark once more; 4) mark the position of the target landmark on a map-like overview of the maze; 5) recall as many landmarks as they could remember, and localise as many as they could on the map of the maze.

Generally, the younger children were found to benefit from the presence of landmarks in an environment when learning a route to criterion, but no more so than older children or adults. Interestingly, no effect of landmark presence on orientation behaviour or on the other measures of spatial knowledge acquisition was found. Younger children were more likely than older children or adults to return to the starting location or to revisit the same segments of the maze, and were significantly less precise than both other age groups in their heading estimations, with or without landmarks present. Younger children were able to recall more landmarks located at correct turn locations compared to older children and adults if such landmarks belonged to the same semantic category (e.g. animals). However, their ability to also localise them correctly was poor. This was taken as evidence of a qualitative difference between the spatial mental representations of children and those of adults. While eight year olds were aware of landmarks and were able to use them during a wayfinding task, their poor spatial recall appeared to indicate only a weak association between landmarks and precise spatial locations. In other words, children did not appear to form reliable survey representations of the maze, but were capable of forming egocentric heading decision-landmark associations necessary to support wayfinding.

A tendency for children to differentiate between salient and non-salient landmarks in a virtual environment was also found by Farran, Courbois, Van Herwegen and Blades (2012), who compared a group of typically developing (TD) children between the ages of six and nine, and a group of children with Williams syndrome (WS). Participants were initially passively guided through the virtual environment by an experimenter who verbally highlighted turns, with or without reference to the landmarks located near it. Participants were then allowed to navigate the environment from start to end until they could do so without errors. Following this training phase, children were once again led on a tour of the environment but with the landmarks removed, and tested both on measures of landmark location and of landmark identity. Results revealed that TD children as young as six were able to differentiate between navigationally salient (junction) and non-salient (path) landmarks, and that junction landmarks were more strongly represented in memory.

However, Lingwood, Blades, Farran, Courbois and Matthews (2015) found that even in the presence of landmarks, the ability of a few six-year-old children to benefit from them was sometimes conditional on landmark cues. They tested a large sample of six-, eight-, and ten-year-old children, as well as adults, on their ability to retrace a route in a virtual environment with six turns when this contained no landmarks, when it contained unlabelled landmarks, and when these landmarks were explicitly pointed out to them by the experimenters and verbally labelled. In the absence of landmarks, six and eight year olds performed poorly, whereas most ten year olds were able to complete the task to criterion (two consecutive, error-free journeys through the environment). It was concluded that the navigational difficulties the younger children encountered were likely due to their inability to implement a directional strategy (e.g. encoding the route in terms of a sequence of left and right turns). In the presence of landmarks, more than 90% of children were able to complete the task to criterion, but six year olds required more trials to reach it. The authors concluded that children may not have been attending to landmarks during the learning phase, and discounted instead the possibility of working memory constraints impacting performance due to the ability of those same children to encode all six turns in the route when landmarks were explicitly cued.

Globally, the literature on the acquisition of landmark knowledge and on the development of landmark salience in children does not appear to have reached a consensus about the developmental milestones underlying these abilities. It appears that the ability to recognise objects as landmarks and the ability to use those objects to aid wayfinding are generally dissociated, and that the latter depends on a number of factors. These include the type and size of the environment, the way it is experienced, the perceptual and cognitive salience of landmarks, and the availability of cues highlighting those landmarks.

5.2. Reference Frames during Development

Another crucial aspect of the development of visuospatial abilities is the acquisition of reference frame processing. Being able to navigate an environment and build lasting memories of it requires the ability to encode not only object identities, but also object locations. These can only be encoded and recalled with respect to other entities in the environment, whether the navigator's own location (i.e. within an egocentric reference frame) or with respect to other landmarks (i.e. within an allocentric reference frame). As described in Section 1.2., considerable research has explored the ways in which adult navigators are able to construct allocentric representations upon exploring novel environments, in contrast with older models (e.g. Piaget & Inhelder, 1967; Siegel & White, 1975) that saw early knowledge of an environment as being limited to egocentric self-object relations. However, one of the goals of Experiments 6-8 was to characterise the ability to construct an allocentric representation of a route experienced egocentrically in a developmental sample. More specifically, Experiments 6-8 were an attempt to determine at what age flexible reference frame use (e.g. reference frame transformation) becomes a stable navigational strategy, and whether children at different developmental stages would favour certain reference frames over others.

The developmental literature already contains evidence of parallel egocentric and allocentric reference frame use early in spatial cognitive development. Nardini, Burgess, Breckenridge and Atkinson (2006) tested children between 3 and 6 years of age on an array rotation paradigm that included a hidden object retrieval task. Participants were shown the array and the target object, and observed the latter being hidden under one of a set of cups. Children were then either walked from the initial to the second viewing point, or half way between them and back. At the same time, the array was either rotated or left in its initial position relative to the room. This resulted in the hiding place being consistent at test either with the participant's body or with the room's distal landmarks.

Contrary to a prediction of an advantage for body consistency on the basis of a sequential model of spatial ability development (which would predict young children to initially favour purely egocentric spatial reasoning), the youngest children showed a remarkable advantage in conditions that maintained viewing consistency with a room-based reference frame. Such a pattern of results could only be explained by the use of spatial representations that accounted either for environmental features or for spatial updating of body-referenced representations via self-motion. The emergence of such allocentric components to spatial reasoning at such an early age starkly contrasted with classic developmental studies (e.g. Piaget & Inhelder, 1967), that placed the critical stage of spatial reference frame transformations well into the school years.

These findings were confirmed by Moraleda, Broglio, Rodríguez and Gómez (2013), who tested children between the ages of 6 and 10 on their ability to locate a goal location inside a radial arm maze within a table-top model of a room containing peripheral cues. Additionally, during training trials a guidance cue marked the target arm of the maze. After reaching criterion, participants were tested in a variety of conditions (removal of the guidance cue, dissociation between guidance cue and target arm, removal of peripheral cues, etc.). Additionally, in one experiment consistency between model-, room- and participant-centred reference frames was maintained, whereas in a second experiment the model was rotated to render these reference frames inconsistent. In agreement with Nardini et al (2006), both 6 and 10 year olds favoured a room-centred reference frame over model- or body-centred ones in the absence of a direct guidance cue. In the second experiment, when the model was rotated and in the absence of a guidance cue, performance by the younger children dropped significantly. Their difficulty was likely caused by an inability to suppress competing reference frames and a tendency to adopt their preferred room-based reference frame.

Ruggiero, D'Errico and Iachini (2015) recently investigated the lifetime development of egocentric and allocentric reference frame use in a large sample of 283 individuals between the ages of 6 and 89 years. Participants were presented with three-object arrays common geometrical shapes of varying perceptual features. The arrays were placed on panels located either in peripersonal space (between 20 and 30 cm from the edge of the desk) or in extrapersonal space (between 50 and 100 cm from the edge of the desk). During each trial, participants were allowed to study the array for 20 seconds. The objects were subsequently removed, and participants were asked to verbally produce eight judgements per object triad: 2 egocentric ("Which object was closest/farthest to/from you?"), 2 allocentric ("Which object was closest/farthest to/from *allocentric target*?"), 2 concerning the size of the objects ("Which object was the tallest/lowest?"), and 2 concerning the colour of the objects ("Which object was the darkest/clearest?"). A handheld stopwatch was used to measure response time (the time between the end of the question and the beginning of the participant's answer). An analysis of response time revealed that the youngest age group (6-7 years old) were equally fast in performing egocentric and allocentric judgements in peripersonal space, but slower in producing allocentric judgements in extrapersonal space. Egocentric and allocentric accuracy also did not differ significantly for the youngest children, although they were generally less accurate in their egocentric judgements than all other age groups (except for the oldest adults). Allocentric accuracy did not significantly differ between age groups.

Summing up, the research presented here has produced evidence for remarkable developmental changes early on in life. In particular, the transition between pre-school and school years appears to be crucial in the maturation of children's spatial abilities. During these years, the use of allocentric reference frames appears to be variable, with children able to use them correctly when these are explicitly prompted or forced, or otherwise depending upon task demands.

In the experiments that follow, I attempted to establish a developmental trajectory in large samples of children from pre-school years to early adolescence in order to study the processes underlying the transformation of spatial reference frames between encoding and test. Concurrently, I also attempted to study possible interactions between this ability and children's determinations of landmark salience, in the hope of complementing the data collected in the adult samples in Experiments 1-5. However, before presenting the experiments themselves, I will provide a brief overview of the analysis of developmental trajectories as used in this body of work.

5.3. Developmental Trajectories

The studies reviewed in Sections 5.1 and 5.2 are obviously only part of a vast literature that has characterised various aspects of the acquisition and development of spatial abilities. Traditionally, many such studies have compared performance in spatial tasks between categorical age groups (e.g. Eight year olds vs Twelve year olds; Heth et al., 1997). Similarly, researchers interested in constructing a behavioural and cognitive profile for intellectual disabilities have traditionally adopted individual or group matching methods, whereby each individual in the disorder group is matched to an individual in a healthy control group by chronological age (CA) or mental age (MA), or the whole groups are matched by mean CA or MA. More recently, however, growing attention has been paid to the changing nature of cognitive and behavioural phenotypes both in disorders and in typical development (e.g. Thomas & Karmiloff-Smith, 2002).

This has led to the development of the developmental trajectory approach. This approach has its roots in the broader current of growth curve modelling methods, a set of statistical analytical approaches broadly concerned with estimating between-subject differences in within-subject change (often referred to as growth curve or latent trajectory; Curran, Obeidat & Losardo, 2010). As such, the developmental trajectory approach has two main goals: to generate a function (e.g. a linear regression function) linking age with performance in an experimental task, and to compare it to that of another task (e.g. in a within-subject design) or to that of a second experimental group (e.g. in a between-subject design). This may be achieved cross-sectionally (collecting data at a single point in time but from a sample of individuals of different ages and abilities), longitudinally (collecting data from an age-matched sample at multiple points over a certain time period), or combining both methods (Thomas, Annaz, Ansari, Scerif, Jarrold & Karmiloff-Smith, 2009). As a result, developmental trajectories have seen wide application at the intersection between developmental psychology and other areas of cognitive science, e.g. in studies attempting to establish different trajectories of juvenile delinquency outcomes in children with different behavioural profiles (Broidy et al., 2003; Nagin & Tremblay, 1999).

However, the experiments presented in this chapter attempted to establish developmental trajectories for two key navigational abilities: salient landmark selection and reference frame translation. As such, they will complement results from recent studies that have also explored the developmental time-course of spatial abilities treating age as a continuous variable. Among them, Buckley, Haselgrove and Smith (2015) have characterised the tendency of children (aged between 5 and 11) and adults to prefer proximal or distal landmarks in order to reorient in a virtual maze and navigate to a hidden goals. In their study (Experiments

2 and 3), participants learnt the location of a hidden goal in a virtual environment presenting both internal and external cues. During test trials, the hidden goal was removed, and internal and external cues were put into conflict by rotating the configuration of the latter. While adults spent, on average, more time exploring the quadrant of the environment where the hidden goal should have been located with respect to the distal landmarks, the developmental trajectory in their sample showed a different pattern. Namely, an initial tendency (i.e. around 5 years of age) to rely on distal landmarks followed by a progressive reliance (i.e. between the ages of 9 and 11) on more proximal landmarks. This was taken as an indication that the preference for distal landmarks observed in adulthood may be a trait developing later in adolescence, but also that its observation may be influenced by task characteristics (e.g. navigation in virtual vs real environments).

In Experiments 6-8, I also investigated developmental changes in the use of landmarks as a function of their location, but in the context of route navigation through a virtual environment. More specifically, the ability of children to prioritise navigationally-salient (decision-point) landmarks during egocentric navigation was investigated in large samples between the ages of 4 and 12. I tested children's ability to make use of environmental features (i.e. landmarks) to 1) construct mental representations of an environment explored egocentrically and 2) transform said mental representations for the purpose of completing an allocentric test phase. The aim of these experiments was twofold. Firstly, I aimed to determine whether navigationally salient landmarks facilitate the acquisition of route knowledge and its transformation into allocentric representations in children. Secondly, I attempted to trace a developmental trajectory of behavioural performances to ascertain the age or developmental stage during which this facilitation can be observed. The specific paradigms employed in the three experiments and the ensuing predictions will be discussed in turn.

Experiment 6: Target Identification after Reference Frame Translation

5.4. Experiment 6: Introduction

To answer these questions, a paradigm was designed drawing inspiration from the extensive work carried out by Janzen and colleagues (see Section 5.1 above, and Section 1.3 for a more extensive discussion). Whereas their experiments involve the encoding of spatial information during exploration of virtual environments followed by a landmark object recall task, the experiments presented here were more concerned with children's ability to construct allocentric mental representations of the routes explored, and whether the presence or absence of landmarks as navigational aids would impact it.

Accordingly, in Experiment 6, participants were first presented with short films detailing egocentric views of routes through a virtual environment. The videos contained landmark objects either at decision- or at non-decision points. After each clip was presented, participants were shown an allocentric map of the environment without landmarks, and asked to point to the final location visited in the video. The hypothesis under test was that videos displaying landmark objects at decision points might be encoded more easily and into more stable mental representations than videos displaying landmark objects at non-decision points. As such, I predicted that participants might be better able to mentally translate an egocentric representation into the allocentric representation required at test following encoding of videos with decision-point landmarks.

5.5. Methods

5.5.1. Participants

A sample of 228 children (119 males, 109 females) was tested during a community outreach event taking place over six days at the University of Nottingham. Ages ranged between four and 12 years (mean age 7.82 ± .11 years). Children were recruited from a broad range of ethnic and socio-economic backgrounds. During the event, children took part in a number of experiments (described as "games") in exchange for tokens to be spent participating in a variety of recreational activities under the supervision of volunteer and academic staff. Parents could be present during testing if they or the child so wished.

5.5.2. Design and Materials

A 3D environmental model was created using Google Sketchup 8. The environment had a tree-like structure, with an origin point common to all routes and four possible destination points. Each route therefore contained two decision points (Figure 5.1). The destination points (the target locations on the allocentric map) were circular, had a bright red floor, and contained a 3D model of a wooden treasure chest. Two versions of the model were created, one with landmarks at decision points and one with landmarks at non-decision points. Accordingly, a total of eight route recordings were made, four in the decision point (D) and four in the non-decision point (ND) condition. The landmarks selected were early-acquisition words (as detailed in Kuperman, Stadthagen-Gonzalez and Brysbaert, 2012) for which high-quality 3D models were available in the Sketchup repository. The objects chosen were: chair, snowman, and umbrella (see Appendix V).

Videos were created collating together individual egocentric snapshots into clips approximately 40 seconds long, and all efforts were made to maintain a constant walking pace and rotation rate when negotiating turns in the route. All videos had a resolution of 720x540 pixels and were shown full-screen. The map used during the test phase was printed in colour onto a standard A4 sheet of paper, and did not contain any landmarks in either condition. A within-subject design was used, and all children watched all eight videos.

5.5.3. Apparatus

The videos were played on a laptop placed at a comfortable distance from the participant, and children were asked to confirm whether they could easily see the screen. The experimental setting consisted of a temporary cubicle created by using three large pin boards as separating walls, and contained a desk and three chairs (one for the experimenter, one for the participant, and one for the parent if present).

5.5.4. Procedure

Participants were each assigned a randomised sequence of the eight videos. The procedure was explained to children as follows: “Now you’re going to watch a few short films. These will look like you are walking through a maze, and you will see a few objects along the way. At the end of the maze there is a treasure chest. Then I will show you a treasure map, and I will ask you to remember where you found the treasure chest.”

After explaining the procedure to the children and having verified they had understood the task, the first video was played. All efforts were made to ensure the children’s attention was directed at the screen, by pointing out landmarks as they appeared on the screen if necessary, and asking the participants “What is that?” This was only required for a minority of the youngest participants, and in these cases both decision- and non-decision-point landmarks were pointed at to ensure the prompting did not systematically affect performance.

Each video was played only once, except for cases when children appeared visibly distracted and failed to look at the screen during playback. In these cases, a second viewing was allowed, but participants were not given a third viewing. After each successful viewing (or after a second viewing when needed), participants were shown the environmental map and asked to point to the room shown as the final location in the video they had just seen ("Can you find that treasure chest?"). No feedback was given to participants until they confirmed their choice. If a participant appeared to ask for guidance, verbally of otherwise, the experimenter responded simply "I don't know, what do you think?" until the participant confirmed their final choice. Parents were also instructed not to reveal or hint at the correct answer to the children. The response was then recorded and positive feedback provided (regardless of whether the answer was correct or not). A new trial was then started by playing the next video. The whole experimental procedure lasted approximately 15 minutes.

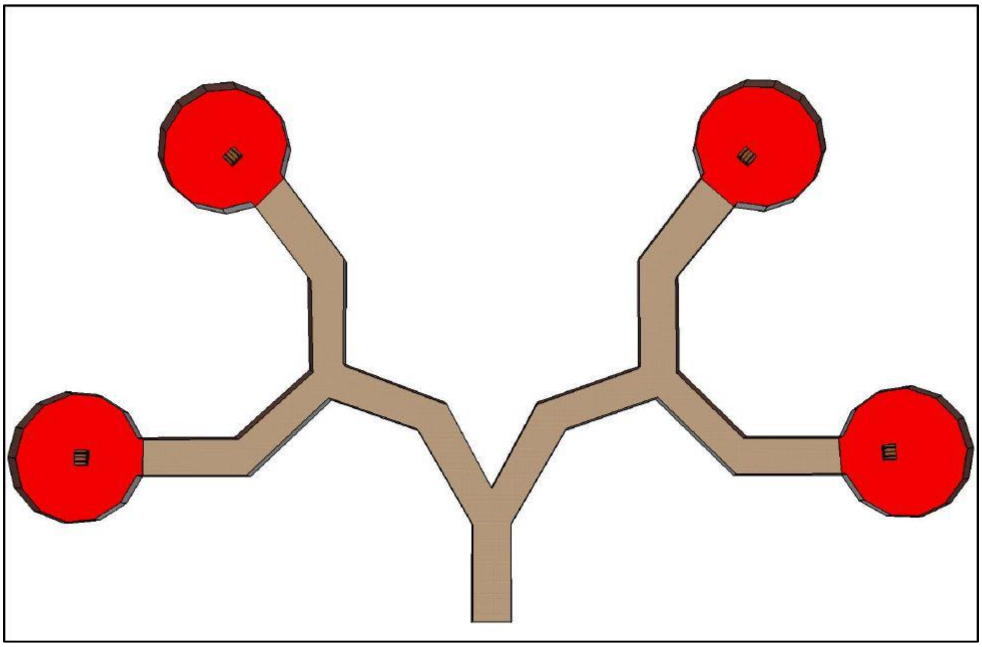


Figure 5.1 – Layout of the virtual environment. The starting location is at the bottom, and the red, circular rooms represent the target locations.

5.5.5. Analysis

In Experiments 6-8, participants' performance was analysed with respect to its tendency to change over development. The methodology used here can be thought of as conceptually similar to standard repeated-measure ANOVA: it compared participants' performance in the two conditions in a within-subject design. However, that performance was not described as a single mean, but as a regression line between performance scores and participants' age. As such, comparisons involved the intercepts and the slopes of the regression lines of the two conditions.

For all three experiments I followed the analysis pipeline presented by Thomas et al (2009). This develops in three stages. First, the developmental trajectories of each condition is characterised individually, by regressing the performance DV against age and computing lower and upper mean confidence intervals. Age is also rescaled, by subtracting the youngest age in the sample from all participants' ages, so that the youngest age tested corresponds to the intercept (i.e. 0) of both regression lines. This does not alter the results, but aids in the interpretation of effects at the earliest age tested

The second step in the analysis involves running a simple repeated-measure analysis of variance to test for the main within-subject effect of condition independent of any effect of or interaction with age. The third and final step involves the inclusion of the rescaled age variable as a covariate in the analysis. A significant interaction between the within-subject effect of condition and the rescaled age covariate in this phase would indicate that performance in the two conditions changes at different rates during development.

The rationale behind running a repeated-measure ANOVA with only condition as a factor, before running an ANCOVA that tests for effects of age, is motivated by the tendency for the latter test to underestimate the main effect of the repeated-measure factor in the presence of a covariate, yielding a more conservative result than a repeated-measure ANOVA run in isolation. For more details, see Delaney and Maxwell (1981) and Thomas et al (2009).

5.6. Results

For each participant I computed an accuracy score as the percentage of correct responses in each condition (D vs ND). Additionally, I rescaled the age of participants so that the intercept of the regression lines corresponded to the youngest age in the sample (Thomas et al., 2009). As a result, while the slope of the regression remained unchanged, the intercept represented the accuracy performance of the youngest participant. Developmental trajectories for performance over time were then constructed in the two conditions by entering Rescaled Age as a predictor in a regression model. Age was found to explain a significant amount of variance in accuracy performance both in the D condition, $R^2 = .19$, $F(1,226) = 53.31$, $p = .001$, and in the ND condition, $R^2 = .30$, $F(1,226) = 99.61$, $p = .001$. See Table 5.1 for the regression coefficients and Figure 5.2 for a plot of the developmental trajectories for both conditions.

I then checked for a main effect of landmark location during encoding, using 2(Condition: D vs ND) as the only factor in a repeated-measures ANOVA. The effect of Landmark presence in the videos on performance was not statistically significant, $F(1,227) = .06$, $p = .801$, $\eta^2_p < .001$. In other words, participants across all ages were globally just as likely to select the correct treasure room at the end of each video, whether it displayed landmark objects at decision points, M = 57.01%, SEM = 2.2%, or at non-decision points, M = 57,56%, SEM = 2.1%. An ANCOVA with rescaled age as a covariate was then run in order to compare the intercepts and slopes of the two developmental trajectories (Annaz et al., 2009; Thomas et al., 2009). This revealed a significant main effect of Age, $F(1,226) = 109.35$, $p = .001$, $\eta^2_p = .326$, indicating an overall performance improvement with age, but also a statistically significant interaction between Condition and Age, $F(1,226) = 5.59$, $p = .019$, $\eta^2_p = .024$. To further explore this interaction, difference scores were computed by subtracting ND percentage scores from D percentage scores. The resulting dependent variable (D_Effect) was intended as a measure of the extent to which a participant benefitted at test from the presence of landmarks at decision points in D videos. This measure was also regressed against Rescaled Age, and the regression coefficients are also included in Table 5.1. These revealed that the overall benefit of landmark presence at decision points tended to decrease with age in this task.

The pattern that emerges from these analyses is one that sees performance in the D condition increase with age at a lower rate compared to the ND condition, B = 7.92% and B = 10.81% respectively. Namely, performance in the D condition increased over the age range tested at 73.26% (i.e. the ratio between the two unstandardised coefficients) the rate of the ND condition (see Thomas et al., 2009). However, performance at the youngest age tested was higher in the D condition, M = 27.30%, SEM = 4.5%, than in the ND condition, M = 15.71%, SEM = 4.5%.

*Table 5.1 - Results of the regression between Rescaled Age and Accuracy. *p < .05, ***p = .001.*

	Decision Point			Non-Decision Point			D Effect		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Age	7.92	1.08	.437***	10.81	1.08	.553***	-2.88	1.21	-.155

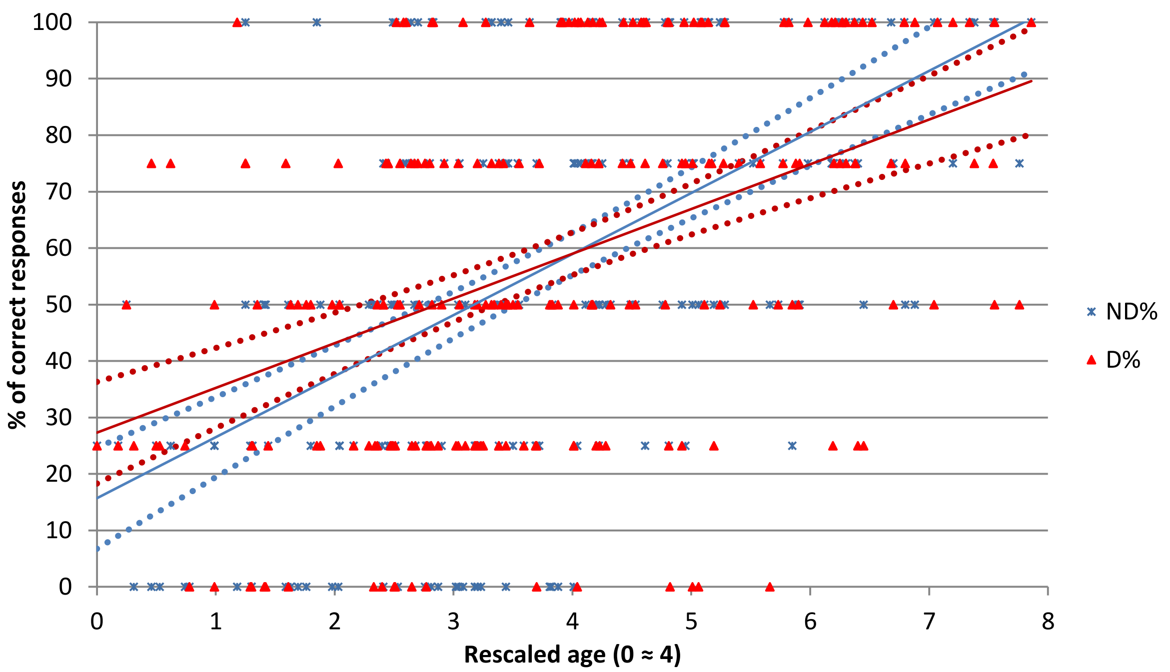


Figure 5.2 – Improvement in participants' accuracy scores as a function of age. Dashed lines represent 95% CIs. Solid lines represent the linear trends.

