DOCTORAL DISSERTATION

Ágoston Török

SPATIAL PERCEPTION AND COGNITION,

INSIGHTS FROM VIRTUAL REALITY EXPERIMENTS

2016

Ágoston Török

SPATIAL PERCEPTION AND COGNITION,

INSIGHTS FROM VIRTUAL REALITY EXPERIMENTS

Doctoral School of Psychology

Head of the Doctoral School: Zsolt Demetrovics, DSc, professor, Eötvös Loránd University

Cognitive Psychology Programme

Head of the Programme: István Czigler, DSc, professor, Eötvös Loránd University

Supervisor: Valéria Csépe, DSc, CMHAS, professor, Brain Imaging Centre, RCNS, HAS

Committee:

Chair: István Czigler, DSc, professor, Eötvös Loránd University

Secretary: Anett Ragó, PhD, senior lecturer, Eötvös Loránd University

Internal opponent: Andrea Dúll, PhD, reader, Eötvös Loránd University

External opponent: Ádám Csapó, PhD, senior lecturer, Széchenyi István University

Members: Attila Krajcsi, PhD, reader, Eötvös Loránd University

Anikó Kónya, PhD, honorary professor, Eötvös Loránd University

Attila Márkus, PhD, MD, chief medical officer, Medical Centre, Hungarian Defence Forces
György Bárdos, PhD, professor, Eötvös Loránd University

Budapest, 2016

ABSTRACT

Several important questions of human spatial perception and cognition can only be answered with the use of virtual reality. Virtual environments enable the manipulation of reality, and their perception provides us insights on how spatial cognition works under normal circumstances. The present dissertation also benefits from this tool in answering how our senses, our body, and our viewpoint affect our spatial representations. In the first study, we investigated how different viewpoints are associated with different reference frames. The results of the tablet PC navigation task showed that when we take a ground-level viewpoint, an egocentric frame of reference is preferred. However, from an aerial viewpoint, using an allocentric frame of reference results in better navigation performance. This difference motivated the second study presented herein. We examined how the lack of constant feedback from our position change affects navigation. In the experiment, participants were searching rewards in the East or the West alleys of a cross-maze. Before each choice, they were teleported randomly either to the South or to the North alley. The teleportation induced reorientation, which resulted in profound topographic ERP differences as early as 100 msec. Furthermore, we found that, here, reward objects were represented in allocentric reference frame. Because both of these studies were primarily visual, in the next study we demonstrated the dominance of vision in spatial perception. We showed that sounds were perceived as coming from the direction of the concurrent visual stimuli in virtual reality. The role of multisensory perception in spatial cognition has been the focus of the last study. In this experiment we showed that object seem farther when we look up to them, and they seem closer when we look down at them. This phenomenon is caused by a multisensory integration between vision and the vestibular sense. The four presented studies support the notion of multisensory and collage-like nature of cognitive maps. The present research, besides of its significance to basic research, holds also important implications for applied fields. Hence, we devote the last chapter to discussing our results from the perspective of virtual reality navigation interface design.

ABSZTRAKT

A human téri észlelés és tájékozódás kutatásában régóta fontos szerephez jut a virtuális valóság. A virtualitás lehetőséget ad arra, hogy a valóságos viszonyokat manipulálva ismerjük meg a téri reprezentációk természetét. Jelen disszertáció is ezen eszköz segítségével keresi a választ a kérdésre, hogy érzékszerveink, testünk helyzete és nézőpontunk hogyan befolyásolják a téri reprezentációinkat. Az első kutatásban egy táblagépen végzett tájékozódási kísérletben azt vizsgáltuk, hogy hogyan kapcsolódik össze a tájékozódás közben felvett nézőpontunk és a preferált téri vonatkoztatási keretünk. Az eredmények szerint, ha a nézőpontunk a tájékozódást végző testtel egy szintben van, akkor az egocentrikus vonatkoztatási keretet részesítjük előnyben. Azonban, ha madártávlatból látjuk magunkat, akkor az allocentrikus vonatkoztatási keret segíti jobban tájékozódásunkat. E nézőpontfüggő preferencia motiválta a második kutatást. Ebben arra kerestük a választ, hogy mi történik, ha nem tudjuk folyamatosan követni a mozgásunkat. Itt a résztvevők a keleti és nyugati szárban jutalmat kerestek egy virtuális keresztlabirintusban úgy, hogy minden választást megelőzően véletlenszerűen az északi vagy a déli szárba teleportáltuk őket. Ez reorientációt idézett elő, melyet már a teleportációt követő 100. ezredmásodpercben meg lehetett figyelni az eseményhez kötött potenciálokban. Továbbá bemutattuk, hogy ebben a helyzetben a tárgyakat allocentrikus referenciakeretben reprezentálják. Mivel a két kísérlet elsődleges vizuális volt, következő kísérletünkben igazoltuk a látás kulcsszerepét a téri észlelésben. Bemutattuk, hogy a hangok helyzetét a vizuális ingerekhez közelinek észleljük virtuális valóságban is. A téri észlelésben szerepet játszó multiszenzoros integrációra hívja fel a figyelmet az utolsó bemutatott kutatás is. Ebben bemutattuk, hogy a tárgyak felfelé nézve távolabbinak, lefelé nézve közelebbinek tűnnek, a vesztibuláris és vizuális rendszer információinak integrációja miatt. A bemutatott négy kutatás eredményei támogatják a kognitív térkép multiszenzoros és kollázs-szerű elképzelését. A kutatás alaptudományos jelentősége mellett fontos gyakorlati következményeket is hordoz. Ezért az utolsó fejezetben az eredményeket a virtuális valóságban történő optimális tájékozódás szempontjából értelmezzük.

PREFACE

I always found it interesting how many ways we can define the position of an object. The same object without changing its position can be in front of us, on top of something, behind something, a part of a compound object, and so on. I first asked what does it depend on how I am going to define the position of an object. I started the scientific exploration of this question using virtual reality. This tool enabled me to apply the scrutiny of psychophysical experiments yet to preserve the most ecological validity. Through the years I became more and more interested in how these results can help the design of virtual reality and what virtual reality is for.

The present work summarizes the exploration and results of my doctoral years. The dissertation is divided into 10 chapters. The first two chapters provide a general introduction to the study of spatial cognition. In these, first, I summarize the results of early exploration on spatial navigation, and then I introduce the topic of reference frames. Then I present four original studies in Chapters 3, 5, 7, and 9. These chapters are linked together with intermediate chapters, which serve as transitions between the studies. The first study deals with the relationship between reference frames and viewpoints. I show that there exists an implicit association between reference frames and viewpoints: allocentric reference frame is preferred from bird's eye view and egocentric is preferred from near navigator perspectives. In the second study, I aimed to show that this implicit association does not simply depend on the position of the camera, but rather on the availability of first-person locomotion experience. Introducing unpredictable teleportation episodes in a cross maze paradigm, I show allocentric coding of object location from 1st person viewpoint. Since these two studies relied primarily on vision, in the third and fourth studies I tested whether other sensory modalities affect spatial perception as well. In the third study, I show that vision captures the perceived location of sounds both on the horizontal and on the vertical planes, supporting the key role of vision in human spatial perception. Then, in the fourth study, I present the results of a distance estimation experiment where the vestibular information modulates the visually perceived distance of the target object. This result shows that spatial perception is indeed a multisensory process. The general discussion is given in Chapter 10, where the focus is shifted from exploratory science to the applied perspectives of the current work.

This work is original, except where references and acknowledgements are made to previous work. Neither this nor any substantially similar dissertation has been or is being submitted for any other degree, diploma, or other qualification at any other university.

A version of Chapter 3 has been published. 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*. I, Zoltán Nádasdy, and Peter Nguyen designed the paradigm, I implemented the tablet PC paradigm in Unity 3D. The experiment was recorded by Mátyás Wollner and myself. I performed the analysis. The text of the chapter is loosely based on the above manuscript.

The OPM model in Chapter 4 was created by me in OPCAT, based on the discussions with Valéria Csépe.

The experiment in Chapter 5 was designed by me, Ferenc Honbolygó, and Andrea Kóbor. The experiment was written by me in XML using the experiment controller extension implemented in Virca by György Persa and Péter Galambos. The experiments were conducted by György Persa, Orsolya Kolozsvári, Gabriella Baliga and Zsuzsanna Kovács. The analysis of the behaviroural data was done by me and Borbála Tölgyesi. The electrophysiological data was analysed by me.

The data of the experiment presented in Chapter 7 has been published. Török, Á., Mestre, D., Honbolygó, F., Mallet, P., Pergandi, J.-M. M., & Csépe, V. (2015). It sounds real when you see it. Realistic sound source simulation in multimodal virtual environments. *Journal on Multimodal User Interfaces, 9(4),* 323–331. The paradigm was designed by Me, Daniel Mestre, Ferenc Honbolygó and Valéria Csépe. It was implemented by Jean-Marie Pergandi. The experiments were conducted by me, Pierre Mallet, and Jean-Marie Pergandi, with equal contributions. The analysis was done by me. The text of the chapter is loosely based on the above manuscript.

The experiment in Chapter 9 was designed by me, Elisa Ferre, David Swapp, and Patrick Haggard. The implementation was done by Elena Kokkinara. The experiments were conducted by Elisa Ferre, Me, and David Swapp. The present analysis was done by me.

ACKNOWLEDGEMENTS

Contents

LIST OF TABLES

LIST OF FIGURES

1 THE HISTORY OF COGNITIVE MAPS

The study of navigation dates back to the earliest days of experimental psychology (Carr & Watson, 1908; Small, 1901; Watson, 1907). Early experimenters already noticed that rats not only learn mazes without reward or instructions (Blodgett, 1929), but they can easily recall them from memory even if sensory cues are absent (Lashley & Ball, 1929). Accumulated evidence indicates that this behaviour cannot be explained by a stored sequence of action-response associations. The term cognitive map was coined by Tolman (Tolman, 1948), who showed that animals learn the general configuration of walls and routes in the environment and thus can make shortcuts if the earlier routes are not available (Tolman, Ritchie, & Kalish, 1946a, 1946b). His notion has provoked a long-standing debate on the nature of this cognitive map (Tversky, 1993). There are three main views on how the cognitive map and spatial knowledge are represented. The first view claims that the cognitive map (Kosslyn, 1981) and spatial knowledge of objects (Shepard & Metzler, 1971; Shepard, 1978) are represented primarily in visual form; the second posits that they are essentially multimodal; whereas the third argues that they are even abstract (Tversky, 1993).

Although these approaches largely differ from each other, they are all plausible considering how spatial information is usually acquired. Navigation is part of our everyday life. While the scale is different from person to person (Gonzalez, Hidalgo, & Barabasi, 2008), a shared feature of all human locomotion patterns is that we experience space from our own perspective. If we think of how we experience space, we first think of how we see the position of objects change in the visual field while we pass by them (Sun, Campos, Young, Chan, & Ellard, 2004). However, we use other modalities as well. We hear the traffic signals or can locate people easily if they call us (Ho, Reed, & Spence, 2007; Koelewijn, Bronkhorst, & Theeuwes, 2010). Furthermore, maybe implicitly, but we use the smell of the bakery (S. Zhang & Manahan-Vaughan, 2015), as well as the vestibular and proprioceptive information from the steepness of the road (Dokka, MacNeilage, DeAngelis, & Angelaki, 2011; Sharp, Blair, Etkin, & Tzanetos, 1995) in representing the environment. Thus, our body, our senses, and our egocentric viewpoint are indispensable parts of the formation of cognitive maps. These support the view that the cognitive map is a concrete, experience based multimodal representation.
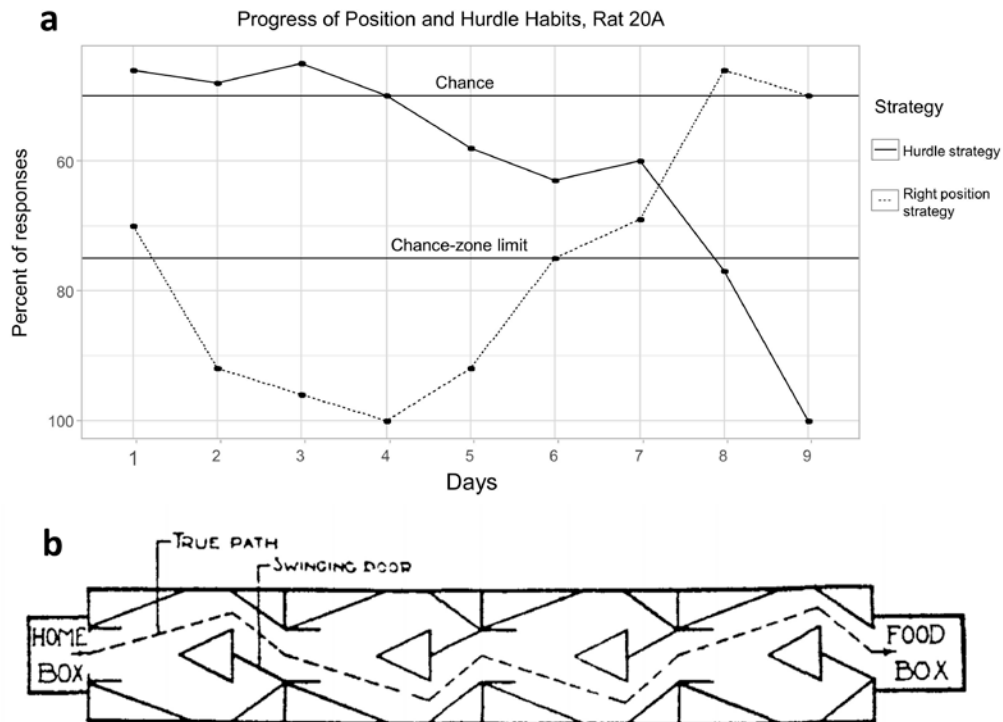
Figure 1 The experiemental design and results of one experiment in Krechevsky (1936). (b) The experimental container contained swinging doors that could either block or let the animal pass through. The doors differed in their position ("left", "right") and whether they had a hurdle in front or not. (a) He found that rats learn to differentiate between the doors first based on their position and only later based on the hurdle. The sketch is taken from I. Krechevsly (later D. Krech) (1932). The genesis of "hypotheses" in rats. *Univ. Calif. Publ. Psychol., vol6/no.4* p46.; the chart is based on the result of one rat taken from the same paper.

Nonetheless, if we are asked to draw a sketch of our journey, we do not draw snapshots of what we have seen, but we make maps. It is even more interesting, that no matter how inherent our own perspective was during our journey, on maps we use a different perspective: the eye of an aerial observer (Bjelland, Montello, Fellmann, Getis, & Getis, 2013). More precisely, this is not even a real perspective because maps are drawn as a perspective-free representation of space (Snyder, 1997). This representation mode is known as the Mercator projection (Monmonier, 2010) and is universal to the human culture throughout continents and ages (Bagrow, 2010; Z. Török, 1993, 2007). This universality might be because maps are close to how we actually *remember* space.

Indeed, hand drawn maps typically contain a number of distortions, some of these is even favoured over reality by the independent viewer (Tversky, 1981). These distortions can be derived from the principles of perceptual organization (Koffka, 1922)

that are already present in the early processing steps of vision (Kovács & Julesz, 1993; Kovács et al., 1994). In Gestalt terms, maps can be viewed as figures in backgrounds (Tversky, 1981). Thus, recalling the absolute position and orientation of parts are difficult, but we easily remember their relative positions. The two most common distortions derive exactly from the motivation to memorize relative positions. We *align* figures in memory if they are close together and *rotate* them if they are slightly tilted from an intrinsic axis (e.g. up- down; Tversky, 1981). In human navigation, the two most important axes are the vertical axis defined by gravity and the horizontal axis defined by the horizon (Howard & Templeton, 1966). Related to this, Stevens and Coupe (Stevens & Coupe, 1978) noticed that people, instead of remembering the relative positions of a great number of cities, remembered the relative position of the countries (in vertical and horizontal terms of a map) where the cities are and used that knowledge to infer the relative position of the cities. For example, people from Budapest may likely agree that Bratislava is northeast of Wien because the relative position of Slovakia and Austria from Hungary suggests so. These results raise the possibility that the cognitive map does not meet the criteria of Euclidean geometry (Spelke, Lee, & Izard, 2010) and is rather a hierarchical, interconnected structure that can easily be an impossible figure (Tversky, 1981, 1993). Indeed, even the earliest studies demonstrated that objects in the environment can be represented in different frames. In an experiment (Krechevsky, 1932), rats had to run through an elongated container with a set of double doors in one direction. The target door - at every choice point either the left or right randomly - had a hurdle in front of it (see Figure 1). Krechevsky found that the rats' behaviour was not described by trying different solutions in each trial. Instead they pursued well defined strategies for several trials before learning the correct solution. Typically, the first strategy was always the choice of either the left or the right door. He interpreted this behaviour as hypothesis testing. Later studies extended these results and showed that what Krechevsky described with the terms "left" and "right" are rather allocentric coordinates in reality. Accordingly, rats choose the identical place (and not the e.g. "left" door again), if they are running in the other direction (Packard & McGaugh, 1996).

In these strategies, the same spatial layout is represented in different frames: doors are defined as West/East, Left/Right, illuminated/dark, or one that has a hurdle in front/one that does not. These frames require the availability of different layers which

could provide the necessary information. As a result, our cognitive map contains information in several layers (Tversky, 1993). This, however, raises the question of whether these layers are learnt sequentially or in parallel.

Unlike mazes, real life environments contain many types of information that could help us learn the structure of our surroundings (Knierim, Kudrimoti, & McNaughton, 1995). Lynch (Lynch, 1960) defined the key elements of spatial mental representations as paths, edges, districts, nodes, and landmarks. Landmarks are stable, often large sized (cf. Lynch, 1960 pp. 48), distinctive objects that are visible from multiple viewpoints and thus are helping our orientation in a novel environment (Chan, Baumann, Bellgrove, & Mattingley, 2012). Indeed, the theoretical framework proposed by Siegel and White (Siegel & White, 1975) describes the knowledge of landmarks as the *initial stage* of spatial knowledge. They claim that landmarks are identified and learned easily; first the mental space between landmarks is empty and only receives scale through repeated experience and traversals. This means that spatial knowledge is initially nonmetric (cf. Hafting et al., 2005; Rowland, Yanovich, & Kentros, 2011).

Route knowledge develops while the animal traverses the environment, (Shemyakin, 1962). This knowledge is acquired from the navigator's perspective and is connected to goal directed navigation (Rossano & Reardon, 1999). The cognitive map stored in route knowledge is narrow (Tolman, 1948); that is, it cannot be used to make novel shortcuts when landmarks are not available (Foo, Warren, Duchon, & Tarr, 2005). The process associated with the development of route knowledge is path integration that is based on our sensory-motor experience (McNaughton, Battaglia, Jensen, Moser, & Moser, 2006).

Once we spent extended period of time in the environment, paths become interrelated. They form a network-like assembly and create a gestalt: the survey knowledge (Siegel & White, 1975). In contrast to route knowledge, survey knowledge is a structured representation of the available space, a *broad* cognitive map (Tolman, 1948). It usually uses an aerial perspective. Environmental axes become important anchors of survey knowledge (Tversky, 1981). Survey knowledge is less related to action and more related to memory (Montello, 2005).

The two kinds of knowledge are different from several other aspects. Route descriptions are analytic, sequential and procedural, whereas survey descriptions

provide a holistic and abstracted representation of space (Brunyé, Gardony, Mahoney, & Taylor, 2012; Schinazi & Epstein, 2010). In survey descriptions the perspective is defined from the outside of the behavioural area. However, this does not necessarily mean that it is defined from above. For example, for a rock climber, it could be a point of view from the ground that provides a view of the whole rock face. Although it is implied that the two types of knowledge are connected to different perspectives, it must be noted that the core of the difference is not the perspective per se but the structure of the representation (Brunyé et al., 2012). Although it might seem that survey is superior to route knowledge, the difference is rather qualitative. For small spaces, route knowledge facilitates performance; survey knowledge is useful when the navigator looks for shortcuts in large spaces (Brunyé et al., 2012). Thus, spatial knowledge changes and broadens, but earlier stages of knowledge will still be active in navigation. The availability of later stages rather extends the capabilities of the organism, and does not substitute the earlier stages.

The use of route or survey knowledge often dynamically switches according to the actual task. Lee and Tversky (2001, 2005) studied how induced perspective change affects comprehension of verbal descriptions. In their experiment participants read sentences. They manipulated the spatial perspective used in the sentences to facilitate the use of either route or survey knowledge. They found that reading times increased when after three sentences from the same perspective the fourth sentence used a different perspective. This result showed that change in perspective likely induces a change in the activated type of knowledge also in spatial language. However, it is still an open question whether this is true for active navigation and whether perspective itself is the underlying factor or it is the amount of information that is available from the different perspectives.

Summarizing this chapter, maze tasks has been of interest from the earliest days of experimental psychology. Researchers were intrigued by the observation how easily rats learnt mazes (Blodgett, 1929) and how flexible this knowledge was when circumstances changed (Lashley & Ball, 1929; Tolman et al., 1946a). These results were integrated by Tolman (Tolman, 1948) into the theory of cognitive map. This representation develops through multiple stages. According to the early (but still widely accepted) theory of Siegel and White (Siegel & White, 1975), first, landmarks are identified; then, path integration develops route knowledge; finally, the interconnected

routes give rise to survey knowledge. The development, however, does not mean that the earlier stages become deprecated; in fact, everyday navigation relies on both and uses them according to the task and the available information.

2 REFERENCE FRAMES IN SPATIAL COGNITION

The studies presented in the previous chapter together with evidence from the neural background of navigation (O'Keefe & Dostrovsky, 1971) pointed towards the argument that navigation in mammals relies on an enduring, comprehensive, and environment centred representation of space. However, subsequent evidence suggested that the results of rodent studies may not be generalized to human cognition, and human navigation relies primarily on a dynamic, egocentric, and limited representation of space (Wang & Spelke, 2002). The core of this debate is whether the position of objects is anchored to our own position and viewpoint (Mou, Fan, McNamara, & Owen, 2008) or if it is defined by landmarks, environmental axes, and other objects (Chan et al., 2012). In short, this is the question of frame of reference.

Cognitive neuroscience distinguishes two frames of reference: allocentric, where the objects and our own heading is defined by the position of other objects in the environment; and egocentric, where the position of objects is dynamically updated when the actor moves (Klatzky, 1998). Allocentric is sometimes also called exocentric (McCormick, Wickens, Banks, & Yeh, 1998; Wickens, Liang, Prevett, & Olmos, 1994) or geocentric (McNamara, Rump, & Werner, 2003). However the latter can also mean a third type of frame of reference, where the global orientation serves as reference (Finney, 1995; Wiltschko & Wiltschko, 2005). While evidence from rodent studies supported the role of an allocentric frame of reference (O'Keefe & Nadel, 1978; Taube, Muller, & Ranck, 1990; Tolman, 1948), research with humans suggested the crucial role of egocentric frame of reference (Wang & Spelke, 2000, 2002).

Wang and Spelke claim (Wang & Spelke, 2002) that although the use of geographic maps led to the widely accepted notion that human navigation relies on allocentric frame of reference; evidence from studies of navigation suggests the contrary: an egocentric frame of reference. Their theory states that three systems underlie human spatial navigation: (1) path integration is used to dynamically update spatial representations during locomotion, (2) place recognition is based on snapshots from experienced viewpoints that are stored in memory, (3) reorientation is based on a geometric module which uses the layout of the surface. This latter system is encapsulated, and thus, can only interact with the other two systems through language. Furthermore, this geometric module represents space in a manner that does not meet all

criteria of Euclidean geometry (Spelke et al., 2010). Interestingly, in their later theory Spelke and colleagues (Spelke et al., 2010) identified two core geometric systems. One is active during navigation and represents length and direction but not the angle of edges, while the other is active during the analysis of visual forms, represents length and angle but not direction (i.e. this is the reason why objects and their mirrored versions are rather hard to distinguish).

The theory of Wang and Spelke is supported by empirical evidence from several studies and provoked intense discussion (Waller & Hodgson, 2006; Wang & Spelke, 2002). One of the most important criticisms came from Burgess (Burgess, 2006) who noted that the studies Wang and Spelke cited do not conclude that allocentric representations have no role to play in navigation. He also argues that their theory not only questions the construction of maps but would lead to a computationally suboptimal navigation. If we updated the location of every object separately during locomotion, our brain would have to cope with increasing amount of information as distances and number of objects increase (Burgess, 2006). He hypothesizes that egocentric and allocentric representations exist in parallel and combine during spatial navigation. Here, we review the four most important results on which Wang and Spelke built on their conclusions (Wang & Spelke, 2002) together with the counterevidence reviewed by Burgess (Burgess, 2006).

In an experiment (Wang & Spelke, 2000) participants were blindfolded and disoriented. Their task was to point to different objects in the experimental room and pointing errors were measured. The analysis showed that disorientation caused increase in the variance of pointing errors; consequently, the location of the objects were individually (i.e. one by one) defined to the participants' own orientation. Wang and Spelke concluded that continuous input is required for path integration, and without that, spatial representations will not result in proper localization of unseen objects. However, a follow-up study by Waller and Hodgson (Waller & Hodgson, 2006) showed that, when participants are asked to make judgements relative to themselves, disorientation leads to increase in pointing errors, but when their task is to make judgements to an object relative to another object, pointing errors actually decrease. Furthermore, they also investigated the role of disorientation angle and found that the 'disorientation effect' appears after a rotation >135°. Based on these, they concluded that two systems underlie spatial navigation, one transient but precise (egocentric) and

one enduring but coarse (allocentric). Disorientation causes a switch from precise but transient egocentric representation to the enduring but coarser allocentric representation. Also, allocentric representations are not coarse by nature, but they require more time in the environment to build up (Golledge, Dougherty, & Bell, 1993; Siegel & White, 1975). Indeed, if the task requires pointing in a familiar environment, pointing errors decrease (Holmes & Sholl, 2005).

The second source of evidence comes from an experiment where the effect of viewpoint was studied on recognition (Diwadkar & McNamara, 1997; Shelton & McNamara, 1997). Here, studies found that the time to recognize another photo from a different viewpoint than a studied one linearly increase as a function of the angle difference between the two viewpoints (Diwadkar & McNamara, 1997). Moreover, if participants are asked to point to an object from an imagined viewpoint, their pointing is faster and more precise if the imagined viewpoint has the same egocentric bearing (Shelton & McNamara, 1997). However, further experiments (Mou & McNamara, 2002) found that it also helps if the task objects' layout contains an intrinsic axis. This effect is even stronger when the intrinsic axis is aligned with the borders of the environment (e.g. the walls of the room). Landmarks also play an important role and help direction judgement even in unexperienced viewpoints (McNamara et al., 2003). Thus, viewpoint is important and is stored in memory; however, we are not just taking and storing mental snapshots for navigation but actively study the environmental layout, landmarks, and borders.

Further evidence for the importance of motion and path integration comes from the experiment of Wang and Simons (1999). In their experiment, the task was to detect if an array has been allocated on a table between the study and the test phase. They manipulated whether the participant, the table, both, or none moved between the two phases. They found that detection was better if the participant moved than when the table, supporting the role of self-motion in path integration. However, in a follow-up experiment (Burgess, Spiers, & Paleologou, 2004) a cue was introduced on the table but outside of the object array. They found that the cue served as a landmark, and detection was better if the table and the cue moved together (but not when only either), meaning that an intrinsic reference point can and will be used when the position of objects is coded into memory.

The fourth source of evidence comes from the early experiments of Cheng (1986). In his experiments, rats were placed in a rectangular room and their task was to look for buried food pellets of Coco Puffs. When they found the food and started to eat, the experimenter interrupted them and took them out of the box. After a 75 s delay, they were put back and the experimenter observed where they went to dig for the remaining food. They found that the rats used the geometry of the environment and searched in the correct and 180° rotated corner for the food. In later manipulations, they manipulated the texture or the brightness of the target wall but found that the rats were primarily looking for the geometrically equivalent corners first. These results were found also when 1.5- to 2- year old children were told to find a hidden toy in a rectangular environment (Hermer & Spelke, 1996). Later experiments repeated these results and generalized it to circular and square rooms and verified that the effect does not stem from the inability to recognize landmarks (Wang, Hermer, & Spelke, 1999). Yet, further investigation of the effect revealed that size of the environment is critical in the task (Hupbach & Nadel, 2005; Learmonth, Nadel, & Newcombe, 2002). If the environment size is increased the texture cue serves as better cue, and children use it better in their reorientation (Burgess, 2006). Supporting evidence for the presence of a slowly built-up but enduring spatial representation comes also from the original study of Cheng (1986). He found that rats choose the correct corner if the pellets were placed in the same corner in repeated trials. This result was replicated with chickens and monkeys, too (Burgess, 2006).

In conclusion, it seems that the egocentric and allocentric systems work in parallel and support different aspects of navigation. On the one hand, the use of an egocentric reference frame provides precise location information at the expense of requiring more cognitive resources. On the other hand, representations in an allocentric reference frame are more enduring and computationally less expensive to use, but they reach the level of precision provided by the egocentric frame of reference more slowly compared to the immediate establishment of representations in the egocentric frame. Results revealed that landmarks and intrinsic axes can help in establishing and using allocentric representations. It was also shown that the viewpoint associated with landmarks or aligned to axes is used more easily than arbitrary viewpoints. This leads to the question whether viewpoints (external and internal) and frames of references are associated in cognition. One can speculate that external viewpoints are associated with

allocentric frame of reference and internal viewpoints - that follow the viewpoint of the actor – are associated with egocentric frame of reference. This is reasonable, since external viewpoints are not (or mildly) affected by the actor's locomotion. However, some evidence suggests that this speculation might not be entirely true. In their experiment, Waller and Hodgson (2006) observed that for small rotations (<90°) participants continued to use the egocentric frame of reference, however, for larger rotations (>135°), they switched to the use of an allocentric frame of reference. Consequently, it is possible that if the external viewpoint is not different enough from the actor's viewpoint, the allocentric switch will not happen; thus, performance will decrease linearly with the angle of mental rotation (Diwadkar & McNamara, 1997).

3 Experiment 1: Implicit association between reference frames and viewpoints[1]

3.1 Introduction and hypotheses

Map-based and direct navigation is different in several aspects (H. Zhang, Copara, & Ekstrom, 2012). First, maps employ a different perspective, taking an aerial point of view instead of a ground level perspective (Z. Török, 1993). Then, maps also offer a wider view of the environment and, hence, easier recognition of landmarks and borders. Finally, since maps typically show the boundary of space, they provide a reliable reference for our current absolute position (Brunyé et al., 2012). All these factors could potentially play a role in biasing performance in map-based versus 1st person-based navigation. In their study, Barra and colleagues (2012) found that a slanted perspective, which provided wider view of the environment, led to better performance in a shortcut finding task. However, they manipulated not only the size of overview but the camera position, as well. Although it is not possible to balance the field of view (FOV) between ground-level and aerial perspectives, it is possible to balance the average visible area. If the FOV from a fixed aerial perspective is constant, then, the effective FOV for ground-level perspective should be controlled, too. This can be achieved by the use of a bounded but open area. In their study, Shelton and Pippitt (2007) followed a similar approach although in their task the navigable area contained several obstacles, rendering the comparison across different visibility conditions ambiguous.

We also have to consider the possibility that while egocentric reference frame is associated with direct navigation, allocentric reference frame is associated more with memorizing (Galati, Pelle, Berthoz, & Committeri, 2010). This predicts the dominance of egocentric reference frame in an active navigation task irrespective of the viewpoint taken. Indeed, people often rotate the map in their hands to match their current heading. This is supported by Wickens and colleagues (1996, see also Eley, 1988), who found that pilots' landed in simulated environments better when the 3D-map was locked to the

---

[1] A version of Chapter 3 has been published. 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*. For author contributions see Preface.

airplane's orientation as opposed to when the view was locked to the north-south axis. However, it is possible that in these studies, it was not the map-based navigation that was associated with egocentric reference frame but the transformation between the active viewpoint (1st person viewpoint from the cockpit) and the supporting system (i.e. the 3D map). Indeed, when this factor is not present, fixed orientation aerial perspectives lead to better configurational knowledge due to the consistency in global orientation over time (Aretz, 1991; McCormick et al., 1998).

One last thing to consider is the core difference between 1st person and aerial viewpoint. There are at least two options. On the one hand, it is possible that the critical difference is whether the viewpoint is inside the actor. In this case, only inside-actor viewpoints will be associated with egocentric reference frame (i.e. the 1st person viewpoint). On the other hand, it is also possible that the difference is how easily the viewpoint taken can be transformed to the viewpoint of the actor. In this case, not only 1st person view but any 3rd person view that dynamically tracks the orientation of the actor will be associated with egocentric reference frame.

Therefore, to answer the question whether certain combinations of viewpoints and reference frames are implicitly associated, one has to design an experiment where three different camera views (map-like, 3rd person, and 1st person views) and two reference frames (egocentric and allocentric) are combined. For this, we implemented a computer game in which we independently varied the camera views (ground-level vs. aerial perspectives) and the orientation of the camera (follows avatar's heading vs. always north). As in the study of Shelton and Pippitt (2007), we counterbalanced the average visible navigable area between conditions. The dependent variables were the navigation time and navigation path length relative to the optimal value for each.

We further introduced a few important constraints: we limited the navigable area with walls, no landmark cues other than the walls were available, and the compartment had a square geometry with visually equivalent corners, making it a less reliable orientation cue (Cheng, 1986; Pecchia & Vallortigara, 2012). In order to compare the accuracy of the cognitive maps stored in memory as opposed to comparing navigation accuracy relative to visible targets, the target objects were only visible at close range. We also provided an avatar during ground-level and aerial navigation; thus, participants were

able to see themselves from an outside perspective. As natural ground-level navigation takes a 1st person perspective, we used this as a baseline condition.

Since both the visible area and the presence of an avatar were counterbalanced across the viewing conditions, differences in navigation accuracy are only attributable to an inherent association between perspective and frame of reference. In our experiment, we dissociated the two factors (view and camera movement) by alternating the reference frames between egocentric and allocentric coordinate systems while also rotating the point of view between 1st person, 3rd person (above and behind the avatar), and an aerial view.

Our main hypothesis was that ground level viewpoint is associated with egocentric frame of reference and aerial view with allocentric reference. This implies the following:

- There is a difference between ground level and aerial viewpoints but not between 1st person and 3rd person viewpoint when the latter follows the orientation of the actor. We hypothesized that, rather than the inner localization of viewpoint being the key factor, the possibility to translate between the avatars viewpoint and the viewpoint taken by the observer would be most important (Ratner, 2016).
- From a ground level perspective, the orientation tracking camera mode leads to better performance because it helps the use of an egocentric frame of reference.
- From an aerial perspective, in contrast, the camera mode that bears a fixed North orientation supports better performance because that helps the use of an allocentric reference frame.

3.2 METHODS

3.2.1 PARTICIPANTS

50 participants (25 female), all university students at ELTE, took part in the experiment. Their age ranged from 18 to 32 years ($M = 22$; $SD = 3$). Forty-six of them were right

handed. We verified that all of the participants could see and hear the stimuli well prior to the experiments. They gave written informed consent and received course bonus points for participating. The study was approved by the research ethical board of the ELTE University and met the principles of the Declaration of Helsinki.

3.2.2 APPARATUS AND STIMULI

The virtual reality game was programmed using the Unity 3D game engine (www.unity3d.com). The game was played on an Asus TF 201 and an Asus TF 301 lightweight tablet PC (NVIDIA® Tegra® 3 Quad Core CPU, 1Gb DDR3 RAM, Android^{TM} 4.x). The devices had a 10.1-inch capacitive multi-touch display with a resolution of 1280 x 800 pixels. We chose the tablet PC as a stimulus presentation interface because we use the same paradigm for testing epileptic patients in clinical settings where the portability of a device and the ease of control are of primary interests (Á. Török, Nguyen, Kolozsvári, Buchanan, & Nadasdy, 2013).

The paradigm was a custom game called "Send Them Back Home". The goal of the game was to collect cartoon-like aliens holding a coloured briefcase and to carry them to the appropriate spaceship. The game's scenario was designed to be comparable to the Yellow Cab game developed by Caplan et al. (2003). As in the Yellow Cab game, the target objects were placed quasi-randomly while the two goal locations bore fixed locations. In this way, both tasks involved visual search during the searching phase and path integration during the delivery phase of the experiment. The target objects were 1.5 unit tall aliens in orange spacesuits and carried either a big yellow or blue briefcase. The two spaceships were simple 3.5 unit diameter and 1.5 unit tall flying saucer-like objects with either a yellow or blue body. To ensure reliance on memory and external spatial cues rather than simple beacon aiming, the spaceships were visible only at the beginning of the game. That is, after the first alien delivery to each spaceship, the spaceships became cloaked (i.e. invisible) except when the avatar was in a 6-unit radius of a ship. Participants were told that the spaceships were using a cloaking machine to hide their location.
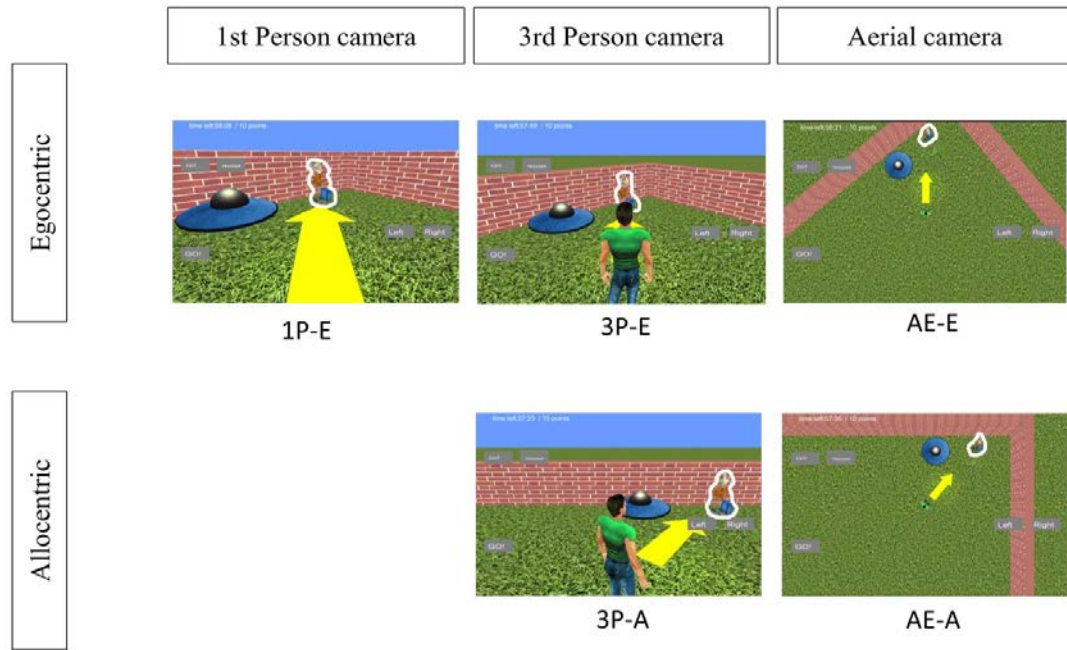
Figure 2 Sample views from the 5 camera modes used. We used 3 different camera modes: 1st person camera was the viewpoint taken in everyday navigation; 3rd person camera was a camera at a fixed 3.5 units distance relative to the avatar and looked down from a 20 degrees slanted perspective; the Aerial viewpoint was a map-like perspective, 16.5 units above the field. For the last two cameras orientation were relative to either the avatar or the environment. The arrow is visible just for presentation purposes. For illustration purposes we outlined the alien figures with a white contour

The virtual environment was a large square-shaped grassy plane enclosed by brick walls. The sky was uniform blue. The size of the environment was 80 x 80 units, and the wall was 5 units tall. There were no landmarks outside the walls, unlike in other experiments (Doeller, Barry, & Burgess, 2010; Doeller & Burgess, 2008; Honbolygo, Babik, & Török, 2014) where these served as distant directional cues. The current scenario, therefore, put more emphasis on the direction bearing only based on the geometry of the enclosure and the optic flow generated by locomotion.

We tested five different camera setups created from combinations of different viewpoints and orientation modes (see Figure 2). The viewpoints consisted of a 1st person view (eye-height 2 units), 3rd person view (3.5 units behind the avatar, 4.5 units above the ground, and slanted 20 degrees downward), and an aerial view (birds-eye

view from 16 units above). The orientation modes were either egocentric (camera turned to follow dynamically avatar's heading) or allocentric (static always-north camera orientation). We excluded 1st person-allocentric combination because there, the orientation of the navigator can only be judged correctly while moving. This resulted in five viewpoint and orientation combinations:

a) a 1st person egocentric camera mode (referred as 1P-E),

b) a 3rd person egocentric camera mode (3P-E),

c) a 3rd person allocentric camera mode (3P-A),

d) an aerial egocentric camera mode (AE-E), and

e) an aerial allocentric camera mode (AE-A).

The average field of view was balanced between camera modes to ~910 m^2 (3P=1P=~908 m^2; AE =~912 m^2).

Motion was controlled by pressing an on-screen "GO" button with the left thumb and a "LEFT" or "RIGHT" button with the right thumb. Simultaneous touch of the "GO" and arrow buttons allowed for continuous steering in the virtual space. The speed of the participant was 6 unit/s, and step sounds were played during forward movement. Turning speed was 80 degree/s. The player's virtual trajectory, including heading, along with the current objective was logged every 50 msec.

3.2.3 PROCEDURE

Participants were sitting in front of a table and were holding the tablet in their hands. According to the instructions, they had to search for aliens landed on Earth and bring them to their spaceships. They were told to deliver as many aliens as they can in 30 minutes. They were also informed that after each delivery the camera mode will switch and that the spaceships will not change their position. Lastly, they were warned of the cloaking mechanism and to make note of spaceship locations at the beginning of the game.
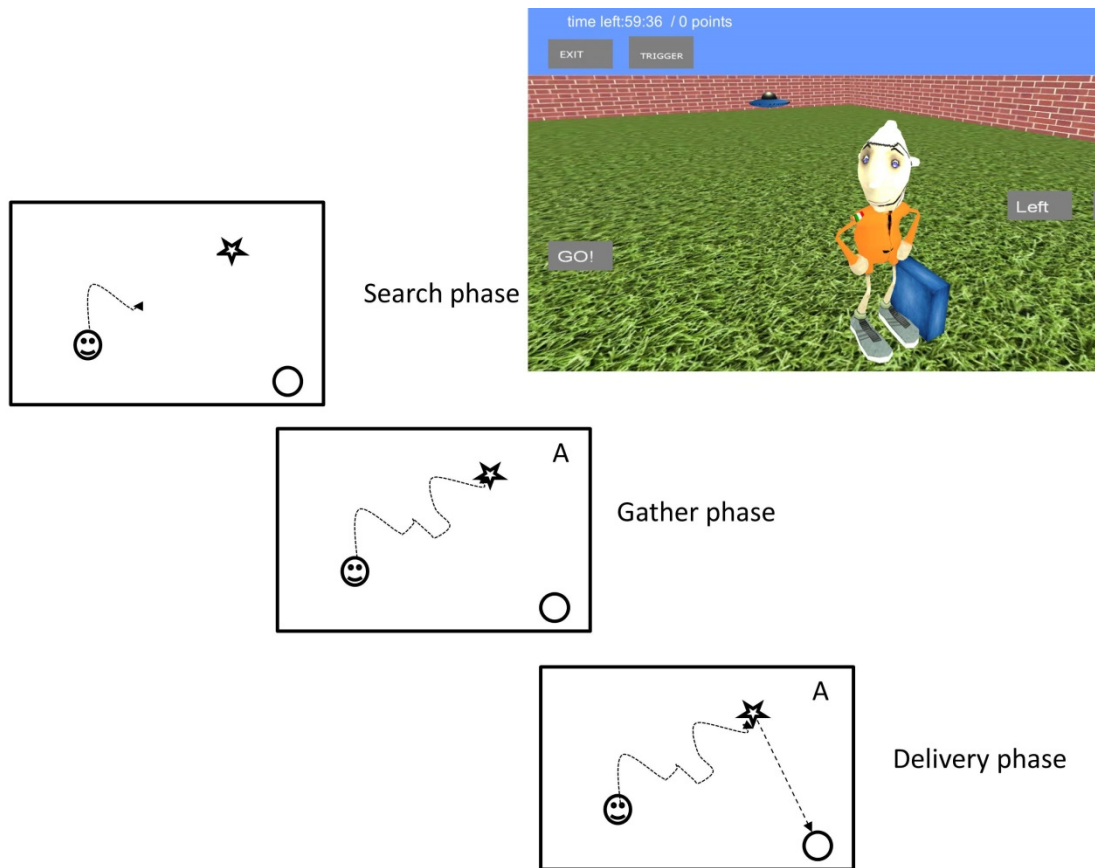
Figure 3 The phases of a trial in Experiment 1. In the search phase participants were searching a space alien. They gathered it by running through it. Carrying of the alien was indicated by a small alien image on the top right corner of the screen (symbolized by a red box here for simplicity). In the delivery phase they carried the alien to its spaceship. Upon contact with the correct spaceship a new alien appeared. The game was controlled by onscreen GO, LEFT, and RIGHT buttons. For illustration purposes we outlined the alien figure with a white contour.