5.7. Discussion

Beyond a general, predictable improvement in task performance over development, the results yielded by Experiment 6 were quite unexpected. The original prediction was that videos portraying an egocentric route through an environment with landmark objects at decision points would produce more stable representations than videos with landmark objects at non-decision points, and that this would translate into an increased ability to translate that egocentric representation into an allocentric model during the test phase. Instead, an analysis of the developmental trajectories revealed that an initial advantage in the D condition translated into an overall slower rate of improvement than in the ND condition. Mean performance in the ND condition appeared to overtake D performance around 8 years of age in this sample, but mean performance in both conditions only increased past 75% after 9 years of age.

Although this would seem, prima facie, to run contrary to findings in the literature of early allocentric abilities (see Section 5.2), there are aspects that warrant consideration. First of all, it is unclear to what extent the observed interaction is a real effect. The statistical significance, but extremely low effect size raises at least the possibility that it may simply be a statistical artefact as a result of the large sample size. Secondly, the studies by Nardini et al. (2006), Moraleda et al. (2013), and Ruggiero et al. (2015) involved the exploration of real environments or the processing of real object arrays. It is possible that the limited perceptual input, and the absence of proprioceptive and vestibular input, normally associated with physical motion might affect children more than adults in their choice of navigational strategies.

This would be in line with results by Jansen-Osman and Fuchs (2006) concerning wayfinding behaviour and spatial knowledge acquisition in virtual environments. More specifically, children may have formed only weak associations between the landmarks in the videos and their respective spatial locations, whether decision- or non-decision points. It is possible that if the task had required participants to retrace the route to the target room in the same virtual environment, they may have done so far more efficiently. However, since the task required them to carry out a spatial reference frame transformation and to then identify locations on a map that provided only global geometric information of the environment, the younger children may largely have been unable to make use of the landmark-turn association knowledge acquired during route encoding. Conversely, because of the relatively short duration of the routes, it is possible that the older children may have accomplished the task relying simply on their sequential knowledge of turn-direction associations, and used it to trace the corresponding route from origin to destination onto the map template. These issues were better explored in Experiment 7.

In Experiment 7, I attempted to address a few of the questions raised by these results. Using a larger model (with eight target rooms rather than four) and longer routes (with three decision points rather than two) I tried to make the task more challenging, in the hope that it would compel participants to rely more strongly on landmarks as a navigational and spatial knowledge aid. Additionally, all videos contained all landmarks, whereas test maps would contain either decision or non-decision point landmarks. This was an attempt to better extricate patterns of performance during test that may be due to landmark salience.

Experiment 7: Target Identification after Reference Frame Translation in Larger Environments

5.8. Experiment 7: Introduction

In Experiment 7, I attempted to address a few of the questions raised by these results. Using a larger model (with eight target rooms rather than four) and longer routes (with three decision points rather than two) I tried to make the task more challenging, in the hope that it would compel participants to rely more strongly on landmarks as a navigational and spatial knowledge aid. Additionally, all videos contained all landmarks, whereas test maps would contain either decision or non-decision point landmarks. This was an attempt to better extricate patterns of performance during test that may be due to landmark salience.

5.9. Methods

5.9.1. Participants

79 children (53 males, 26 females) took part in the study during a subsequent community outreach event taking place over six days at the University of Nottingham. Ages ranged between four and 12 years (mean age 7.99 ± .20 years). Children were recruited from a broad range of ethnic and socio-economic backgrounds. During the event, children took part in a number of experiments (described as "games") in exchange for tokens to be spent participating in a variety of recreational activities under the supervision of volunteer and academic staff. Parents could be present during testing if they or the child so wished.

5.9.2. Design and Materials

An environment similar to the one used in the previous experiment was created in Google Sketchup 8. The number of destination points was increased from four to eight, and landmarks were present both at decision and non-decision points in the route. Eight route recordings were produced, and the speed of locomotion was increased compared to Experiment 6, so as to keep the length of the videos around 40 seconds and in order to increase task difficulty. All videos were created with a resolution of 720x540 pixels and were displayed full-screen. The maps used during the test phase were printed in colour onto a standard A4 sheet of paper: one only contained the landmarks that had appeared at decision points during the videos (Figure 5.3), and the other only contained the landmarks that had appeared at non-decision points in the videos (Figure 5.4). A within-subject design was used, and all children watched all eight videos twice: for each video they were tested once on the D map and once on the ND map.

5.9.3. Apparatus

The experimental setup was the same used in Experiment 6.

5.9.4. Procedure

Participants were each assigned a randomised sequence of sixteen videos (each of the eight videos was viewed once per condition). The children were told to pay very close attention even when a video felt familiar to them, and they were told that there may be small differences between them (although that was not actually the case). Task instructions were the same as in Experiment 6. However, due to the longer videos, the overall duration of the experimental procedure was longer than in the previous experiment, at approximately 20 minutes.

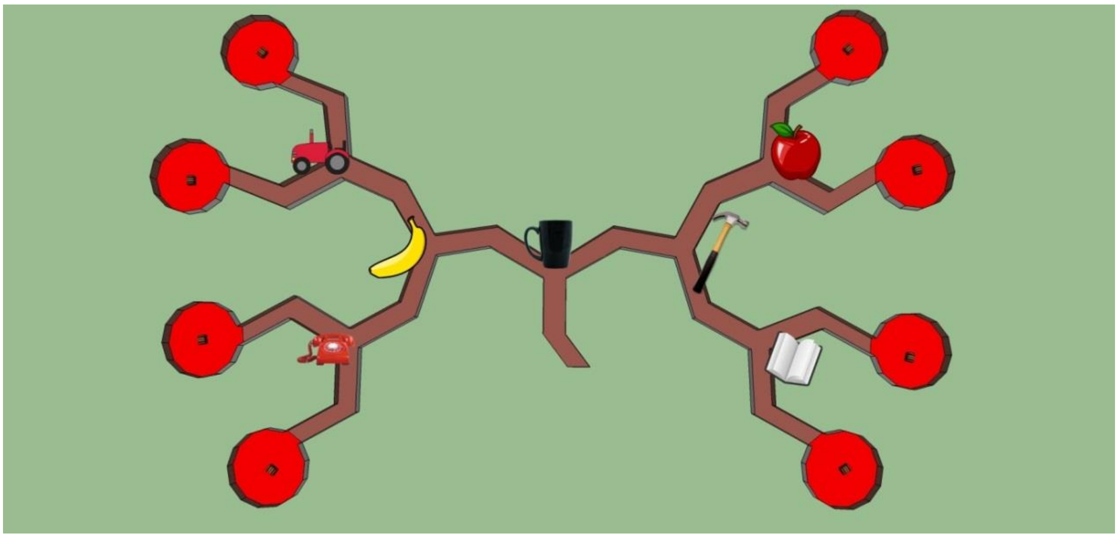


Figure 5.3 – Test map with landmarks at decision points.

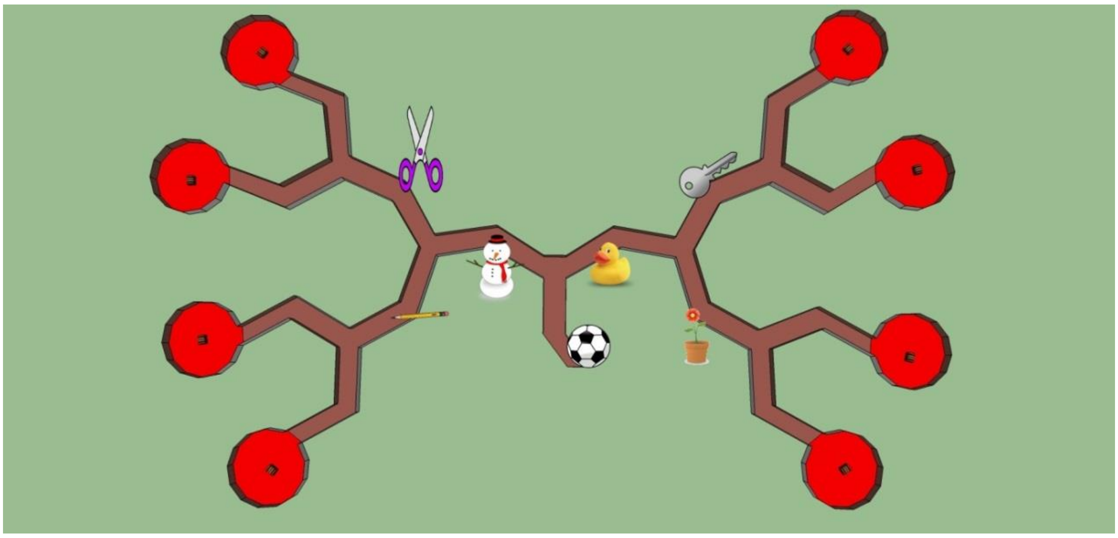


Figure 5.4 – Test map with landmarks at non-decision points.

5.10. Results

Accuracy scores were computed and the age variable rescaled as in Experiment 6. Two distinct developmental trajectories were constructed by entering the rescaled age as a predictor in a regression model. Age was found to explain a significant amount of variance in accuracy performance both in the D condition, R^2 = .21, F(1,77) = 21.12, p = .001, and in the ND condition, R^2 = .38, F(1,77) = 48.61, p = .001. See Table 5.2 for the regression coefficients and Figure 5.5 for a plot of the developmental trajectories for both conditions.

As in Experiment 6 (and following the guidelines of Thomas et al., 2009), the main effect of salience of the landmarks included on the test maps was tested for in isolation (i.e. across all levels of Age), using 2(Condition: D map vs ND map) as the only factor in a repeated-measures ANOVA. The effect was marginally significant, F(1,78) = 3.50, p = .065, η^2_p = .065, indicating slightly higher performance when children were asked to point to treasure chests on the D map, M = 63.76%, SEM = 2.9%, compared to when they had to locate them on the ND map, M = 58.54%, SEM = 2.8%.

An ANCOVA with rescaled age as a covariate was then run in order to compare the two developmental trajectories. This revealed a significant main effect of the Age covariate, F(1,77) = 46.96, p = .001, η^2_p = .379, but no statistically significant interaction between Condition and Age, F(1,77) = 1.76, p = .188, η^2_p = .022. In other words, although age produced an overall performance improvement, this was comparable in both conditions.

*Table 5.2 - Results of the regression between Rescaled Age and Accuracy. ***p =.001.*

	Decision Point			Non-Decision Point		
	B	*SE B*	β	*B*	*SE B*	β
Age	6.61	1.43	.464***	8.61	1.23	.622***

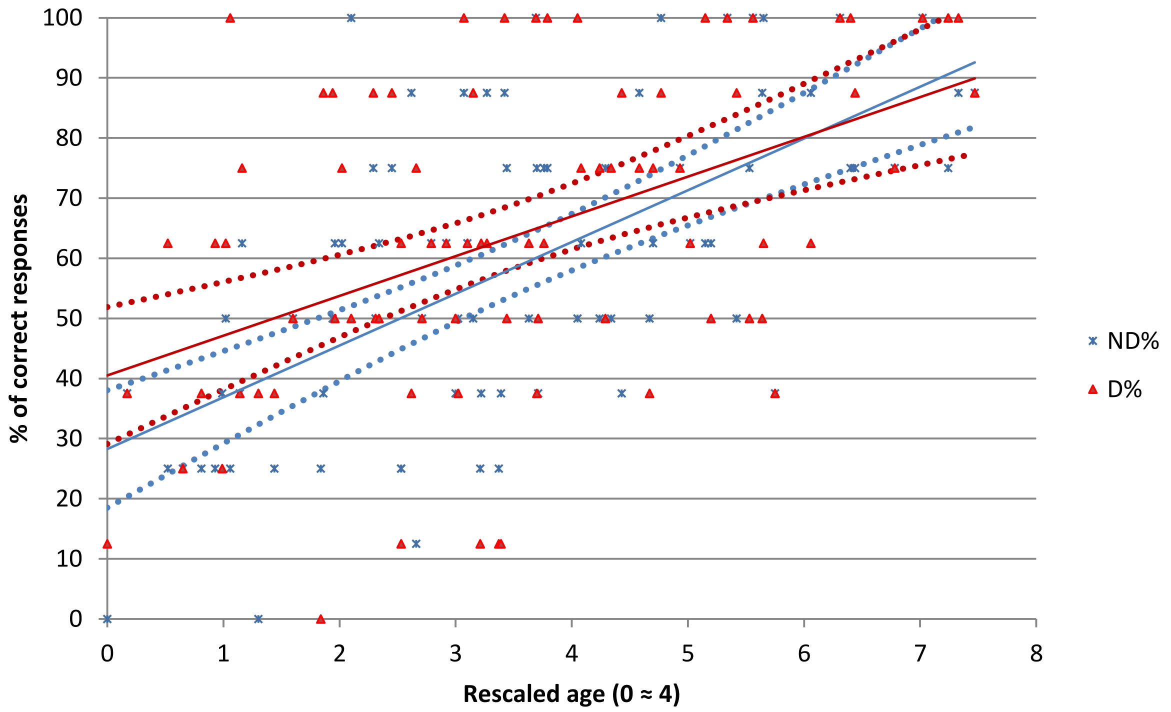


Figure 5.5 – Improvement in participants' accuracy scores as a function of age. Dashed lines represent 95% CIs. Solid lines represent the linear trends.

5.11. Discussion

Comparing performance in the two conditions in this experiment did not reveal significant differences between them as a function of age. An overall task improvement as children got older was observed, and performance was globally higher in the D condition when controlling for age (63.76%) compared to the ND condition (58.54%). Age appeared to be a more significant predictor of ND performance compared to D performance, and again what seems to be a slower development of the ability to use decision-point landmarks effectively was observed, but the developmental trajectories for the two conditions did not differ significantly. Interestingly, the intercepts of the two regression lines (i.e. performance scores in the youngest children) were higher, and led to generally higher performance across development, in this experiment compared to Experiment 6. However, it is not clear whether this was due to the presence of both types of landmarks in the videos, or to the presence of landmarks on the two test maps.

On the one hand, this is a possible indication that even the youngest children may have generally benefitted from the presence of landmarks on the test maps, but decision and non-decision point landmarks improved performance equally. Alternatively, it is possible that displaying more landmarks during the videos created a more challenging task, leading participants to employ more efficient navigational strategies and thereby paradoxically improving performance. Conversely, it is also possible that the environment may still not be large enough, and that the ND landmarks may be sufficiently close to junctions in the path to also act as association cues for changes in direction.

Although increasing the length of the journey and the overall size of the path legs might not be feasible due to the time constraints set on the experimental session duration, a different approach might be more practical. It is possible that simple response accuracy might not be sufficiently sensitive as a measure to detect differences in developmental changes in this task and with these environments. Experiment 8 employed a landmark recognition task allowing for the acquisition of reaction time measures, in the hope that these would be sufficiently sensitive to modulations of landmark salience. By introducing a landmark object recognition task, Experiment 8 was a more faithful replication of the classical paradigm used by Janzen and colleagues (see Section 5.11 below), but it also still included a reference frame transformation (partly in line with Experiments 1-3 in Chapter 3) by asking participants to trace on a map-like representations the paths taken in the videos.

Experiment 8: Landmark Recall and Path Verification

5.12. Experiment 8: Introduction

Experiments 6 and 7 have shown a trend indicating a slight facilitation effect of decision-point landmarks. When the youngest children in the tested samples attempted to encode an egocentric route from video, and transformed that representation to identify its final destination on a map, the presence of decision-point landmarks, either in the learning phase (on the route videos) or during the test phase (on the maps) seemed to slightly improve performance. However, this advantage manifested as a statistically significant interaction between the developmental trajectories of the two conditions only in Experiment 6. However, this had a small effect size and thus accounted for very little variance, despite a large sample size. In Experiment 7, the main effect of salient landmark presence was only marginally significant, failing to address the question of whether children were actually attending to the landmarks during encoding, prioritising the navigationally salient ones, and making use of them during testing.

Accordingly, Experiment 8 was an attempt to dissociate the encoding and recall processes of landmarks and routes. The former have recently been explored within a developmental context by van Ekert, Wegman and Janzen (2015), whose earlier work in adults (see Section 1.3) provided much of the inspiration and theoretical foundation for the focus on landmark salience in Experiments 1-5 in this thesis. In their recent study, van Ekert et al. (2015) presented volunteers between the ages of 8 and 18 with videos of an egocentric tour through a virtual environment, under instructions to assume that they would be asked to guide fellow students through the same environment later that day. The route videos were approximately 12 minutes long, and 120 3D models of objects were chosen as landmarks and distributed across the environment so that 30 occurred only once at decision points (1DP), 30 once at non-decision points (1NDP), 30 twice at decision points (2DP), and 30 twice at non-decision points (2NDP). The encoding phase was followed by a recognition memory task performed inside an fMRI scanner, in which each trial consisted of the brief presentation (500 ms) of a landmark object (and foils) removed from the spatial context in which they may have appeared during encoding. Participants were tasked with indicating whether each object had occurred in the film sequence they had observed.

Across all ages tested, van Ekert and colleagues observed no effect of spatial context (whether the landmark object had appeared at decision or non-decision points, and how many times) on either recognition accuracy or response times. Additionally, they found no evidence of an increase in object recognition performance as a function of age, indicating that simple object memory may be already mature by the age of eight (the youngest age in their sample). However, they did observe a linear decrease in response times as a function of age. Furthermore, analysis of BOLD responses showed an age-related increase in parahippocampal (PHG – previously implicated in landmark processing. See Section 1.3.) and anterior cingulate (ACC – an area involved in cognitive control. Luna, Padmanabhan & O'Hearn, 2010.) activation in response to decision-point landmarks (contrasting responses to 1DP and 1NDP landmark objects), but not for objects associated with ambiguous spatial contexts (i.e. 2DP vs 2NDP). Thus, their study established an age-related trajectory of changes in medial-temporal function in response to navigationally relevant object throughout adolescence.

In Experiment 8, I used a modification of this paradigm to try and push back that developmental trajectory to early childhood, and to attempt to tease out the starting point of the ability to differentiate navigationally salient features from less salient ones. Additionally, Experiment 8 used two distinct tasks to isolate the processes involved in the recall of landmarks and of routes. Participants watched videos of routes through virtual environments containing four target rooms (and two decision points per route), as well as both D and ND landmarks. Each video presentation was followed by a landmark recognition task, and then by a route recall task. In the former, children were presented with a series of images representing both the landmarks found on the route (both D and ND), as well as control landmarks that were not present in the environment. Their goal was to judge whether they remembered seeing each landmark in the preceding video or not. Children's response accuracy and response times were recorded.

In formulating predictions for this experiment, I attempted to address questions raised both by van Ekert et al (2015) and by the results obtained in Experiments 6 and 7. In particular, while van Ekert and colleagues observed no effect of landmark salience on behavioural measures, I attempted to determine whether that would be the case also with children younger than eight years of age and with the types of stimuli used in these experiments. Experiments 6 and 7 have so far shown hints of an effect of spatial context on behavioural measures, albeit a small and inconsistent one. This has raised the possibility that, due to task characteristics, children may not be approaching the objects encountered as navigationally relevant. To better explore this possibility, in Experiment 8, percentage recall measures were complemented by response time measures obtained via an object recognition task akin to that used by van Ekert and colleagues, to determine whether RTs might be more sensitive to this manipulation than response accuracy. The hypothesis being tested with this task was that, if D landmarks were recognised as salient, they should then have been better represented in memory following encoding, and their verification should be faster and more accurate. If, on the other hand, children were not relying on landmarks to successfully encode the route, then landmark recall accuracy and RTs should not differ between decision- and non-decision-point landmarks. In that case, we should have observed a main effect of age in decreasing response times. In the latter task, participants were shown a map layout of the virtual environment that did not contain any of the landmarks, and were asked to trace on it the route they saw in the preceding video. The percentage of routes correctly drawn was recorded for each participant, and decomposed into the percentage of first and second turns correctly recalled. In light of Experiments 6 and 7, I predicted a significant overall improvement in performance as a function of age.

5.13. Methods

5.13.1. Participants

111 children (60 males, 51 females) took part in the study during a community outreach event taking place over six days at the University of Nottingham. Ages ranged between four and 12 years (mean age 8.00 ± .18 years). Children were recruited from a broad range of ethnic and socio-economic backgrounds. During the event, children took part in a number of experiments (described as "games") in exchange for tokens to be spent participating in a variety of recreational activities under the supervision of volunteer and academic staff. Parents could be present during testing if they or the child so wished.

5.13.2. Design and Materials

The environment used was a four-chambered version of the one used in Experiment 7, with landmarks present both at decision and non-decision points in the route. Four route recordings were produced, with a resolution of 720x400 pixels and a frame rate of 30 fps. All videos lasted approximately 36 seconds. The 3D models used as landmarks in the environment and imported from the Sketchup library were exported as 2D images, and used as stimuli during the landmark verification task (see Figure 5.6 for examples and Appendix VI for all the images.). The images had an initial resolution of 1916x969 pixels, but were downsized during presentation in the landmark task to better match their apparent sizes during the route videos.

A PsychoPy (Peirce, 2007; 2008) script was built to automate video playback and to randomise the presentation of landmarks to be recognised. A keyboard mask was used to conceal all keys not used during the landmark recognition task. The keys pressed by participants to confirm whether they recalled the landmarks or not were the "m" and "z" keys respectively, so as to be distant from each other. However, children did not see the letters, which were covered using adhesive stickers of different colours, green for "yes" (on the "m" key) and red for "no" (on the "z" key). These corresponded to a green tick icon presented on the right-hand side of the screen and a red cross icon presented on the left-hand side of the screen, both below the picture of the particular landmark object being probed (Figure 5.7). Landmark objects were presented on a grey background, and remained on screen until a response was given. The maps used during the route test phase were printed in colour onto standard, white A4 sheets of paper (Figure 5.8). Participants used felt-tip pens to trace the routes on the maps, and a separate map template was used for each trial.

5.13.3. Apparatus

The experimental setup was the same used in Experiments 6 and 7.

5.13.4. Procedure

The procedure for participant setup and route presentation was similar to that used in Experiments 6 and 7. After each video playback, a pause was included in the script to ensure participants were focused and ready for the landmark recognition test, and the script did not progress until the experimenter was satisfied that was the case and pressed a button. The message “Do you remember these?” appeared on screen, and participants were reminded about the goal of the task. They were instructed to keep their hands on the keyboard and their index fingers on the two response keys. There was no time limit for the presentation of the landmark images, but participants were instructed to answer as quickly and as accurately as they could. After the last trial of the landmark recognition phase, the message “Can you find the way?” appeared on screen, and prompted the presentation of the map template. Participants were then given a pen, and asked to trace the path they had taken in the video. No time limit was imposed for this task. Once the route had been drawn, participants were encouraged to focus on the screen once more and asked “Are you ready to watch the next film?”. A new route video was then presented.

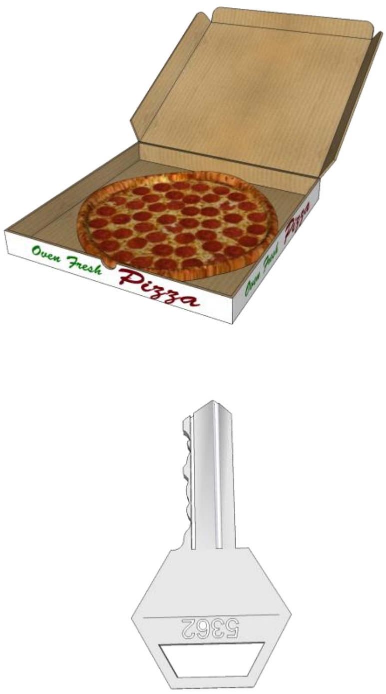


Figure 5.6 - Examples of the object models used as landmarks.

Figure 5.7 – The appearance of the screen during the presentation of an object probe.