Each trial started with an alien appearing somewhere in the environment (see Figure 3). The placement of the aliens followed a quasi-random design to guarantee optimal coverage of the whole field. The participants searched for the alien and gathered it by walking over it. When they gathered the alien, a small alien figure appeared in the top right corner with a text indicating the target spaceship's colour. At the same time, the alien gave vocal instructions about the updated objective in the task by saying "Now take me to my spaceship". Delivery of the alien to the appropriate spaceship was signalled by the alien saying "Thank you very much". Each delivery was worth one point; and immediately after the delivery a new alien appeared on the field. The camera modes alternated in a random order after each delivery but without returning to a previous camera mode until all 5 of the possible modes had been experienced. The

game started with one alien in the view along with the appropriate spaceship. Each experiment lasted for cca. 45 minutes including instructions, practice, the main task, and debriefing. The practice was done in a separate environment.

3.2.4 DATA ANALYSES

Because we were interested in how different viewpoints and reference frame combinations affect spatial memory and path integration, we analysed only the delivery phases, where participants had to navigate to a (not visually signalled) location in space from memory. Hence, we did not compare the search parts, where visual search is inherently easier in some combinations of viewpoints and reference frames (e.g. in 3P-E a simple turn reveals the location of the alien. Other studies verified that visual guided search and approach of the target location does not lead to enduring spatial representations (Waller & Lippa, 2007). For this reason, we excluded trials where the destination spaceship was not visible at the time of gathering. Following this criterion, on average we excluded 2.02 delivery trials (*Min* = 0; *Max* = 4). Furthermore, we excluded all first visits to each spaceship as the cloaking mechanism only activated afterwards.

Performance was scored both in terms of route efficiency and time efficiency. The former is defined as the percentage of the player's actual trajectory (Δd) based on the shortest possible route (d_{ideal}) being 100%. Since there were no obstacles, d_{ideal} was taken as the absolute distance between the alien pick-up point and the target spaceship:

$$E_{route} = \frac{\Delta d}{d_{ideal}} \cdot 100\% \ \Big| \ d_{ideal} = \sqrt{x^2 + y^2}$$

Where E_{route} is the route efficiency and *x* and *y* are the two coordinates of the spaceship relative to the current position.

The other measure is referred to as time efficiency and is defined as the percentage of observed delivery time (Δt) based on the shortest possible delivery time (t_{ideal}) being 100 %. The ideal phase completion time was calculated by the equation below, where *x* and

34

y are the relative coordinates for the absolute distance, α is the minimum angle needed to turn from the current heading to the spaceship, v_{forw} is the speed of forward motion and v_{turn} is the speed of turning (both speeds were constant).

$$E_{time} = \frac{\Delta t}{t_{ideal}} \cdot 100\% \quad \Bigg| \quad t_{ideal} = \sqrt{\left(\frac{\sqrt{x^2 + y^2}}{v_{forw}}\right)^2 + \left(\frac{\alpha}{v_{turn}}\right)^2}$$

Where E_{time} is the time efficiency and *x* and *y* are the two coordinates of the spaceship relative to the current position, and α is the minimum angle to turn.

Although path length and path time are closely related, they are not necessarily proportional, except when the avatar is continuously moving toward the target in a straight line. All other times, either when turning without moving or when turning simultaneously and advancing, what creates a curved trajectory, the two are disproportionate. Therefore, both parameters were used in the analyses.

In some trials, participants did not simply take suboptimal routes but completely lost directions. Because these trials were not artefacts per se, we decided not to exclude them. Instead, we winsorized (Dixon & Yuen, 1974) the upper 5% of all data (0 to 7 data points for every person; *M*: 2.90). Therefore, we did not analyse the extreme values as they were, nevertheless, were able to include those trials in analysis. Please note that trimming instead of winsorization did not change the main results.
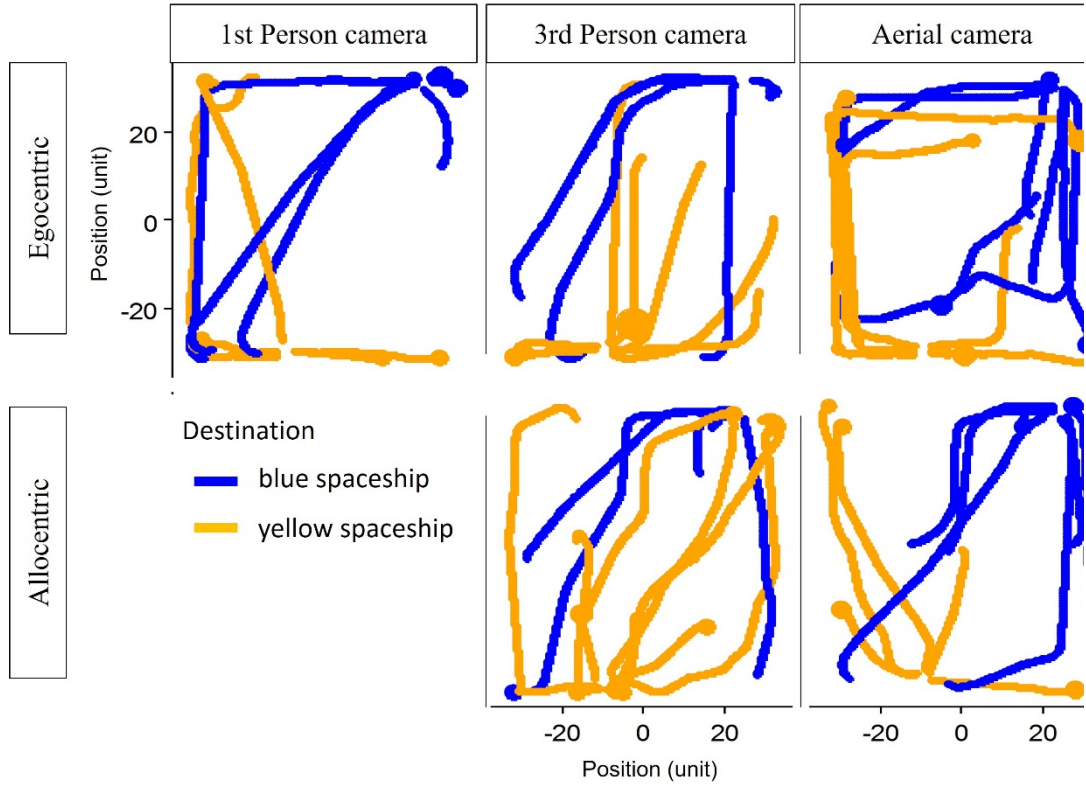
Figure 4 Raw trajectory of one participant in the five camera conditions. It is visible that, although the spaceships were invisible during delivery, the trajectories seem to be close to optimal. Also, trajectories in the aerial egocentric (top right plot) and in the 3rd person allocentric condition seem to be less optimal in some cases. Blue and Orange denotes trajectories to the different spaceships, the thickness of the line is proportionate to the time spent in the given position

3.3 RESULTS

3.3.1 OVERALL PERFORMANCE

Before analysing the efficiency metrics, we examined the overall performance in the task. Across the 30-minute runs, participants collected 57.34 (*SD* = 9.08) aliens on average. This means that they experienced each of the five viewpoints at least in 10 trials during the game. To note, male subjects collected significantly more aliens than
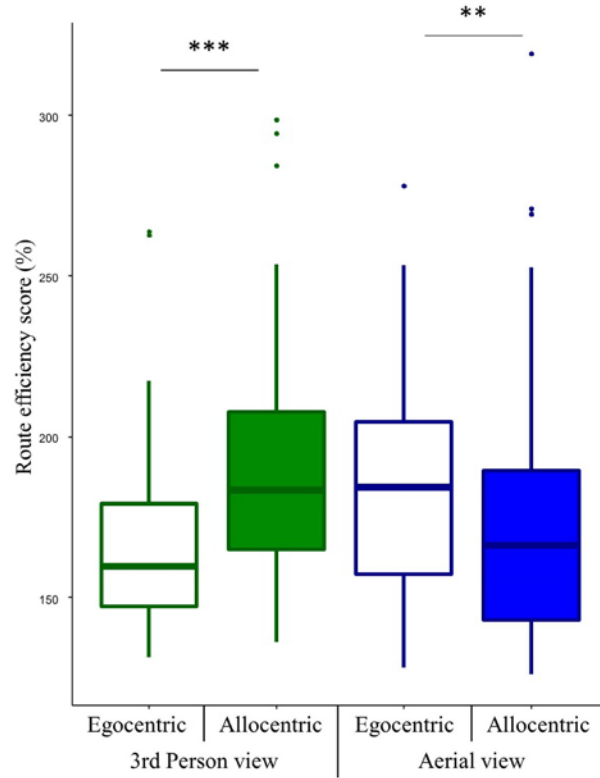
Figure 5 Route efficiency scores according to viewing conditions and reference frames. A significant interaction was found between point of view and frame of reference. In the 3rd person view egocentric frame of reference, while in the aerial view allocentric frame of reference was preferred. Boxplot displays median, first and third quartile ("*bottom and top of boxes*"), and the 95% confidence interval of median ("*whiskers*"). *** : p < .001; ** : p < .01

female subjects (60.24 (*SD* = 9.00) > 54.4 (*SD* = 8.35); *t*(1,48) = 2.378; *p* =0.021, *Cohen's d* = 0.69).

3.3.2 ANALYSIS OF ROUTE EFFICIENCY

We first analysed route efficiency scores (see raw trajectories of one participant on Figure 4). We compared 1P-E and 3P-E viewing conditions to see whether the egocentric 1st person and egocentric 3rd person point of view produced consistently different performance. A paired sample *t*-test showed no significant difference (*t*(1,49)

= 0.280, p = .781, 95% CI[5.8079, -4.3867]). This suggests that the 3P-E point of view is not better or worse for virtual navigation than the natural 1st person egocentric perspective.

We continued by comparing route efficiency for the different viewing conditions in a 2 (*point of view*) by 2 (*frame of reference*) repeated measures mixed ANOVA, using Gender as a grouping variable. Results showed a main effect of point of view ($F(1,48)$ = 8.472, p = .006, η_p^2 = 0.150) indicating that route estimations were better from the ground-level (3P-E, 3P-A) than from aerial point of view (AE-A, AE-E) (see Figure 5). Furthermore, we found a strong interaction effect between *frame of reference* and *point of view* ($F(1,48)$ = 34.178, p < .001, η_p^2 = 0.416, Figure 5). Post hoc comparison, using Tukey HSD test, showed (p < .05) that 3P-A performance (M = 134.59, SD = 14.41) was inferior to 3P-E (M = 124.53, SD = 13.73) performance. Therefore, from the ground-level point of view, an egocentric frame of reference provided better route estimation than an allocentric frame of reference did. Meanwhile, the difference between AE-A (M = 129.80, SD = 15.80) and AE-E (M = 139.22, SD = 19.64) showed that from the aerial point of view, the allocentric frame of reference was preferred (p = .002). The effect of gender on the interaction reached significance ($F(1,48)$ = 4.445, p = .040, η_p^2 = 0.089): female participants showed slightly stronger *frame of reference* and *point of view* interaction.

3.3.3 ANALYSIS OF TIME EFFICIENCY

After the comparison of route efficiency scores, we examined time efficiency scores. Starting again with the baseline comparison between 1P-E and 3P-E conditions, we did not find significant difference ($t(1,49)$ = 0.609, p = 0.545, 95% CI[12.4416, -6.6551]) similarly to route efficiency scores. We then compared time efficiency scores in a 2 by 2 (*Point of view* by *Frame of reference*) mixed ANOVA, using gender as the grouping variable. We found that male participants' time efficiency was closer to optimal than that of female participants ($F(1,48)$ = 4.873, p =0.0321, η_p^2 = 0.0922). Furthermore, results showed an interaction between *point of view* and *frame of reference* ($F(1,48)$ = 48.221, p <0.0001, η_p^2 = 0.5011; see Figure 6). Post hoc analyses of means by Tukey HSD test showed (p <0.001) that 3P-A performance (M = 191.19, SD = 37.77) was again inferior to 3P-E performance (M = 165.54, SD = 29.08). This suggests that in the ground-level point of view, an egocentric frame of reference leads to faster route

planning and execution. Post hoc test also showed (p = 0.022) that, again, AE-A performance (M =174.84, SD = 39.82) was better than AE-E (M = 186.11, SD = 34.04). This provides further evidence that an allocentric frame of reference is preferred when using an aerial point of view. Time efficiency was significantly better (p = 0.029) in 3P-E than in the AE-A condition, but the AE-A condition was better than the 3P-A (p = 0.0005). Gender did not modulate the point of view and frame of reference interaction.
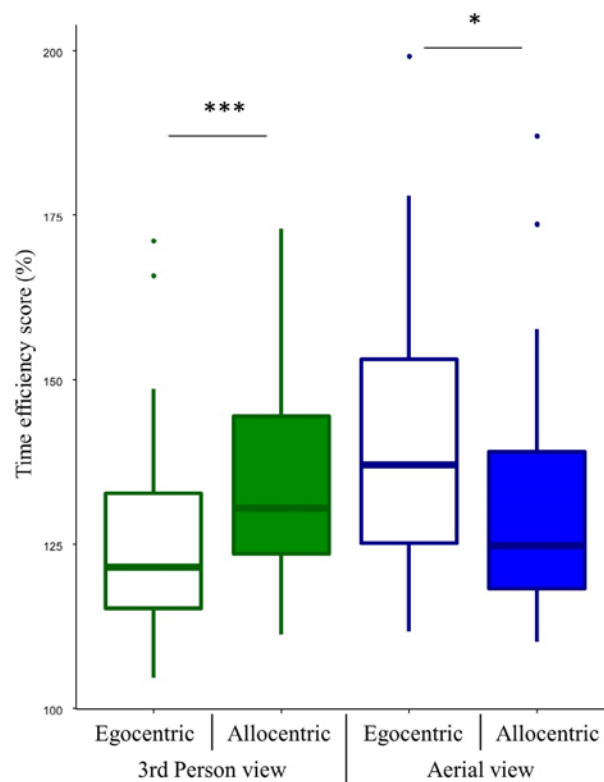
Figure 6 Time efficiency scores according to viewing conditions and reference frames. Significant interaction was found between point of view and frame of reference. In the 3rd person view, egocentric frame of reference was preferred. In the aerial view a preference was present for an allocentric frame of reference. Boxplot displays median, first and third quartile ("*bottom and top of boxes*"), and 95% confidence interval of median ("*whiskers*"). *** : p < .001; * : p < .05

3.4 DISCUSSION

In the present study, we examined the effect of viewpoint and frames of reference on performance in a virtual navigation task. We found that a ground level perspective led to better performance if it was associated with an egocentric, as opposed to allocentric, frame of reference. Meanwhile, when given an aerial point of view, the use of an allocentric frame of reference led to superior performance over an egocentric one. Overall, the ground-level/egocentric combination and the aerial-view/allocentric combination provided users with the best performance conditions; and the former was slightly superior. The results also showed that men performed slightly better in general by collecting more targets in the game. This was partly attributable to the fact that men chose time-optimal routes more often than women and that the interaction between frame of reference and point of view was stronger for women.

Our results are in line with earlier theories suggesting that ground level navigation activates egocentric frames of reference (Linde & Labov, 1975; Siegel & White, 1975). It also agrees with results that the use of orientation fixed maps leads to increased performance (Aretz, 1991; McCormick et al., 1998). Earlier results showed that perspective and frame of reference both affect navigation performance, but the current study provides the first direct evidence that an egocentric reference frame is more effective in ground-level navigation than allocentric and that an allocentric reference frame allows for more accurate navigation in map-like aerial perspectives.

A possible explanation of such association between viewpoint and frame of reference is implied in the study of Waller and Hodgson (2006). In their disorientation study, they found that subjects maintain egocentric localization in blindfolded pointing tasks after less than 135 degrees of rotation but switch to allocentric localization after larger rotations. From ground level perspectives, mental rotations are small so it is easier to match our 3rd person viewpoints with the viewpoint of the avatar. In contrast, an aerial perspective requires larger mental rotations with large potential errors, thus, leaving the allocentric frame as a better option.

We found that the navigation performance did not differ between 1st person and 3rd person viewpoints. Most studies to date have used a 1st person viewpoint for navigation

experiments (e.g. Bird et al., 2010; Caplan et al., 2003; Ekstrom et al., 2003) even though this gives little to no feedback during the task about the position of body parts. However, proprioceptive, vestibular and visual inputs of our own body in space are important for spatial navigation (Ravassard et al., 2013). A possible way to give feedback about the position of body parts during the task could be an external perspective that lets the participant to visually observe them (Marton, 1970). In fact, seeing actions taken on human-like avatars can induce tactile and posture related illusions (Lenggenhager, Tadi, Metzinger, & Blanke, 2007). To note, the current and other studies show that navigationally relevant aspects (e.g. distance) of the environment are equally accurately perceived from both 1st person and 3rd person viewpoint (Lin et al., 2011; Mohler et al., 2010).

An important question derived from our study is to determine which feature of the camera's position caused the switch between ego- and allocentric reference frames. We can consider at least two explanations based on the differences between the aerial and 3rd person cameras used in the current study. One could argue that if the angular difference between the camera view and the avatar exceeds a given value; then, an allocentric reference frame is preferred, which is consistent with the above mentioned finding of Waller and Hodgson (2006). It is also conceivable that simply the change in distance between the camera and the avatar may cause the switch itself. In this case, it would be interesting to see how reference frame use works in a guided navigation situation (e.g. radio controlling a mini car/plane/drone). Further studies are necessary for addressing these questions, for example, by systematically manipulating the distance or the angular difference between the camera and the avatar.

Results related to the role of external perspective bear practical importance from the perspective of urban navigation too (Ball, 2015). Large-scale urban environments are characterized by rich sensory stimulation, high time pressure, and increased levels of stress (Lederbogen et al., 2011; Tranter, 2010). It has been showed that time pressure causes a shift in navigational strategies; from a configurational allocentric to a route-based egocentric one (Brunyé, Wood, Houck, & Taylor, 2016). In a related study, Barra et al. (2012) found that increasing the eye level during navigation (slanted perspective) led to increased activation of allocentric reference frame related areas. These two results suggest that the perspective may play a beneficial role in stressful, time pressure situations. Nonetheless, they did not control the FOV, hence we cannot decide if the

effect is attributable to the more distant perspective, to the more overview, or to the combination of the two.

We found significant gender differences in performances as males overall earned more points in the task and also planned routes faster than women. This result is in line with earlier findings showing that males tend to rely on geometry and path integration, whereas women tend to rely more on landmarks (Andersen, Dahmani, Konishi, & Bohbot, 2012; C.-H. Chen, Chang, & Chang, 2008). To note, we did not find difference in strategy use between genders that is consistent with the results of larger scales studies too (Goeke et al., 2015).

A limitation of the current study is that it involved egocentric controls (left, right) that may also bias performance in favour of egocentric navigation. Thus, further studies should validate the present results in a scenario where allocentric controls are used.

The method of the current study is also novel because, to our knowledge, it is the first implementation of a spatial navigation paradigm for an Android-based tablet PC. Participants were able to control their movements with a multi-touch screen. Although tablet PCs are not yet optimized for neuroscience research, they have an increasing potential for the adaptation of current paradigms. These devices provide a high-resolution display, powerful graphical rendering and are light-weight and able to operate for up to eight hours on their built-in batteries. Relying on battery power is ideal for research because it does not generate AC artefacts and is easy to handle in clinical environments. We believe that multi-touch user interfaces, gesture control, and motion control through webcam are viable alternatives for current keyboard control applications.

In conclusion, we found evidence for default associations between perspectives and frame of reference. First, we found that an egocentric frame of reference was preferred when the perspective was close to the eye level of the navigator and the transformation between our viewpoint and the avatar's was effortless. Second, we found that an allocentric frame of reference is preferred if the perspective is outside of the navigable area (in our case in the air) where viewpoint matching is hard but path integration relative to environmental cues was effortless. Furthermore, we found that 1st person and 3rd person perspectives do not differ regarding navigation performance when the only difference is the presence or absence of an avatar in view. Lastly, we found that men

performed better in our task. The significance of the current results is that they provide the first direct verification for the default frame of reference and point of view for spatial navigation.

4 THE NEURAL UNDERPINNINGS OF NAVIGATION[2]

So far, we have focused our investigation to the behavioural level. However, the neural background of the cognitive map or maps has been of interest from the earliest days of research on navigation (Lashley & McCarthy, 1926; Lashley, 1943, 1950). Furthermore, behavioural evidence suggested the existence of multiple types of cognitive maps in the brain (Howard & Templeton, 1966; Siegel & White, 1975; Stevens & Coupe, 1978; Tolman, 1948; Tversky, 1981). Therefore, in this chapter we summarize results on multiple levels of spatial information processing in the brain from various fields of neuroscience. Because of these various approaches, instead of a unified theory our current understanding of the neural background of navigation is large and complex knowledge network. To facilitate the formulation of valid research questions, evidence from behavioural, cognitive, computational, and systems neuroscience needs to be integrated with cases studies of neurology and results of developmental neuroscience. Thus, we developed a conceptual model to help the understanding of the function and connectivity of brain structures related to spatial navigation (Á. Török, Csépe, et al., 2015) .

We used the ISO 19450 certified framework of Object-Process Methodology (Dori, 2011) for this purpose. Our choice was motivated by two major reasons. First, Object-Process Methodology (OPM) provides a holistic graphical modelling language and methodology. Complex hierarchical models can be created by the recursive use of a minimal set of generic, universal concepts. An important feature of OPM is that the conceptual models created are represented as an Object-Process Diagram (OPD) and as a set of natural English sentences (Object-Process Language, OPL). Second, OPM has been successfully used in systems biology in the understanding of mRNA transcription cycle (Somekh, Choder, & Dori, 2012). Our model is based on several comprehensive reviews (such as Aminoff, Kveraga, & Bar, 2013; Bird & Burgess, 2008; Hartley, Lever, Burgess, & O'Keefe, 2014; Hitier, Besnard, & Smith, 2014; James J Knierim, Neunuebel, & Deshmukh, 2014; A. M. P. Miller, Vedder, Law, & Smith, 2014; Nadel & Hardt, 2011; Nelson, Powell, Holmes, Vann, & Aggleton, 2015; Pennartz et al., 2009). The full model in OPD and OPL format can be found in Appendices 1-8.