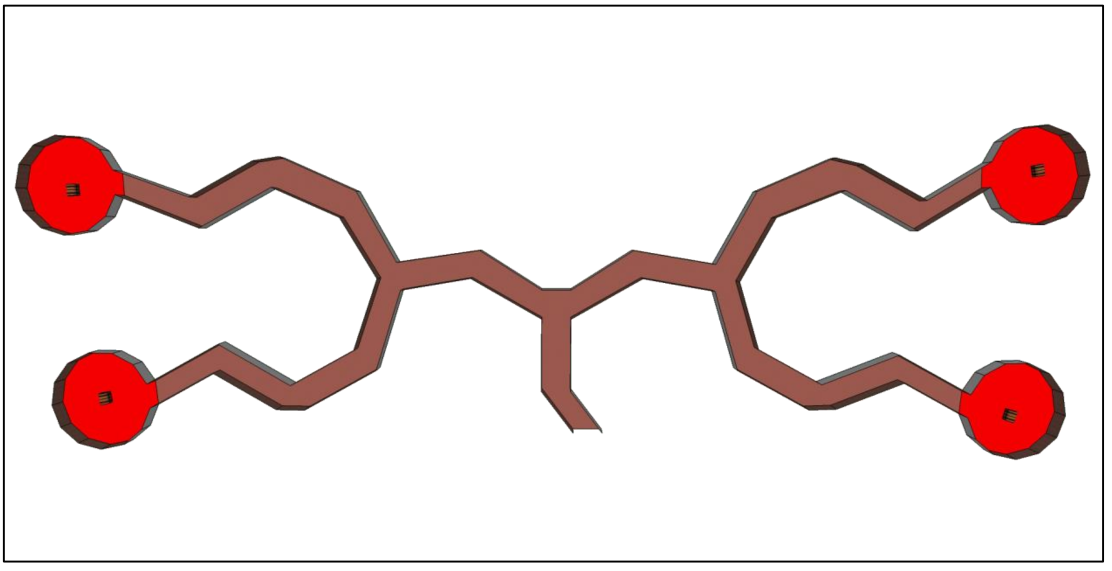


Figure 5.8 – Map of the virtual environment used.

5.14. Results

Landmark Recall Accuracy

A mean landmark recall accuracy score was computed for each participant (i.e. percentage of landmarks correctly identified as present or absent in the videos). These mean scores were then entered in a within-subject repeated-measure ANOVA with 2(Landmark type: D vs ND) as the only factor, in order to determine whether a significant main effect was present. However, accuracy in this task was generally very high (Table 5.3), and did not differ significantly as a function of landmark type, $F(1,110) = .030$, $p = .863$, $\eta^2_p < .001$. Namely, children were just as likely to correctly remember decision-point landmarks as they were non-decision-point landmarks (or to correctly recognise foils as novel objects).

Recall accuracy for D and ND landmarks was then regressed against rescaled age to establish the two developmental trajectories. However, age was not found to be a significant predictor of landmark recognition accuracy for either D or ND landmarks, R^2 = .002, F(1,110) = .229, p = .633, and, R^2 = .001, F(1,110) = .023, p = .879 respectively.

Table 5.3 – Mean percentage of landmark objects correctly recognised as present or absent in the videos.

	Decision-Point	Non-Decision-Point	Foils
% correctly identified	92.68	92.45	95.89
SEM	.89	1.06	1.44

Landmark Recall RT

Response times (correct responses) for D, ND, and Foil objects were tested using the Kolmogorov-Smirnov test of normality, and found to significantly deviate from it, D(111) = .145, D(111) = .143, and D(111) = .159, all ps < .001. Visual inspection of the histograms for the three dependent variables confirmed a significant skew in the distribution of results. Accordingly, response times were log transformed for the subsequent analyses.

By regressing the Log-transformed RTs against Rescaled Age, the latter was found to be a significant predictor of the former during landmark recognition both for D and ND landmarks, R^2 = .409, F(1,110) = 75.54, p < .001, and, R^2 = .324, F(1,110) = 52.25, p < .001 respectively. See Figure 5.9 for the regression plot and Table 5.4 for the regression coefficients. The Log-transformed RTs were then entered into an ANOVA with 2(Landmark type: D vs ND) as the only within-subject factor, but this was found to have no statistically significant effect, F(1,110) = .66, p = .418, η^2_p = .006. In other words, participants (across all ages tested) were just as quick to verify the landmark objects that had appeared at decision points in the videos as they were to verify the landmark objects that had appeared at non-decision points.

Log-transformed RTs were then entered into an ANCOVA to test for a main effect of the rescaled age covariate (i.e. an overall performance improvement over development), as well as for an interaction between landmark type and rescaled age (i.e. a difference between the developmental trajectories for the two landmark types). Rescaled age was found to have a highly significant effect on landmark verification times, F(1,109) = 69.24, p < .001, η^2_p = .388, with older children responding more quickly on correct trials than younger children, in line with the regression analysis (see coefficients in Table 5.4).

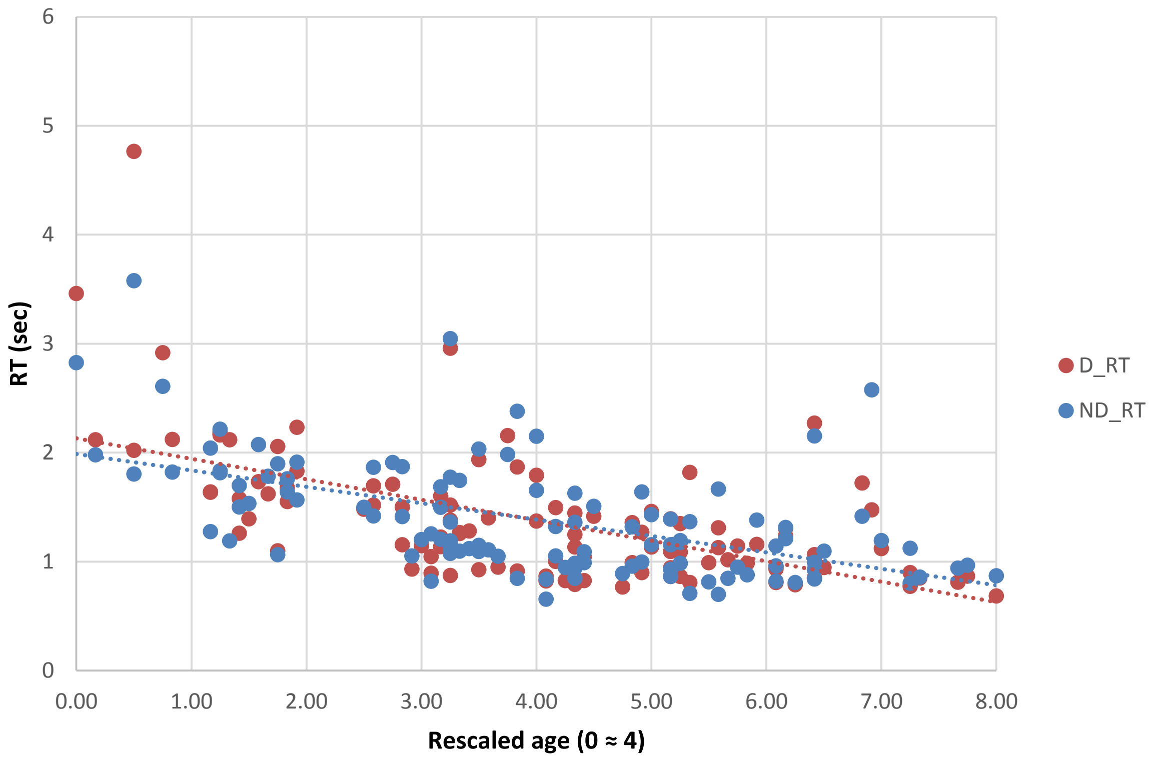


Figure 5.9 – Decrease in mean landmark recognition RT (not transformed) as a function of age. Dotted lines represent the linear trends in RT for their respective landmark types.

A statistically significant interaction was also found between rescaled age and landmark type, $F(1,109) = 4.10$, $p = .045$, $\eta^2_p = .036$, indicating a difference between the two developmental trajectories. Globally, the pattern emerging from this analysis stands in contrast with what was found for performance measures in Experiments 6 and 7. Whereas those experiments revealed slightly improved performance (i.e. better treasure location recall) in the D condition at intercept (i.e. earliest age tested), the RT trends in the landmark recognition task used here revealed slightly slower responses at intercept to landmark objects that had appeared at decision points during encoding. RTs to decision-point landmarks, however, appeared to decrease at a greater rate over the developmental ages tested compared to RTs to non-decision-point landmarks.

*Table 5.4 – Coefficients for the regression between Rescaled Age and (log) RT. ***p < .001*

	Decision Point			Non-Decision Point		
	B	*SE B*	β	*B*	*SE B*	β
Age	-.053	.006	-.640***	-.045	.006	-.569***

Table 5.5 – Mean recognition RT (not transformed) for the different landmark types.

	Decision-Point	Non-Decision-Point	Control
RT (sec)	1.37	1.38	1.50
SEM	.05	.05	.05

Route Recall Accuracy

I subsequently recorded participants' ability to correctly recall the direction of turns when they were asked to retrace the routes from starting point to target on the map templates. On average, participants recalled M = 60.13% of routes correctly, SEM = 2.68%, and this ability was found to be significantly predicted by rescaled age, R^2 = .210, F(1,110) = 28.89, p < .001. I then proceeded to compute distinct turn recall measures for the two turns contained in each route, sequentially ordered from the origin point (i.e. Junction 1 vs Junction 2). Participants were found to recall the direction of the change in heading at the first junction in the route more accurately than at the second junction, M = 72.74%, SEM = 2.88% and M = 47.52%, SEM = 3.06% respectively.

To quantify the main effect of turn position on participants' turn recall abilities, I entered their turn recall scores in a repeated-measure ANOVA with 2(Turn position: J1 vs J2) as the only factor. The main effect of turn position was significant, F(1,110) = 94.90, p = < .001, η^2_p = .463. A subsequent ANCOVA including rescaled age also revealed an interaction between turn position and age affecting children's directional turn recall, F(1,110) = 4.05, p = .046, η^2_p = .036, indicating different parameters for the two developmental trajectories (Figure 5.10). The most significant difference was between the intercepts of the two regression lines at the youngest age tested, 51.01%, SEM = 6.3%, and 14.95%, SEM = 6.3%, for first-junction and second-junction turns respectively. However, this difference may have been partly due to a number of 100% recall scores for J1 turns for some of the youngest participants.

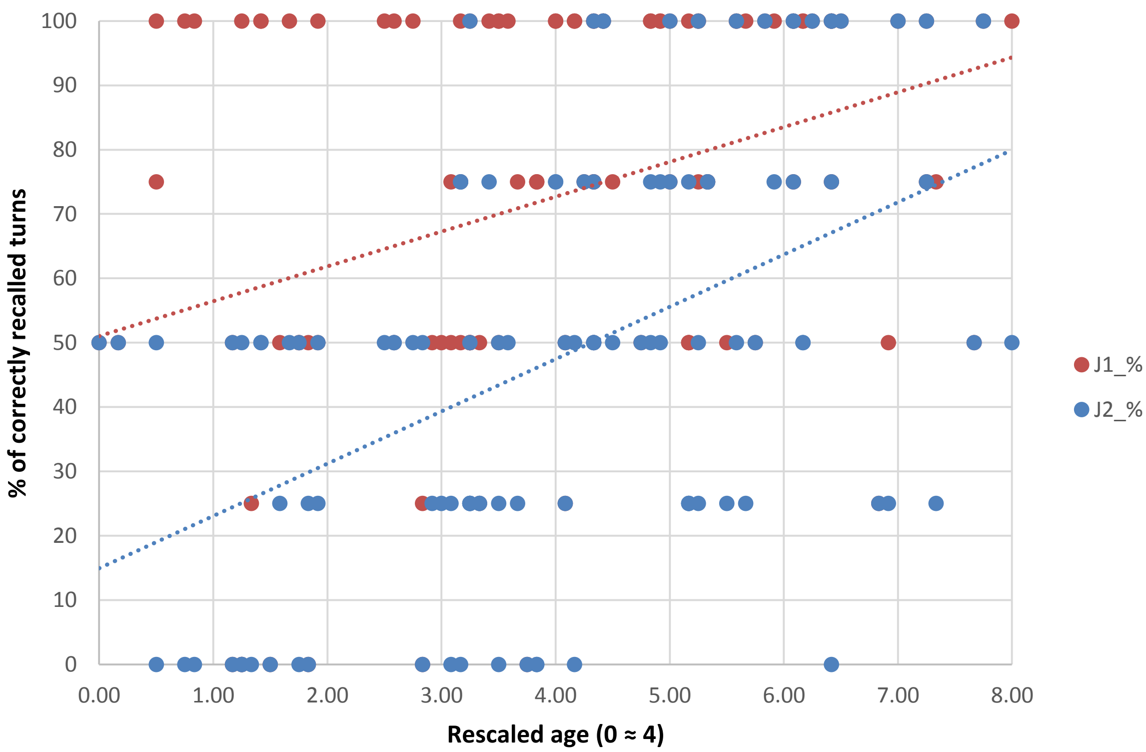


Figure 5.10 - Increase in correct recall of direction changes after Junction 1 and Junction 2 in the route.

5.15. Discussion

These results seem to be in accordance with those from Experiments 6 and 7. The small and inconsistent facilitation effect of decision-point landmarks in the previous two experiments is confirmed here by a very small difference between the regression lines in the landmark recognition task. Although children were attending to the objects and were able to correctly determine, across all ages, whether they had seen them in the video or not, they did not seem to think of them as landmarks or to rely on them as navigational aids. This is evident from the analysis of both response times in the landmark recognition task, and of route recall accuracy. The former revealed that decision-point landmarks were not responded to significantly faster than non-decision-point ones, indicating that they were not more strongly represented in memory. Additionally, response times were found to decrease with age, and at broadly comparable rates for D and ND landmark objects. Although a significant interaction was found between landmark type and age in RT measures, this was of low statistical significance and associated with a small effect size (in line with Experiments 6 and 7). Globally, these results are in accord with those yielded by behavioural measures in van Ekert and colleagues' object recognition task.

On the other hand, analysis of route memory (i.e. the route drawing task) revealed a significant modulation of turn location. Children struggled to recall the correct direction to turn at the second junction significantly more than at the first, effectively indicating that task complexity grew as a function of distance travelled. It was only around the age of nine that children began to more reliably recall the direction to turn at the second junction. Since both junctions had landmarks located at them, one might expect their recall to be on par. It is possible that children may have been employing a sequential encoding of directional turns independent of landmarks, thereby taxing working memory resources, and that the first turn of a route may have been better represented in memory due to a primacy effect. However, given that the routes used in this study contained only two turns, and in light of studies showing that children as young as six can correctly navigate virtual environments with six turns (Farran, Courbois, van Herwegen, Cruickshank & Blades, 2012), this appears to be an unlikely explanation.

What seems more likely is that, after having encoded the route egocentrically, errors may have occurred during either the retrieval of children's representations or their transformation into allocentric ones. Considering children were aware from the onset that the task involved retracing the route they had watched onto a map-like representation of the environment, this finding seems at odds with reports of allocentric abilities in young children (Moraleda et al., 2013; Nardini et al., 2006; Ruggiero et al, 2013), and of parallel computation of multiple reference frames. However, as discussed in Section 5.7, one crucial aspect that distinguishes the experiments presented here from other studies of spatial abilities during navigation of both real and virtual environments is the lack of interactivity. Participants here merely watched video playbacks of routes through a virtual environment from an egocentric perspective. This could feasibly have hampered their ability to solidly bind landmark objects to their locations, thus preventing landmark-turn associations from becoming more strongly represented than landmark-path associations.

Future studies should more carefully characterise the differences between passive and active navigation of virtual environments with respect to the development of landmark salience perception. This would be particularly interesting when exploring the developmental timecourse of spatial language processing. With an eye to Experiments 1-3 presented in Chapter 3, it would be interesting to explore a number of issues from a developmental perspective. For example, the extent to which children might be more susceptible to the absence of perceptual, proprioceptive, and vestibular input during the processing of route directions. Additionally, future experiments should explore the developmental trajectory of the top-down modulation of landmark (word and region) salience by mental imagery observed in Experiments 2 and 3. In that same context, studying any possible interactions between presentation modality and the complex relationship between spatial imagery and the vestibular system (see Section 1.10) may prove both interesting and informative. For example, a mental representation following active navigation in a virtual environment may be found to be more resistant to disruption caused by caloric vestibular stimulation compared to a mental representation built on the basis of a passive exposure (e.g. a video, or a guided tour) to the same environment. Lastly, although time constraints prevented this under the experimental settings described here, future experiments should be run with longer routes and a greater number of landmarks, bringing them in line with other studies in the literature.

Ultimately, the pattern observed in recall accuracy measures in Experiments 6 and 7 remains unexplained. While the possibility remains that the hint of decision-point facilitation observed for the youngest children and the ensuing lower slope over development compared to the ND trend may be a statistical artefact of the task used in these experiments, it warrants further investigation. With respect to children's ability to identify a spatial location experienced egocentrically on an allocentric representation of the same environment, reliably above-chance accuracy (i.e. >75%) appears to emerge, purely on the basis of the data provided by Experiments 6 and 7, between the ages of 8 and 10. However, considerable between-subject variance can clearly be observed, reliable predictors of which must still be identified. This, too, should be one of the areas of focus for future research.

CHAPTER 6
General Discussion

6.1. Key Issues: Spatial Language and Mental Imagery

Humans are linguistic beings with a penchant for action – or inherently spatial and interactive beings, whose linguistic abilities are a specific case of the myriad complex movements we have evolved to produce, and whose brains are probabilistic machines dedicated to simulating and planning motion in a noisy and uncertain world (e.g. Wolpert, 2007). Regardless of one's own stance on what the human species' most distinguishing feature is, the importance of language to understanding human cognition cannot be overstated. As we move in the world, we interact with parts of it and with other human agents in it, sharing information and often operating on the basis of information provided through language. This is not surprising. Although most organisms can loosely be said to communicate with other members of their – and, occasionally, other – species, human natural languages could easily be defined as the richest yet most flexible forms of communication we are aware of. Using language, we can communicate information regarding all aspects of our reality, from the deeply personal realm of introspection, to the mundane, the practical, and the philosophical. Despite the everyday nature of the phenomenon, quite how this is achieved is still a matter of investigation, and the life goal of many a researcher. The aim of this thesis was certainly not so ambitious, but the past five chapters have nevertheless tackled a few complex issues that require an integration.

The specific focus of this work was on untangling some of the processes involved in the communication and processing of spatial language, the sort of language used to describe environments and routes, and to direct action in the world. A number of scholars have explored the power of language to convey spatial information in a way comparable to sensory modalities such as vision, hearing, or hapsis (e.g. Avraamides et al., 2004), and the general trends that seem to be dominant in the way humans output spatial information through language (e.g. Denis et al., 1999). However, many questions remain unanswered concerning the key processes involved. Namely, how does propositional content provide input to the disparate cognitive systems that must be involved in driving action in the world? Although seemingly narrow, this area of investigation actually sits at the potential confluence of work exploring thus far compartmentalised areas of linguistic and cognitive research. The solution to bridging these gaps lies, I contend, at least partly in a more systematic understanding of mental imagery. The claim is not particularly novel; Jackendoff (2012) has quite convincingly argued for the need to bridge the gaps between action, perception, and language in order to develop a coherent theory of spatial understanding. The key to this, Jackendoff explains, is in understanding how information from different sources and modalities is encoded within mental representations (and in the underlying brain states), and how it is translated between different formats in order for it to be input and output for different cognitive systems.

The work presented in this thesis was driven by this overarching goal, but predominantly concerned with developing paradigms that might help produce these answers. While operating on the (empirically motivated) assumption that mental imagery mediates language processing (see Sections 1.9 and 1.10), methodological difficulties still complicate the study of mental representations. Owing to their internal nature, they are inherently difficult to control and leave the interpretation of behavioural data resting on assumptions as to the format and content of the representations active during a number of tasks. This is especially true when studying the construction of mental representations on the basis of linguistic input, the latter being often less amenable to experimental control than other stimuli. Accordingly, Experiments 1-3 represented an attempt to design a simple paradigm that may constrain language encoding processes and the construction of mental representations, in order to make the formulation and testing of hypotheses easier.

In this respect I relied on several key theoretical notions, manipulations, and methodologies. Among them, the manipulation of reference frames and imagined perspectives was an essential component, given their central role in mental representations of space. As we physically move through space, we have little choice but to experience that environment from the confines of our own body. Someone engaging in mental navigation, however, is afforded considerably more freedom. When exploring an imagined location we may mentally visualise a portion of said environment far larger than what we would normally be able to see from a single vantage point. In that case we can imagine the locations of various landmarks in the environment with respect to each other. Alternatively, we might imagine mentally locomoting through the same environment from a more familiar ground-level perspective, thus visualising the locations of various landmarks with respect to our current imagined location. These are examples of, respectively, an allocentric and an egocentric mental representation.

The two types of representations have been studied extensively in a number of paradigms within spatial cognition, and are thought to rely on largely different neural structures. However, our ability to construct both of them in parallel and to switch between them more or less seamlessly (see Section 1.2), as well as the limits of the egocentric-allocentric dichotomy highlighted by studies of language production (see Section 2.3), constitute a source of considerable uncertainty in the interpretation of behavioural results unless we are willing to make significant assumptions as to the nature of the tasks being used. One such assumption might be that encoding spatial texts that describe spatial relations (or motion) from an egocentric perspective (i.e. using “left” and “right”) will produce an egocentric representation, and a text describing an environment from a survey perspective (i.e.

using cardinal terms) will result in an allocentric mental representation (see Section 1.10). Experiment 1 tested this hypothesis by presenting participants with route directions written using egocentric or cardinal relational terms, and then testing their spatial knowledge in an allocentric task. The prediction that encoding a set of cardinal route directions would prompt an allocentric mental representation of the route, and thus lead to better performance during an allocentric test (see Section 2.5), proved incorrect. This indicated that the sequential nature of a route may lead readers to construct an egocentric representation of it regardless of the relational terms used and irrespective of optimal task strategy.

Experiments 2 and 3 improved upon the paradigm by introducing explicit imagery instructions, intended to further constrain the mental representations constructed during the processing of spatial texts, in order to (better) extricate egocentric and allocentric mental representations. That is, participants were asked to imagine an egocentric or an allocentric perspective while encoding the two types of route descriptions. This extended the manipulation of congruence between encoding and test to the manipulation of congruence between aspects of encoding itself. Behavioural and self-report measures pointed to the success of this manipulation, and to the resulting creation of four experimental conditions representing different degrees of congruence (or interference) between relational terms and imagined perspective. While this was an important step in the direction of establishing experimental control over mental representations, it also raised two interesting theoretical questions: what, exactly, is interfering with what? And what is the relationship between mental imagery and the phenomenological experience of it during language processing? I will attempt to provide answers, or to present possible ways to obtain them, in Section 6.6. Nevertheless, the manipulation of congruence in this fashion provided a way to better interpret not only behavioural data, but also encoding processes by putting them in context. These processes were explored by recording participants' eye movements during reading and, in Experiment 3, map inspection.

Key results from Experiments 1-3 showed that eye movements during reading and during the processing of sketch maps are susceptible to manipulations of reference frame and imagined perspective. This was observed in the specific case of landmark salience, pointing to distinct salience profiles for landmark words and landmark regions depending on whether participants were reading an egocentric or a cardinal description, and on whether they were imagining a first-person or a bird's-eye view of the described route. This not only replicates and extends recent results (Piccardi et al., 2016), but also confirms the role of eye movements as a potential source of information for certain aspects of mental representations that would otherwise be difficult to probe. By extension, it also provides good cause to improve our models of eye movement control to account for top-down effects and individual proclivities. Lastly, Experiments 2, 4 and 5 also highlighted the potential importance of investigating phenomenal experiences to ascertain the effectiveness of key manipulations and avoid risky assumptions concerning task demands and strategy. More generally, these experiments have demonstrated the existence of a hidden layer of complexity between linguistic input and behavioural output, and that investigating it in terms of the forms of mental representations that might be active during a task might be a fruitful research approach. In this sense, models of mental representations will need to coalesce into a more coherent theory of mental imagery in order to better investigate its relevance and connections to language and other non-linguistic domains. A particularly intriguing issue is that of the embodiment of mental representations, discussed in Chapter 4. What are the factors – linguistic and non-linguistic – that contribute to embodiment? Do mental representations exist on a continuum of embodied states, and does this continuum intersect with the continuum of possible imagined perspectives? What are the individual difference factors that determine the tendency to construct embodied and quasi-perceptual representations as opposed to more abstract and schematic ones? These remain open questions, but in this thesis I have presented a possible approach to tackling them that relies on constraints set at different stages of a task (encoding, processing, response) in order to better interpret behavioural and physiological data.

In the following sections I will discuss the key results that emerged from these experiments and their possible theoretical implications in more detail. Their relevance to potential practical applications will be then discussed in Section 6.6.

6.2. Eye Movements, Attention, and Imagery

Devising linguistic and imagery manipulations that could systematically constrain the mental representational processes at play during language processing and behavioural tests was only part of the challenge posed by this research. Equally important was finding ways to take advantage of that experimental control in order to try and explore the nature of the mental representations involved. In Experiments 1-3, the choice fell on the study of eye movements for a number of reasons.

Eye-tracking has a long history in cognitive science (see Section 2.6 and Section 2.6). It has been used extensively to inform models of reading (e.g. Reichle et al., 1998), models of attention (e.g. Hoffman, 1998), and models of scene processing (e.g. Itti & Koch, 2000). It has also found application in the study of eye movements during map processing (e.g. Caster & Eastman, 1985), and, more recently, during visual imagery (e.g. Johansson et al., 2012) and spatial navigation (e.g. Wiener et al., 2011b,c). Experiments 1-3 attempted to tie these distinct lines of research together, while also providing the first example (to my knowledge) of an eye-tracking study of spatial language. That is, the experiments presented in Chapter 3 employed eye-tracking to study the allocation of attention to different regions of interest (RoIs) in a spatial description or on a map-like representation, and its modulation as a result of the linguistic and imagery manipulations used. The benefit of an eye-tracking methodology is a significant spatial and temporal resolution, allowing to study the allocation of attention and processing difficulties much more precisely than with the more coarse measures of reading used in a few previous studies of spatial language processing (e.g. Tom & Tversky, 2012). For the purposes of Experiments 1-3, particular focus was placed on landmarks, key navigational features the salience perception of which is an area of active investigation (see Section 1.3).