---

[2] The chapter and OPM model (including the diagrams presented here) are made in OPCAT by Ágoston Török, based on the discussions with Valéria Csépe.
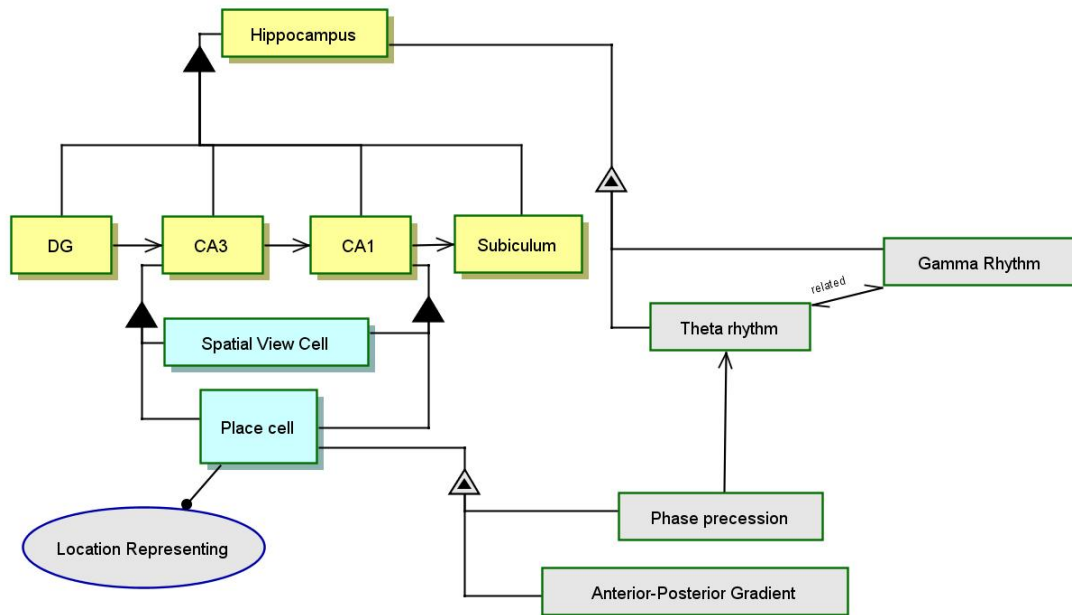
Figure 7 The model of the spatial function of the hippocampus. Information processing inside the hippocampus goes from the dentate gyrus (DG) to the cornu ammonis layer 3 (CA3) then to the CA1 and to the subiculum. Place cells and spatial view cells are most prevalent in the CA regions. Place cells are responsible for representing our location and they show anterior-posterior gradient. Theta and gamma are the primary oscillations in the hippocampus, and single cell firings precess to earlier phases of the theta cycle while the animal moves through the cell's receptive field. Physical object have shades, informatical objects do not. Parts of the hippocampal formation is coloured yellow for convenience.

The first successful attempt to localize navigation related activity in the brain was the exploration of place responsive cells in the rat hippocampus by O'Keefe and Dostrovsky (1971). The locus of their exploration was motivated by earlier results showing defect of maze learning in rats after hippocampal lesion (Hughes, 1965). In their first report, O'Keefe and Dostrovsky recorded activity from the hippocampi of 23 rats. They found a number of cells which fired only when the rat was at certain places of the enclosure. Further study of the spatial selectivity of these cells revealed that most of them fire independent of sensory stimulation, and a substantial amount fires independent of the direction of movement (O'Keefe & Nadel, 1978). Consequently, hippocampal place cells are representing space in allocentric coordinates (see Figure 7).

These results encouraged exploration on the function of the hippocampus for more than forty years. Place cells have been identified in mice (Harvey, Collman, Dombeck, &

Tank, 2009), primates (Hori et al., 2005; Ono, Nakamura, Fukuda, & Tamura, 1991), and direct evidence was found for their existence even in humans (Ekstrom et al., 2003). Research showed that they are most prevalent in the CA1 and in the CA3 region (Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Mizuseki, Royer, Diba, & Buzsáki, 2012). The key interest of these studies was to explore how cell assemblies in the hippocampus code the environment. Two interesting features of place cells have been revealed; these are phase precession and the lack of topographic organization.

Theta oscillations in the hippocampus are regulated by GABAergic interneurons (Freund & Buzsáki, 1998; Klausberger et al., 2003) and are thought to provide the time frame in which single cell firing can be integrated into a same event (Buzsáki & Moser, 2013). An interesting interaction was found between the hippocampal theta rhythm and the activity of place cells (O'Keefe & Recce, 1993). O'Keefe and Recce observed that the individual spikes of a cell advance to earlier phases of the theta cycle as the animal passes through the cell's place field. This (together with results on the gamma oscillations) provides a neural basis of phase-coding of information in the brain (Nadasdy, 2009, 2010).

Functional explorations of the hippocampus structure revealed that the size of place fields exhibit a gradient on the posterior-anterior axis (dorsal-ventral axis in rats), and while cells in the posterior part have small place fields, cells in the anterior end can have place fields of size >1m (Jung, Wiener, & McNaughton, 1994). Further studies showed that their relative size can change with experience. The seminal study of Maguire and colleagues (2000) showed that there is a striking difference in the relative size of their anterior and posterior hippocampi of London taxi drivers compared to controls. Further investigation suggests that while the posterior part is likely responsible for highly accurate position coding, the anterior part is more involved in context coding (Nadel, Hoscheidt, & Ryan, 2013; Zeidman & Maguire, 2016).

In spite of the anterior-posterior gradient, neighbouring cells do not seem to code neighbouring places (M. A. Wilson & McNaughton, 1993). Moreover, although often the same cells are active in different environments, the relationship between their firing fields changed from one environment to the next (O'Keefe & Nadel, 1978). Interestingly, recent analysis of fMRI activation patterns showed that patterns are more

similar if the places are close in physical space, as well (Sulpizio, Committeri, & Galati, 2014).

These results suggest that the cognitive map in the hippocampus is unlike real maps we know. Moreover, primate (Rolls, Robertson, & Georges-François, 1997) and human studies (Ekstrom et al., 2003) found cells in hippocampus and in the parahippocampus, whose activity was not place specific but view specific (e.g. fired if a store was visible). Thus, hippocampal activity might not be bound to the current place, but with phylogenetic development it is increasingly less constrained to the present position and more sensitive to mental traveling between places (Dragoi & Tonegawa, 2011; Kardos, da Pos, Dellantonio, & Saviolo, 1978).

One of the most interesting questions about place cells is how they acquire their location-specific responses. It was a widely held assumption that the required computations occur inside the hippocampus (Brun et al., 2002), until results showed that
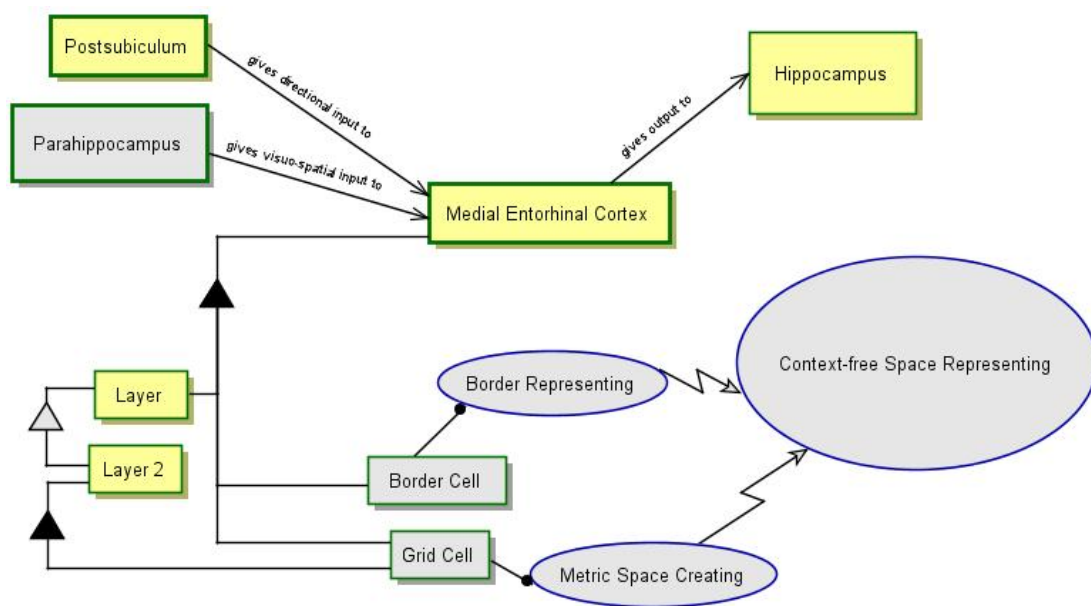
Figure 8 The model of the spatial functions of the medial entrohinal cortex (MEC). MEC receives input from the postsubiculum and the parahippocampus, and output to the hippocampus. It has four layers, of which Layer 2 contains the most grid cells. Grid cells responsible for representing metric space, border cells (also widespread in the MEC) responsible for processing the borders of the space. These two types of cells are underlying context-free space representation. Physical object have shades, infromatical objects do not. Parts of the hippocampal formation is coloured yellow for convenience

place cells preserve their firing field even if the intrahippocampal input is removed. This observation led to the exploration of other areas of the hippocampal formation (Hartley et al., 2014). Since path integration (McNaughton et al., 2006) is required for a location-specific firing pattern, researchers searched for a multisensory area. The entorhinal cortex receives visual (from the parahippocampus/postrhinal cortex) and proprioceptive (from the postsubiculum) inputs, but for a long time, it was believed that the entorhinal cortex contains place cells only with less specific and multiple firing fields (Fyhn, Molden, Witter, Moser, & Moser, 2004; Quirk, Muller, Kubie, & Ranck, 1992). However, when the experimenters used larger experimental environment, a surprising hexagonal pattern emerged from the multiple fields (Hafting et al., 2005). This grid-like pattern tessellated the whole environment, and each cell had a unique phase and grid-size. Moreover, they observed the same anterior-posterior gradient, and, unlike in the hippocampus, in the entorhinal cortex neighbouring cells had similar phases (Hafting et al., 2005). Despite the gradient and similar phases of nearby cells, deeper examination of the grid cell network showed that they do not form a unified map of the environment but likely group into a self-organizing assembly of different orientation and scale (Stensola et al., 2012).

Grid cells quickly develop their firing pattern and preserve it even in darkness, showing that motion related path integration cues are enough to maintain the grids (Hafting et al., 2005). This, however, does not mean that they rely only on proprioceptive cues. Grid cells anchor their orientation to external cues (Hafting et al., 2005; Parron, Poucet, & Save, 2004) and expand their firing field if the compartment size changes (Barry, Hayman, Burgess, & Jeffery, 2007). The representation of geometric borders (by so called border cells) is also associated with the medial entorhinal cortex (Solstad, Boccara, Kropff, Moser, & Moser, 2008). The existence of grid cells has been verified recently both in primates (Killian, Jutras, & Buffalo, 2012) and in humans (Doeller et al., 2010; Jacobs et al., 2013; Nadasdy et al., 2015).

The medial entorhinal cortex receives input from the postsubiculum and the parahippocampus (see Figure 8). The former has been shown to contain spatial view cells (Rolls et al., 1997) and head-direction cells (Taube et al., 1990). Head-direction cells fire whenever the animal is looking in a certain direction in the environment, and they are abundant in the postsubiculum, in the anterodorsal thalamus, and in the mammillary nuclei (Yoder, Peck, & Taube, 2015). Their firing is driven by

environmental landmarks, and removal of those lead to random angular shifts in their preferred orientation (Yoder et al., 2015). However, as we have seen with other cell types earlier, head direction cells also receive proprioceptive and vestibular input (Hitier et al., 2014). Head direction information has been shown to be important also for some higher level areas, like the posterior parietal cortex and the retrosplenial cortex (J. N. Epstein et al., 2011). Heading direction coding is critical for path integration and, thus, for establishing stable spatial firing in grid and place cells. It also seems that head-direction cells are not the same across regions. For example, a special kind of direction specific activity was found in the retrosplenial cortex, which maintained its directional preference through different buildings (in this case museum halls) in the environment (Marchette, Vass, Ryan, & Epstein, 2014).

The other important source of information that reaches the entorhinal cortex is the parahippocampus (PHC). Its function in humans probably is best understood if contrasted with that of the retrosplenial cortex (RSC, see Figure 9) since similar experimental manipulations led to increased activity in both areas (Park & Chun, 2009; Sulpizio, Committeri, Lambrey, Berthoz, & Galati, 2013). Neuroimaging studies of navigation (E. Maguire et al., 1998; Sulpizio et al., 2013) and viewing spatial scenes (Auger, Mullally, & Maguire, 2012; R. Epstein & Kanwisher, 1998) mostly found activity in both places. Neurological data demonstrates that cerebellar strokes often affect these areas and lead to severe orientation deficits (Aguirre & D'Esposito, 1999; Farrell, 1996).

They are both related to first person navigation, and activation is greater in both regions after direct experience compared to studying a map (H. Zhang et al., 2012), which, in contrast, leads to increased activation in the inferior frontal gyrus. They are both active when viewing landmarks; however, the PHC tends to be more related to landmarks that are associated with an action (Chan et al., 2012; Ekstrom et al., 2003; Janzen & van Turennout, 2004). The RSC, on the other hand, is more related to general processing of large, distal landmarks that can serve orientation (Chan et al., 2012). An interesting difference is that while the RSC is sensitive to familiarity of the scene shown in a photo, the PHC is not (R. A. Epstein, Parker, & Feiler, 2007). The same study showed that the RSC activity depends on the question the experimenter asks about the picture; being strongest when a place-related question is asked (R. A. Epstein et al., 2007). In another study, Sulpizio and colleagues (2013) showed that only the RSC activation is modulated

by the amount of viewpoint change relative to the landmark. This is consistent with the single cell electrophysiology results. In the PHC spatial view, cells were found that are active when the target landmark is visible, but their activity loosely depends on the viewpoint (Ekstrom et al., 2003). In contrast to this, the RSC contains head-direction cells which - in principle - are related to the viewpoint (R. A. Epstein, 2008). Furthermore, the landmark's permanence (i.e. whether it is movable or not) is a critical factor only in the RSC (Auger et al., 2012).

These indicate that the RSC plays an important role in processing one's own orientation changes in a known environment. This updating function requires processing spatial relations in both egocentric and allocentric reference frames (C.-T. Lin, Chiu, &
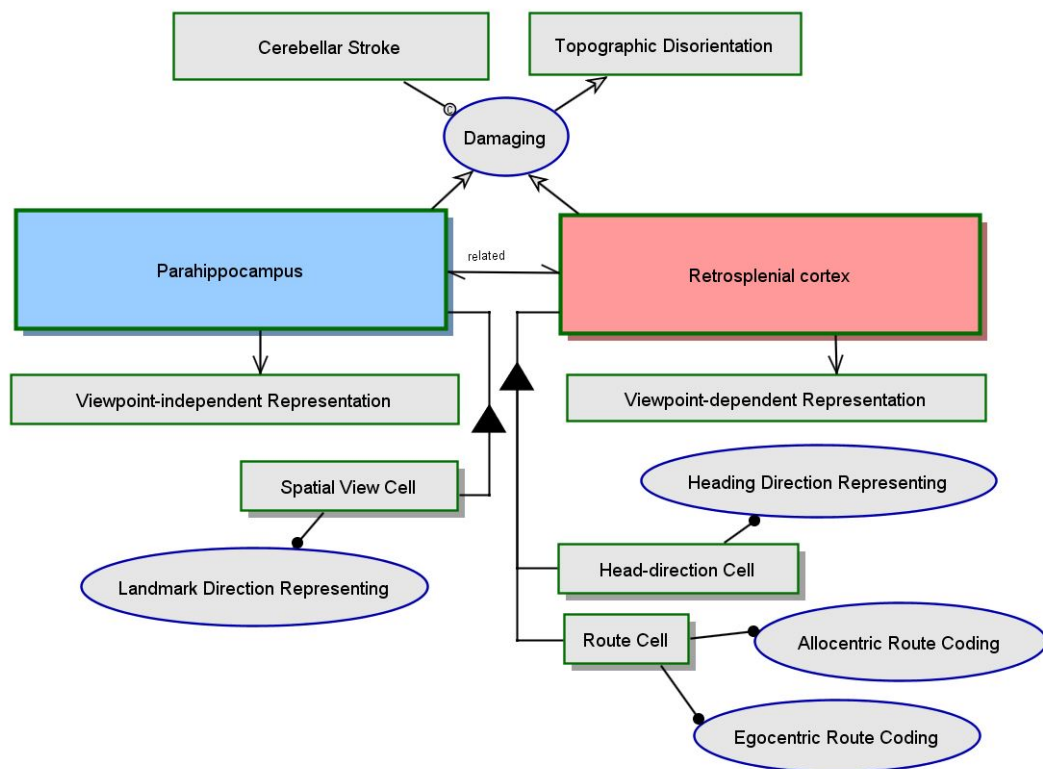
Figure 9 The model of the parahippocampal (PHC) and retrosplenial cortices (RSC) spatial functions. Damage to both areas cause topographical disorientation. Their role is similar but important differences also exist. While the PHC contains spatial view cells, the RSC contains head direction and route cells. This way the PHC is more related to processing of landmarks, irrespective of from where we look at them. The RSC on the other hand is more related to viewpoint dependent coding, and so to path integration. Physical object have shades, infromatical objects do not. The PHC is coloured blue and the RSC is coloured red for convenience.

50

Gramann, 2015). A recent rodent study found evidence for route-cells in the RSC that code in both ego- and allocentric reference frames (Alexander & Nitz, 2015; Nitz, 2006). These pieces of evidence make the RSC (and not the PHC) a more likely candidate for object location coding when a task requires both path integration and reorientation.

Importantly, several other areas contribute to navigation. We summarize them on a higher level of the conceptual model of spatial processing on Figure 10. One area that received attention lately is the striatum, the main part of basal ganglia system (Márkus, 2006). This research is motivated by the observation that while the hippocampus is responsible for incidental spatial learning, the striatum shows increased activity when reinforcement learning is involved in the task (Doeller & Burgess, 2008; Doeller, King, & Burgess, 2008). A recent fMRI study validated these results and found that memory-guided attention is quicker by the hippocampus in a visual search task (Goldfarb, Chun, & Phelps, 2016).

From the perspective of the current thesis, we should note the significant contribution of the frontal, parietal, and occipital cortices to spatial navigation. While, the hippocampal formation codes spatial locations mostly in allocentric frame, the sensory experience leading to these representations is primarily egocentric. The visual stream to the lateral geniculate nucleus and further to the V1-V2 areas of the visual cortex define space in retinotopic coordinates (Tootell, Hadjikhani, Mendola, Marrett, & Dale, 1998). Neuronal representations of space along the dorsal stream (Goodale & Milner, 1992) become progressively independent from the retinal coordinates and increasingly body centred in the parietal and premotor areas of the frontal cortex (Galati et al., 2010).
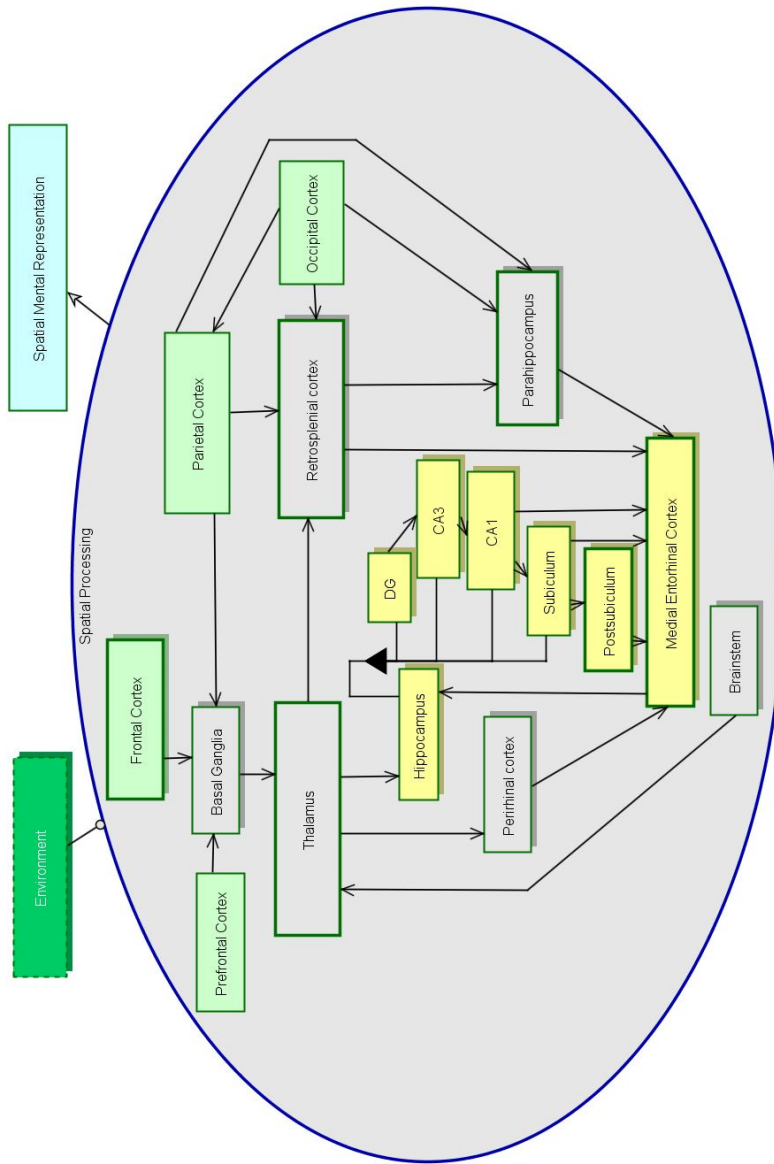
Figure 10 First level of the conceptual model of spatial perception and navigation. Spatial mental representations are constructed through the process of spatial processing, which is based on input from the environment. The model's topology loosely reflects the relative topography of structures; however the size of the Objects has no relation to either the actual size or the importance of function. Structures outside the temporal lobe and midbrain are coloured green, the hippocampal formation (Hartley et al., 2014) is coloured yellow. The direction of the arrows does not exclude the presence of backward connectivity. The black triangle denotes whole-part relation in the OPM terminology

Summarizing this chapter, we found that the structures responsible for spatial location processing receive increased attention for almost a hundred years already. Studies explored the functions of the hippocampal formation, the parahippocampus, and the retrosplenial cortices in navigation. Single cell recordings both in human and in rodents identified different cell types, whose firing activity showed complex spatial specific patterns. Recently, increasing attention is given to the cortical areas in human studies, and most importantly to the parietal and occipital cortices. These results contributed not only to our low-level understanding of the brain but also to a better understanding of spatial deficits, proper target medications, and more successful rehabilitation of diseases and age-related changes affecting these areas (S. L. Bates & Wolbers, 2014; Chouliaras et al., 2013; Fjell et al., 2014; Kunz et al., 2015). Importantly, despite having extensive knowledge on the cortical and subcortical regions involved in spatial computations, the temporal dynamics of location processing of spatial navigation and object location processing are still not well understood. One candidate method to target this question is EEG and event-related potentials (ERPs). Therefore, in our investigation, we used this method to study when we decode the spatial location of objects and how much time it takes to reorient ourselves in a familiar environment.

5 EXPERIMENT 2: THE TEMPORAL ASPECTS OF WAYFINDING[3]

5.1 INTRODUCTION

In the previous chapter, we concluded that while much is known about where spatial processing occurs, less is known about when it occurs. The study of temporal navigation answers important questions, such as when do we process the location information of an object (van Hoogmoed, van den Brink, & Janzen, 2012), when does the identity and location information of an objects bind in perception (Simon-Thomas, Brodsky, Willing, Sinha, & Knight, 2003), or, in general, how the cognitive map is organized in space and time (cf. Lynch, 1960). Related to these questions, in the present study we studied not only when we process spatial information, but we explored the nature of the information we process. For this purpose, we designed a virtual reality paradigm where participants searched rewards in the side alleys of a cross-maze. The critical manipulation was that they randomly started each trial in either the upper or the lower alley of the maze. This way, we were able to explore whether the location of the reward object is processed in an allocentric ('one' or the 'other' alley) or in an egocentric (alley on the left or on the right) reference frame.

Only a limited number of studies used ERP to investigate the temporal aspects of spatial processing (Mollison, 2005; Simon-Thomas et al., 2003; van Hoogmoed et al., 2012; Weidemann, Mollison, & Kahana, 2009). In their seminal study, Baker and Holroyd (2009) used a virtual T-shaped maze, in which participants had to make consecutive left and right choices to maximize the rewards found in the alleys. They identified an ERP component, termed the topographical N170 (referred to as Nt170), which was found to be sensitive to the egocentric location of an object. Their main finding was that the latency of the P1 - Nt170 complex was shorter when the object was in the right alley as compared to the left alley. They verified that this effect was related to the spatial

---

[3] The experiment in Chapter 5 was designed by me, Ferenc Honbolygó, and Andrea Kóbor. The experiment was written by me in XML using the experiment controller extension implemented in Virca by György Persa and Péter Galambos. The experiments were conducted by György Persa, Orsolya Kolozsvári, Gabriella Baliga and Zsuzsanna Kovács. The analysis of the behaviroural data was done by me and Borbála Tölgyesi. The electrophysiological data was analysed by me.

location of the object and not to the reward value (Baker & Holroyd, 2009) and also that it could only be recorded if the task was done in a navigation context location (Baker & Holroyd, 2013).

Nonetheless, the location of an object can be represented in at least two ways (Klatzky, 1998). In allocentric reference frame objects are defined in relation to other objects, while in egocentric reference frame the observer's position serves as a reference frame, and objects are defined relative to the observer (Klatzky, 1998). Previous studies suggested that the preferred reference frame in a task depends on personal preferences (Gramann et al., 2010), viewpoint taken (Á. Török, Nguyen, Kolozsvári, Buchanan, & Nadasdy, 2014), scale of disorientation (Waller & Hodgson, 2006), and the instructions (Iglói, Zaoui, Berthoz, & Rondi-Reig, 2009; P. U. Lee & Tversky, 2001; Taylor & Tversky, 1996). For example, in one study Waller and Hodgson (2006) found that while after small rotations, participants made pointing errors consistent with the use of an egocentric reference frame, after more severe disorientation, they showed a switch to an allocentric strategy. Therefore, in the current paradigm we used severe disorientation, and participants randomly started in either the South or North alley of the cross-maze and had to reorient themselves in the beginning of each trial. We hypothesized that this manipulation will favour the use of an allocentric reference frame, of which ERP correlates have not been investigated yet.

The current study is also the first ERP study to investigate spatial reorientation. The neural correlates of spatial reorientation has been already studied in rodents (Jezek, Henriksen, Treves, Moser, & Moser, 2011) but only recently in humans (Vass et al., 2016). In their study, Vass and colleagues (2016) introduced teleportation gates in a virtual town. They found that when participants teleported from a distant place, hippocampal oscillations did not diminish as compared to the case when they travelled smoothly between distant locations, suggesting that spatial processing is maintained. Moreover, they found that it is possible to classify short distance and long distance teleportations solely based on the measured oscillation. In the current study, we investigated how participants identify their current orientation after teleportation events.

Previous rodent studies showed that the cross-maze design activates mainly allocentric processing (Botreau & Gisquet-Verrier, 2010; Chang & Gold, 2003; Packard & McGaugh, 1996). However, they also showed that whether rodents learn egocentric or

allocentric strategy depends on whether the experiment rewards one or the other strategy. Previous studies showed egocentric location coding of objects in similar tasks where none of the alleys was rewarded more than the other (Baker & Holroyd, 2009, 2013). This could imply that, although rodents prefer allocentric strategy in the cross-maze, humans may use egocentric strategies by default (Spelke et al., 2010; Wang et al., 1999). Therefore, we did not reinforce either alleys in the current paradigm; consequently, results would show whether the frequent and unpredictable reorientation events present in the cross-maze paradigm favour the use of allocentric reference frame or not.

Based on these results, we hypothesized that (1) participants reorient themselves in each trial, and the electrophysiological correlates of this process can be measured with EEG, (2) participants not only use primarily allocentric strategies in the task, but (3) event related potentials (ERP) time-locked to the appearance of the feedback objects would differ for object appearing in the East vs. West alleys. Though it is worth noting that we use the terms left, right, East, and West here for convenience, this does not necessarily mean that participants used exactly these labels during task solving.

5.2 MATERIALS AND METHODS

5.2.1 PARTICIPANTS

EEG data was collected from 38 participants. Four participants' data was later excluded because of failing to meet the inclusion criteria (see below) or recording error. Of the remaining 34 participants, 18 were females. Participants were naïve to the aims of the study, and all of them were right-handed. Their mean age was 22 years (SD = 2.26, Min = 19, Max = 29). They had normal hearing and normal or corrected-to-normal vision. Participants were neither colour nor stereoblind. They were university students from either the Budapest University of Technology and Economics or the Eötvös Loránd University and received payment or course credits for their participation. They gave informed consent prior to the experimental session. The study was approved by the Ethical Review Committee for Research in Psychology (EPKEB).

5.2.2 APPARATUS AND STIMULI

The experiment was run at a CAVE virtual reality environment (Cruz-Neira, Sandin, & DeFanti, 1993) of the 3DICC Laboratory, MTA SZTAKI. Participants sat in a

comfortable chair in the centre of the virtual environment they were surrounded by three screens (3(width) by 2(height) m each) on the front and the two sides. They wore stereoglasses (Infitec), and three-dimensional stereopsis was generated by two projectors on each screen (passive stereo projection). Motion was controlled by the *Left* and *Right* arrow keys on a keyboard placed in the lap of the participant.

The virtual reality environment was a cross shaped maze (see Figure 11). The maze consisted of 4 alleys, each of which had different textures on the walls. The maze's diameter was 7 m, and alleys were 3 m wide. The maze was rotated between participants so we were able to counterbalance the effect of physical difference between textures. There was a platform with a 1 m diameter in each alley, and reward objects were presented floating over it. A 0.5 m tall and 0.5 m wide yellow (golden) apple and a similar sized blue (magic) plum were used as feedback objects. The scenario was programmed in NeuroCogSpace, a custom xml interface built in the VIRCA environment (Galambos & Baranyi, 2011; Persa et al., 2014).

EEG was recorded from 62 sites placed according to the 10/20 system. Recording was done with BrainAmp amplifiers and MOVE system (Brain Products GmbH) with 1000 Hz sampling rate. An online 0.1 – 70 Hz bandpass filter was applied during acquisition.

5.2.3 PROCEDURE

Before starting the task, the experimenter explained the task to the participant with a video presentation. Participants were placed in a cross-maze where they had to collect as many points as they could. They were told that the appearance of reward and non-reward objects follow a complex rule (in reality they occured randomly). They started each trial in one of the vertical alleys. First, they saw the intersection for 800 msec; then, a double arrow sign appeared at the centre of the maze. They were told to choose one horizontal alley when the sign appears. We did not limit their time for the choice. After they made their decision, they were translated and rotated to face the chosen alley in 550 msec. In pilot experiments we made sure that the speed of the translation and rotation was not too fast and/or caused nausea. 500 msec after they arrived to the alley, the reward stimulus was presented for 800 msec. For half of the participants, the golden apple valued 5 points; for the other half, the magic plum was the reward. The non-reward object valued 0 points. After the feedback stimulus disappeared, they were teleported (white screen for 300 msec) to either of the vertical alleys to start the next

trial. Participants were told that the teleportation follows a random order. Figure 11 shows the timing of each part of a trial.
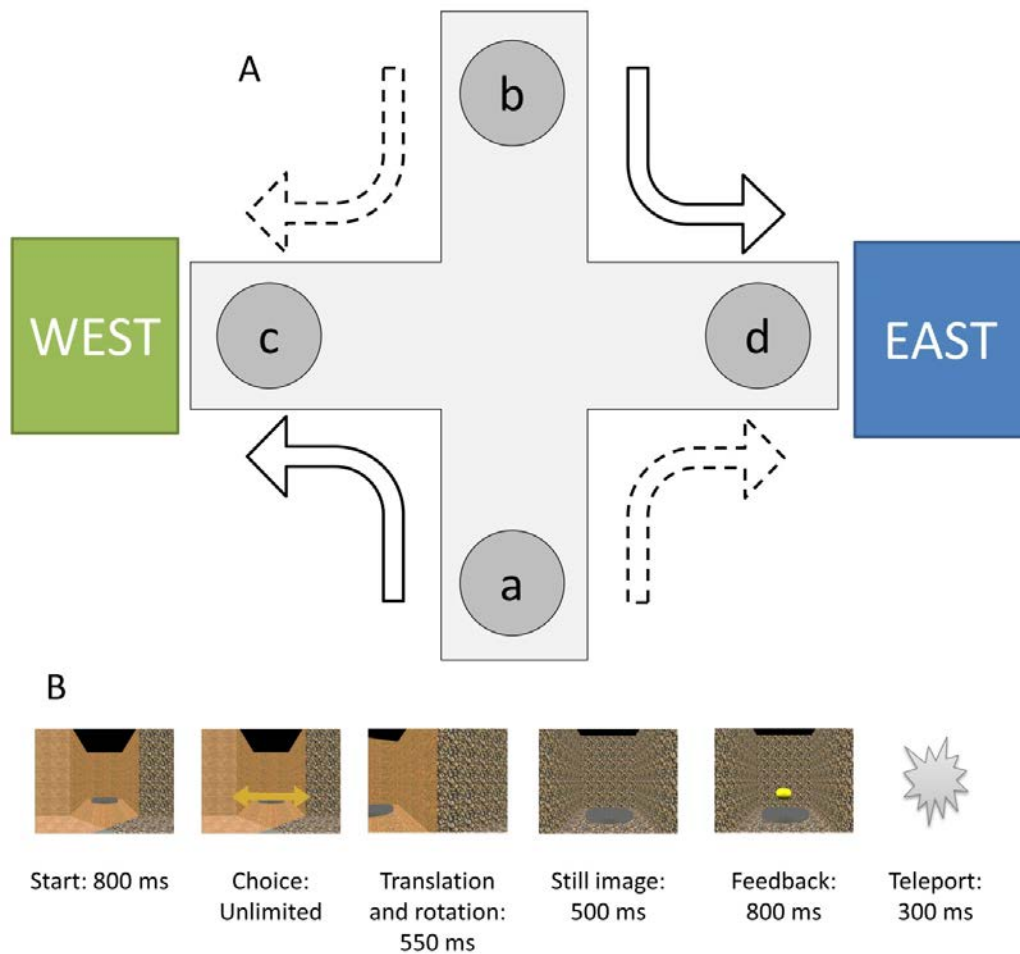
Figure 11 The layout of the cross-maze and the trial timeline. A. Participants started either in point a or b and were told to choose between the two horizontal alleys. After they chose, they were translated and rotated to look into the chosen alley where the reward object appeared (point c and d). In the turn choice analysis we considered only trials where after a rewarded trial the next trial started in the opposite alley. For example, if the first trial started in alley a and the participant chose alley d (and the reward was in this alley) the next trial was analysed only if it started in alley b, where egocentric reference frame predicts alley c and allocentric predicts alley d. For an easier interpretation of the consecutive figures, hereafter left turns are marked with continuous and right turns with dashed lines, whereas turns that led to the West alley are with green color and turns that led to East alley are with blue color. B. Illustration of a trial's timeline. First, participants saw the opposite starting alley with the intersection; after 800 ms, a green arrow sign appeared in the intersection. After participants made their choice, they were virtually translated and rotated (550 ms) to face the chosen alley. After they arrived to the chosen alley, they watched the alley for 500 ms, and then the feedback stimulus was presented. The feedback was visible for 800 ms, then, the screen turned white; participants were teleported into one of the possible starting alley, and the

The experiment started with a practice phase of 130 trials, where participants always started from the lower alley. After the practice phase, 4 blocks of 100 trials were recorded. Trials were presented in pseudorandom order in each block, where no more than three of the same starting alley followed each other; however, reward and non-reward trials followed each other in random order. That is, in contrast to previous rodent studies but in line with the studies of Baker and Holroyd (Baker & Holroyd, 2009, 2013), we did not reinforce one specific alley, but each alley was rewarded with equal probability (50%). The experiment lasted cca. 90 minutes with the electrode cap mounting and debriefing.

5.2.4 EEG AND STATISTICAL ANALYSES

Preanalysis of the electrophysiological data was done using Matlab and EEGLAB (Delorme & Makeig, 2004). First, data was re-referenced to average reference (Bertrand, Perrin, & Pernier, 1985; Doise, Mugny, & Perret-Clermont, 1982), and the original reference was retained (FCz). Then, we filtered the data with a 0.2-30 Hz band-pass FIR filter according to the directions of Rousselet (Rousselet, 2012). Continuous EEG was epoched using a - 100 msec and + 500 msec window relative to the appearance of (a) trial starts and (b) feedback objects in the side alleys. Data were then decomposed by independent component analysis (Delorme & Makeig, 2004). For each participant, the ICA returned 63 components. We rejected components carrying eye blinks and muscle artefacts, then, recomposed the channel based data. Moreover, we rejected every epoch where the EEG signal exceeded a +/- 100 μV limit within the -100 to 500 msec time window. Baseline potential was calculated using the -100 – 0 msec window.

The analysis consists of two parts. First we analysed the trial starts. Because each trial started with the participant randomly placed in either the South or North alley (see Figure 11), they had to reorient themselves every time. Therefore, we looked at whether the ERPs time-locked to the start events differ for the two starting positions. Second, we analysed ERPs time-locked to the appearance of the feedback objects. Previous studies showed that ERPs relative to feedback object appearance do show processing of spatial information besides the feedback value related cognitive processing (Baker & Holroyd, 2009, 2013). The location of the feedback object is an important aspect of the task because it helps finding the strategy to maximize reward. Location of the feedback

object can be processed in two reference frames: it can be either in the 'left' or 'right' alley or it can be in the 'East' or 'West' alley. Due to the cross-maze design, the current paradigm made it possible to differentiate between these two reference frames.

Although there exists previous ERP literature on the topic (Baker & Holroyd, 2009, 2013; Simon-Thomas et al., 2003; van Hoogmoed et al., 2012), our knowledge is still in an early stage of when and how the spatial location of objects can be measured using ERPs. Moreover, the study of spatial reorientation with ERP is without precedence (for ERSP evidence see Gramann et al., 2010, 2006; C.-T. Lin et al., 2015). Therefore, we started both parts of the analysis by conducting exploratory analysis time point by time point from 0 to 300 msec. Point-by-point tests on all electrodes can inflate the possibility of false positive results due to the multiple comparisons problem (Murray, Brunet, & Michel, 2008). To avoid this, we computed randomization statistics in Ragu (Koenig, Kottlow, Stein, & Melie-García, 2011) with a significance threshold of 0.05 and 1000 randomization runs (Koenig et al., 2011). Duration thresholds were established based on global duration tests. Time points with inconsistent scalp topography between subjects were excluded from further analysis (Koenig & Melie-García, 2010). Randomization statistics were calculated for global field power (GFP) and for topographic dissimilarity (TD) (Koenig & Melie-Garcia, 2009; Wirth et al., 2008). These two measures provide a reference free measure of change in the strength (GFP) and distribution/topography (TD) of event related EEG scalp dynamics (Lehmann & Skrandies, 1980; Murray et al., 2008). After the topographic analysis, differences in topography were further explored on the electrodes where the difference scalp topography was greatest. Here the results on the analyses are reported with False Discovery Rate (FDR) and Cluster method corrections applied (Maris & Oostenveld, 2007).

5.3 Results

5.3.1 Behavioural results

We analysed the participants' alley choices to see whether they show preference to any of the two reference frames. During the debriefing, participants reported several

complex strategies (or strategy snippets); therefore, a simple preference of one alley may not be an adequate measure of reference frame preference.

Instead, we followed two complementary strategies. We looked at (1) how rewards affect the participants' choices and (2) whether complex patterns can be identified in their choice sequences. We hypothesized that if rewarding affects their strategy, earning reward in one alley would lead to the choice of the same alley again in the next trial. Furthermore, the introduction of the teleportation made it possible to differentiate between preference for egocentric and allocentric reference frame use. Due to this, we considered only those trials where after a rewarded trial, the next trial started from a different starting alley. Here, because the egocentric and allocentric preference predicts different alleys of choice, we were able to identify which frame of reference was dominant. Using a binomial regression we tested whether participants prefer one alley over the other. According to the results, on average participants preferred the use of an allocentric (61.23%) over an egocentric reference frame (38.76%; β =.45, z (33) = 13.95, p < .001) in their choices.

Next, we used knowledge discovery strategy (Han, Kamber, & Pei, 2011) to identify frequent patterns in the choice sequences. Based on the participants' reports, we assumed that favoured strategy snippets would occur multiple times throughout the task. Further, because the choice sequences can be defined both in egocentric (sequence of Left and Right responses) and in allocentric form (sequence of West and East responses), we were able to identify frequent patterns in both types of strategies and were able to contrast which sequence coding scheme predicts more complex strategies. We mined frequent patterns using the generalized sequential patterns (GSP) method (Srikant & Agrawal, 1996). In this method, repeated scans of the whole sequence are run. Starting with the set of atomic sequences, in each run, a one element longer candidate sequence is generated from the found frequent patterns. We defined the minimum support threshold as 3; thus, only patterns repeated at least three times were considered. Our choice of method was motivated by its easy implementation, and because of the dichotomic nature of sequences, relatively few number of candidates were generated in each run (Pei et al., 2004). Further, given that unsupervised learning was done, the method generated probable frequent patterns that are included in each other. For example, in case of a nine times repeating pattern of [East, West, West, East] and a ten times repeating pattern of [East, West, West] it is reasonable to say that in

reality the longer pattern was the actual frequent pattern and the shorter is only a subpattern of it. Hence, we applied lossy compression by δ clustering using a cover threshold of 0.3 (Xin, Han, Yan, & Cheng, 2005) and identified only one representative for δ-covered pattern sets. For this, we defined the distance between a pattern and its subpattern using the following equation:

$$D(P, P_r) = 1 - \frac{|S(P) \cap S(P_r)|}{|S(P) \cup S(P_r)|} = 1 - \frac{|S(P_r)|}{|S(P)|}$$

where P is subpattern, P_r is the pattern, D is their distance, and $S(P)$ is the set of sequences containing pattern P, and $S(P_r)$ is the sequences containing pattern P_r. According to our criteria when $D(P, P_r) \leq \theta_{threshold} \mid \theta_{threshold} = 0.3$, the subpattern is dropped and pattern P_r is kept to represent the given δ cluster.
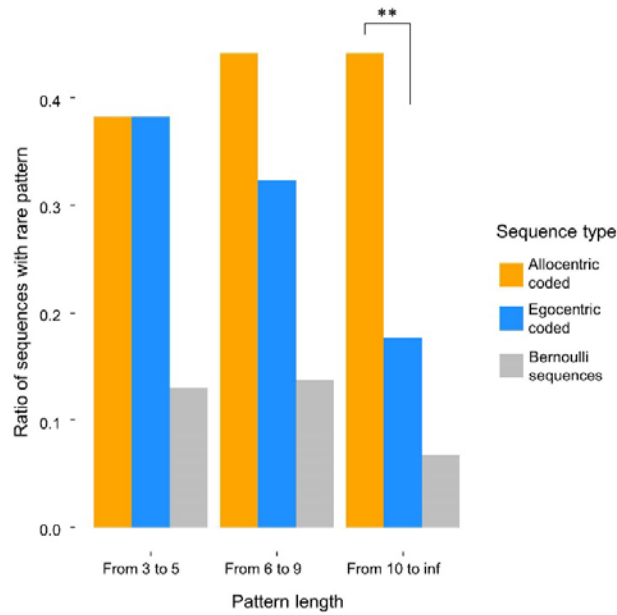
Figure 12 Ratio of sequences with rare pattern in the different conditions. Patterns were binned to short (length 3-5), medium (length 6-9), and long patterns (length > 10). Ratios in the participants' choice sequences differed from random estimated based on 5000 Bernoulli sequences of the same length (p < .001). The ratios between egocentric and allocentric coding schemes differed significantly for the long sequences, but egocentric ratios are visually smaller even in the medium pattern bin. ** : *p* < .01

Next, because in a long Bernoulli sequence frequent patterns emerge by chance, we estimated the probability of the found frequent patterns under the hypothesis of a Markov process with transition probability of .5. For this, we simulated 5000 sequences using a Monte Carlo (MC) method and calculated the probability of a pattern (defined only by its length and support) to be present in a sequence. To note, this method is only sensitive for supports higher than expected. Then, using these probabilities, we identified those patterns in the participants' choice sequences that were unlikely to emerge by chance ($p < .05$). These patterns were defined as rare patterns. We found rare patterns in 22/34 participants' allocentric coded choice sequences and 16/34 participants' egocentric coded choice sequences, both of these ratios were significantly different from the expected (Binomial test, $p(Y = 22 \mid N = 34, P_{expected} = 0.23) < .0001$, 95% CI[0.46; 0.80]; $p(Y = 16 \mid N = 34, P_{expected} = .2334) = .003$, 95% CI[0.30; 0.65]; expected probabilities are based on the MC simulation). This means that it is unlikely that participants just followed an incidental strategy when making their choices. Next, we grouped these rare patterns into three bins: short patterns (length: 3-5), medium patterns (length: 6-9), and long patterns (length: 10+). These bins were chosen based on the expectation the short patterns may easily emerge if the participant is trying out different types of responses; medium patterns, however, require more mental effort as their length reaches the limits of our short term memory span (G. A. Miller, 1956). Any pattern longer than that may require some sort of help in memorization, be it external (e.g. choosing the same well-identifiable place) or internal (e.g. choosing the same button).