In this sense, eye movement measures in Experiments 1-3 have provided a fairly consistent result. Namely, the salience profile of landmark words in spatial descriptions is modulated by the type of relational term used, with egocentric terms leading to longer dwell times on navigationally salient landmark words (i.e. landmarks described as being at turn locations) compared to non-salient landmark words, and cardinal terms leading to a reduction of this difference. Similarly, the salience profile of landmark regions during a map verification task (Experiment 3) appears to be modulated top-down by the mental imagery manipulation used, so that allocentric imagery appears to reduce the navigational salience of turn-location landmark regions. These results partly replicate existing results concerning the allocation of attention to visual stimuli (e.g. sketch maps, as in Piccardi et al., 2016), while also generalising them to the processing of language. In this sense, they are also potentially informative both with respect to the role of mental imagery during language processing and navigation, and with respect to models of reading and scene processing. Models of eye movement control have not traditionally accounted for discourse-level effects in reading or top-down effects more generally, focusing instead on word-level lexical and semantic factors, and bottom-up perceptual influences. This appears to be changing – perhaps more so in scene processing research than in reading (e.g. Torralba et al., 2006) – and interest is growing in the effects of high-level cognitive and external factors on eye movements and attentional control. These results suggest that mental imagery may play a significant role, further strengthening the link between imagery, perception, and action. This link has been implied to varying degrees at different stages of the imagery debate (see Section 1.8), but has become more prominent in discussions surrounding enactive and embodied models of cognition and imagery (e.g. Moulton & Kosslyn, 2009. See Section 1.10.). It was also relevant in Chapter 4 of this thesis, in which Experiments 4 and 5 were presented. Those results will be reviewed in the next section.

6.3. Mental Imagery and the Embodiment of Imagined Perspectives

If establishing a degree of experimental control over participants' mental representations during encoding was the methodological challenge that characterised Experiments 1-3, a similar challenge stood in the way of interpreting the pattern of behavioural results obtained at test in Experiments 4 and 5. The experiments were designed to test the effects of different degrees of encoding-test congruence when the test phase probes egocentrically-represented spatial knowledge. However, unlike a, reasonably, unambiguously allocentric map-based task, tasks intended to test egocentric representations may not be unambiguously egocentric in nature. More specifically, tasks involving bearing estimates such as the triangle completion task (e.g. Klatzky et al., 1998) may be completed adopting an allocentric strategy (e.g. Gramann et al., 2005; Wiener, Berthoz & Wolbers, 2011), and considerable overlap may exist between egocentric pointing tasks (e.g. Wang & Spelke, 2000) and the judgements of relative directions (JRD) traditionally used to study allocentric knowledge (e.g. Kozhevnikov & Hegarty, 2001. See Section 1.11.). This warrants particularly cautious task analyses and devising methods to try and constrain test performance (and its underlying mental representations) (Ekstrom et al., 2014).

Accordingly, Experiments 4 and 5 maintained the same factorial design used in Experiments 2 and 3, but also contrasted the effects of two different response methods on the accuracy of egocentric bearing estimates. These were a 2D template used in one type of JRD task, and a body-referred response performed via a pointer. Globally, the experiments provided interesting result. On the one hand they provided evidence that spatial updating is possible in the absence of motion-related proprioception, vestibular input, and optic flow (contra Klatzky et al., 1998). On the other, they also challenged the idea that spatial updating during imagined movement (such as during the processing of described routes) is only observed if the response is performed verbally (Wraga, 2003). On the contrary, the use of a physical pointer appeared to increase bearing accuracy in the left/right-egocentric condition (what we might refer to as the purely egocentric condition of this design) compared to the use of a paper template. Although the difference in question was only marginally statistically significant (likely as a result of a small sample size), a reduction in error rate with the use of a physical pointer appears in line with embodied models of mental imagery (e.g. Barsalou, 1999; Byrne et al., 2007).

However, open questions remain. A specific one is with respect to the possible sources of variability in the consistency (or lack thereof) of self-report measures. While the questionnaire, designed to probe participants' phenomenological experiences of their internal representations, has provided good consistency between Experiments 2 and 3, consistency between them and Experiment 4, and between Experiments 4 and 5, was lower. Although differences in stimulus and task complexity may have played a role (i.e. the shorter routes used in Experiments 4 and 5 may have decreased the perceived encoding difficulty in the cardinal-egocentric condition), together with sample size differences, the questionnaire itself probably requires further validation. In Section 6.6, I will discuss future research possibilities with an eye to exploring the phenomenological aspects of imagery, together with a discussion of why these may be important. Furthermore, I will explore future research ideas of clear embodied and enactive inspiration, including ideas to better characterise the cognitive systems differentially active during experiential navigation in real environments, active navigation of virtual environments, passive exposure to visual routes, and mental navigation of described routes. This might not only contribute to understanding other sources of variability in spatial updating studies, but it might also be revealing from a developmental perspective. This possibility is motivated, in part, by the results of Experiments 6-8, described in the next section.

6.4. The Development of Reference Frame Translation and Landmark Salience

While the core of this research was concerned with characterising spatial language encoding and imagery processes, and the way they might influence the spatial performance of typical adults, it is also important to consider the developmental implications. For example, determining whether (and to what extent) the individual differences in reference frame strategy selection observed in adults have a developmental origin, and whether (and to what extent) they are environmentally reinforced through development. While such an ambitious longitudinal study was not the goal of Experiments 6-8, having access to large cross-sectional samples of children between early childhood and adolescence presented an opportunity to characterise the continuous development of certain spatial abilities in children. In these experiments I focused on the emergence of landmark salience perception and of the ability to transform egocentric representations of routes (i.e. derived from videos of routes through a virtual environment) into allocentric representations of those routes.

Although the paradigm used was inspired by the available literature, both developmental (e.g. van Ekert et al., 2015) and in adults (e.g. Janzen & van Turennout, 2004), questions remain as to whether the design and tasks adopted were sensitive enough to capture the developmental changes of interest, and whether part of the observed effects were simply statistical artefacts. Children's performance in identifying, on an allocentric representation, the terminus location of a route experienced egocentrically revealed a particularly puzzling trend of results. In Experiments 6 and 7, this measure displayed a performance advantage for the youngest children when decision-point landmarks were available either during route encoding or on the test map. However, performance in these conditions then revealed a slower rate of improvement over development compared to performance when non-decision-point landmarks were available. The exact reason for this pattern of results remains unclear, and the small effect size of the observed interactions warrants caution.

Two key issues emerge from a comparison between these experiments and developmental studies of reference frame use and landmark processing: the use of shorter routes with fewer landmarks than is usual in the literature, and the use of passive exploration of visual routes rather than allowing participants to actively explore real or virtual environments. Both of these factors could have feasibly affected children's ability to form stable associations between object identities and their spatial contexts. They also raise the possibility that children may be particularly susceptible to the lack of allothetic input that characterises non-experiential navigation, and may therefore have adopted non-navigational and/or non-imaginal strategies to carry out these particular tasks, e.g. storing and recalling propositional sequences of relational terms rehearsed during encoding. Related to this, the maturational aspects of enactive and embodied mechanisms in imagery should be more closely explored in future studies, with an eye to teasing out factors that might significantly contribute to individual differences. These will be important in delineating the developmental aspects of a model whose goal is to integrate perception and action with imagery.

6.5. Modelling Spatial Cognition: Missing Pieces

The impetus for this research came largely from the need to better understand and characterise the relationship between spatial language and spatial cognition more generally. This need, and the status of spatial language as a potential source of information on spatial cognition, was convincingly argued by Jackendoff (2012) and articulated in terms of informational transfer between different mental representational formats (Section 2.2); more specifically, as the translation of linguistic conceptual structures into a spatial structure (and vice versa), wherein words contained in the mental lexicon are linked to an imaginal component (Figure 2.4). Jackendoff understands this spatial structure as an amodal or multimodal (in that it is generated by the confluence of input from potentially all sensory modalities), viewpoint-independent volumetric representation of objects and environments that supports all the processes involved in spatial understanding.

As such, it seems, a theory of spatial understanding is ultimately a theory of the representations that are generated in different cognitive systems and of the information they encode (computational theory); of how these representations are generated (algorithmic theory); and of how they are implemented in the biological substrate of the brain (implementation theory). Such a conceptualisation effectively transforms the deceptively simple interface problem in Figure 2.4, into a much more complex web of cognitive interactions (Figure 6.1) that must be disentangled before a viable model of spatial cognition can be constructed. Although a number of different approaches have been attempted (see Jackendoff, 2015), no model has so far been proposed comprehensive enough to satisfy all three of Marr's (1982) levels of analysis with respect to the entirety of spatial understanding. However, attempts have been made to model key aspects of spatial cognition. These may serve as inspiration for a model that can similarly address the key processes through which spatial language provides input for mental representations that can support spatial behaviour. In light of the results presented in this thesis, such a model must be able to account for the creation of both egocentric and allocentric mental representations (and, crucially, anything in between), and integrate both top-down and bottom-up effects. Additionally, it should be able to provide a biologically plausible platform onto which to develop a model of spatial language processing (as per Marr's implementation theory).

The BBB model (so named after the authors who formulated it; Byrne, Becker & Burgess, 2007) could be a starting point in this sense. It is a model of spatial memory that covers the encoding and retrieval of spatial scenes, key processes involved in spatial navigation (e.g. spatial updating), and certain aspects of mental imagery and route planning. Its key components (Figure 6.2) include different neuronal populations tuned to fire under different conditions. A neuronal population in the precuneus (Parietal Window, or PW) is involved in generating transient egocentric representations of environments. These neurons integrate sensory information between head- and body-centred reference frames on a polar grid centred on the navigator's location, and each neuron in the grid is tuned to respond more strongly to object and landmarks at a certain distance away from and direction relative to the origin point (i.e. the navigator's location). Conversely, a set of medial-temporal populations is involved in generating allocentric representations (hippocampal Place Cells in area CA3, and Boundary Vector Cells, or BVCs), wherein each neuron on the allocentric grid is tuned to respond more strongly to objects, landmarks, and environmental boundaries at a particular distance from the organism's location. The navigator's heading, however, is fixed in an allocentric direction with a fixed alignment to the environment.

A third component (identified with the Retrosplenial cortex) drives the translation between egocentric and allocentric representations under modulation of a Head Direction component (Head Direction cells) and of idiothetic input during locomotion in general. This way, the model can generate a representation of the relative positions of environmental features visible from a visited location while maintaining a certain head direction. Crucially, information in the model can flow both bottom-up, building transient representations from sensory input and translating them into viewpoint-independent representations for long-term storage, and top-down, allowing for memory recall of object locations from imagined viewpoints. During imagined navigation, the efference-proprioceptive-vestibular signals that would be generated by actual motion can be simulated and used to support spatial updating during mental navigation. Lastly, a perirhinal Object Identity module simulates the binding of object features (ventral visual pathway) with locations within allocentric space.

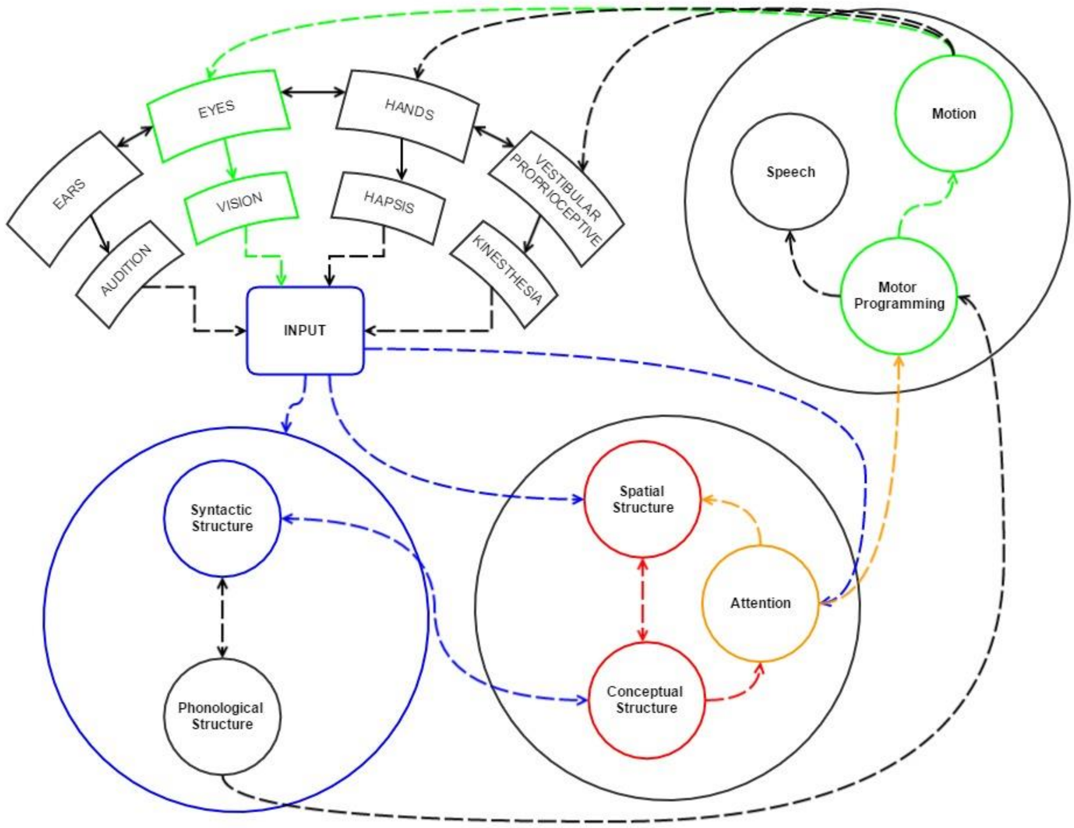


Figure 6.1 – Schematic representation of some of the interactions that must be explained by a theory of spatial language. Experiments 1-3, in particular, focused on the ways manipulating linguistic input and task instructions (blue) influenced discourse-level processing and the construction of the resulting mental representations (red). The latter process was studied indirectly by measuring both behavioural output and the allocation of attention (orange) during reading as measured via eye movements (green).

The model displayed considerable flexibility during training and validation on simple environmental representations (Byrne et al., 2007), and has received empirical support from brain-imaging studies, particularly with respect to hippocampal and retrosplenial activation in tasks involving mental transformation of reference frames (e.g. Dhindsa, Drobinin, King, Hall, Burgess & Becker, 2014; Lambrey, Doeller, Berthoz & Burgess, 2012). However, it also has important limitations that are relevant to this thesis. First of all, it was not intended to model the construction of mental representations of space on the basis of non-sensory input, such as during language processing. While it is reasonable to compartmentalise cognitive models to the extent that it is necessary to fully understand individual aspects of cognition, a more comprehensive model of the type envisioned by Jackendoff must attempt to model aspects of language processing. One option for achieving this may be found within enactive and embodied models of cognition (Section 1.10), in which cognition is grounded in and influenced by the physical properties of the world, our experience of which is similarly mediated and constrained by the physical characteristics of our bodies (Pezzulo, Barsalou, Cangelosi, Fischer, McRae & Spivey, 2013). Crucially, this principle extends to the acquisition of language. Thus, during development, we accrue episodic memories that allow us to populate our conceptual structure with linguistic labels. These are associated with multimodal inputs as we have personally experienced them, and whose reactivation can be modulated in a situated fashion depending on circumstances (e.g. task demands). Such a conceptualisation of cognition (Figure 6.3), and its extension to language, could be extremely useful in modelling individual differences in spatial tasks (Section 1.11), individual differences in performance in spatial language tasks like the ones presented in this thesis (Chapters 3 and 4), or cross-linguistic differences in the expression and use of concepts (Section 2.2). In this way, it may also inform a global theory of mental imagery more generally.

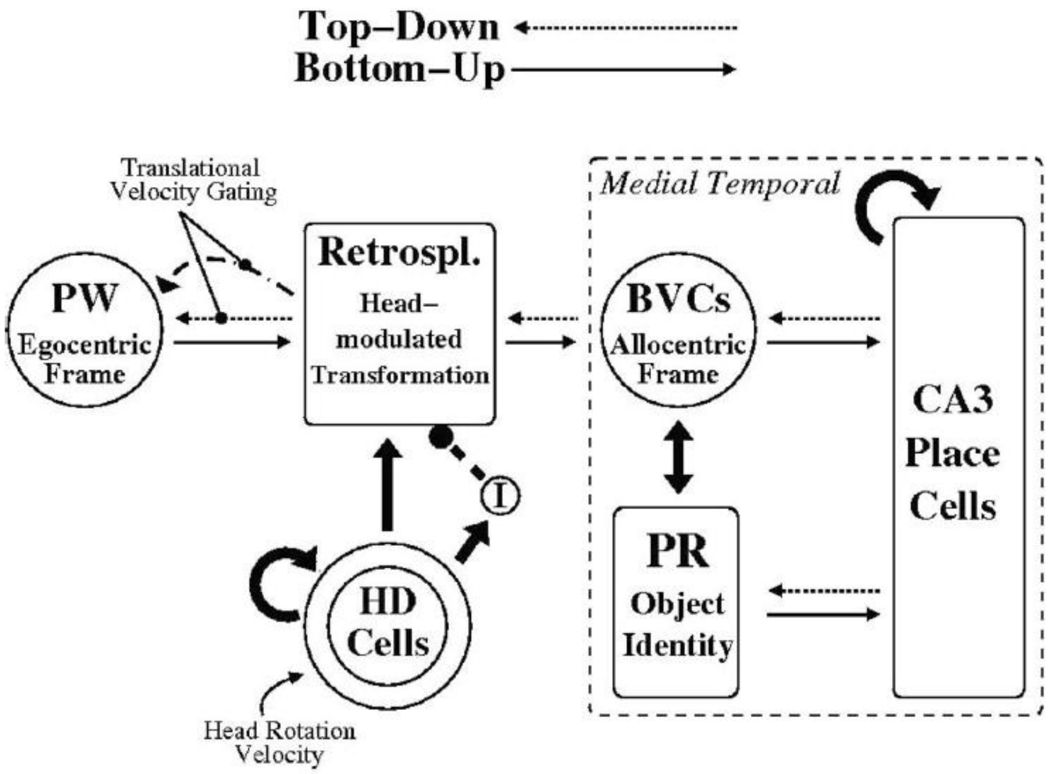


Figure 6.2 – Schematic representation of the BBB model (Byrne et al., 2007), including bottom-up and top-down information flow between its key components.

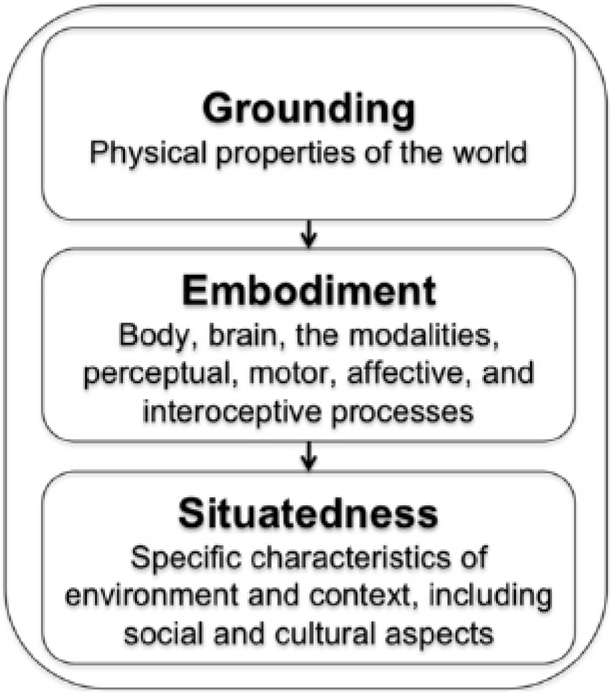


Figure 6.3 – Cascade of effects on cognition during the acquisition and use of conceptual structures (Pezzulo et al., 2013).

The experiments presented in this thesis can contribute to this development by pointing to additional elements that a model of spatial language comprehension and imagery should consider. More specifically, Experiments 2 and 3 (and, to an extent, 4 and 5) suggest that complex interactions exist between linguistic reference frame and imagined perspective, and that they may not necessarily coincide during naturalistic reading. While the type of relational term used in a description seemed to significantly determine the allocation of attention to different elements of the text's surface level (e.g. landmark words), results indicate that a top-down imagery modulation resulted in the creation of different mental models. This was evidenced by participants' self-reports (i.e. representational difficulties in Cardinal-Egocentric conditions), behavioural responses (i.e. increased performance in conditions with congruent imagery), and eye movement measures during map processing (i.e. shift in landmark salience between description encoding and map verification in Left/Right-Allocentric and Cardinal/Egocentric conditions). To begin modelling these interactions, here I will argue for the use of a construct already presented in Section 2.6 in the context of top-down effects on scene processing. That is, the idea of a spatial priority map. This notion is an extension of established bottom-up models of vision, in which attentional capture is the result of purely perceptual features (Figure 1.2). Within spatial priority maps (Figure 6.4), perceptual salience in the visual field is modulated by non-perceptual factors such as individual differences and preferences, expectations, behavioural goals and task demands (Ptak, 2012).

In light of the analogue nature of imagery, the notion of priority map might lend itself to being applied to the preferential activation of certain features of a spatial representation of a described space. In this model, linguistic factors and imagined spatial perspective would compete to establish different salience profiles during the construction of a mental representation and for control of motor programs (e.g. eye movement control). In Experiment 3, egocentric relational terms appeared to prioritise landmarks that would be navigationally relevant during egocentric navigation, whereas cardinal relational terms appeared to reduce this difference, regardless of imagined perspective (consistent with recent results in the literature. See Piccardi et al., 2016.). However, while linguistic factors seemed to drive eye movements during reading (Figure 3.37), eye movements during map verification (Figure 3.38) showed evidence that, in conditions involving an encoding conflict (Left/Right-Allocentric and Cardinal-Egocentric), the top-down imagined spatial perspective determined the priority profile of landmarks, resembling their "purer" imagery conditions (Cardinal-Allocentric and Left/Right-Egocentric, respectively). This might be taken to indicate a situated nature of eye movement control, wherein motor programs during reading tasks are less susceptible to top-down effects, even when these are influencing the way the linguistic information is being mentally represented.

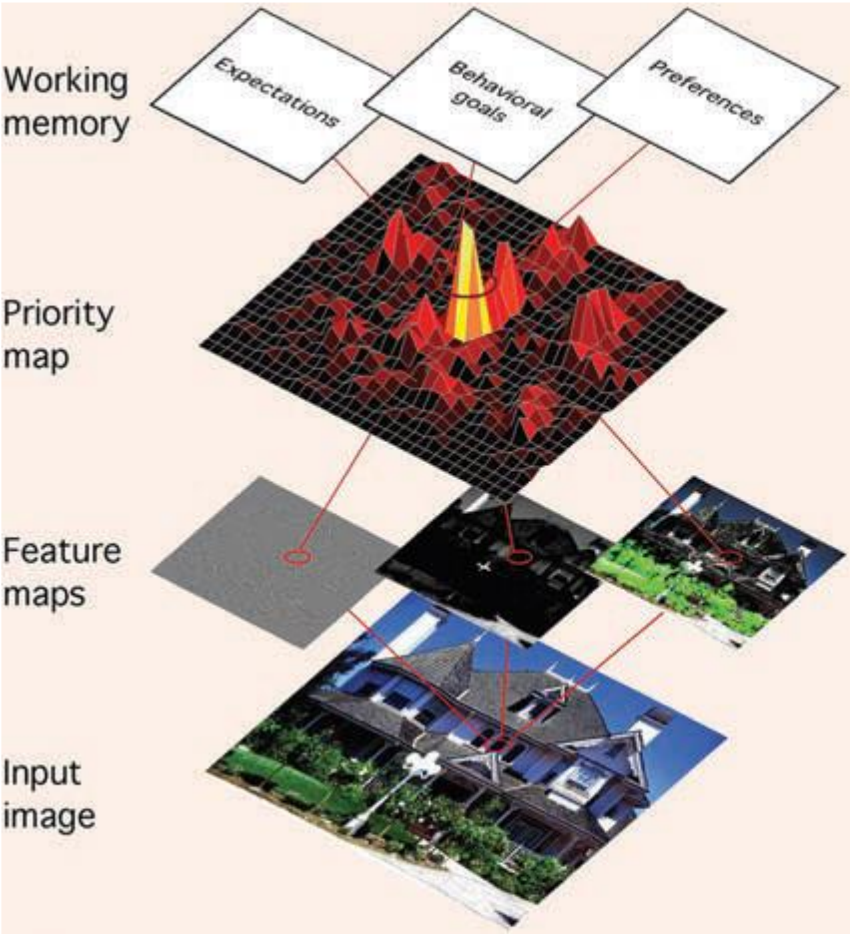


Figure 6.4 – An example of priority map, in which bottom-up salience and top-down relevance are combined. The resulting activation peaks correspond to the areas of the visual field with the highest priority.

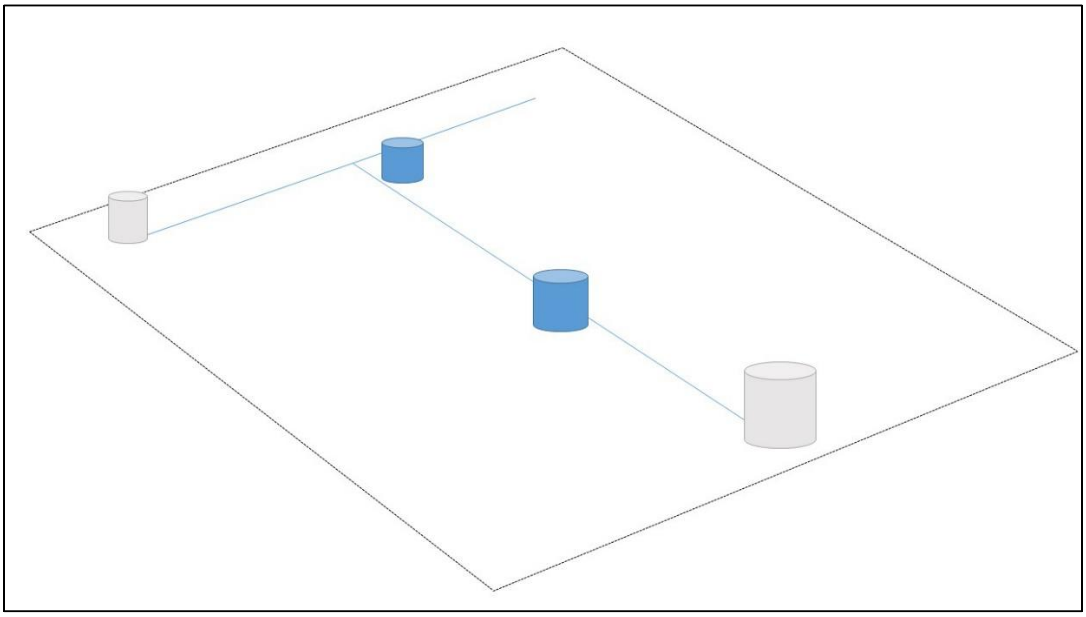


Figure 6.5 – Visual representation of a simple route through an environment, containing an origin and a destination landmark (grey) and two route landmarks (blue), one at a turn location and one along a path segment.

Accordingly, encoding written descriptions of the route in Figure 6.5 in the four conditions used in Experiments 2-5 would yield the pattern of results described in Figure 6.6 and Figure 6.7. The use of egocentric relational terms (L/R) prompts the allocation of attention to landmark words described as being at turn points over other landmark words (+). Cardinal terms yield the opposite pattern of eye movements during reading (-). While the type of relational term determines the allocation of attention to text regions during reading, the top-down instructions determine the imagined perspective of the representation being built and its priority map. The latter, represented here as a grid-like structure superimposed on the allocentric mental representation participants would have used during an allocentric test phase, reflects the landmark salience profile created by the imagined viewpoint maintained during description encoding, rather than that generated at the surface level of the text. That is, the degree of salience of each landmark is encoded in the quasi-perceptual properties of mental imagery during encoding, and is inherited by the mental representation used at test even through an intervening reference frame translation.

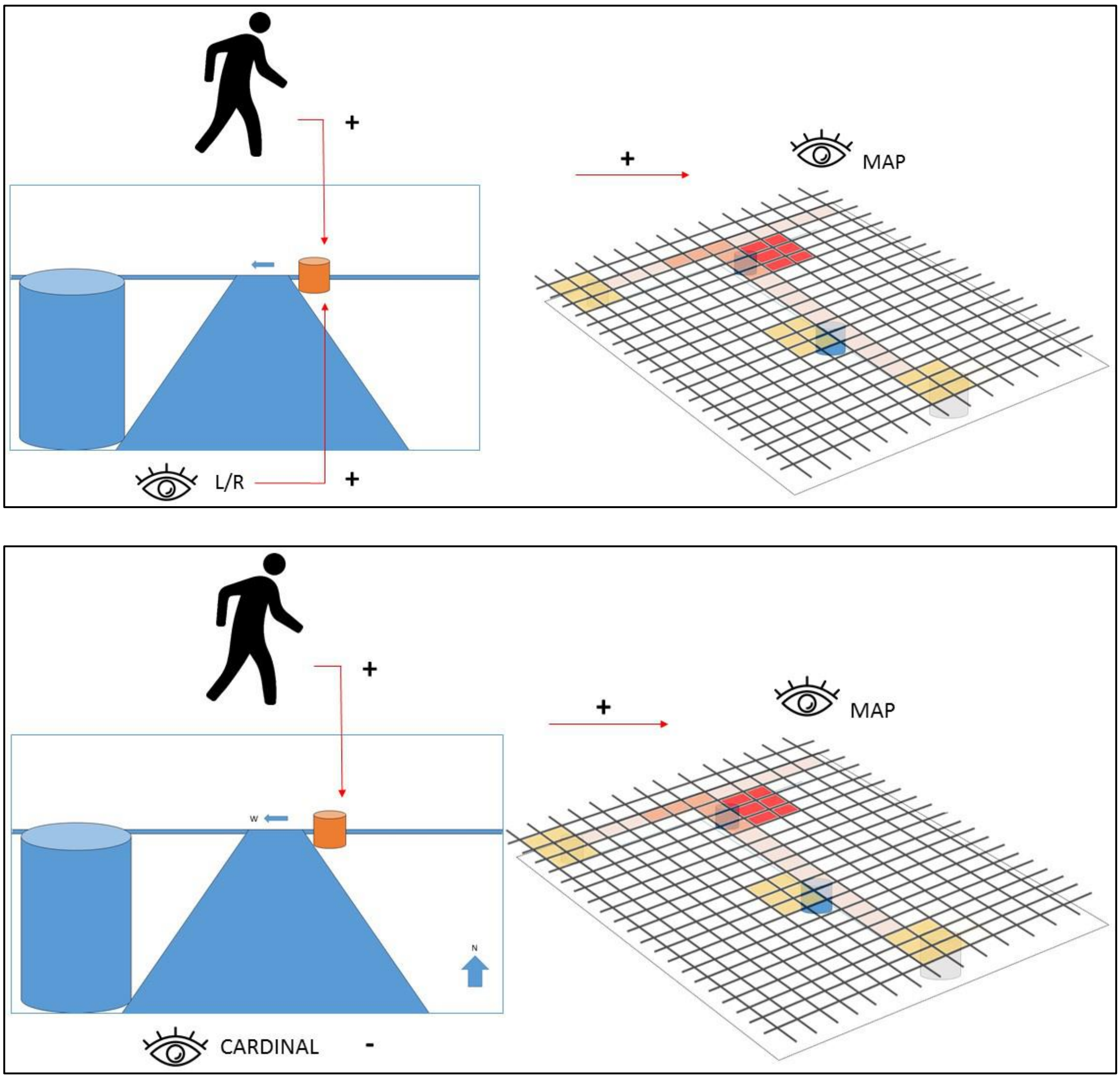


Figure 6.6 – Schematic representation of the pattern of eye movements observed during description encoding and map verification in the L/R-Egocentric (top) and Cardinal-Egocentric (bottom) conditions. A + symbol indicates greater attention allocation to salient landmark words or map regions. The blue arrows in the egocentric view panels indicate the direction of the turn being encoded. Additionally, the egocentric view of the Cardinal-Egocentric condition displays the cardinal reference directions that a participant would need to encode at the start of the route and update after every turn.

While the model being presented here is tentative and in need of considerable empirical validation, it nevertheless suggests that the link between mental imagery and eye movements, already attested in the literature (Section 2.6 and Section 2.7), warrants further research in order to better model its underlying mechanisms. Similarly, it provides a theoretical reason to better explore and control participants' phenomenological experiences of imagery, as these might account for a significant amount of variance in key aspects of spatial behaviour such as landmark salience processing. Section 6.6 will present ideas for future research aimed at further exploring the interactions described thus far, and possibly developing practical applications.

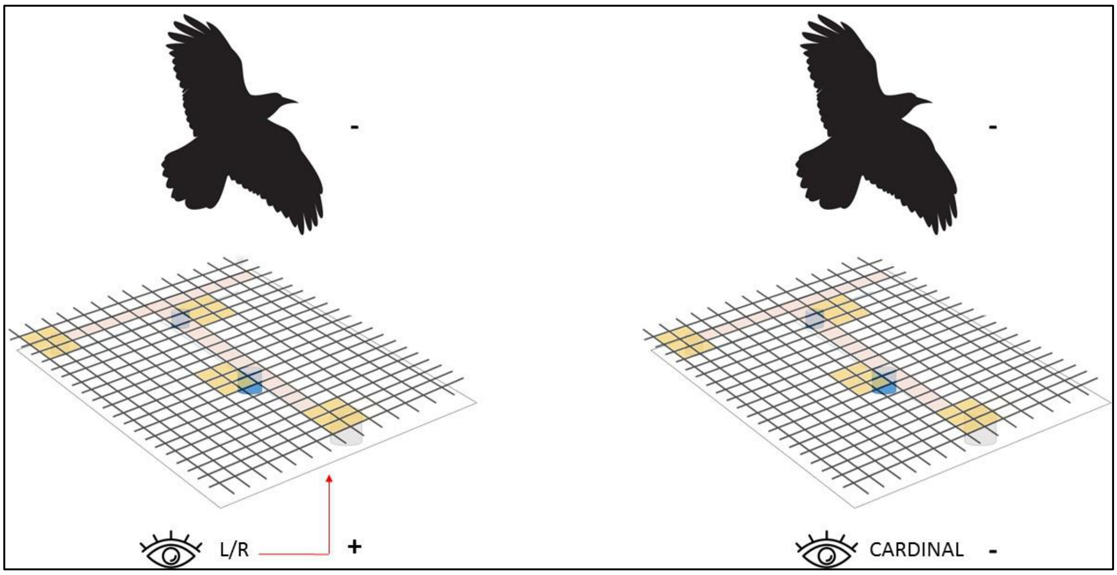


Figure 6.7 – Schematic representation of the pattern of eye movements observed during description encoding and map verification in the L/R-Allocentric (left) and Cardinal-Allocentric (right) conditions. A + symbol indicates greater attention allocation to salient landmark words or map regions.

6.6. Future Research and Conclusions

In Section 6.5 I have presented an embryonic extension to the BBB model (Byrne et al., 2007) with the goal of better characterising the processes involved in the construction of mental representations of space on the basis of linguistic input. Motivated both by the available literature and by the results obtained in Experiments 1-5, I have made the following proposals.

I have suggested that spatial language processing might, at least in part, find its theoretical home in grounded models of cognition based on perceptual symbols and embodiment (e.g. Barsalou, 1999). While the extent to which cognition is based on modal or multimodal symbols is unclear and the subject of considerable debate (see Barsalou, 2016), I believe a research programme investigating modulations of embodiment could be greatly informed by an investigation of spatial language processing and imagery. In this sense, vestibular input and proprioception might provide intriguing targets, given the involvement of the vestibular system and vestibular dysfunction in visuospatial and motor imagery (Section 1.10), in hippocampal atrophy and spatial memory impairments in the absence of active motion (e.g. Brandt et al., 2005), as well as in embodied perspective-taking (e.g. Deroualle, Borel, Devèze & Lopez, 2015; Gardner, Stent, Mohr & Golding, 2016). As such, it appears that a better characterisation of the processes and neural structures involved in the top-down simulation of motion-related signals (a possibility also under the BBB model) is required. Brain imaging studies comparing real and imagined locomotion have provided a wealth of potential neural targets (see la Fougère et al., 2010), and these should be studied systematically to determine what imaginal (e.g. imagined perspective) and linguistic (e.g. implied reference frame, described mode of locomotion, route features, etc.) factors can modulate their activation or deactivation.

A second proposition put forward in Section 6.5 involves the notion of spatial priority maps (e.g. Ptak, 2012). I have suggested that these might be invoked to account for the observed top-down modulation of landmark salience within mental representations. Previous research has already provided neural targets for putative spatial priority maps (e.g. occipital, parietal and frontal cortices; Sprague & Serences, 2013). Although these maps have been predominantly explored within the context of the visual attention system and using psychophysical paradigms, activity patterns in the fronto-parietal attention network (FPAN) should be better characterised using the type of paradigm adopted in Experiments 1-3. The main reason for this is the degree of interconnectivity between a number of functional regions that the results from these experiments appear to point to. These regions would have to be implicated in imagery, e.g. visual cortex (Kosslyn, 1994), intraparietal sulcus (Just, Newman, Keller, McEleney & Carpenter, 2004), posterior parietal cortex (Mast et al., 2006), and in oculomotor control, e.g. Frontal Eye Field (FEF) (Ptak, 2012), lateral intraparietal area (LIP) (Mirpour, Arcizet, Ong & Bisley, 2009), and superior colliculus (SC) (Johansson, 2013).

The FPAN's diffuse nature (see Scolari, Seidl-Rathkopf & Kastner, 2015) produces this degree of functional overlap. In this sense, exploring structures that might modulate the allocation of attention during spatial language processing and imagery appears intrinsically tied to an exploration of how these processes influence eye movements, whose role in attentional processes is well established (see Section 2.6 and Section 2.6). In particular, research in this area might help explain why eye movements during reading appear to be less susceptible to top-down effects than during scene processing in Experiments 2 and 3. Additionally, other networks (e.g. Dorsal Attention Network, Default Mode Network) might interface the FPAN with hippocampal and parahippocampal regions, involved in the encoding and retrieval of information (e.g. Kim, 2015) and, as discussed previously, also in the construction of viewpoint-invariant, allocentric spatial representations (e.g. Spiers & Barry, 2015).

As a result, investigating the neural correlates of the effects observed in Experiments 1-3 is likely to reveal intricate patterns of activations that will require better theoretical frameworks to be extricated. Nevertheless, this line of research seems suited to the eventual use of co-registration paradigms employing both eye-tracking and brain-imaging components, in order to investigate how the key linguistic and top-down manipulations used here influence activity in the many areas involved in visuospatial imagery, spatial memory, navigation, and in oculomotor control (Figure 6.8) during reading and map verification. For example, future studies might employ EEG, MEG or fMRI to investigate changes in activity in and functional connectivity between these regions as participants fixated landmark words (salient and non-salient) in route descriptions (left/right and cardinal) while imagining different spatial perspectives (egocentric and allocentric).

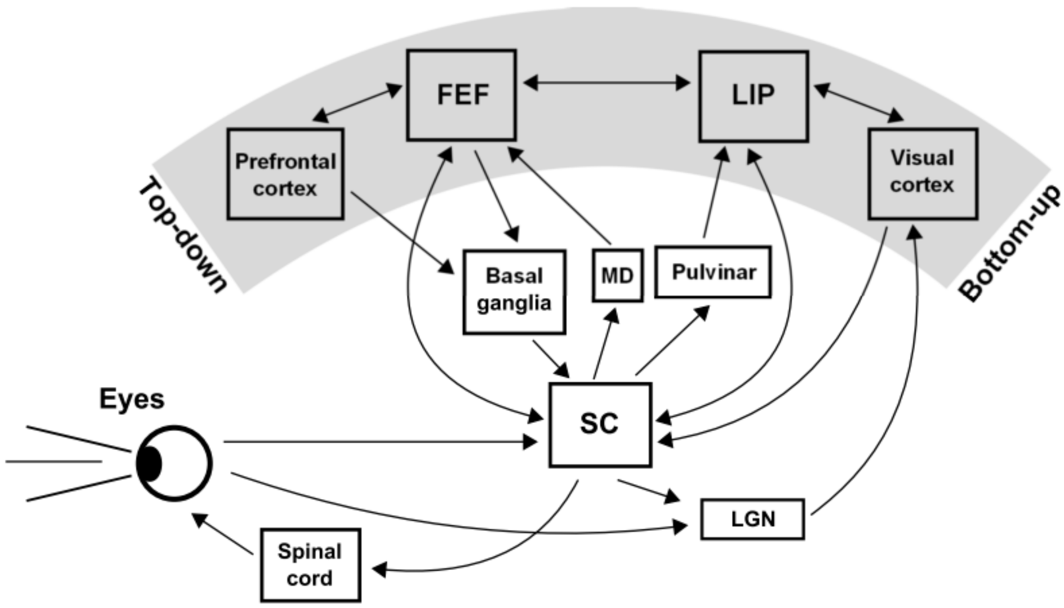


Figure 6.8 – The key areas involved in oculomotor control as a function of bottom-up perceptual and top-down factors (Johansson, 2013).

A further issue raised in this thesis is that of phenomenology. This was presented explicitly in Chapter 4 (Section 4.1), when discussing the importance attributed by some scholars to egocentric representations in defining consciousness (see also Briscoe, 2009 and McClelland & Bayne, 2016). However, to the extent that both egocentric (e.g. imagined locomotion) and allocentric (e.g. environmental rotation) mental representations are subjectively experienced, phenomenology in this case can be considered as the study of the qualitative aspects of mental representations. As seen in Experiments 4 and 5, probing aspects of participants' subjective imagery experiences (e.g. constancy of imagined perspective, perspective switching, distance visualisation, etc.) showed limitations. However, these may have been largely methodological, in that the use of a five-point Likert scale may have limited the usefulness of the self-report questionnaire and failed to capture finer nuances in the data. While the questionnaire itself will need to be restructured and validated, it also provided both evidence of conscious awareness of these qualitative aspects, and a sanity check to determine the effectiveness of the key experimental manipulation (i.e. relational term-imagined perspective mis/match). As such, the general approach shows promise, and might allow to address interesting questions. For example, whether qualitative differences in phenomenology correlate with qualitative differences both in behaviour and brain activity, to what extent visual and spatial imagery are distinct, or if the various types of mental representations discussed in Chapter 1 are best characterised as existing on a continuum along different measures (e.g. vividness, complexity, reference frame, etc.).

The possibility of objectively studying individual differences in mental imagery vividness has been empirically demonstrated during imagination of visual scenes (e.g. Cui, Jeter, Yang, Montague & Eagleman, 2006), but the principle could be extended to probe, for example, visual vividness during spatial description encoding, or to compare self-reported differences in embodiment as a function of reference frame and imagined perspective with physiological measures (e.g. ventilation and blood pressure. See Section 1.10.). In fact, studying individual differences in subjectively reported imagery experiences (and correlating them with objective measures and behavioural performances) might allow to address a more fundamental question: is imagery phenomenology necessary or epiphenomenal? Although the experiments presented in this thesis appear to argue for a functional role of at least some phenomenal aspects of imagery, the question remains open. Zeman and colleagues (Zeman, Della Sala, Torrens, Gountouna, McGonigle & Logie, 2010; Zeman, Dewar & Della Sala, 2015; See also Zeman, Dewar & Della Sala, 2016) have recently identified a (so far) small group of individuals who report no significant imagery phenomenology (as measured via VVIQ; Marks, 1973). The first such individual, MX (Zeman et al., 2010), reported an abrupt loss of "mind's eye" phenomenology following coronary angioplasty, but was found to have an otherwise broadly preserved cognitive profile (including ability to navigate around familiar environments, describing familiar routes, or drawing pictures from memory). Despite an ability to provide detailed descriptions of visual details of scenes and landmarks, MX stated: *"I can remember visual details, but I can't see them…I can't explain that…From time to time I do miss being able to see"* (Zeman et al., 2010, p. 147). Zeman and colleagues have termed this particular state aphantasia, thus providing a challenge for current theories of mental imagery. Future studies should attempt to fully characterise the imagery profile of individuals with congenital or acquired aphantasia. With respect to spatial language and navigation, aphantasic individuals might be the ideal population to extricate visual and spatial imagery, and the experiments presented in this thesis offer a good paradigm to attempt just that. The use of skeletal route descriptions without any vividly visual content, and the manipulation of a spatial element (reference frame/imagined perspective), might allow to determine whether aphantasia extends to all forms of imagery or affects only its most quasi-perceptual visual aspects. If participants are able to correctly translate a spatial representation between two imagined perspectives without actually reporting a phenomenological experience of the process, then it would be interesting to see if they are also immune to the reference frame-imagined perspective conflict experienced by participants in the Cardinal-Egocentric condition in Experiments 2 and 3.

All theoretical interests aside, it is also important to consider what practical applications an understanding of mental imagery, and its integration with spatial language within a broader theory of spatial cognition, might produce. A number of applied (and clinical) research possibilities were listed in Section 2.8. Among them, the possibility of developing an intuitive imagery-based brain-computer interface (BCI) is certainly among the most intriguing. A BCI is a device (usually involving non-invasive EEG interfaces or more invasive intracranial and intracortical electrodes, e.g. Homer, Nurmikko, Donoghue & Hochberg, 2013; Miller, Schalk, Hermes, Ojemann & Rao, 2016; sometimes fMRI, e.g. Emmerling, Zimmerman, Sorger, Frost & Goebel, 2016) that allows a human being (occasionally a non-human animal, e.g. Rajangam et al., 2016) to control a machine (i.e. a computer or assistive device). Once the raw signal is obtained from the neural source of interest (usually in response to specific training stimuli or task instructions), it is processed to determine which of its components best differentiate it from other signals, a process called feature extraction. Once sufficient amounts of different signals are obtained, they are used to build a classifier, an algorithm that can then be used to classify new data. By associating distinct signal types, each with distinct features, to distinct interaction events (e.g. cursor movements or clicking), control of a system can be achieved; at least in principle. The idea of developing an imagery-based BCI is not new, and motor imagery has been frequently used to operate direct BCIs (e.g. imagining the appropriate arm and hand movements to move a cursor and click on an icon on the screen) (Donoghue, 2008).

To determine the most efficient way of using motor imagery to control a BCI, Pfurtscheller et al. (2006) compared the performance of a learning classifier in categorising signals produced by kinaesthetic (i.e. egocentric) motor imagery and visuo-motor (i.e. third-person) imagery. The former was found to produce more reliable patterns of oscillatory brain activity that could less ambiguously be categorised as representing specific movements. However, patterns of cortical activation also revealed significant between-subject variability, a common problem in creating flexible BCIs. Over the years, substantial progress has been made on multiple fronts, but most significantly in the development of better neural classifiers. The use of machine-learning-based statistical techniques has allowed to significantly reduce the time required for users to train on a BCI (e.g. Blankertz, Dornhege, Lemm, Krauledat, Curio & Müller, 2006), while also building classifiers flexible enough to adapt, to an extent, to between-subject and inter-session variability in the underlying signals (e.g. Reuderink, Farquhar, Poel & Nijholt, 2011). Nevertheless, challenges remain, with too many studies using sample sizes too small to meaningfully inform our understanding of how information is encoded in the brain signals being recorded (Donoghue, 2008). Additionally, the field of brain-computer interaction is currently faced with one of the key issues that emerge in behavioural studies of mental representations: experimental control.

Without underestimating the substantial methodological and logistical difficulties of recruiting participants with electrocorticographic grids implanted, it is possible that the field might benefit from significantly increasing the average sample size used in studies adopting less invasive methodologies (e.g. EEG). This, coupled with systematic behavioural manipulations aimed at controlling the content and properties of the mental representations whose corresponding brain signals are being recorded, should provide plentiful training sets that can both account for within- and between-subject variability and take full advantage of the computational power afforded by the machine learning methods available today. This, in turn, might get us closer to a seamless imagery-based BCI that can be widely adopted and used without significant training. Such a device would not only be of substantial importance as an assistive device for individuals with limited mobility (e.g. Truccolo, Friehs, Donoghue & Hochberg, 2008), but also become an everyday item for healthy individuals, by complementing or replacing traditional interaction methods in areas such as gaming (e.g. Ahn, Lee, Choi & Jun, 2014). Intriguingly, eye-tracking could also find use as one component of similar human-computer interaction systems (e.g. Allison, Brunner, Kaiser, Müller-Putz, Neuper & Pfurtscheller, 2010; Belkacem et al., 2015; Lee, Woo, Kim, Whang & Park, 2010). Until such time, however, imagery research must continue to explore not only the mechanisms and factors that give rise to imagery, but also the factors (task-dependent and individual) that determine its format, properties, and specific content (e.g. Emmerling et al., 2016). To do this, far more, and far more in-depth, data must be collected on participants' subjective experiences, with mental representations being scored on a number of important dimensions such as visual vividness, spatial perspective, the perception of volumes, motion, embodiment, multimodal experiences, and many others. Collecting these rich, multidimensional data will allow researchers to better identify and make sense of patterns in behavioural data that might otherwise remain unexplained variance.

The research presented in this thesis attempted to do just that by systematically manipulating key elements of spatial representations of environments: landmark salience, imagined perspective, and the reference frame implied in linguistic descriptions. The results were complex, and confirmed the view of language as a potential treasure trove of information concerning broader spatial cognition, but only if appropriately accompanied by behavioural and self-report data. Although eye movements during reading appear immune to top-down modulation (at least in this paradigm), the linguistic reference frame influences the allocation of attention to landmark words in the text. Conversely, top-down imagined perspective determines the salience profile of the landmarks within the representation. While these results will require replication with a wider array of spatial texts, they seem to justify the use and sensitivity of eye-tracking measures as proxies of attention during both spatial language processing and visual processing, providing a window onto the processes that mediate it. Future research should build on these results to address unanswered questions concerning these complex interactions.