For each bin and each coding scheme, we calculated the percentage of participants having at least one pattern for the given bin and coding scheme. We expected that longer patterns would be more common in the sequences which are coded according to the preferred reference frame. Indeed, the number of participants differed only in the bin containing the long patterns (Likelihood tests with Williams' correction; Short patterns: G (1) = 0, $p = 1$, medium patterns: G (1) = 0.976, $p = .323$; G (1) = 5.568, $p = .018$; see Figure 12). Thus, allocentric coding of the sequences contains more long patterns, that is, strategy snippets. Importantly, this does not mean that all participants followed an allocentric strategy, on the contrary, the analysis showed that for a small proportion of them the egocentric coding revealed more rare patterns. Consequently, this method is capable of identifying preferences for strategies in one reference frame at the individual

level and can be, thus, used for clustering. Yet, we found that long rare patterns were significantly more frequent (more than 40 percent of the participant).

Summarizing the analysis of the behavioural data, we found that the two analysis strategies converged and showed that participants were relying more on allocentric strategies rather than on egocentric when solving the cross-maze task.

5.3.2 EEG ANALYSIS

5.3.2.1 ANALYSIS OF TRIAL STARTS

To investigate whether participants reorient themselves in the task, we analysed global field power and topographic dissimilarity time locked to the trial starts. The analysis with global duration threshold revealed difference in global field power from 182 to 262 msec ($p < .05$), where the scalp field power was stronger for North starting position. Difference in global field power means stronger presence of the same scalp topography in one condition. Here it showed stronger activation in the right lateralized parieto-occipital processing with a peak over the O2 electrode (see Figure 13). The difference was also greatest over this electrode. Analysis of the waveforms was done using parametric testing (with FDR correction) and non-parametric testing (with Cluster method correction) on this electrode and showed significant differences between 123 msec and 152 msec and between 175 and 300 msec.
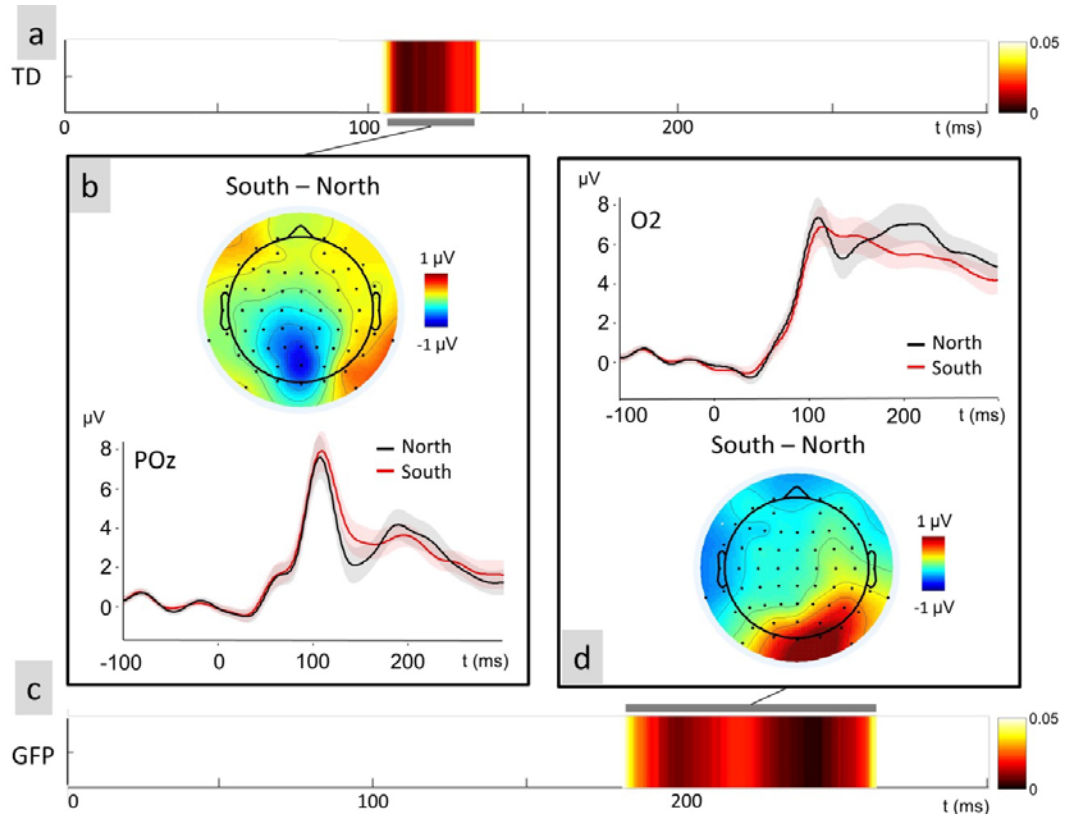
Figure 13 Reorientation at trial starts. (a) TANOVA revealed significant difference between scalp topographies from 103 to 134 ms, (b) the difference had parieto-occipital maxima. On the POz electrode it is visible that reorientation is more pronounced in the North alley and is accompanied with a negative going waveform. (c) The topographic difference is followed by a global field power difference from 182 to 262 ms. (d) The maximum of this difference was centered on occipital electordes and was slightly right localized. The ERP wavefrom on the O2 electrode shows a positive going shift after 175 ms.

The topographic dissimilarity analysis (TANOVA) revealed differences in scalp topographies in the two conditions. Before the analysis, scalp topographies were normalized by the intensity of the signals at each time point; thus, significant results reflect pure topographic differences, probably driven by the involvement of new generators or change in the existing generators. We found a difference in scalp topographies between 103 and 134 msec ($p < .05$, see Figure 13). This activity may refer to early reorientation related activity, when the starting position was the new position that they did not experience during the long learning phase. The difference of the scalp topographies was maximal over the parieto-occipital midline. Therefore, we analysed waveforms on the POz electrode using both parametric testing (with FDR correction) and non-parametric testing (with Cluster method correction). Differences

were found between 114 and 163 msec, where a negative deflection is visible on the waveforms when participants started in the North alley (see Figure 13). Additional differences were found between 179 and 197 msec.

Because both of these activities were maximal on electrode sites where the visual evoked potential maxima are observed, we conducted a control analysis. Namely, despite we deliberately rotated the order of the wall textures for each participants, one could still argue that mere visual differences of the textures caused the effect. This can be tested by comparing the texture related ERP differences regardless of the starting position. Because four textures were used, we created two groups: in one group texture 1 and texture 2 were seen in the beginning of each trial, in the other group texture 3 and 4 were seen. We ran separate analysis of topographic dissimilarity and global field
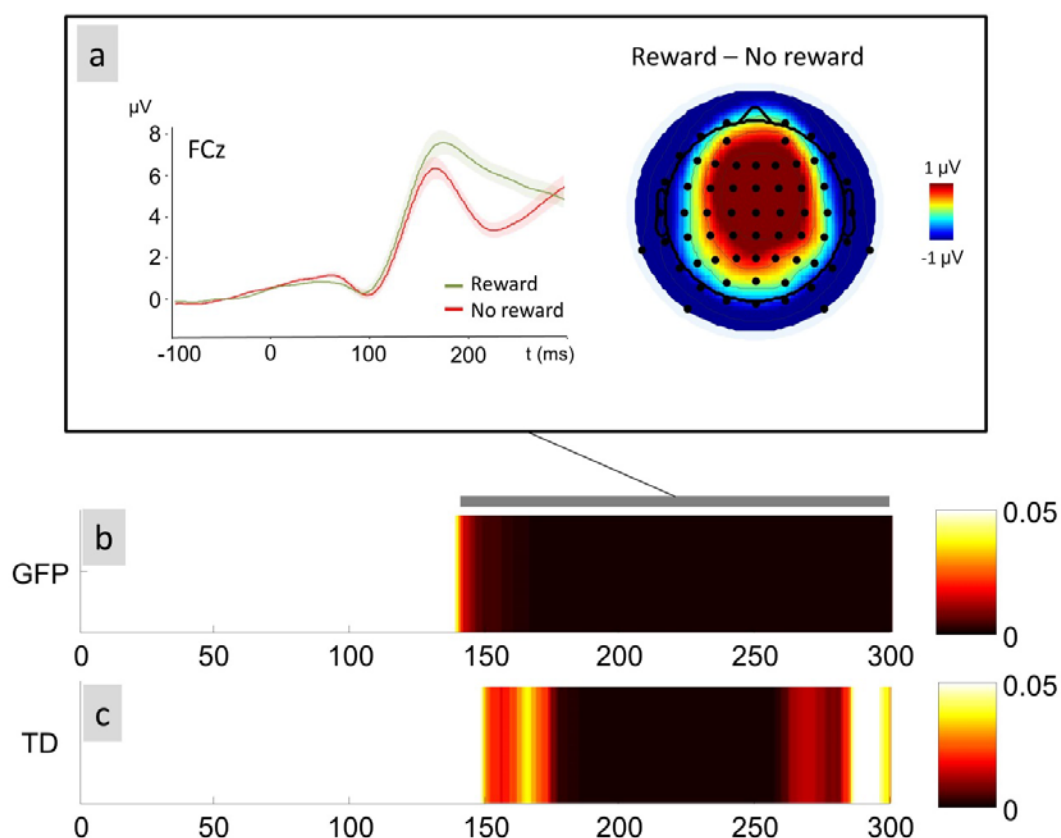
Figure 14 Effect of reward value. (a) The feedback value related activity was a long and pronounced difference between "reward" and "no reward" trials after 150 ms, maximal over the FCz electrode. The difference was apparent both in the (b) GFP and (c) topographic dissimilarity analysis.

power on both groups using the same parameters that we used during the main analysis. This analysis did not yield significant results in neither group (Appendix 9). Therefore, we concluded that the differences found are related to spatial reorientation and not visual texture processing per se.

5.3.2.2 ANALYSIS OF THE FEEDBACK OBJECT PROCESSING

Although the focus of the current study is the better understanding of spatial processing related activity, the feedback objects had rewarding values, and reward value is known to modulate ERPs. Among the most studied ERP correlates of feedback processing, the feedback-related negativity (FRN) is a frontocentral or medial frontal negative deflection occurring 250-270 msec after the onset of a negative (unfavourable) outcome (Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997; Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). The FRN is thought to mirror the rapid evaluation of external feedback and phasic dopaminergic changes in activity between the basal ganglia and the anterior cingulate cortex, as proposed by the reinforcement learning theory (Holroyd & Coles, 2002). The reward positivity (occurring within a similar time window and over similar electrode positions as the FRN) is a positive ERP deflection following rewarding feedback, and its amplitude is larger for unexpected than expected rewards; therefore, it could be regarded as a reward prediction error signal (Foti, Weinberg, Dien, & Hajcak, 2011; Holroyd, Krigolson, & Lee, 2011). Therefore, we started by analysing whether reward and non-reward objects elicit FRN. According to our analysis method, we explored differences in global field power and topographic dissimilarity using randomization statistics. Significant differences were found from 150 -300 msec in the global field powers and between 158 and 288 msec in the topographies. These long lasting differences signalled the processing of reward information. Consistent with our expectations the elicited negativity in the non-reward condition was maximal over frontocentral electrodes (see Figure 14).

Because the feedback related activity appeared to strongly affect ERPs after 150 msec, we included feedback value as an additional factor in the analysis of spatial position related activity. According to the behavioural analysis of the task, participants followed mostly allocentric strategies during the task. Unlike with the explorative analysis of the

68
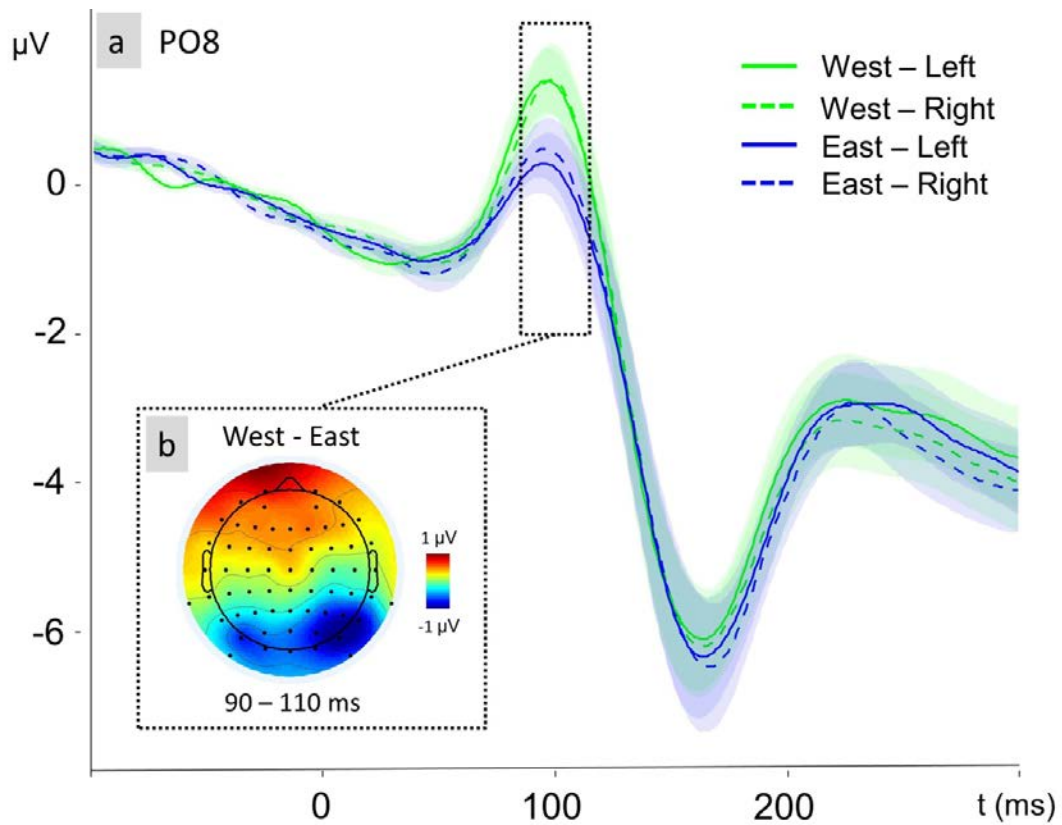
Figure 15 Processing of the location of feedback objects in allocentric reference frame. (a) The topographic dissimilarity analysis showed difference between West and East object locations (but not between Left and Right) in the P1 time window (between 90-110 ms). We show the difference in waveforms on the PO8 electrode, as (b) difference scalp map shows that the topographic difference was caused by an activity with parieto-occipital maxima.

reorientation events, we had prior knowledge of where to expect ERP differences caused by object location. Earlier studies found ERP differences related to object location processing in the time window of P1 an NT170 (Baker & Holroyd, 2009, 2013; Simon-Thomas et al., 2003); therefore, we conducted additional analyses in the time window of these two components. Both GFP analysis and TANOVA were run in the predefined window for the P1 (90-110 msec) and for the NT170 (160-180 msec). The analysis showed significant difference in scalp topographies in the P1 time window (TANOVA, $p = .003$, GFP, n.s.) but not in the NT170 (TANOVA, n.s., GFP, n.s.). The difference was greatest over parieto-occipital sites, consistent with earlier studies (Baker & Holroyd, 2009, 2013; Simon-Thomas et al., 2003). Analysis on the PO8 electrode found difference of ERPs between 74 and 115 msec (see Figure 15). The P1 was more positive when the object appeared in the West alley than when it appeared in the East

alley. The interaction with the feedback value was not significant in any of the two time windows ($p > .5$).

Next, although the choices and the participants' reports reflected allocentric strategies, we examined whether egocentric processing related ERP difference also occurs in the task. Neither the global field power nor the topographic dissimilarity timewise analysis yielded significant result exceeding the duration threshold. As with the analysis of allocentric position related ERP differences, we examined whether there is difference in the time intervals of P1 or NT170 (see Figure 15). We did not find difference in any of the two components, suggesting that robust egocentric processing did not occur in the first 300 msec after the feedback object appeared.

5.4 DISCUSSION

In the present study we sought deeper understanding of the temporal dynamics of object location processing. We designed a cross-maze task where participants started either in the South or North alley and searched for rewards in the side alleys. Using this design, we were able to observe how the human cognitive system processes the current orientation in each trial and decodes object location in allocentric reference frames.

We found that spatial reorientation correlates with topographic difference as early as 100 msec. The focus of this activity was maximal over parieto-occipital sites and signalled extra processing when participants started in the North alley. Importantly, because participants during the long (140 trials) learning phase always started in the South alley, that became their 'usual' starting point. Additional activity related to the processing of the 'unusual' starting point is compatible with this notion. This increased reorientation effort was signalled by subsequent global field power differences between the two starting points after 180 msec. Several earlier studies using EEG found activity on parietal (Chiu et al., 2012; Snider, Plank, Lynch, Halgren, & Poizner, 2013) and parieto-occipital sites (Baker & Holroyd, 2009, 2013; Baker, Umemoto, Krawitz, & Holroyd, 2015; Simon-Thomas et al., 2003) in navigation-related tasks. Importantly, parietal activity was often found to be related to path integration (Snider et al., 2013) and heading direction (Chiu et al., 2012). Although the present study did not use source estimation, subsequent explorations should reveal the possible neural generators underlying this effect.

70

Multiple analyses on strategies of turn choices converged to show a strong behavioural preference for using allocentric reference frame in the task. This finding is in line with previous results of rodent studies that also showed allocentric preference first and a shift to egocentric strategy use only after prolonged training (> 8days) (Botreau & Gisquet-Verrier, 2010; Chang & Gold, 2003). Although we hypothesized a pivotal role for unpredictable teleportation in the strategy choice, there were other motivational, textural, and individual factors to take into account as they might play a role(for a review see Packard & Goodman, 2013). Further investigation is required to identify which factors are of key importance in human VR navigation.Importantly, the knowledge discovery strategy used in the analysis may be used for identifying possible strategies also in other tasks, such as in studies of categorization strategies (Quinn, Doran, Reiss, & Hoffman, 2009; Spencer, Quinn, Johnson, & Karmiloff-Smith, 1997).

Consistently with the behavioural analysis, we found that the peak of the P1 ERP component was sensitive to coding of object location in allocentric reference frames. Similarly to earlier results, the P1 was maximal over right parieto-occipital electrode sites (Baker & Holroyd, 2013; Simon-Thomas et al., 2003). We identified profound activity related to feedback processing. It occurred only after 150 msec over central medial electrode sites. Importantly, we found that the spatial location related activity in the P1 window was not affected by this reward value. Our results supplemenet the interpretation of Baker and Holroyd (2009) stating that the egocentric encoding of object location is conveyed in the latency effect of the Nt170 component. Here we provide evidence that the allocentric encoding of an object is reflected in the difference in the amplitude of the P1 component.

Importantly, none of the ERP differences found in the current study can be attributed to a simple association between textures and reward objects because reward objects were present in both alleys with equal probability. Furthermore, because the orientation of the cross-maze also varied randomly between participants and we summed ERPs according to Left/Right and East/West alley turns, ERPs cannot reflect any texture related cognitive process. Note that earlier studies (Baker & Holroyd, 2013) did not reveal topographical modulation of the egocentric Nt170 component when the task was presented in a non-spatial context. This suggests that the presented effects are indeed related to spatial processing. We also analysed whether there were reward related

changes coinciding with the spatial differences and found that (1) reward based processing starts only later in time and (2) with fronto-central topography.

The simplicity and intuitiveness makes this paradigm a promising candidate for neuropsychological testing with elderly individuals. For instance, impaired navigation ability is one of the first signs of Alzheimer's disease (Kunz et al., 2015; Lithfous, Dufour, Blanc, & Després, 2014; Lithfous, Dufour, & Després, 2013). ERP could be a powerful tool to recognize signs of Alzheimer's disease and other dementia even before the appearance of behavioural symptoms. Furthermore, the availability of consumer virtual reality displays (e.g., Oculus Rift) and EEG headsets (e.g., Emotiv Epoc) make it even easier to use paradigms like the cross-maze in clinical research in the near future.

6 PERCEIVING SPACE THROUGH MULTIPLE SENSES

In the previous chapter, we presented the results of a virtual reality experiment. The results showed that participants were immersed in the virtual reality scenario as they reoriented themselves in the beginning of each trial in spite of having not moved physically between trials. This finding is not unique to the current setup. Human recordings of place cells (Ekstrom et al., 2003), grid cells (Doeller et al., 2010; Jacobs et al., 2013), and direction selective neurons (Chadwick, Jolly, Amos, Hassabis, & Spiers, 2015) were all done in a virtual reality setup where actual locomotion was not present. Actually, this is a limitation of currently available technology that is frequently used in the study of spatial cognition. On the one hand, single cell recordings in humans are only possible in pharmacologically intractable epilepsy patients having electrode implantation to better localize the seizure generators for the sake of the surgery (Clemens et al., 2013; Ekstrom et al., 2008; Á. Török, Nguyen, et al., 2013). These patients typically are able to sit or lie in their bed but are often dizzy and are on strong painkiller medication. Hence, any experiment involving active locomotion is not feasible. On the other hand, participants in fMRI experiments are often healthy humans with unconstrained locomotion ability. Here the problematic aspect is the MRI techinque that requires shielding and is not movable due to its size. EEG does not suffer from any of these limitations and, indeed, has been used to record brain activity during active locomotion (Á. Török, Sulykos, et al., 2014). One study (Snider et al., 2013), for example, recorded path integration related activity in the parietal cortex during a task that involved active motion and visual search inside a deck of a virtual ship.

Spatial perception is essentially multimodal, meaning that it relies on other sensory modalities besides vision and proprioception. Indeed, when only visual and proprioceptive cues are present, but auditory and olfactory cues are missing, a significant number of place cells remain silent (Ravassard et al., 2013). Thus, the generalizability of insights from navigation experiments in virtual reality will be limited until simulation is extended to all sensory modalities (Janzen & van Turennout, 2004; Lloyd, 2014; Slater, Usoh, & Steed, 1994; Ziemer, Plumert, Cremer, & Kearney, 2009). In order to understand why multisensory perception is so important for spatial navigation, we describe what multisensory integration is and how it works.

One of the most well-known multisensory phenomenon is the McGurk effect (McGurk & MacDonald, 1976; Huhn, Szirtes, Lorincz, & Csépe, 2009). In the original experiment, a heard *'ba'* syllable integrates with the seen lip movements of a *'ga'* syllable leading to the percept of a *'da'* syllable. However, multisensory integration events usually start at an even more basic level. In their seminal series of experiments, Meredith and Stein (1983, 1986) described how individual cells in the superior colliculus (SC) react to simultaneous stimulation of different modalities. They observed that certain cells multiply their firing rate if they receive input from more than one modality. This facilitation effect exceeded what would have been predicted from the linear summation of the unimodal responses (J. Miller, 1982; Stein & Meredith, 1993). They manipulated several features of the multisensory input that led to the description of three principles in multisensory integration. These were the principles of (1) temporal and (2) spatial congruity, which means that the unisensory constituents of the multisensory percept have to originate close both temporally and spatially, and (3) inverse effectiveness that states response enhancement is strongest when the individual sensory inputs are weak.

Later studies proved that multisensory integration is not restricted to the superior colliculus (Ghazanfar & Schroeder, 2006; Morgan, Deangelis, & Angelaki, 2008). For example, using scalp EEG, ERP evidence was found for multisensory integration as early as 30-50 msec after stimulus onset (Giard & Peronnet, 1999). Interestingly, these ERP responses also follow the principle of inverse effectiveness (Senkowski, Saint-Amour, Höfle, & Foxe, 2011; Á. Török, Tóth, Honbolygó, & Csépe, 2013). The role of inverse effectiveness is to enhance the perception of near-threshold stimuli and allow us to evaluate stimuli around us more effectively (Ohshiro, Angelaki, & DeAngelis, 2011).

The spatial and the temporal principle define the requirement of coincidence. This is useful because if it weren't the case that only coincidental stimuli could be integrated, we would perceive two separate stimuli (e.g. one of which might be a threat) as only one compound stimulus. When both the temporal and spatial requirements are met, each unimodal stimulus has to be processed by a dedicated sensory area that gives output to the multisensory areas. In order for multisensory integration occur, the different unisensory constituents have to reach the multisensory integration areas roughly at the same time. Colonius and Diederich (Colonius, Diederich, & Steenken, 2009; Colonius & Diederich, 2004; Diederich & Colonius, 2004) introduced the time window of

integration (TWIN) model to describe how the coincidence principles contribute to multisensory response facilitation. They state that the neural and behavioural response to a multimodal object depend on two stages of stimulus processing. In the first stage, multisensory signals are processed in separate and encapsulated sensory areas. The processing times of this stage are well characterized by an exponential probability distribution. Completion of this stage opens a TWIN. Two sensory signals are integrated only if their TWINs overlap. Integration is an active process which binds together the different modalities (Senkowski, Schneider, Foxe, & Engel, 2008) or stimulus features (Csibra, Davis, Spratling, & Johnson, 2000). This way a unified percept will reach the second stage which consists of all higher level, temporally overlapping processes, like evaluation, preparation, and execution of a response. Because a multitude of independent factors affect these processes, the processing times of this stage are well characterized by a Gaussian probability distribution. Multisensory integration can either increase (multisensory inhibition) or decrease (multisensory facilitation) the processing times of this second stage (Diederich & Colonius, 2015; Á. Török, Kolozsvári, Virágh, Honbolygó, & Csépe, 2014). Summarizing the TWIN model formally:

$$\widehat{RT} = \frac{1}{\lambda} + \mu - P(i)\Delta$$

Where RT is the expected reaction time, λ is the parameter of the exponential distribution, μ is the parameter of the Gaussian distribution, *P(i)* is the probability of integration and Δ is the multisensory facilitation effect.

However, while coincidence can be easily defined on a physical level, its definition on the neural level is not straightforward. Sensory modalities differ in their respective temporal and spatial localization accuracy. For example, while auditory events are well localized in time, they are only moderately localized in space. The temporal localization accuracy of hearing is so accurate that it can lead to perceptual phenomena, such as the illusory flash illusion (Shams, Kamitani, & Shimojo, 2000). Here, the experimenter shows the participants one single flash and at the same time two beep sounds. Surprisingly, what the participants perceive is two flashes *with* beep sounds.

In the spatial domain vision dominates not only hearing (Vroomen & Gelder, 2004) but also touch (Ho, Santangelo, & Spence, 2009; Kóbor, Füredi, Kovács, Spence, &

Vidnyánszky, 2006), and probably chemical senses (Gottfried & Dolan, 2003). This visual capture effect gives rise to the ventriloquism phenomenon (Howard & Templeton, 1966; Thurlow & Jack, 1973). Here, simultaneously presented but spatially discordant auditory and visual stimuli are perceived as one unified percept originating from the position of the visual stimulus. This phenomenon is part of our everyday life: we experience it in the cinema, in vehicle warning systems, and there is even a puppeteer art form dedicated to it. Although ventriloquism illusion is created at higher level of disparities than the one the subcortical multisensory neurons can tolerate (Bertini, Leo, Avenanti, & Làdavas, 2010), it is still unconscious (Bertelson & Aschersleben, 1998) and preattentive (Stekelenburg, Vroomen, & de Gelder, 2004). Interestingly, when in a sound oddball paradigm the oddballs have different locations but are ventriloquized to the same location as the standards, no mismatch negativity is recorded (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002). This result suggests that ventriloquism occurs on an early cortical processing level.

The question emerges why ventriloquism happens at an early processing stage but not on the level of single neurons in the superior colliculus. In their seminal study, Alais and Burr found that ventriloquism results from the near-optimal integration of sensory inputs in the brain (2004). In their experiment, they presented audiovisual stimuli pairs and asked participants to decide if the second presentation was left to the first. Randomly, either in the first or in the second part of the pair they manipulated the offset between the sound and visual stimulus and also the degradation of the visual stimulus. The visual stimulus was a Gaussian blob; thus, changing the deviation of the Gaussian envelope led to an increasingly blurry image with less discrete focus. They found that when the blob's centre is easily localized, standard ventriloquism happens, and the perceived location is the location of the blobs centre; however, when the blobs centre is severely degraded, reverse ventriloquism occurs, and the perceived location is the sound's position. Additionally, in case of medium degradation, the perceived location will be somewhere between the blob and the sound showing that integration is actually working in a near-optimal way. The experiment of Alais and Burr showed not only that ventriloquism is a result of how multisensory integration works, but they also pointed out that the system works in a dynamic way. Thus, vision does not capture audition under any circumstances, but the brain constantly monitors how well perception and reality matches and is able to fine tune the integration weights (Fetsch, Pouget,

DeAngelis, & Angelaki, 2012). Indeed, several experiments showed that if 'reality is manipulated', the neural system can learn to accept offsets as natural. For example, if in an experiment, the auditory and the visual stimulus has a constant offset, after some trials, perception will be biased to process the offsets as coincident. This phenomenon is called ventriloquism aftereffect (Bertelson, Frissen, Vroomen, & de Gelder, 2006), that can occur after a single exposure (Wozny & Shams, 2011) and persist for minutes (Bertelson et al., 2006).

The importance of the ventriloquism illusion from the perspective of the current work is that multisensory stimulation in virtual reality rarely fulfils the criteria of spatial coincidence. This happens because the reason for using auditory stimulation is mostly to increase the scenario's immersiveness, and not to provide spatialized audio stimulation (Hendrix & Barfield, 1996; Serafin & Serafin, 2004). If spatial audio is required researchers often use a 5.1 surround system, which seemingly creates good spatial audio experience (Skalski & Whitbred, 2010). As we have seen, however, vision not always captures the relative location of sounds and possibly not always dominates other senses even in the spatial domain. Moreover, in a virtual reality setup vertical simulation of audio sources is almost always lacking, hence, in the vertical axis (higher or lower than the location of the speakers), perception of unified stimulus can only be achieved through ventriloquism. Nevertheless, the effectiveness of ventriloquism in the vertical axis is a relatively unexplored in virtual reality. Thus, in the next chapter, we investigated how visual stimuli effectively capture the location of sounds in virtual reality in both the horizontal and the vertical axes.

7 Experiment 3: Vision captures sound in virtual reality[4]

7.1 Introduction

For a long time, virtual reality designers and researchers have been using auditory stimulation to support visualization (Nacke, Grimshaw, & Lindley, 2010; Zhou, Cheok, Yang, & Qiu, 2004). While even a single loudspeaker is sufficient to change the quality of experience, this type of setup has at least one important limitation: it does not provide any spatial information other than its own position. Spatial audio requires, therefore, a more complex approach, but the question emerges what is the minimal setup complexity that will provide good virtual audio environment without noticeable mismatch to the conceived visual content. Several experiments investigated which sound generation technique provides the most reliable spatial information (e.g. Hu, Zhou, Ma, & Wu, 2008; Seeber, Kerber, & Hafter, 2010; Wenzel, Arruda, Kistler, & Wightman, 1993). Here, instead of focusing on proper spatialization of sounds, we investigated whether a horizontal surround speaker setup was capable of creating the illusion of a three-dimensional audio environment *when* perceived in the presence of visual objects.

The human brain uses binaural and monaural cues to localize sound sources (Middlebrooks & Green, 1991). Binaural cues are based on the fact that our ears are placed on the two sides of the head. Therefore, they receive auditory information from the same sound source at slightly different times (Interaural Time Difference, ITD) and at different levels (Interaural Level Difference, ILD). The duplex theory of hearing (Middlebrooks & Green, 1991) states that sound localization is based on ITDs for low frequency sounds (under 1500 Hz) when phase differences are big enough to be perceived. For higher intensities, the shadowing effect of the skull serves as the basis

---

[4] The data of the experiment presented in Chapter 7 has been published. Török, Á., Mestre, D., Honbolygó, F., Mallet, P., Pergandi, J.-M. M., & Csépe, V. (2015). It sounds real when you see it. Realistic sound source simulation in multimodal virtual environments. *Journal on Multimodal User Interfaces, 9(4),* 323–331. For author contributions see Preface.

for sound localization, attenuating the sound while it spreads through it (ILD). These two cues allow good localization in the horizontal plane (azimuth).

In the vertical plane, however, sound localization is more difficult. Sounds on the medial plane cause no ITDs because they are at the same angle and distance from both ears (Middlebrooks & Green, 1991). Vertical sound localization relies on the characteristics of the pinna of which shape and structure modifies the sound's spectrum as it reaches the inner ear. This basis enables only poorer sound source localization in the vertical plane than in the horizontal plane (Middlebrooks & Green, 1991; Thurlow & Jack, 1973).

As we have seen in the experiment of Alais and Burr (2004), the perceptual system takes into account the relative reliability of the senses for optimal multisensory integration. A potential implication of this argument is that the integration does not solely depend on the degradation of the stimuli or the general reliability of a sensory modality but could also be affected by the position where the stimulus appears. Thus, on the vertical axis where sound localization is less reliable, one would expect to get a stronger ventriloquism effect (Hartnagel, Bichot, & Roumes, 2007; Thurlow & Jack, 1973; Werner, Liebetrau, & Sporer, 2013). Taking this into account, one may assume that visual stimuli catch the vertical location of sounds in virtual environments. If this is true, a surround system setup could be a golden mean between highly accurate binaural and easy-to-install one speaker solutions, especially because surround speaker systems are easy to install and broadly available in the consumer market. Moreover, this kind of audio stimulation is readily available in most VR labs.

Our hypotheses were the following:

1. Sounds can be ventriloquized in the vertical plane; therefore, it is not necessary to provide vertical auditory spatial cues, and a sound system of good horizontal resolution is enough to provide a realistic audio-visual environment.
2. Sounds can be ventriloquized in the horizontal plane, thus, a small mismatch or slight scarcity of sound simulation (e.g. because of asymmetric room reverberation characteristics) does not lead to measurable changes in perception.
3. In multimodal situations, sound source localization in environments using surround systems is as good as in environments using free field speakers.

4. The ventriloquist effect differs in the horizontal and vertical plane when using surround speakers.

In order to test these hypotheses, we designed two experiments in a CAVE (Cave Automatic Virtual Environment) setup, installed in the Mediterranean Virtual Reality Center (www.crvm.eu). Participants were asked to locate sound sources occurring with or without simple visual stimuli (Gaussian blobs). The paradigm was based on the study of Besson et al (2010). Sound sources were either free field speakers (Free field condition) or their simulated copies delivered through a stereo speaker set (Surround condition). We used left, middle, and right sound directions to test whether there was any difference in the ventriloquism effect depending on the location from which the participants heard the sound. Visual stimuli were placed on the vertical plane in the first experiment and on the horizontal plane in the second experiment.

7.2 METHODS

7.2.1 PARTICIPANTS

Participants were recruited as volunteers from Aix-Marseille University, Marseilles, France. Six participants (1 female, *M*: 32.4yrs, *Min*: 25yrs *Max*: 48yrs) took part in the first and five (1 female, *M*: 27yrs; *Min*: 21yrs, *Max*: 41yrs) in the second experiment. Participants were tested for normal hearing and had normal or corrected to normal vision. The experiment used stereoscopic virtual reality, therefore we also measured the participants' stereo vision using stereoscopic random dot figures (Randot Stereotests, Stereo Optical Co.). Stereo vision was adjusted for each participant based on their interocular distance. Each participant took part in only one experiment. They did not receive any compensation for the experiments. Written informed consents were collected prior to the experiments. The study involved exclusively non-invasive perceptual measurements and was approved by the Institute of Movement Science Laboratory Review Board. The experiment was conducted in accordance with the Declaration of Helsinki.

7.2.2 APPARATUS

The experiment took place in a dimly lit hall designed for the virtual reality equipment. The walls were painted black and the hall had no windows. The Cave Automatic Virtual Environment, using its common abbreviation CAVE (Cruz-Neira et al., 1993) had a set of four screens. Three 3 x 4 m displays (one frontal and two lateral) and the ground, a 3 x 3 m fiberoptic floor, was also illuminated. Participants sat in a comfortable chair at a 1.2 m distance from the frontal screen with their eye-level at about 1.15 m from the ground level of the CAVE. For the experiments, we defined our setup in a way that all visual stimuli were on the frontal screen; thus, we avoided any bias caused by brightness transitions on the edges of the screens. The frontal screen's resolution was 1400x1050 pixels. Visual stimuli were light-blobs (visual angle 7.6°) with a Gaussian grading. Blobs were presented for 16.67 msec (one frame). The baseline luminance of the screen was 0.006 cd/m2, and the luminance of the visual stimuli was 0.35 cd/m2. Participants wore passive stereo-glasses (Infitec) and the projectors used static stereo image rendering. At the beginning of each trial, a fixation cross appeared on the screen in the centre at 1.1 m height.

The acoustic stimuli were broadband noises of 16.67 msec duration, high pass filtered at 250 Hz. Sounds were delivered via 7 identical speakers of a 7.1 surround system (Creative Inspire p7800); sound pressure level at the participants position was 65 dB. Speakers were placed on a 2.99 m radius circle with its center at 0.76 m from the ground level of the CAVE in the participants' position. The 7 speakers were placed 10.5° from each other with speakers 2, 4, and 6 as free field speakers, and speakers 1 and 7 for the surround condition. Speakers 3 and 5 were not used in the current experiment. We used panning (inter-speaker sound level differences) to create the stereo sounds, and sound levels were matched between the free field and surround conditions. The participants used a flystick (ART Flystick 2) to respond. The flystick's position was logged by infrared cameras (ART) with high precision; in this way, participants could easy and naturally locate sounds. The three dimensional orientation of the flystick was used as an indication of the perceived sound direction.

7.2.3 PROCEDURE

Each trial began with a fixation cross. Participants were asked to move the cursor of the flystick to the fixation cross. This way we ensured that at the start of each trial their hand was in the same position and also that they fixated the central cross. The fixation

cross disappeared after 1000 msec, and the test stimulus occurred with a 50 msec onset delay. Each trial consisted of either a single auditory or an audio-visual stimulus presented for 16.67 msec followed by 420 msec blank screen. Then, the cursor appeared, and the participants had to respond as accurately as possible by moving it to the location of the auditory stimuli. A new trial started after the response. The participants were asked to locate the auditory stimuli but were asked to always keep their eyes open until the end of the experiment to ensure that they see the visual stimuli, too.

In Experiment 1, visual stimuli had vertical offset relative to sounds. There were 3 sound positions (left, middle, right), 2 sound types (free field, surround), and 6 visual positions (no visual, -21°, -10.5, 0° (same position as the sound), 10.5°, 21° relative to the sounds on the vertical axis).
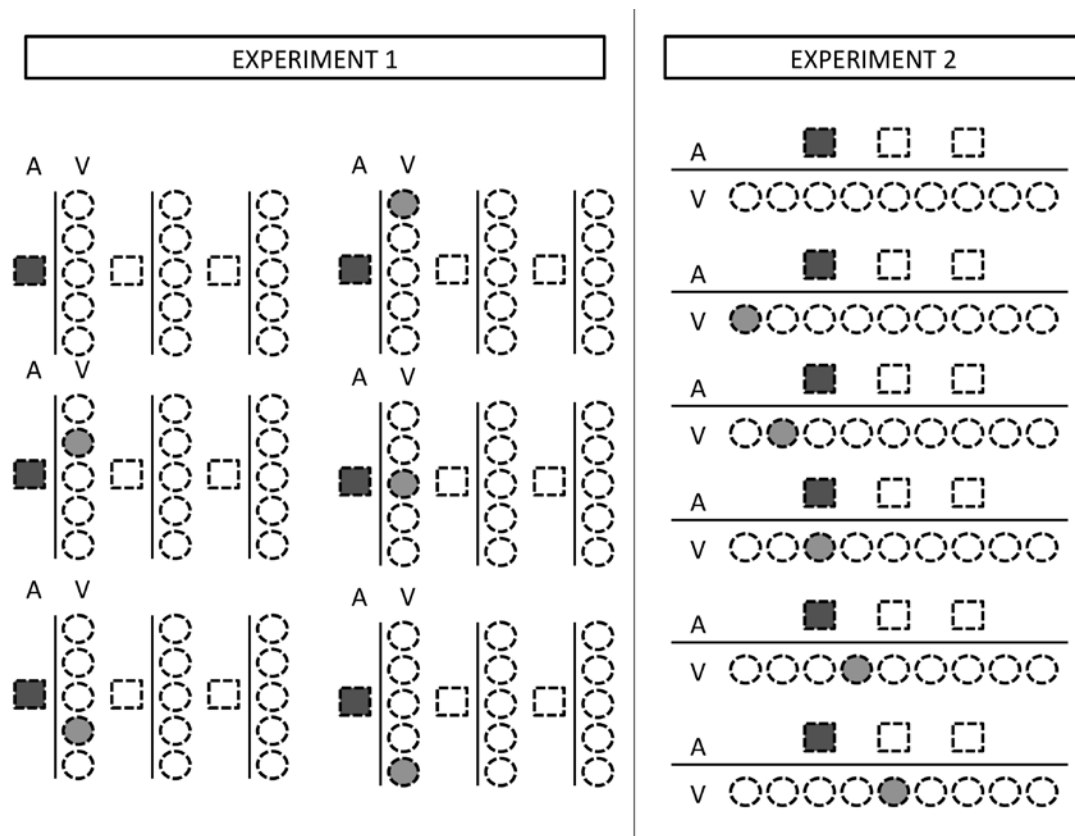
Figure 16 Possible stimulus presentation sets for a left sound. Dark squares indicate the sound positions and grey circles mark the place of the synchronously presented visual stimuli. In the figure, audio and visual stimuli positions are presented with a separator line for illustrative purposes.

In Experiment 2 a similar design was used, only the 6 visual positions were spread out horizontally, not vertically.

Figure 16 illustrates the possible stimulus presentations for the left sound position in the two experiments. In both experiments, each trial was repeated 15 times, resulting in a total of 750 presentations. The experiments lasted one hour with one or, if participants needed, two 5 minutes long breaks.

7.2.4 DATA ANALYSIS

We first inspected the data for outliers. We rejected every response where RT was less than 300 msec or more than 4000 msec. Due to significant time uncertainty (variable delay) caused by the computer cluster system, we used response times just for filtering. After the removal of outliers, on average 92% of each participant's data in Experiment 1 and 82% in Experiment 2 remained and were entered in the analyses. Errors were calculated as response bias in the direction of the distractor stimulus, where the reference was the average unimodal response direction in each condition. We used this approach because some participants tended to mislocalize sounds on the vertical plane; thus, analysing relative bias to veridical sound positions would distort our results (Wozny & Shams, 2011). Since our data were collected from a relatively small sample (six and five participants) and sphericity was violated, our data structure is not well suited for standard ANOVA analyses (Hoffman & Rovine, 2007). Instead, a mixed-effects modelling approach allowed us to explicitly model the sample specific and population general effects (D. M. Bates, 2010). We modelled the Sound type (Real, Surround), Sound direction (Left, Center, Right), and Visual stimulus direction ($\pm 21°$, $\pm 10.5°$, $0°$) as general effects (fixed effects) and participants' ID as sample specific effects (random effects) in the model as random intercept. Mixed-effects modelling was done in R using lme4 (D. M. Bates, 2008); visualization was done with ggplot2 (Wickham, 2008).
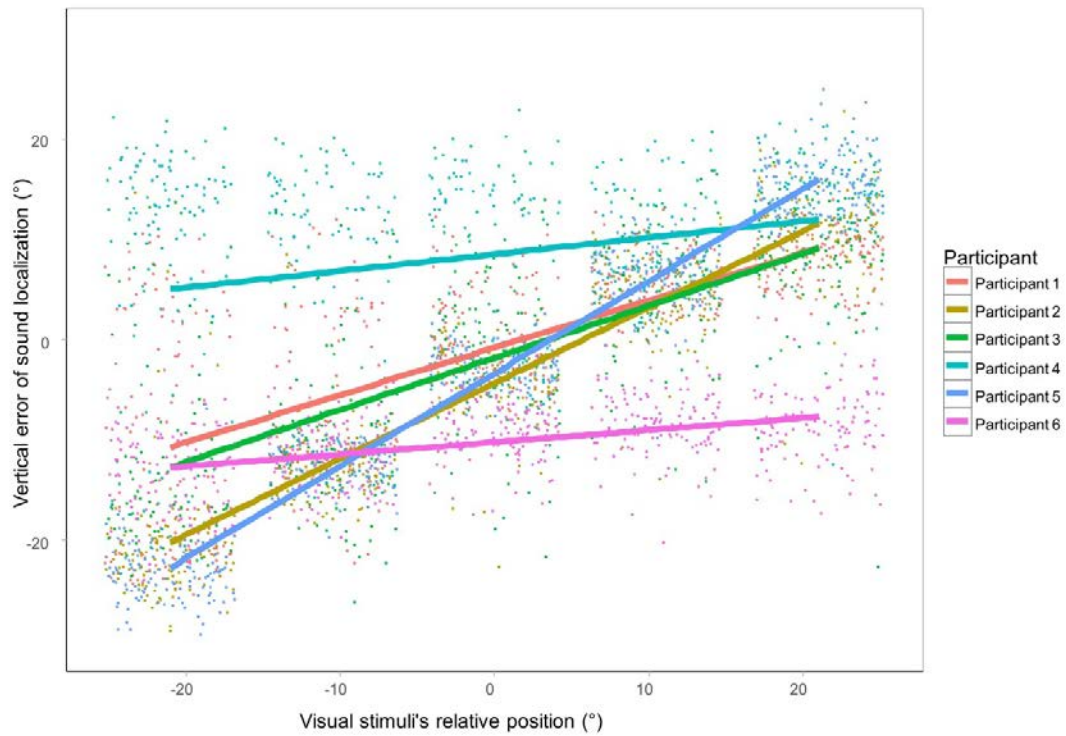
Figure 17 Visual capture effect for each participant in Experiment 3.1. Clear visual capture effect is visible for all participants. The strength of the effect visibly varies for each subject, which means it is reasonable to model the effect on the group and on the individual level together. On the figure different colours denote different participants; dots denote the individual responses.