References

Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 529–543. doi:10.1037/0096-1523.15.3.529

Ahn, M., Lee, M., Choi, J., & Jun, S. (2014). A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors*, *14*(8), 14601–14633. doi:10.3390/s140814601

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000). Mental imagery and perception in Hallucination-prone individuals. *The Journal of Nervous and Mental Disease*, 188(12), 830–836. doi:10.1097/00005053-200012000-00007

Allen, G. L., Kirasic, K. C., Siegel, A. W., & Herman, J. F. (1979). Developmental issues in cognitive mapping: The selection and utilization of environmental landmarks. *Child Development*, *50*(4), 1062. doi:10.2307/1129332

Allison, B. Z., Brunner, C., Kaiser, V., Müller-Putz, G. R., Neuper, C., & Pfurtscheller, G. (2010). Toward a hybrid brain–computer interface based on imagined movement and visual attention. *Journal of Neural Engineering*, *7*(2), 026007. doi:10.1088/1741-2560/7/2/026007

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79–B87. doi:10.1016/j.cognition.2004.02.005

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi:10.1016/s0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi:10.1016/j.jml.2006.12.004

Anderson, J. R. (1983). *The architecture of Cognition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Anderson, J. R. (1985). *Cognitive psychology and its implications*. New York: W.H.Freeman & Co.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Avillac, M., Denève, S., Olivier, E., Pouget, A., & Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience*, *8*(7), 941–949. doi:10.1038/nn1480

Avraamides, M. N. (2003). Spatial updating of environments described in texts. *Cognitive Psychology*, *47*(4), 402–431. doi:10.1016/s0010-0285(03)00098-7

Avraamides, M. N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (2004). Use of cognitive versus perceptual heading during imagined Locomotion depends on the response mode. *Psychological Science*, *15*(6), 403–408. doi:10.1111/j.0956-7976.2004.00692.x

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*(4), 801–814. doi:10.1037/0278-7393.30.4.804

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). New York: Academic Press.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Lawrence Erlbaum Associates.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(04). doi:10.1017/s0140525x99002149

Barsalou, L. W. (2002). Being there conceptually: Simulating categories in preparation for situated action. In N. L. Stein, P. J. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 1–15). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. doi:10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, 23(4), 1122–1142. doi:10.3758/s13423-016-1028-3Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols and Embodiment: Debates on Meaning and Cognition*. New York: Oxford University Press.

Bartolomeo, P. (2002). The relationship between visual perception and visual mental imagery: A reappraisal of the Neuropsychological evidence. *Cortex*, *38*(3), 357–378. doi:10.1016/s0010-9452(08)70665-8

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332. doi:10.1073/pnas.1306572110

Baylor, G. W. (1973). Modelling the mind's eye. In A. Elithorn & D. Jones (Eds.), *Artificial and human thinking*. Amsterdam: Elsevier.

Belkacem, A. N., Saetia, S., Zintus-art, K., Shin, D., Kambara, H., Yoshimura, N., Berrached, N., Koike, Y. (2015). Real-time control of a video game using eye movements and two temporal EEG sensors. *Computational Intelligence and Neuroscience*, *2015*, 1–10. doi:10.1155/2015/653639

Bellugi, U., Bihrle, A., Neville, H., Jernigan, T., & Doherty, S. (1992). Language, cognition, and brain organization in a developmental disorder. In M. Gunnar & C. Nelson (Eds.), *Developmental behavioral neuroscience: The Minnesota symposium* (pp. 201–232). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. In S. Broman & J. Grafman (Eds.), *Atypical Cognitive Deficits in Developmental Disorders: Implications for Brain Function*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bense, S., Stephan, T., Yousry, T. A., Brandt, T., & Dieterich, M. (2001). Multisensory cortical signal increases and decreases during Vestibular galvanic stimulation (fMRI). *Article*, *85*(2), 886–899. Retrieved from http://jn.physiology.org/content/85/2/886

Bexton, W. H., Heron, W., & Scott, T. H. (1954). Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *8*(2), 70–76. doi:10.1037/h0083596

Billinghurst, M., & Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings of the Virtual Reality Annual International Symposium (VRAIS'95)*. Washington, DC: IEEE Computer Society.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the Parietal lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. doi:10.1146/annurev-neuro-060909-152823

Blackburn, H. L., & Benton, A. L. (1957). Revised administration and scoring of the digit span test. *Journal of Consulting Psychology*, *21*(2), 139–143. doi:10.1037/h0047235

Blades, M. (1990). The reliability of data collected from sketch maps. *Journal of Environmental Psychology*, *10*(4), 327–339. doi:10.1016/s0272-4944(05)80032-5

Blankertz, B., Dornhege, G., Krauledat, M., Muller, K.-R., Kunzmann, V., Losch, F., & Curio, G. (2006). The Berlin Brain–Computer interface: EEG-Based communication without subject training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(2), 147–152. doi:10.1109/tnsre.2006.875557

Borel, L., Redon-Zouiteni, C., Cauvin, P., Dumitrescu, M., Devèze, A., Magnan, J., & Péruch, P. (2014). Unilateral Vestibular loss Impairs external space representation. *PLoS ONE*, *9*(2), e88576. doi:10.1371/journal.pone.0088576

Bourlon, C., Oliviero, B., Wattiez, N., Pouget, P., & Bartolomeo, P. (2011). Visual mental imagery: What the head's eye tells the mind's eye. *Brain Research*, *1367*, 287–297. doi:10.1016/j.brainres.2010.10.039

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 529–533. doi:10.1016/s0022-5371(70)80096-2

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, MA: Cambridge University Press.

Brandt, T. (2005). Vestibular loss causes hippocampal atrophy and impaired spatial memory in humans. *Brain*, *128*(11), 2732–2741. doi:10.1093/brain/awh617

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38. doi:10.1162/jocn.1997.9.1.27

Brewster, S., & Blades, M. (1989). Which Way to go? Children's Ability to Give Directions in the Environment and from Maps. *Journal of Environmental Education and Information*, *8*, 141–156.

Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, *79*(2), 423–460. doi:10.1111/j.1933-1592.2009.00284.x

Broadbent, H. J., Farran, E. K., & Tolmie, A. (2014). Egocentric and allocentric navigation strategies in Williams syndrome and typical development. *Developmental Science*, *17*(6), 920–934. doi:10.1111/desc.12176

Broadbent, H. J., Farran, E. K., & Tolmie, A. (2015). Sequential egocentric navigation and reliance on landmarks in Williams syndrome and typical development. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00216

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., … Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, *39*(2), 222–245. doi:10.1037/0012-1649.39.2.222

Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2010). Moving through imagined space: Mentally simulating locomotion during spatial description reading. *Acta Psychologica*, *134*(1), 110–124. doi:10.1016/j.actpsy.2010.01.003

Bryant, D. J., Tversky, B., & Lanca, M. (2001). Retrieving spatial relations from observation and memory. In E. van der Zee & U. Nikanne (Eds.), *Conceptual structure and its interfaces with other modules of representation.* (pp. 116–139). Oxford: Oxford University Press.

Buckley, M. G., Haselgrove, M., & Smith, A. D. (2015). The developmental trajectory of intramaze and extramaze landmark biases in spatial navigation: An unexpected journey. *Developmental Psychology*, *51*(6), 771–791. doi:10.1037/a0039054

Burenhult, N., & Levinson, S. C. (2008). Language and landscape: A cross-linguistic perspective. *Language Sciences*, *30*(2-3), 135–150. doi:10.1016/j.langsci.2006.12.028
Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. doi:10.1016/j.tics.2006.10.005
Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences*, *1124*(1), 77–97. doi:10.1196/annals.1440.002
Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, *35*(4), 625–641. doi:10.1016/s0896-6273(02)00830-9
Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, *16*(2), 130–138. doi:10.1038/nn.3304
Byrne, R. M., & Johnson-Laird, P. (1989). Spatial reasoning. *Journal of Memory and Language*, *28*(5), 564–575. doi:10.1016/0749-596x(89)90013-2
Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375. doi:10.1037/0033-295x.114.2.340
Caduff, D., & Timpf, S. (2007). On the assessment of landmark salience for human navigation. *Cognitive Processing*, *9*(4), 249–267. doi:10.1007/s10339-007-0199-2
Campbell, F. W., & Wurtz, R. H. (1979). Saccadic omission: Why we do not see a gray-out during a saccadic eye movement. *Vision Research*, *18*, 1297–1303.
Castner, H. W., & Eastman, R. J. (1984). Eye-movement parameters and perceived map Complexity—I. *Cartography and Geographic Information Science*, *11*(2), 107–117. doi:10.1559/152304084783914768
Castner, H. W., & Eastman, R. J. (1985). Eye-movement parameters and perceived map Complexity—II. *Cartography and Geographic Information Science*, *12*(1), 29–40. doi:10.1559/152304085783914712
Castner, H. W., & Lywood, D. W. (1978). Eye movement recording / Some approaches to the study of map perception. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *15*(2), 142–150. doi:10.3138/9136-gh77-5330-2v78
Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., & Pietrini, P. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1346–1360. doi:10.1016/j.neubiorev.2008.05.002
Chan, E., Baumann, O., Bellgrove, M. A., & Mattingley, J. B. (2012). From objects to landmarks: The function of visual location information in spatial navigation. *Frontiers in Psychology*, *3*, . doi:10.3389/fpsyg.2012.00304
Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of Vestibular and Proprioceptive inputs to path integration. *Presence: Teleoperators and Virtual Environments*, *7*(2), 168–178. doi:10.1162/105474698565659
Chase, W. G. (1983). Spatial representations of taxi drivers. *The Acquisition of Symbolic Skills*. doi:10.1007/978-1-4613-3724-9_43
Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83–121. doi:10.1016/0010-0277(91)90033-z
Choi, S., & Hattrup, K. (2011). Relative contribution of perception/Cognition and language on spatial Categorization. *Cognitive Science*, *36*(1), 102–129. doi:10.1111/j.1551-6709.2011.01201.x
Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, *14*(2), 241–268. doi:10.1016/s0885-2014(99)00004-0
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: M.I.T. Press.

Clearfield, M. W. (2004). Infants’ enumeration of dynamic displays. *Cognitive Development*, *19*(3), 309–324. doi:10.1016/j.cogdev.2004.03.003

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. doi:10.1016/s0022-5371(69)80069-1

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43. doi:10.1016/0010-0285(75)90003-1

Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics*, *19*(4), 296–302. doi:10.3758/bf03204234

Cooper, L. A., & Shepard, R. N. (1978). Transformations on representations of objects in space. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception, Volume VIII: Perceptual Coding*. New York: Academic Press.

Coppens, L. C., Gootjes, L., & Zwaan, R. A. (2012). Incidental picture exposure affects later reading: Evidence from the N400. *Brain and Language*, *122*(1), 64–69. doi:10.1016/j.bandl.2012.04.006

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A. E., … Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans. *Neuron*, *44*(5), 877–888. doi:10.1016/j.neuron.2004.11.022

Cornell, E. H., Heth, C. D., & Broda, L. S. (1989). Children’s wayfinding: Response to instructions to use environmental landmarks. *Developmental Psychology*, *25*(5), 755–764. doi:10.1037/0012-1649.25.5.755

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, *279*(5355), 1347–1351. doi:10.1126/science.279.5355.1347

Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain and Language*, *112*(3), 202–213. doi:10.1016/j.bandl.2009.06.001

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478. doi:10.1016/j.visres.2006.11.013

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*(2), 121–136. doi:10.1080/15248371003699969

Çöltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17. doi:10.1559/1523040

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology*, *17*(1), 77–95. doi:10.1080/09541440340000529

DeLoache, J. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*(4833), 1556–1557. doi:10.1126/science.2446392

Decety, J. (1996). Do imagined and executed actions share the same neural substrate?. *Cognitive Brain Research*, *3*(2), 87–93. doi:10.1016/0926-6410(95)00033-x

Decety, J., Jeannerod, M., Germain, M., & Pastene, J. (1991). Vegetative response during imagined movement is proportional to mental effort. *Behavioural Brain Research*, *42*(1), 1–5. doi:10.1016/s0166-4328(05)80033-6

Delaney, H. D., & Maxwell, S. E. (1981). On using analysis of covariance in repeated measures designs. *Multivariate Behavioral Research*, *16*(1), 105–123. doi:10.1207/s15327906mbr1601_6

Deloache, J. S., & Marzolf, D. P. (1995). The use of dolls to interview young children: Issues of symbolic representation. *Journal of Experimental Child Psychology*, *60*(1), 155–173. doi:10.1006/jecp.1995.1036

Dement, W., & Kleitman, N. (1957). The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology*, *53*(5), 339–346. doi:10.1037/h0048189

Denis, M. (1996). Strategies for the description of spatial networks: Extending Levelt's paradigm to new situational contexts. *Psychologische Beiträge*, *38*, 465–483.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, *16*, 409–458.

Denis, M., & Zimmer, H. D. (1992). Analog properties of cognitive maps constructed from verbal descriptions. *Psychological Research*, *54*(4), 286–298. doi:10.1007/bf01358266

Denis, M., Pazzaglia, F., Cornoldi, C., & Bertolo, L. (1999). Spatial discourse and navigation: An analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, *13*(2), 145–174. doi:10.1002/(sici)1099-0720(199904)13:2<145::aid-acp550>3.0.co;2-4

Deroualle, D., Borel, L., Devèze, A., & Lopez, C. (2015). Changing perspective: The role of vestibular signals. *Neuropsychologia*, *79*, 175–185. doi:10.1016/j.neuropsychologia.2015.08.022

Deutschländer, A., Bense, S., Stephan, T., Schwaiger, M., Brandt, T., & Dieterich, M. (2002). Sensory system interactions during simultaneous vestibular and visual stimulation in PET. *Human Brain Mapping*, *16*(2), 92–103. doi:10.1002/hbm.10030

Deutschländer, A., Stephan, T., Hüfner, K., Wagner, J., Wiesmann, M., Strupp, M., Brandt, T., Jahn, K. (2009). Vestibular cortex activation during locomotor imagery in the blind. *Annals of the New York Academy of Sciences*, *1164*(1), 350–352. doi:10.1111/j.1749-6632.2009.03863.x

Dhindsa, K., Drobinin, V., King, J., Hall, G. B., Burgess, N., & Becker, S. (2014). Examining the role of the temporo-parietal network in memory, imagery, and viewpoint transformations. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00709

Donoghue, J. P. (2008). Bridging the brain to the world: A perspective on neural interface systems. *Neuron*, *60*(3), 511–521. doi:10.1016/j.neuron.2008.10.037

Ehrich, J. F. (2006). Vygotskian inner speech and the reading process. *Australian Journal of Educational and Developmental Psychology*, *6*, 12–25.

Ekstrom, A. D., Arnold, A. E., & Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: Toward a network-based perspective. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00803

Emmerling, T. C., Zimmermann, J., Sorger, B., Frost, M. A., & Goebel, R. (2016). Decoding the direction of imagined visual motion using 7T ultra-high field fMRI. *NeuroImage*, *125*, 61–73. doi:10.1016/j.neuroimage.2015.10.022

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, *42*(5), 621–636. doi:10.1016/s0042-6989(01)00301-7

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*(4), 659–675. doi:10.1016/j.jecp.2011.06.009

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396. doi:10.1016/j.tics.2008.07.004

Falconer, C. J., & Mast, F. W. (2012). Balancing the mind: Balancing the mind: Experimental psychology: Vol 59, no 6. *Experimental Psychology*. doi:10.1027/1618-3169/a000161

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*(3), 307–317. doi:10.1037/0033-295x.95.3.307

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, *20*(4), 439–462. doi:10.1016/0010-0285(88)90012-6

Farah, M. J., Soso, M. J., & Dasheiff, R. M. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 241–246. doi:10.1037/0096-1523.18.1.241

Farran, E. K., Courbois, Y., Van Herwegen, J., & Blades, M. (2012). How useful are landmarks when learning a route in a virtual environment? Evidence from typical development and Williams syndrome. *Journal of Experimental Child Psychology*, *111*(4), 571–586. doi:10.1016/j.jecp.2011.10.009

Farran, E. K., Courbois, Y., Van Herwegen, J., Cruickshank, A. G., & Blades, M. (2012). Colour as an environmental cue when learning a route in a virtual environment: Typical and atypical development. *Research in Developmental Disabilities*, *33*(3), 900–908. doi:10.1016/j.ridd.2011.11.017

Farrell, M. J., & Robertson, I. H. (1998). Mental rotation and automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(1), 227–233. doi:10.1037/0278-7393.24.1.227

Feldman, A., & Acredolo, L. (1979). The effect of active versus passive exploration on memory for spatial location in children. *Child Development*, *50*(3), 698. doi:10.2307/1128935

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for Allocentric, object-centered, or world-centered reference frames. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00648

Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., Papadakis, N., Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature Neuroscience*, *4*(10), 1043–1048. doi:10.1038/nn733

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fogassi, L., & Luppino, G. (2005). Motor functions of the parietal lobe. *Current Opinion in Neurobiology*, *15*(6), 626–631. doi:10.1016/j.conb.2005.10.015

Foos, P. W. (1980). Constructing cognitive maps from sentences. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 25–38.

Fourtassi, M., Hajjioui, A., Urquizar, C., Rossetti, Y., Rode, G., & Pisella, L. (2013). Iterative fragmentation of cognitive maps in a visual imagery task. *PLoS ONE*, *8*(7), e68560. doi:10.1371/journal.pone.0068560

Frankenstein, J., Mohler, B. J., Bulthoff, H. H., & Meilinger, T. (2011). Is the map in our head oriented north? Psychological Science, 23(2), 120–125. doi:10.1177/0956797611429467

Gagliano, M., Renton, M., Duvdevani, N., Timmins, M., & Mancuso, S. (2012). Out of sight but not out of mind: Alternative means of communication in plants. *PLoS ONE*, *7*(5), e37382. doi:10.1371/journal.pone.0037382

Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by the human brain for spatial perception and memory. *Experimental Brain Research*, *206*(2), 109–120. doi:10.1007/s00221-010-2168-8

Gallistel, C. R. (1990). Representations in animal cognition: An introduction. *Cognition*, *37*(1-2), 1–22. doi:10.1016/0010-0277(90)90016-d

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. doi:10.1016/j.cogbrainres.2004.02.012

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first-person approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1413), 1351–1361. doi:10.1098/rstb.2001.0955

Gardner, M. R., Stent, C., Mohr, C., & Golding, J. F. (2016). Embodied perspective-taking indicated by selective disruption from aberrant self motion. *Psychological Research*. doi:10.1007/s00426-016-0755-4

Gentner, D., & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition*, *2*(02), 261–283. doi:10.1515/langcog.2010.011

Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: Bradford Books.

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In E. Amsel & J. P. Byrnes (Eds.), *Language, Literacy, and Cognitive Development: The Development and Consequences of Symbolic Communication* (pp. 89–125). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2-3), 263–297. doi:10.1016/s0010-0277(98)00002-x

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*(3), 621–634. doi:10.1037/a0022331

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*(2-3), 231–262. doi:10.1016/s0010-0277(97)00047-4

Gomez, A., Rousset, S., & Baciu, M. (2009). Egocentric-updating during navigation facilitates episodic memory retrieval. *Acta Psychologica*, *132*(3), 221–227. doi:10.1016/j.actpsy.2009.07.003

Gottschling, V. (2006). Visual imagery, mental models, and reasoning. In C. Held, M. Knauff, & G. Vosgerau (Eds.), *Mental Models and the Mind - Current Developments in Cognitive Psychology, Neuroscience, and Philosophy of Mind* (pp. 211–235). doi:10.1016/s0166-4115(06)80037-3

Grabherr, L., Cuffel, C., Guyot, J.-P., & Mast, F. W. (2011). Mental transformation abilities in patients with unilateral and bilateral vestibular loss. *Experimental Brain Research*, *209*(2), 205–214. doi:10.1007/s00221-011-2535-0

Gramann, K. (2013). Embodiment of spatial reference frames and individual differences in reference frame Proclivity. *Spatial Cognition & Computation*, *13*(1), 1–25. doi:10.1080/13875868.2011.589038

Gramann, K., Onton, J., Riccobon, D., Mueller, H. J., Bardins, S., & Makeig, S. (2010). Human brain dynamics accompanying use of egocentric and Allocentric reference frames during navigation. *Journal of Cognitive Neuroscience*, *22*(12), 2836–2849. doi:10.1162/jocn.2009.21369

Gramann, K., Müller, H. J., Eick, E.-M., & Schönebeck, B. (2005). Evidence of separable spatial representations in a virtual navigation task. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1199–1223. doi:10.1037/0096-1523.31.6.1199

Gramann, K., Müller, H. J., Schönebeck, B., & Debus, G. (2006). The neural basis of ego- and allocentric reference frames in spatial navigation: Evidence from spatio-temporal coupled current density reconstruction. *Brain Research*, *1118*(1), 116–129. doi:10.1016/j.brainres.2006.08.005

Grèzes, J., & Decety, J. (2000). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*(1), 1–19. doi:10.1002/1097-0193(200101)12:1<1::aid-hbm10>3.0.co;2-v

Gunzelmann, G., Douglass, S., & Khooshabeh, P. (2008). Learning to Orient Using a Map Display: Evidence from Eye Tracking. In C. Holscher (Ed.), *Spatial Cognition 2008: Poster Proceedings* (pp. 85–88). Bremen, Germany: Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition.

Gyselinck, V., Beni, R., Pazzaglia, F., Meneghetti, C., & Mondoloni, A. (2007). Working memory components and imagery instructions in the elaboration of a spatial mental model. *Psychological Research*, *71*(3), 373–382. doi:10.1007/s00426-006-0091-1

Hale, B. D. (1982). The Effects of Internal and External Imagery on Muscular and Ocular Concomitants. *Journal of Sport Psychology*, *4*(4), 379–387.

Harris, D. V., & Robinson, W. J. (1986). The effect of skill level on EMG activity during internal and external imagery. *Journal of Sport Psychology*, *8*(2), 105–111.

Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The well-worn route and the path less traveled. *Neuron*, *37*(5), 877–888. doi:10.1016/s0896-6273(03)00095-3

Haun, D. B. M., Rapold, C. J., Janzen, G., & Levinson, S. C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition*, *119*(1), 70–80. doi:10.1016/j.cognition.2010.12.009

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2015). On forward inferences of fast and slow readers. An eye movement study. *Scientific Reports*, *5*, 8432. doi:10.1038/srep08432

Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, *55*(1), 39–84. doi:10.1016/0010-0277(94)00643-y

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1084–1102. doi:10.1037/0278-7393.18.5.1084

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, *8*(6), 280–285. doi:10.1016/j.tics.2004.04.001

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191. doi:10.1016/j.intell.2003.12.001

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental Animations and external Animations in understanding mechanical systems. *Cognition and Instruction*, *21*(4), 209–249. doi:10.1207/s1532690xci2104_1

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176. doi:10.1016/j.intell.2005.09.005

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425–447. doi:10.1016/s0160-2896(02)00116-2

Herman, J. F., & Siegel, A. W. (1977). The development of spatial representations of large-scale environments. *University of Pittsburg: Learning Research and Development Center*

Heth, C. D., Cornell, E. H., & Alberts, D. M. (1997). Differential use of landmarks by 8- and 12-year-old children during route reversal navigation. *Journal of Environmental Psychology*, *17*(3), 199–213. doi:10.1006/jevp.1997.0057

Hoffman, J. E. (1998). Visual attention and eye movements. In H. Pashler (Ed.), *Attention* (pp. 119–153). London: University College London Press.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721–742. doi:10.1037/0278-7393.25.3.721

Holt, R. R. (1964). Imagery: The return of the ostracized. *American Psychologist*, *19*(4), 254–264. doi:10.1037/h0046316

Holyoak, K. J., & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, *14*(3), 328–352. doi:10.1016/0010-0285(82)90013-5

Homer, M. L., Nurmikko, A. V., Donoghue, J. P., & Hochberg, L. R. (2013). Sensors and Decoding for Intracortical brain computer interfaces. *Annual Review of Biomedical Engineering*, *15*(1), 383–405. doi:10.1146/annurev-bioeng-071910-124640

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology*, *26*(6), 842–847. doi:10.1016/j.cub.2016.01.042
Huey, E. B. (1908). *The psychology and pedagogy of reading.* New York: Macmillan.
Huttenlocher, J., & Presson, C. C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, *4*(2), 277–299. doi:10.1016/0010-0285(73)90015-7
Huttenlocher, J., & Presson, C. C. (1979). The coding and transformation of spatial information. *Cognitive Psychology*, *11*(3), 375–394. doi:10.1016/0010-0285(79)90017-3
Iachini, T. (2011). Mental imagery and embodied cognition: A multimodal approach. *Journal of Mental Imagery*, *35*(3-4),
Irwin, D. E., & Brockmole, J. R. (2000). Mental rotation is suppressed during saccadic eye movements. *Psychonomic Bulletin & Review*, *7*(4), 654–661. doi:10.3758/BF03213003
Iseki, K., Hanakawa, T., Shinozaki, J., Nankaku, M., & Fukuyama, H. (2008). Neural mechanisms involved in mental imagery and observation of gait. *NeuroImage*, *41*(3), 1021–1031. doi:10.1016/j.neuroimage.2008.03.010
Ishikawa, T. (2013). Retention of memory for large-scale spaces. *Memory*, *21*(7), 807–817. doi:10.1080/09658211.2012.758289
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 places. *Cognitive Psychology*, *52*(2), 93–129. doi:10.1016/j.cogpsych.2005.08.003
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489–1506. doi:10.1016/s0042-6989(99)00163-7
Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. doi:10.1038/35058500
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. doi:10.1109/34.730558
Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.
Jackendoff, R. (2012). Language as a source of evidence for theories of spatial representation. *Perception*, *41*(9), 1128–1152. doi:10.1068/p7271
Jackendoff, R. (2015). In Defense of theory. *Cognitive Science*. doi:10.1111/cogs.12324
Jansen-Osmann, P., & Fuchs, P. (2006). Wayfinding behavior and spatial knowledge of adults and children in a virtual environment. *Experimental Psychology*, *53*(3), 171–181. doi:10.1027/1618-3169.53.3.171
Janzen, G., & Jansen, C. (2010). A neural wayfinding mechanism adjusts for ambiguous landmark information. *NeuroImage*, *52*(1), 364–370. doi:10.1016/j.neuroimage.2010.03.083
Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, *7*(6), 673–677. doi:10.1038/nn1257
Janzen, G., Jansen, C., & van Turennout, M. (2007). Memory consolidation of landmarks in good navigators. *Hippocampus*, *18*(1), 40–47. doi:10.1002/hipo.20364
Janzen, G., Wagensveld, B., & van Turennout, M. (2006). Neural representation of navigational relevance is rapidly induced and long lasting. *Cerebral Cortex*, *17*(4), 975–981. doi:10.1093/cercor/bhl008
Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, *33*(11), 1419–1432. doi:10.1016/0028-3932(95)00073-c
Johansson, R. (2013). Tracking the mind's eye: Eye movements during mental imagery and memory retrieval. (Doctoral Thesis, Lund University, Lund, Sweden). In *Lund University Cognitive Studies* 155. Retrieved from http://lup.lub.lu.se/record/4000957

Johansson, R., & Johansson, M. (2013). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*(1), 236–242. doi:10.1177/0956797613498260

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1289–1314. doi:10.1037/a0026585

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason* (1st ed.). Chicago: University of Chicago Press.

Johnson-Laird, P., & Wason, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, *1*(2), 134–148. doi:10.1016/0010-0285(70)90009-5

Johnson-Laird, P. N. (1975). Models of deduction. In R. J. Falmagne (Ed.), *Reasoning: representation and process in children and adults*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (6th ed.). Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250. doi:10.1073/pnas.1012933107

Johnson-Laird, P. N., Legrenzi, P., & Sonino-Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*(3), 395–400. doi:10.1111/j.2044-8295.1972.tb01287.x

Jolicoeur, P., & Kosslyn, S. M. (1985). Is time to scan visual images due to demand characteristics?. *Memory & Cognition*, *13*(4), 320–332. doi:10.3758/bf03202500

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. doi:10.1037/0033-295x.87.4.329

Just, M. A., Newman, S. D., Keller, T. A., McEleney, A., & Carpenter, P. A. (2004). Imagery in sentence comprehension: An fMRI study. NeuroImage, 21(1), 112–124. doi:10.1016/j.neuroimage.2003.08.042

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. doi:10.1111/j.1749-818x.2008.00072.x

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. doi:10.1016/s0749-596x(03)00023-8

Karmiloff-Smith, A. (1992). *Beyond Modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences*, *12*(3), 92–98. doi:10.1016/j.tics.2007.12.004

Kiefer, P., Staub, F., & Raubal, M. (2012). Location-aware mobile eye tracking for the explanation of wayfinding behavior. In J. Gensel, D. Josselin, & D. Vandenbroucke (Eds.), *Proceedings of the AGILE'2012 International Conference on Geographic Information Science*

Kiefer, P., Straub, F., & Raubal, M. (2012). Towards location-aware mobile eye tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications ETRA '12* (pp. 313–316). New York, NY

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin*, *85*(3), 532–554. doi:10.1037/0033-2909.85.3.532

Kim, H. (2015). Encoding and retrieval along the long axis of the hippocampus and their relationships with dorsal attention and default mode networks: The HERNET model. Hippocampus, 25(4), 500–510. doi:10.1002/hipo.22387

Kitchin, R. M. (1994). Cognitive maps: What are they and why study them?. *Journal of Environmental Psychology*, *14*(1), 1–19. doi:10.1016/s0272-4944(05)80194-x

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. *Lecture Notes in Computer Science*. doi:10.1007/3-540-69342-4_1

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual Locomotion. *Psychological Science*, *9*(4), 293–298. doi:10.1111/1467-9280.00058

Klatzky, R. L., Pellegrino, J., McCloskey, B. P., & Lederman, S. J. (1993). Cognitive representations of functional interactions with objects. *Memory & Cognition*, *21*(3), 294–303. doi:10.3758/bf03208262

Klein, J. P., & Moritz, S. (2014). On the relevance of mental imagery beyond stress-related psychiatric disorders. *Frontiers in Psychiatry*, 5. doi:10.3389/fpsyt.2014.00077

Knauff, M., & Johnson-Laird, P. N. (2000). Visual and spatial representations in spatial reasoning. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 759–765). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *The Quarterly Journal of Experimental Psychology*, *59*(1), 161–177. doi:10.1080/17470210500149992

Knauff, M., Fangmeier, T., Ruff, C. C., & Johnson-Laird, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, *15*(4), 559–573. doi:10.1162/089892903321662949

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics*, *14*(1), 90–94. doi:10.3758/bf03198621

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, *7*(3), 341–370. doi:10.1016/0010-0285(75)90015-8

Kosslyn, S. (1988). Aspects of a cognitive neuroscience of mental imagery. *Science*, *240*(4859), 1621–1626. doi:10.1126/science.3289115

Kosslyn, S. M. M. (1994). *Image and brain: The resolution of the imagery debate* (4th ed.). Cambridge, MA: MIT Press.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3-4), 333–347. doi:10.1080/02643290442000130

Kosslyn, S. M., & Shwartz, S. P. (1977). A simulation of visual Imagery. *Cognitive Science*, *1*(3), 265–295. doi:10.1207/s15516709cog0103_2

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 47–60. doi:10.1037/0096-1523.4.1.47

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., … Buonanno, F. S. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, *5*(3), 263–287. doi:10.1162/jocn.1993.5.3.263

Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology*, *35*(2), 151–161. doi:10.1111/1469-8986.3520151

Kosslyn, S. M., Holyoak, K. J., & Huffman, C. S. (1976). A processing approach to the dual coding hypothesis. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(3), 223–233. doi:10.1037/0278-7393.2.3.223

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A Positron emission tomography study. *NeuroImage*, *6*(4), 320–334. doi:10.1006/nimg.1997.0295

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery, Vol. 39*. Oxford: Oxford University Press.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), 745–756. doi:10.3758/bf03200477

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*(3), 397–417. doi:10.1002/acp.1192

Kraeutner, S., Gionfriddo, A., Bardouille, T., & Boe, S. (2014). Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations. *Brain Research*, *1588*, 81–91. doi:10.1016/j.brainres.2014.09.001

Kuipers, B. (1978). Modeling spatial knowledge. *Cognitive Science*, *2*(2), 129–153. doi:10.1016/s0364-0213(78)80003-2

Kuipers, B. (1982). The 'map in the head' metaphor. *Environment and Behavior*, *14*(2), 202–220. doi:10.1177/0013916584142005

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30, 000 English words. *Behavior Research Methods*, *44*(4), 978–990. doi:10.3758/s13428-012-0210-4

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi:10.1126/science.7350657

La Fougère, C., Zwergal, A., Rominger, A., Förster, S., Fesl, G., Dieterich, M., Brandt, T., Strupp, M., Bartenstein, P., Jahn, K. (2010). Real versus imagined locomotion: A [18F]-FDG PET-fMRI comparison. *NeuroImage*, *50*(4), 1589–1598. doi:10.1016/j.neuroimage.2009.12.060

Laing, E., & Jarrold, C. (2007). Comprehension of spatial language in Williams syndrome: Evidence for impaired spatial representation of verbal descriptions. *Clinical Linguistics & Phonetics*, *21*(9), 689–704. doi:10.1080/02699200701541441

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lambert, W. E., & Pavio, A. (1956). The influence of noun-adjective order on learning. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *10*(1), 9–12. doi:10.1037/h0083277

Lambrey, S., Doeller, C., Berthoz, A., & Burgess, N. (2011). Imagining being somewhere else: Neural basis of changing perspective in space. *Cerebral Cortex*, *22*(1), 166–174. doi:10.1093/cercor/bhr101

Landau, B., & Hoffman, J. E. (2005). Parallels between spatial cognition and spatial language: Evidence from Williams syndrome. *Journal of Memory and Language*, *53*(2), 163–185. doi:10.1016/j.jml.2004.05.007

Landau, B., & Jackendoff, R. (1993). 'What' and 'where' in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*(02), 217. doi:10.1017/s0140525x00029733

Landau, B., & Zukowski, A. (2003). Objects, motions, and paths: Spatial language in children with Williams Syndrome. *Developmental Neuropsychology*, *23*(1-2), 105–137. doi:10.1080/87565641.2003.9651889

Langacker, R. W. (1987). *Foundations of cognitive grammar: Volume I: Theoretical prerequisites*. United States: Stanford University Press.

Langacker, R. W. (1991). *Foundations of cognitive grammar: Volume II: Descriptive application*. United States: Stanford University Press.

Lee, E. C., Woo, J. C., Kim, J. H., Whang, M., & Park, K. R. (2010). A brain–computer interface method combined with eye tracking for 3D interaction. *Journal of Neuroscience Methods*, *190*(2), 289–298. doi:10.1016/j.jneumeth.2010.05.008

Levine, D. N., Warach, J., & Farah, M. (1985). Two visual systems in mental imagery: Dissociation of 'what' and 'where' in imagery disorders due to bilateral posterior cerebral lesions. *Neurology*, *35*(7), 1010–1010. doi:10.1212/wnl.35.7.1010

Levinson, S. C., & Schmitt, B. (1993). Animals in a row. In *Cognition and Space Kit Version 1.0. Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics* (pp. 65–69). Nijmegen

Levinson, S. C., Kita, S., Haun, D. B. ., & Rasch, B. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition*, *84*(2), 155–188. doi:10.1016/s0010-0277(02)00045-8

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge, U.K.: Cambridge University Press.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265–294. doi:10.1016/s0010-0277(02)00009-4

Lingwood, J., Blades, M., Farran, E. K., Courbois, Y., & Matthews, D. (2015). Encouraging 5-year olds to attend to landmarks: A way to improve children's wayfinding strategies in a virtual environment. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00174

Livingstone-Lee, S. A., Murchison, S., Zeman, P. M., Gandhi, M., van Gerven, D., Stewart, L., Livingston, N. J., Skelton, R. W. (2011). Simple gaze analysis and special design of a virtual Morris water maze provides a new method for differentiating egocentric and allocentric navigational strategy choice. *Behavioural Brain Research*, *225*(1), 117–125. doi:10.1016/j.bbr.2011.07.005

Lobel, E., Kleine, J. F., Leroy-Willig, A., van de Moortele, P.-F., Le Bihan, D., Grusser, O.-J., & Berthoz, A. (1999). Cortical areas activated by bilateral galvanic Vestibular stimulation. *Annals of the New York Academy of Sciences*, *871*, 313–323. doi:10.1111/j.1749-6632.1999.tb09194.x

Lobel, E., Kleine, J. F., Bihan, D. L., Leroy-Willig, A., & Berthoz, A. (1998). Functional MRI of galvanic Vestibular stimulation. *Article*, *80*(5), 2699–2709. Retrieved from http://jn.physiology.org/content/80/5/2699.long

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far Mappings. *Journal of Cognition and Development*, *2*(2), 189–219. doi:10.1207/s15327647jcd0202_4

Loomis, J. M., Klatzky, R. L., & Giudice, N. A. (2012). Representing 3D space in working memory: Spatial images from vision, hearing, touch, and language. *Multisensory Imagery*. doi:10.1007/978-1-4614-5879-1_8

Loomis, J. M., da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 906–921. doi:10.1037/0096-1523.18.4.906

Lucy, J. A. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lukács, Á., Pléh, C., & Racsmány, M. (2007). Spatial language in Williams Syndrome: Evidence for a special interaction?. *Journal of Child Language*, *34*(02), 311. doi:10.1017/s0305000906007926

Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence?. *Brain and Cognition*, *72*(1), 101–113. doi:10.1016/j.bandc.2009.08.005

Lynch, K. (1960). *The image of the city*. Cambridge, MA: M.I.T. Press.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. doi:10.1016/j.tics.2004.01.003

Mandler, J. M., & Pagán Cánovas, C. (2014). On defining image schemas. *Language and Cognition*, 6(04), 510–532. doi:10.1017/langcog.2014.14

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: Coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, *17*(11), 1598–1606. doi:10.1038/nn.3834

Mark, D. M., Freksa, C., Hirtle, S. C., Lloyd, R., & Tversky, B. (1999). Cognitive models of geographical space. *International Journal of Geographical Information Science*, *13*(8), 747–774. doi:10.1080/136588199241003

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. doi:10.1111/j.2044-8295.1973.tb01322.x

Marozzi, E., & Jeffery, K. J. (2012). Place, space and memory cells. *Current Biology*, *22*(22), R939–R942. doi:10.1016/j.cub.2012.10.022

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman and Co.

Mast, F. W., Merfeld, D. M., & Kosslyn, S. M. (2006). Visual mental imagery during caloric vestibular stimulation. *Neuropsychologia*, *44*(1), 101–109. doi:10.1016/j.neuropsychologia.2005.04.005

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. doi:10.1037/h0037368

Mayer, A., & Orth, J. (1901). Zur qualitativen Untersuchung der Association. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*. Retrieved from http://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/permanent/vlp/lit25188/index.meta

McClelland, T., & Bayne, T. (2016). Concepts, contents, and consciousness. *Neuroscience of Consciousness*, *2016*(1), niv012. doi:10.1093/nc/niv012

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology*, *46*(3), 229–259. doi:10.1016/s0010-0285(02)00514-5

McNamara, T. P., & Diwadkar, V. A. (1997). Symmetry and asymmetry of human spatial memory. *Cognitive Psychology*, *34*(2), 160–190. doi:10.1006/cogp.1997.0669

McNamara, T. P., & Shelton, A. L. (2003). Cognitive maps and the hippocampus. *Trends in Cognitive Sciences*, *7*(8), 333–335. doi:10.1016/s1364-6613(03)00167-0

McNamara, T. P., Rump, B., & Werner, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin & Review*, *10*(3), 589–595. doi:10.3758/bf03196519

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, *14*(11), 1256–1265. doi:10.1093/cercor/bhh087

Mellet, E., Petit, L., Mazoyer, B., Denis, M., & Tzourio, N. (1998). Reopening the mental imagery debate: Lessons from functional anatomy. *NeuroImage*, *8*(2), 129–139. doi:10.1006/nimg.1998.0355

Meneghetti, C., De Beni, R., Gyselinck, V., & Pazzaglia, F. (2011). Working memory involvement in spatial text processing: What advantages are gained from extended learning and visuo-spatial strategies?. *British Journal of Psychology*, *102*(3), 499–518. doi:10.1111/j.2044-8295.2010.02007.x

Meneghetti, C., De Beni, R., Gyselinck, V., & Pazzaglia, F. (2013). The joint role of spatial ability and imagery strategy in sustaining the learning of spatial descriptions under spatial interference. *Learning and Individual Differences*, *24*, 32–41. doi:10.1016/j.lindif.2012.12.021

Meneghetti, C., Gyselinck, V., Pazzaglia, F., & De Beni, R. (2009). Individual differences in spatial text processing: High spatial ability can compensate for spatial working memory interference. *Learning and Individual Differences*, *19*(4), 577–589. doi:10.1016/j.lindif.2009.07.007

Mervis, C. B., & John, A. E. (2008). Vocabulary abilities of children with Williams Syndrome: Strengths, weaknesses, and relation to visuospatial construction ability. *Journal of Speech Language and Hearing Research*, *51*(4), 967. doi:10.1044/1092-4388(2008/071)

Mervis, C. B., Robinson, B. F., & Pani, J. R. (1999). Visuospatial construction. *The American Journal of Human Genetics*, *65*(5), 1222–1229. doi:10.1086/302633

Milgram, S., & Jodelet, D. (1976). Psychological maps of Paris. In H. Proshansky, W. Ittelson, & L. Rivlin (Eds.), *Environmental psychology* (2nd ed.) (pp. 104–124). New York: Holt, Reinhart and Winston.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortext function. *Annual Review of Neuroscience*, *24*(1), 167–202. doi:10.1146/annurev.neuro.24.1.167

Miller, K. J., Schalk, G., Hermes, D., Ojemann, J. G., & Rao, R. P. N. (2016). Spontaneous Decoding of the timing and content of human object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. *PLOS Computational Biology*, *12*(1), e1004660. doi:10.1371/journal.pcbi.1004660

Mintzer, M. Z., & Snodgrass, J. G. (1999). The picture superiority effect: Support for the distinctiveness model. *The American Journal of Psychology*, *112*(1), 113. doi:10.2307/1423627

Mirpour, K., Arcizet, F., Ong, W. S., & Bisley, J. W. (2009). Been there, seen that: A neural mechanism for performing efficient visual search. *Journal of Neurophysiology*, *102*(6), 3481–3491. doi:10.1152/jn.00688.2009

Money, J., Alexander, D., & Walker, H. T. (1965). *Manual: A standardized road-map test of direction sense*. Baltimore: Johns Hopkins Press.

Monson, C. K. (1993). A comment to suspend the introspection controversy. In R. T. Hurlburt (Ed.), *Sampling Inner Experience in Disturbed Affect* (pp. 15–26). doi:10.1007/978-1-4899-1222-0_2

Montello, D. R. (1993). Scale and multiple psychologies of space. *Spatial Information Theory A Theoretical Basis for GIS*. doi:10.1007/3-540-57207-4_21

Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In M. J. Egenhofer & R. G. Golledge (Eds.), *Spatial and temporal reasoning in geographic information systems.* (pp. 143–154). New York: Oxford University Press.

Moraleda, E., Broglio, C., Rodríguez, F., & Gómez, A. (2013). Development of different spatial frames of reference for orientation in small-scale environments. *Psicothema*, *25*(4), 468–475. doi:10.7334/psicothema2013.29

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. doi:10.1016/s0022-5371(77)80016-9

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: Mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273–1280. doi:10.1098/rstb.2008.0314

Mullally, S. L., & Maguire, E. A. (2011). A new role for the Parahippocampal cortex in representing space. *Journal of Neuroscience*, *31*(20), 7441–7449. doi:10.1523/jneurosci.0267-11.2011

Mullally, S. L., & Maguire, E. A. (2013). Exploring the role of space-defining objects in constructing and maintaining imagined scenes. *Brain and Cognition*, *82*(1), 100–107. doi:10.1016/j.bandc.2013.02.013

Mulligan, N. W., & Osborn, K. (2009). The modality-match effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 564–571. doi:10.1037/a0014524

Nagin, D., & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and Hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, *70*(5), 1181–1196. doi:10.1111/1467-8624.00086

Nardini, M., Burgess, N., Breckenridge, K., & Atkinson, J. (2006). Differential developmental trajectories for egocentric, environmental and intrinsic frames of reference in spatial memory. *Cognition*, *101*(1), 153–172. doi:10.1016/j.cognition.2005.09.005

Newcombe, N. S., & Huttenlocher, J. (2003). *Making space: The development of spatial representation and reasoning*. United States: Bradford Books.

Newcombe, N., Huttenlocher, J., Drummey, A. B., & Wiley, J. G. (1998). The development of spatial location coding: Place learning and dead reckoning in the second and third years. *Cognitive Development*, *13*(2), 185–200. doi:10.1016/s0885-2014(98)90038-7

Noordzij, M. L., Zuidhoek, S., & Postma, A. (2006). The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition*, *100*(2), 321–342. doi:10.1016/j.cognition.2005.05.006

Nori, R., & Giusberti, F. (2006). Predicting cognitive styles from spatial abilities. *The American Journal of Psychology*, 119(1), 67. doi:10.2307/20445319

Oertel, V., Rotarska-Jagiela, A., van de Ven, V., Haenschel, C., Grube, M., Stangier, U., … Linden, D. E. J. (2009). Mental imagery vividness as a trait marker across the schizophrenia spectrum. *Psychiatry Research*, 167(1-2), 1–11. doi:10.1016/j.psychres.2007.12.008

O'Keefe, J., & Burgess, N. (1996). Geometric determinants of the place fields of hippocampal neurons. *Nature*, *381*(6581), 425–428. doi:10.1038/381425a0

Ooms, K., De Maeyer, P., Fack, V., Assche, E. V., & Witlox, F. (2012). Investigating the effectiveness of an efficient label placement method using eye movement data. *The Cartographic Journal*, *49*(3), 234–246. doi:10.1179/1743277412y.0000000010

Otten, L. J., Sveen, J., & Quayle, A. H. (2007). Distinct patterns of neural activity during memory formation of nonwords versus words. *Journal of Cognitive Neuroscience*, *19*(11), 1776–1789. doi:10.1162/jocn.2007.19.11.1776

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 32–38. doi:10.1016/s0022-5371(65)80064-0

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, Affective and social Neurosciences. *Frontiers in Human Neuroscience*, *9*, . doi:10.3389/fnhum.2015.00660

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1179–1187. doi:10.1016/j.neubiorev.2012.01.008

Pasqualotto, A., Spiller, M. J., Jansari, A. S., & Proulx, M. J. (2013). Visual experience facilitates allocentric spatial representation. *Behavioural Brain Research*, *236*, 175–179. doi:10.1016/j.bbr.2012.08.042

Paterson, K. B., McGowan, V. A., White, S. J., Malik, S., Abedipour, L., & Jordan, T. R. (2014). Reading direction and the central perceptual span in Urdu and English. *PLoS ONE*, *9*(2), e88358. doi:10.1371/journal.pone.0088358

Pazzaglia, F., De Beni, R., & Meneghetti, C. (2006). The effects of verbal and spatial interference in the encoding and retrieval of spatial and nonspatial texts. *Psychological Research*, *71*(4), 484–494. doi:10.1007/s00426-006-0045-7

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*(6), 1108–1114. doi:10.1080/17470210802633255

Pederson, E., & Schmitt, B. (1993). Eric's maze task. In *Cognition and Space Kit Version 1.0. Cognitive Anthropology Research Group at the Max Planck Institute for Psycholinguistics* (pp. 73–76). Nijmegen.

Peebles, D., Davies, C., & Mora, R. (2007). Effects of Geometry, Landmarks and Orientation Strategies in the 'Drop-Off' Orientation Task. In S. Winter, M. Ducham, L. Kulik, & B. Kuipers (Eds.), *COSIT 2007 LNCS 4736*. Melbourne, Australiia.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. doi:10.3389/neuro.11.010.2008

Pelekanos, V., & Moutoussis, K. (2011). The effect of language on visual contrast sensitivity. *Perception*, *40*(12), 1402–1412. doi:10.1068/p7010

Penfield, W. (1958). Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proceedings of the National Academy of Sciences*, *44*(2), 51–66. doi:10.1073/pnas.44.2.51

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language*, *24*(5), 503–518. doi:10.1016/0749-596x(85)90042-7

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2013). Computational grounded Cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00612

Phillips, C., Jarrold, C., Baddeley, A., Grant, J., & Karmiloff-Smith, A. (2004). Comprehension of spatial language terms in Williams Syndrome: Evidence for an interaction between domains of strength and weakness. *Cortex*, *40*(1), 85–101. doi:10.1016/s0010-9452(08)70922-5

Piaget, J., & Inhelder, B. (1967). *The child's conception of space*. New York: W. W. Norton & Company, Inc.

Piccardi, L., De Luca, M., Nori, R., Palermo, L., Iachini, F., & Guariglia, C. (2016). Navigational style influences eye movement pattern during exploration and learning of an environmental map. *Frontiers in Behavioral Neuroscience*, 10. doi:10.3389/fnbeh.2016.00140

Picucci, L., Gyselinck, V., Piolino, P., Nicolas, S., & Bosco, A. (2013). Spatial mental models: The interaction of presentation format, task requirements and availability of working memory components. *Applied Cognitive Psychology*, *27*(3), 314–327. doi:10.1002/acp.2909

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *1*, 3–25.

Presson, C. C., & Montello, D. R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception*, *23*(12), 1447–1455. doi:10.1068/p231447

Ptak, R. (2012). The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *The Neuroscientist*, *18*(5), 502–515. doi:10.1177/1073858411409051

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*(1), 1–24. doi:10.1037/h0034650

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, *88*(1), 16–45. doi:10.1037/0033-295x.88.1.16

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(02), . doi:10.1017/s0140525x02000043

Pylyshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the brain?. *Trends in Cognitive Sciences*, *7*(3), 113–118. doi:10.1016/s1364-6613(03)00003-2

Pylyshyn, Z. W. (2003). Explaining mental imagery: Now you see it, now you don't. *Trends in Cognitive Sciences*, *7*(3), 111–112. doi:10.1016/s1364-6613(03)00004-4

Péruch, P., Lopez, C., Redon-Zouiteni, C., Escoffier, G., Zeitoun, A., Sanjuan, M., Devèze, A., Magnan, J., Borel, L. (2011). Vestibular information is necessary for maintaining metric properties of representational space: Evidence from mental imagery. *Neuropsychologia*, *49*(11), 3136–3144. doi:10.1016/j.neuropsychologia.2011.07.026

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. doi:10.1007/s00426-008-0173-3

Rajangam, S., Tseng, P.-H., Yin, A., Lehew, G., Schwarz, D., Lebedev, M. A., & Nicolelis, M. A. L. (2016). Wireless cortical brain-machine interface for whole-body navigation in primates. *Scientific Reports*, *6*, 22170. doi:10.1038/srep22170

Rao, R., & Ballard, D. (1998). Development of localized oriented receptive fields by learning a translation-invariant code for natural images. *Network: Computation in Neural Systems*, *9*(2), 219–234. doi:10.1088/0954-898x/9/2/005

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi:10.1080/17470210902816461

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. doi:10.1037/0033-295x.105.1.125

Reuderink, B., Farquhar, J., Poel, M., & Nijholt, A. (2011). A subject-independent brain-computer interface based on smoothed, second-order baselining. *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. doi:10.1109/iembs.2011.6091139

Ridderinkhof, R. K., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447. doi:10.1126/science.1100301

Riecke, B. E., Cunningham, D. W., & Bülthoff, H. H. (2006). Spatial updating in virtual reality: The sufficiency of visual information. *Psychological Research*, *71*(3), 298–313. doi:10.1007/s00426-006-0085-z

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1157–1165. doi:10.1037/0278-7393.15.6.1157

Rowen, R. B., & Hardwick, D. A. (1983). Effects of landmark Salience and direction of travel on young children's memory for spatial location. *The Journal of Psychology*, *113*(2), 271–276. doi:10.1080/00223980.1983.9923586

Rowland, D. C., Yanovich, Y., & Kentros, C. G. (2011). A stable hippocampal representation of a space requires its direct experience. *Proceedings of the National Academy of Sciences*, *108*(35), 14654–14658. doi:10.1073/pnas.1105445108

Ruggiero, G., Errico, O. D', & Iachini, T. (2015). Development of egocentric and allocentric spatial representations from childhood to elderly age. *Psychological Research*, *80*(2), 259–272. doi:10.1007/s00426-015-0658-9

Ruggiero, G., Iachini, T., Ruotolo, F., & Senese, V. P. (2010). Spatial memory: The role of egocentric and allocentric frames of reference. In J. B. Thomas (Ed.), *Spatial memory: Visuospatial processes, cognitive performance and developmental effective* (pp. 51–75). United States: Nova Science Publishers.

Sack, A. T., van de Ven, V. G., Etschenberg, S., Schatz, D., & Linden, D. E. J. (2005). Enhanced vividness of mental imagery as a trait marker of schizophrenia? *Schizophrenia Bulletin*, 31(1), 97–104. doi:10.1093/schbul/sbi011

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 516–528. doi:10.1037/0278-7393.6.5.516

Schuett, S., Heywood, C. A., Kentridge, R. W., & Zihl, J. (2008). The significance of visual information processing in reading: Insights from hemianopic dyslexia. *Neuropsychologia*, *46*(10), 2445–2462. doi:10.1016/j.neuropsychologia.2008.04.016

Scolari, M., Seidl-Rathkopf, K. N., & Kastner, S. (2015). Functions of the human frontoparietal attention network: Evidence from neuroimaging. Current Opinion in Behavioral Sciences, 1, 32–39. doi:10.1016/j.cobeha.2014.08.003

Selen, L. P. J., & Medendorp, W. P. (2011). Saccadic updating of object orientation for grasping movements. *Vision Research*, 51(8), 898–907. doi:10.1016/j.visres.2011.01.004

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27. doi:10.1037/0096-3445.125.1.4

Sharma, N., & Baron, J. (2013). Does motor imagery share neural networks with executed movement: A multivariate fMRI analysis. *Frontiers in Human Neuroscience*, *7*. doi:10.3389/fnhum.2013.00564

Shepard, R. N., & Metzler, J. (1971). Mental rotation of Three-Dimensional objects. *Science*, *171*(3972), 701–703. doi:10.1126/science.171.3972.701

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior Volume 10*. doi:10.1016/s0065-2407(08)60007-5

Sima, J. F., Schultheis, H., & Barkowsky, T. (2013). Differences between spatial and visual mental representations. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00240

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, *25*(3), 321–332. doi:10.3758/bf03211288

Sinha, C., & López, K. J. de (2000). Language, culture, and the embodiment of spatial cognition. *Cognitive Linguistics*, *11*(1-2). doi:10.1515/cogl.2001.008

Spiers, H. J., & Barry, C. (2015). Neural systems supporting navigation. Current Opinion in Behavioral Sciences, 1, 47–55. doi:10.1016/j.cobeha.2014.08.005

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. doi:10.1038/nn.3574

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156. doi:10.1111/1467-9280.00326

Staresina, B. P., Gray, J. C., & Davachi, L. (2008). Event congruency enhances episodic memory encoding through semantic elaboration and relational binding. *Cerebral Cortex*, *19*(5), 1198–1207. doi:10.1093/cercor/bhn165

Strasser, A. (2010). A functional view towards mental representations. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 15–25). doi:10.1007/978-1-4419-5662-0_2

Sugimoto, M., & Kusumi, T. (2014). The effect of text continuity on spatial representation: Route versus survey perspective. *Cognitive Processing*, *15*(1), 65–75. doi:10.1007/s10339-013-0582-0

Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., & Galati, G. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioural Brain Research*, *242*, 62–75. doi:10.1016/j.bbr.2012.12.031

Szameitat, A. J., Shen, S., & Sterr, A. (2007). Motor imagery of complex everyday movements. An fMRI study. *NeuroImage*, *34*(2), 702–713. doi:10.1016/j.neuroimage.2006.09.033

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate statistics* (6th ed.). Boston: Pearson Education.

Talmy, L. (2000). *Toward a cognitive semantics: Volume 1: Concept structuring systems and volume 2: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Taube, W., Mouthon, M., Leukel, C., Hoogewoud, H.-M., Annoni, J.-M., & Keller, M. (2015). Brain activity during observation and motor imagery of different balance tasks: An fMRI study. *Cortex*, *64*, 102–114. doi:10.1016/j.cortex.2014.09.022

Taylor, H. A., & Tversky, B. (1992). Spatial mental models derived from survey and route descriptions. *Journal of Memory and Language*, *31*(2), 261–292. doi:10.1016/0749-596x(92)90014-o

Tenbrink, T. (2015). Cognitive discourse analysis: Accessing cognitive representations and processes through language data. *Language and Cognition*, 7(01), 98–137. doi:10.1017/langcog.2014.19

Tenbrink, T., & Salwiczek, L. H. (2016). Orientation and metacognition in virtual space. *Journal of Experimental Psychology: Human Perception and Performance*, 42(5), 683–705. doi:10.1037/xhp0000178

Thomas, N. J. T. (2016). Mental imagery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 ed.). Retrieved from http://plato.stanford.edu/archives/spr2016/entries/mental-imagery/

Thomas, M. S. C., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences*, *25*(6), 727-750. doi:10.1017/s0140525x02000134

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental Trajectories to understand developmental disorders. *Journal of Speech Language and Hearing Research*, *52*(2), 336. doi:10.1044/1092-4388(2009/07-0144)

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. doi:10.1037/h0061626

Tom, A. C., & Denis, M. (2003). Referring to Landmark or Street Information in Route Directions: What Difference Does It Make? In W. Kuhn, M. F. Worboys, & S. Timpf (Eds.), *COSIT 2003 LNCS 2825*.

Tom, A. C., & Denis, M. (2004). Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology*, *18*(9), 1213–1230. doi:10.1002/acp.1045

Tom, A. C., & Tversky, B. (2012). Remembering routes: Streets and landmarks. *Applied Cognitive Psychology*, *26*(2), 182–193. doi:10.1002/acp.1805

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. doi:10.1037/0033-295x.113.4.766

Truccolo, W., Friehs, G. M., Donoghue, J. P., & Hochberg, L. R. (2008). Primary motor cortex tuning to intended movement Kinematics in humans with Tetraplegia. *Journal of Neuroscience*, *28*(5), 1163–1178. doi:10.1523/jneurosci.4415-07.2008

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. doi:10.1146/annurev.psych.53.100901.135114

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, *13*(3), 407–433. doi:10.1016/0010-0285(81)90016-5

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, *23*(2), 131–138. doi:10.1016/0016-7185(92)90011-r

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory A Theoretical Basis for GIS*. COSIT'93, Marciana Marina, Elba Island, Italy. doi:10.1007/3-540-57207-4_2

Tversky, B. (2000). Levels and structure of spatial knowledge. In R. Kitchin & S. Freundschuh (Eds.), *Cognitive mapping. Past, present and future* (pp. 24–42). London: Routledge.

Török, Á., Nguyen, T. P., Kolozsvári, O., Buchanan, R. J., & Nadasdy, Z. (2014). Reference frames in virtual spatial navigation are viewpoint dependent. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00646

Uchiyama, H., Mitsuishi, K., & Ohno, H. (2009). Random Walker Test: A computerized alternative to the road-map test. Behavior Research Methods, 41(4), 1242–1253. doi:10.3758/brm.41.4.1242

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. doi:10.1162/089892999563724

Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Ekert, J., Wegman, J., & Janzen, G. (2015). Neurocognitive development of memory for landmarks. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.00224

Vandeberg, L., Eerland, A., & Zwaan, R. A. (2012). Out of mind, out of sight: Language affects perceptual vividness in memory. *PLoS ONE*, *7*(4), e36154. doi:10.1371/journal.pone.0036154

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. doi:10.2466/pms.1978.47.2.599

Van der Ham, I. J. M., Faber, A. M. E., Venselaar, M., van Kreveld, M. J., & Löffler, M. (2015). Ecological validity of virtual environments to assess human navigation ability. *Frontiers in Psychology*, *6*, 637. doi:10.3389/fpsyg.2015.00637

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do?. *Nature Reviews Neuroscience*, *10*(11), 792–802. doi:10.1038/nrn2733

Vass, L. K., Copara, M. S., Seyal, M., Shahlaie, K., Farias, S. T., Shen, P. Y., & Ekstrom, A. D. (2016). Oscillations go the distance: Low-frequency human Hippocampal Oscillations code spatial distance in the absence of sensory cues during Teleportation. *Neuron*, *89*(6), 1180–1186. doi:10.1016/j.neuron.2016.01.045

Viaene, P., Ooms, K., Vansteenkiste, P., Lenoir, M., & De Maeyer, P. (2014). The Use of Eye Tracking in Search of Indoor Landmarks. In P. Kiefer, I. Giannopoulos, M. Raubal, & A. Krüger (Eds.), *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research*. Vienna, Austria:

Vicari, S., Brizzolara, D., Carlesimo, G. A., Pezzini, G., & Volterra, V. (1996). Memory abilities in children with Williams Syndrome. *Cortex*, *32*(3), 503–514. doi:10.1016/s0010-9452(96)80007-4

Vingerhoets, G., Lannoo, E., & Bauwens, S. (1996). Analysis of the money road-map test performance in normal and brain-damaged subjects. Archives of Clinical Neuropsychology, 11(1), 1–9. doi:10.1016/0887-6177(95)00055-0

Voermans, N. C., Petersson, K. M., Daudey, L., Weber, B., van Spaendonck, K. P., Kremer, H. P. H., & Fernández, G. (2004). Interaction between the human Hippocampus and the Caudate nucleus during route recognition. *Neuron*, *43*(3), 427–435. doi:10.1016/j.neuron.2004.07.009

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences*, *7*(1), 38–42. doi:10.1016/s1364-6613(02)00003-7

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, *16*(5), 817–827. doi:10.1162/089892904970799

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: M.I.T. Press.

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception*, *1*(2), 33–68. doi:10.1068/i0389

Waller, D., & Lippa, Y. (2007). Landmarks as beacons and associative cues: Their role in route learning. *Memory & Cognition*, *35*(5), 910–924. doi:10.3758/bf03193465

Wang, Y., & Morgan, W. P. (1992). The effect of imagery perspectives on the psychophysiological responses to imagined exercise. *Behavioural Brain Research*, *52*(2), 167–174. doi:10.1016/s0166-4328(05)80227-x

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–250. doi:10.1016/s0010-0277(00)00105-0

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

Wassenburg, S. I., & Zwaan, R. A. (2010). Readers routinely represent implied object rotation: The role of visual experience. *The Quarterly Journal of Experimental Psychology*, *63*(9), 1665–1670. doi:10.1080/17470218.2010.502579

Wegman, J., & Janzen, G. (2011). Neural Encoding of objects relevant for navigation and resting state correlations with navigational ability. *Journal of Cognitive Neuroscience*, *23*(12), 3841–3854. doi:10.1162/jocn_a_00081

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, *15*(4), 269–280. doi:10.3758/bf03197030

Wenzel, R., Bartenstein, P., Dieterich, M., Danek, A., Weindl, A., Minoshima, S., Ziegler, S., Schwaiger, M., Brandt, T. (1996). Deactivation of human visual cortex during involuntary ocular oscillations. *Brain*, *119*(1), 101–110. doi:10.1093/brain/119.1.101

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: M.I.T. Press.

Wiener, J. M., Berthoz, A., & Wolbers, T. (2011). Dissociable cognitive mechanisms underlying human path integration. Experimental Brain Research, 208(1), 61–71. doi:10.1007/s00221-010-2460-7

Wiener, J. M., de Condappa, O., & Hölscher, C. (2011). Do you have to look where you go? Gaze behaviour during spatial decision making. In *CogSci 2011: Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston, MA, USA: Cognitive Science Society.

Wiener, J. M., Hölscher, C., Büchner, S., & Konieczny, L. (2011). Gaze behaviour during space perception and spatial decision making. *Psychological Research*, *76*(6), 713–729. doi:10.1007/s00426-011-0397-5

Wilson, A. D., & Golonka, S. (2013). Embodied Cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58. doi:10.3389/fpsyg.2013.00058

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (1971). *A manual for the embedded figures tests*. Palo Alto, California: Consulting Psychologists Press.

Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities?. *Trends in Cognitive Sciences*, *14*(3), 138–146. doi:10.1016/j.tics.2010.01.001

Wolpert, D. M. (2007). Probabilistic models in human sensorimotor control. Human Movement Science, 26(4), 511–524. doi:10.1016/j.humov.2007.05.005

Wraga, M. (2003). Thinking outside the body: An advantage for spatial updating during imagined versus physical self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 993–1005. doi:10.1037/0278-7393.29.5.993

Wundt, W. M. (2013). *Elements of folk psychology - outline of A psychological history of the development of mankind* [Original work published 1916]. Retrieved from http://www.forgottenbooks.com/books/Elements_of_Folk_Psychology_1000099041

Yaden, D. B. (1984). Inner speech, oral language, and reading: Huey and Vygotsky revisited. *Reading Psychology*, *5*(1-2), 155–166. doi:10.1080/0270271840050118

Yartsev, M. M., & Ulanovsky, N. (2013). Representation of Three-Dimensional space in the Hippocampus of flying bats. *Science*, *340*(6130), 367–372. doi:10.1126/science.1235338

Zeman, A., Dewar, M., & Della Sala, S. (2015). Lives without imagery – congenital aphantasia. *Cortex*. doi:10.1016/j.cortex.2015.05.019

Zeman, A., Dewar, M., & Della Sala, S. (2016). Reflections on aphantasia. *Cortex*, *74*, 336–337. doi:10.1016/j.cortex.2015.08.015

Zeman, A., Della Sala, S., Torrens, L. A., Gountouna, V. E., McGonigle, D. J., & Logie, R. H. (2010). Loss of imagery phenomenology with intact visuo-spatial task performance: A case of 'blind imagination'. *Neuropsychologia*, *48*(1), 145–155. doi:10.1016/j.neuropsychologia.2009.08.024

Zhao, M., Gersch, T. M., Schnitzer, B. S., Dosher, B. A., & Kowler, E. (2012). Eye movements and attention: The role of pre-saccadic shifts of attention in perception, memory and the control of saccades. *Vision Research*, *74*, 40–60. doi:10.1016/j.visres.2012.06.017

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*(12), e51382. doi:10.1371/journal.pone.0051382

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. doi:10.1037/0033-2909.123.2.162

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171. doi:10.1111/1467-9280.00430

Appendices

Appendix I – Examples of route descriptions used in Experiment 1 and list of landmark words used.

EGOCENTRIC DESCRIPTION	ALLOCENTRIC DESCRIPTION
Leave the ___.	Leave the ___.
Turn left/right at the ___.	Take the second road heading n/s/e/w.
Take the first/second ___.	At the ___, head n/s/e/w.
Walk past the ___ on your left/right.	Walk past the ___ on your left/right.
Turn left/right at the ___.	Take the first road heading n/s/e/w.
Take the second left/right.	Walk past the ___ on your left/right.
Walk past the ___ on your left/right.	At the ___, head n/s/e/w.
You have reached the ___.	You have reached the ___.

LANDMARK WORDS	
House	Pet store
Gym	Bank
Pub	Cinema
Train station	Town hall

Appendix II – Examples of route descriptions used in Experiment 2 and list of landmark words used.

EGOCENTRIC DESCRIPTION	ALLOCENTRIC DESCRIPTION
Leave the ___.	Leave the ___.
Turn left/right at the ___.	Take the first road heading n/s/e/w.
Take the second left/right.	At the ___, head n/s/e/w.
Walk past the ___.	Walk past the ___.
Turn left/right at the ___.	Take the first road heading n/s/e/w.
Take the first left/right.	Walk past the ___.
Walk past the ___.	At the ___, head n/s/e/w.
You have reached the ___.	You have reached the ___.

LANDMARK WORDS			
House	Park	Zoo	Station
Gym	Pub	Theatre	Barber
Clinic	Bank	Mall	Hotel
School	Florist	Bakery	Cemetery
Library	Aquarium	Hospital	Café
Cinema	Dentist	Church	Museum

Appendix III – Examples of route descriptions used in Experiment 3 and list of landmark words used.

EGOCENTRIC DESCRIPTION	**ALLOCENTRIC DESCRIPTION**
Leave the ___.	Leave the ___.
Turn left/right at the ___.	At the ___ head n/s/e/w.
Take the first left/right.	Take the second road heading n/s/e/w.
Walk past the ___.	Walk past the ___.
Turn left/right at the ___.	Take the first road heading n/s/e/w.
Take the second left/right.	Walk past the ___.
Walk past the ___.	At the park head n/s/e/w.
You have reached the ___.	You have reached the ___.

LANDMARK WORDS		
University	Newsagent	Library
Bank	Estate agent	Indian restaurant
Toy store	Karaoke bar	Hospital
Science museum	Cemetery	Café
Butcher	Pharmacy	Pond
Car dealer	Hotel	Basketball court
Opera house	Skating arena	Italian restaurant
Florist	Pub	Botanical garden
Archaeological museum	Animal shelter	Football field
Christmas market	Japanese restaurant	Bakery
Nursing home	Theatre	School
Chinese restaurant	Solicitor’s office	Computer store
House	Veterinary	Cinema
Park	Sushi bar	Aquarium
Zoo	Turkish restaurant	Church
Station	Nightclub	Videogame store
Playground	Barber	Greenhouse
Square	Bookstore	Vintage shop
Beauty salon	Mall	Tailor store
Paintball arena	Bowling alley	Noodle bar
Fish market	Thai restaurant	Greek restaurant
Hockey field	Auction house	Dentist
Pet store	Print shop	Caribbean restaurant
Upholstery store	Steakhouse	Mexican restaurant

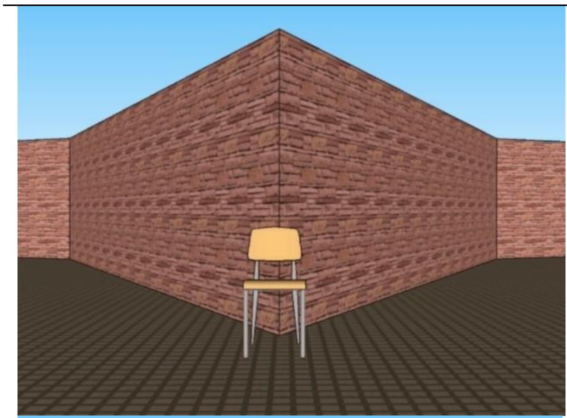
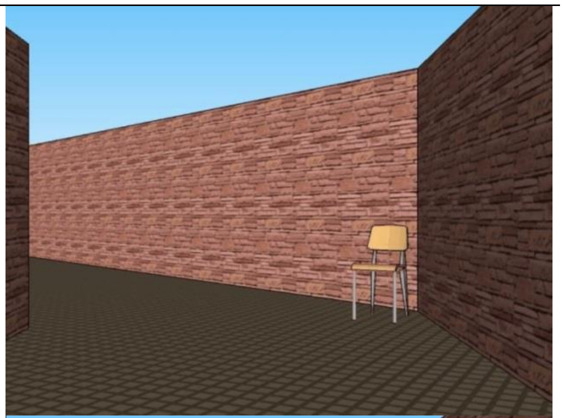
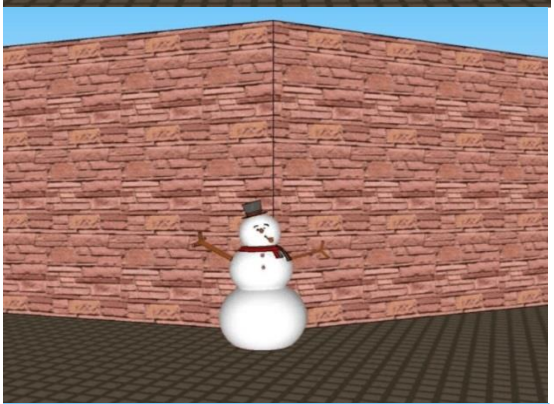
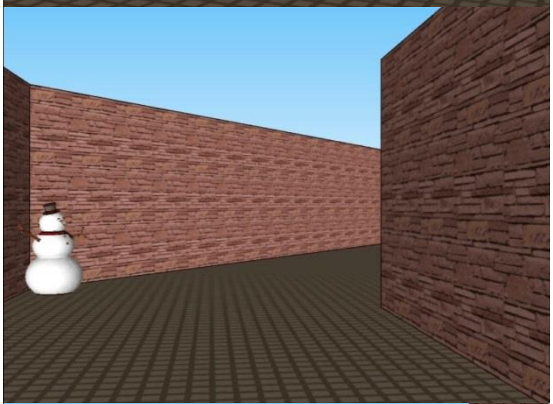
Appendix IV – Examples of route descriptions used in Experiments 4 and 5, and lists of landmark words and distances used.

EGOCENTRIC DESCRIPTION
Leave the ___ and turn left/right. Walk for ___ metres. Turn left/right at the ___ and walk for ___ metres. You have reached the ___.
ALLOCENTRIC DESCRIPTION
Leave the ___ and head n/s/e/w. Walk for ___ metres. At the ___ head n/s/e/w and walk for ___ metres. You have reached the ___.

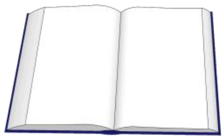
METRIC DISTANCES	
50	100
150	200

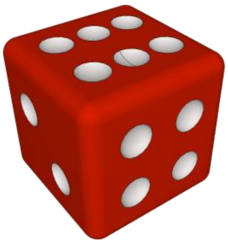
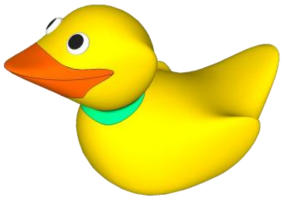
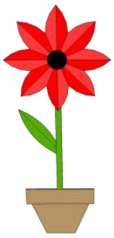
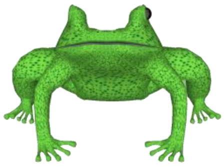
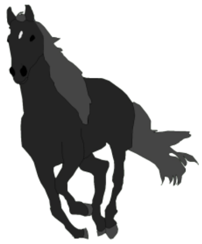
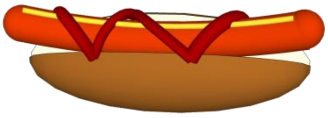
LANDMARK WORDS		
archaeological museum	playground	botanical garden
fish market	football field	Caribbean restaurant
solicitor's office	nursing home	vintage shop
skating arena	print shop	Italian restaurant
Christmas market	estate agent	hospital
hockey field	butcher	opera house
animal shelter	noodle bar	Japanese restaurant
sushi bar	beauty salon	science museum
basketball court	car dealer	Chinese restaurant
paintball arena	fish and chips stand	cemetery
bridal wear store	tailor store	zoo
Indian restaurant	veterinary	aquarium
auction house	upholstery store	videogame store
karaoke bar	bowling alley	Greek restaurant
fancy dress shop	hockey field	Thai restaurant
miniature model store	nightclub	Turkish restaurant

Appendix V – Screenshots of the landmark objects used in Experiment 6 as they appeared in the virtual environment.

DECISION POINT	NON-DECISION POINT
	
	
	

Appendix VI – Object models used as environmental landmarks and as foils in the landmark recognition task in Experiment 8.


Oven Fresh
Pizza


brussels airlines