7.3 RESULTS

7.3.1 RESULTS OF EXPERIMENT 3.1

In the first experiment, visual stimuli had vertical offset relative to the sounds. We manipulated Sound type, Sound direction, and Visual stimulus direction. Theoretically, if these factors could affect the responses on several levels, it is possible, that – as we expect it – visual stimuli attract the perceived location of sounds, and hence, responses would show consistent misplacement in the direction of visual stimuli. However, it could also happen that we do not see any systematic effect in the grand averages, yet the visual stimuli affects the responses, just differently for each participant. For example,

while one participant responds with mislocalization towards the visual stimulus, another will follow a different strategy and will try to separate the audio and visual components, resulting in responses slightly misplaced in the other direction. Often the factors affect both levels; in this case, estimation of subject-level effects could help us discerning sample related variability and population general effects (Barr, Levy, Scheepers, & Tily, 2013).

Our main interest was the effect of Visual stimulus direction; thus, we inspected the

Table 1 Summary of the Mixed-effects model in Experiment 3.1

	Dependent variable:
	Localization error
Visual stimulus direction	0.419[**]
	(0.178)
Sound direction:Left	2.633[***]
	(0.347)
Sound direction:Right	1.450[**]
	(0.706)
Sound type:Surround	3.156[***]
	(0.626)
Sound direction:Left* Sound type:Surround	-3.349[***]
	(0.613)
Sound direction:Left* Sound type:Surround	-5.477[***]
	(0.936)
Constant (Sound direction:Centre, Sound type:Free-field)	-5.599[***]
	(1.992)
Observations	3,449
Log Likelihood	-11,053.650
Akaike Information Criterion	22,279.300
Bayesian Information Criterion	22,807.840
Note:	[*]$p<.05$[**]$p<.01$[***]$p<.001$

effect of this factor for each subject individually. Figure 17 shows that there is variability in the strength of visual capture for subjects; therefore, the mixed-effects modelling is adequate.

In contrast to an earlier analysis of this dataset (Á. Török, Mestre, et al., 2015), instead of forward model building, we built the maximal model on the data that provides a more reliable estimate of fixed effects (Barr et al., 2013). The model was estimated with Restricted Maximum Likelihood estimation (REML); optimization was done using Bound Optimization by Quadratic Approximation (Powell, 2009).

The model with the full fixed and random effect matrix provided good information criteria (AIC: 22316.160, BIC: 22875.430). Wald chi-square tests (type 3) were used to test the fixed effects, which showed a significant effect of Visual stimulus direction (χ^2 (1) = 15.99, p < .001), Sound direction (χ^2 (2) = 32.50, p < .001) and Sound type (χ^2 (1) = 5.40, p = .02). We also found a significant interaction between Sound direction and Sound type (χ^2 (2) = 11.54, p = .003). Importantly, we did not find interaction with Visual stimulus direction, indicating that while there is variance in the localization of different sounds, the effect of visual stimulus is consistent and does not depend on the localization of different types of sounds. Therefore, in the final model we removed all interaction terms from the fixed effect structure that contained Visual stimulus direction. This model provided better fit compared to the full model (AIC = 22279.300, BIC = 22807.840). The estimated parameters of the final model can be seen on Table 1.

Summarizing the results of Experiment 1, we found strong and consistent visual capture for both sound positions. We found also variability in the localization of individual sound sources and positions. Thus, ventriloquism on the vertical axis was verified in the current experiment.

7.3.2 RESULTS OF EXPERIMENT 2

Analysis of the data of Experiment 2 was done using the same factors as in Experiment 1. Similarly, a forward modelling approach of the dataset is available in the study of Török et al. (2015). Here, we followed a backward building approach and built a full model first. This contained all possible effect combinations of Visual stimulus direction, Sound type, and Sound direction on the sample level (fixed effects) and on the level of individual participants (random effects). To understand the subject level variability of

our factor of interest, we visualized the effect of Visual stimulus direction for each subject. Figure 18 shows that the strength of visual capture varies from participant to participant.

The full model contained the full random covariance matrix (Barr et al., 2013) and provided good information criteria (AIC: 17441.550, BIC: 17973.690). We used Wald chi-square tests (type 3) to test the fixed effects, which showed a significant effect of Visual stimulus direction ($\chi^2(1) = 5.71$, $p = .017$), Sound direction ($\chi^2(2) = 6.26$, $p = .044$) and Sound type ($\chi^2(1) = 6.02$, $p = .014$). We also found a significant interaction between Sound direction and Sound type ($\chi^2(2) = 9.24$, $p = .010$). We run the model with removing the non-significant interactions, but the simpler model did not show
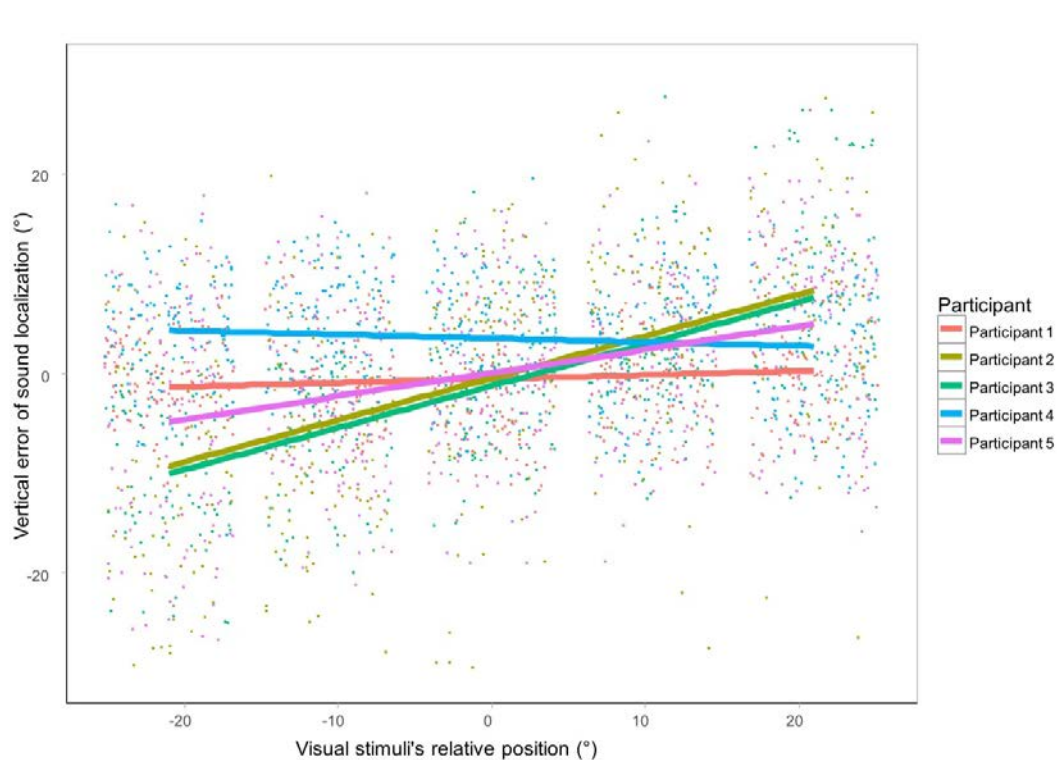
Figure 18 Visual capture effect for each participant in Experiment 3.2. Clear visual capture effect is visible for most participants. The strength of the effect is visible varies for each subject, which means it is reasonable to model the effect on the group and on the individual level together. Two participants have almost flat linear fits, meaning that they were able to escape almost entirely the visual capture. On the figure different colours denote different participants; dots denote the individual responses

improvement on the information criteria (AIC = 17540.030, BIC = 18042.940); hence, we present the parameter estimates for the full model (see Table 2).

Summarizing the results of Experiment 2, we found consistent visual capture of sound localization. We found variance in sound localizations for different sound types and positions. Overall, the results obtained in Experiment 2 show similar patterns to the ones we observed in Experiment 1.

Table 2 Summary of the Mixed-effects model in Experiment 3.2

	Dependent variable:
	Localization error
Visual stimulus direction	0.191***
	-0.074
Sound direction:Left	-4.900*
	-2.603
Sound direction:Right	-6.211***
	-0.833
Sound type:Surround	-3.352***
	-1.288
Sound direction:Left* Sound type:Surround	5.698***
	-1.884
Sound direction:Left* Sound type:Surround	2.381*
	-1.278
Constant (Sound direction:Centre, Sound type:Free-field)	3.351***
	-0.343
Observations	2,560
Log Likelihood	-8,684.02
Akaike Information Criterion	17,540.03
Bayesian Information Criterion	18,042.94
Note:	*p<.05**p<.01***p<.001

7.4 Discussion

In the presented experiments, we investigated the audio source localization ability of 11 subjects by measuring their performance in multimodal situations. We aimed to evaluate the usability of surround systems in supporting visualization and creating realistic perceptual situations. In the two experiments, we looked at how participants localized sound sources when they occur with synchronous but displaced visual distractors. In the first experiment, visual stimuli had vertical offsets to sounds; in the second experiment, visual stimuli had horizontal offsets to sounds. We found similar pattern of effects in both experiments. In both cases, visual distractor positions greatly affected the participants' localization judgments. Besides the visual capture we found variance in the localization of different sound sources. Importantly this did not interact with the visual stimuli's capturing effect.

The results are converging with forward modelling approach presented in Török et al. (2015). One important difference is that while Török et al. found stronger visual capture in both experiments for centrally presented sounds; here we did not find that. This is due to the more conservative modelling technique followed in the present analysis (Barr et al., 2013). Here we modelled the full random effect matrix with all possible slopes and intercepts, in this way limiting the possibility of false positive results in the fixed effects (optimizing alpha level). The lack of interaction, therefore, indicates that it may not generalize to the population. Importantly, the effect of interest of both analyses was the visual capture effect which was present on both the vertical and the horizontal axes.

We observed slight differences between surround and free field speakers in both experiments. Besides perceptual mechanisms, there were some other possible contributing factors. One likely explanation for the variance is that there was some difference in the speakers' characteristics. Alternatively, the asymmetry in the reverberation structure of the experimental hall could alter the reverberation properties of the sounds. Because we used identical speakers and sound levels were measured for each speaker separately, it is more likely that the asymmetry of the experimental hall contributed to the differences in localization. This further highlights the importance of

visual capture and multimodal stimulation to prevent perceptual changes caused by imperfection of sound source modelling and rendering.

The observed effects are comparable to those of Besson et al (2010), Besson, Bourdin, & Bringoux (2011) and Hartnagel et al. (2007). Although our methodology was based on these earlier studies, important differences exist. Besson et al (2010, 2011) used only one movable near sound source (cca. 50 cm) in a soundproof chamber, whereas in our experiments sound sources were much farther (3 m) away from the viewpoint in a reverberating hall. This difference is even more important since near and far sound sources are localized differently (Moore & King, 1999). This could be also important when we consider why the effects were different for sound directions. Another important difference is that while Besson et al (2010) used a led array as visual stimuli positioned at the distance of the sound sources, in our case, Gaussian blobs were projected to the frontal screen at a distance of 1.5 m from the participant's viewpoint. Moreover, the screen was not curved, but the blobs were stereo-projected to a virtual sphere at 2.99 m from the viewpoint. The last important difference was that in contrast to the earlier studies, we allowed participants to respond both horizontally and vertically simply by moving their hands. In this way, we could avoid artefacts caused by unnatural response methods, such as choice from a button array or button rotation.

We decided not to compare the data of the experiments in one analysis because the random effects in the models were different for the two experiments indicating sample variability. However, indirect comparison is possible. The fact that MLM showed more consistent effect of visual stimuli for vertical arrangements indicates that visual capture is stronger in the vertical plane. Earlier, with different methodology Thurlow and Jack (1973) reached very similar conclusions.

A limitation of the current study is that based on the methodology we used, we cannot decide whether the sounds were really perceived close to the visual stimuli or the effect was caused by post perceptual response strategies. After the experiments, the participants reported that they felt sometimes that sounds and flashes were coming from elsewhere. In fact those responses fell between the visual and sound positions (especially in Experiment 2). This means that the participants did try to locate the sounds and not simply chose the position of the visual stimuli. Our methodology was

based on standard ventriloquism paradigms, which were also affected by this criticism (Vroomen & Gelder, 2004).

Nonetheless, there are other studies showing that the ventriloquist effect occurs in non-transparent (i.e. where the discrepancy is so little that it is not possible to differentiate consciously the audio and visual signal's location) paradigms, as well (Alais & Burr, 2004; Bertelson & Aschersleben, 1998). It is also important to note that the brain responses elicited by ventriloquized and non-ventriloquized sounds differ even at early cortical processing stages (Bonath et al., 2007). A preattentive brain response, the mismatch negativity observed in EEG studies, is sensitive to the ventriloquist effect (Colin et al., 2002; Stekelenburg et al., 2004).

Our study fits well within the scope of cognitive infocommunications (Baranyi & Csapó, 2012). Cognitive infocommunications is the level of the development of infocommunications where cognitive science and infocommunication technologies converge (Baranyi, Csapó, & Sallai, 2015). One important aspect of this convergence is that it shows new ways to expand both the capabilities of humans and of artificial systems. The current study shows an example where a perceptual illusion serves as a leap through the current technological barriers of widespread VR technologies. Similar approaches demonstrated how perceptual illusions can benefit multimodal user interfaces (Colonius & Diederich, 2011; J.-H. Lee & Spence, 2009; Á. Török, Kolozsvári, et al., 2014).

To know more about how multimodal integration works in virtual reality, further studies are needed, utilizing brain imaging and electrophysiological methods. The question of how the brain perceives virtual environments is already a major topic in neuroscience research (Haans & IJsselsteijn, 2012; Kober, Kurzmann, & Neuper, 2012). However, studies involving recordings of brain activity in interactive conditions are mostly lacking (cf. Snider et al., 2013; Á. Török, Sulykos, et al., 2014).

To sum up, in the present experiments we found that 1) the ventriloquist effect works in virtual reality 2) sounds can be ventriloquized both vertically and 3) horizontally, and 4) there is a slight deterioration in the sound source position judgments when using surround system and free field speakers. In conclusion, researchers and virtual reality designers should use surround systems to support visualization and increase presence in

VR (Slater et al., 1994). The human perceptual system is well adapted to the experienced mismatches in audio and visual positions.

8 THE BODY IN SPACE

The last experiment showed that vision easily captures the perceived location of sounds even if participants are explicitly told to ignore the visual stimuli. Several earlier studies reached similar results (Bonath et al., 2007; Slutsky & Recanzone, 2001; Thurlow & Jack, 1973; Vroomen & Gelder, 2004). This phenomenon is reliably present under well-defined circumstances to such an extent, that it is widely used in applied scenarios, such as emergency warning systems (Csapó & Wersényi, 2013; Patterson, 1990; Politis, Brewster, & Pollick, 2014; Spence & Santangelo, 2009; Steenken, Weber, Colonius, & Diederich, 2014). From this line of research we can draw the conclusion that for human navigation visual input is of primary importance (cf. rodents see Diamond, von Heimendahl, Knutsen, Kleinfeld, & Ahissar, 2008). This notion is strengthened by the results of studies that found place and grid cell studies in humans, relying primarily on the visual modality (Doeller et al., 2010; Ekstrom et al., 2003; Jacobs et al., 2013). Moreover, processing of spatial location is not disrupted by even virtual teleportation (Baker & Holroyd, 2009); the neural oscillations may encode the path between the views at the two ends of the teleportation wormhole (Schnapp & Warren, 2007; Vass et al., 2016).

Thus, vision seems to dominate other senses in the spatial domain; nevertheless, there is one important compound modality we should yet to talk about in detail. This is the sensation of the position of our own body that is based on (1) proprioception, the sense of the relative position of body parts and on (2) the vestibular sense, which is the sense of balance and of gravitational up and down.

Our own body and motoric actions play crucial roles in the development of spatial vision (Marton, 1970). The visual-postural body-model (Marton, 1970) posits that seeing our actions and having internal feedback of the motion leading to them is integrated and serves as the basis of differentiating us from the environment. Supporting evidence comes from the seminal study of Hein and Held (1963) where pairs of kittens were placed in a circular treadmill apparatus. Of each pair, both kittens wore a neck yoke and a body clamp, but while one was able to move freely and turn the treadmill, the other was restrained to a gondola and only passively experienced the locomotion. After exposure to this task for three hours each day for several weeks, the restrained kitten showed impaired performance on visually-guided behaviour tasks which required

the visual estimation of distances. The same authors later showed that if only one eye sees that self-initiated action, depth perception will develop normally only for that eye but not for the other (Hein, Held, & Gower, 1970).

Normal visual depth perception requires intact retina, colliculus superior, and the primary visual cortex (Hein et al., 1970; Hubel & Wiesel, 1959; Kuffler, 1953; Roland & Gulyás, 1995; Sprague, 1966). Similarly to kittens, in human neonates, depth perception develops after birth and requires self-initiated movements (Wexler & Van Boxtel, 2005). The cues that we use to perceive depth can be classified as either monocular or binocular cues. Monocular cues are motion parallax, relative size/form, absolute size/form, aerial perspective, accommodation and occlusion (Servos, 2000). These cues typically require experience about the outside world. Binocular cues, such as disparity and convergence, on the other hand, do not depend on familiarity (Julesz, 1964, 1971).

Concluding the last paragraphs, bodily feedback of actions plays an important role in the development of the visual system. However, increasing number of studies suggest that after the sensitive early period, vision is starting to dominate proprioception, too. Indeed, a successful pain relief therapy for patients with amputated limb is based on a simple vision-induced somatosensory illusion (Ramachandran & Rogers-Ramachandran, 1996). In this method, the neurologist shows the patient an open box that has entry for both hand/arms. Inside the box there is a mirror in which the patient is able to see the healthy hand mirrored to the position of the missing limb. This way the patient not only sees the missing hand but observe its motion when the healthy hand moves. This therapy successfully ease their pain in a number of cases (Ramachandran & Hirstein, 1998).

The mirror box is not the only vision-induced body illusion. A few years after the introduction of the mirror box, the rubber hand illusion was described (Botvinick & Cohen, 1998). In the original paradigm, the participants sat with their left arm resting on a table. The experimenter covered this arm and put an artificial arm in front of the participant in the same angle as the real arm. Then, the participants are asked to focus on the rubber hand while the experimenter applies synchronous brushing strokes to both the real and the rubber hand. Interestingly, after ten minutes of exposure the participants report the felt sensory stimulation on the rubber hand. Moreover, Tsakiris and Haggard

showed that although synchronous sensory stimulation is necessary condition for the illusion, it is not sufficient (2005). The illusion does not develop when the fake object is not the artificial version of the covered arm but a wooden stick, for example.

Vision-induced somatosensory illusions are not limited to the arms. A study showed that under the right circumstances it is possible to evoke even a whole body illusion (Lenggenhager et al., 2007). In a virtual reality experiment, participants wore a head-mounted display (cf. Sutherland, 1968). On the display they saw a human-like doll in a position as it was them filmed from the behind. The experimenter applied synchronous strokes to the back of the participant and the doll. After one minute of stimulation, the participants already tended to report feeling the doll's body was their own body. Just after the stimulation finished, they displaced the participants and asked them to return to their initial position. Intriguingly, the participants showed a drift towards the position of the doll. This drift exceeded what they observed without the experimental stimulation and was not present when the doll was replaced by a human-size box. Later, the same group showed that the illusion is reflected by an activity change in the temporo-parietal junction (Blanke, 2012; Ionta et al., 2011) which was previously associated with out-of-body experiences (Blanke, Ortigue, Landis, & Seeck, 2002). In conclusion, the sensory experiences related to our own body can be and in fact are affected by vision.

Finally, we discuss the role of vestibular sense in spatial perception. The vestibular system is responsible for the sense of balance (Barany, 1906). It interfaces the environment through two distinct structures. The otolith organs and the semicircular canals both contain endolimphatic fluid and are sensitive to linear and angular acceleration, respectively (Ferrè, Longo, Fiori, & Haggard, 2013). Unlike other senses, the vestibular system remained an evolutionary primitive system in the human brain, and afferent projections from its sensory epithelia are distributed widely in the brain (Bottini et al., 1994; Ferrè, Bottini, Iannetti, & Haggard, 2013). The core region of the vestibular network is the parietal-insular vestibular cortex, where multisensory neurons were found that received input from not only the vestibular system but other sensory modalities responsible for posture control (Grüsser, Pause, & Schreiter, 1990; Guldin & Grüsser, 1998).

Amongst other spatial functions (Ferrè, Longo, et al., 2013), the vestibular system is responsible for the most basic form of spatial knowledge: the feeling of earth-vertical

(Angelaki, Klier, & Snyder, 2009). It is primarily based on the work of the otolith organs, but distinguishing between self-initiated motion and opposite direction tilt requires the system combining information from the semicircular canals, as well (Angelaki, Shaikh, Green, & Dickman, 2004). This contribution is so strong that it leads to a strange sensation of tilting when pilots are accelerated in a centrifuge (Peters, 1969).

The importance of knowing the earth-vertical becomes apparent when people either permanently lose their vestibular sense (Dix & Hallpike, 1952; Ménière, 1861) or are in a place that does not affect the endolimphatic fluid (Balázs, Barkaszi, Czigler, & Takács, 2015). Weightlessness during a spaceflight cause various changes in cognitive, perceptual, and motoric abilities (Lackner & DiZio, 1993). One interesting effect of zero-gravity is the altered perception of perspective, which can be measured by the lack of illusion building on our strong concept of linear perspective (Villard, Garcia-Moreno, Peter, & Clément, 2005). Moreover, astronauts were reported to underestimate distances when they are in space (Clément, Lathan, & Lockerd, 2008; Clément, Skinner, & Lathan, 2013). These results suggest that the perception of gravity and hence the vestibular system might have an effect on visual distance estimation even in terrestrial conditions.

In fact, several studies showed that visual perception is affected by gravity and more specifically, the position of the body relative to vertical (Di Cesare, Sarlegna, Bourdin, Mestre, & Bringoux, 2014; Fouque, Bardy, Stoffregen, & Bootsma, 1999; Harris & Mander, 2014). In one study (Di Cesare et al., 2014), participants were asked to reach for an object in virtual reality from a tilting chair. The chair's pitch was adjustable and in three of the five conditions the chair was slowly tilted forward (i.e. participants were looking downwards); additionally, the authors manipulated the angle of the virtual environment, too. They found systematic errors in the reaching movements: tilting caused underestimation in all conditions, except when only the virtual environment was tilted forward. The authors found that their results better fit the predictions of a gravity based model than a body-centred one. To note, in this experiment the target location in the visual field also changed with the scene tilt (but not with the body).

Similar experiment was done by Harris and Mander (2014) with two important differences. First, unlike in the virtual reality experiment of Di Cesare and colleagues

(2014), here the authors used an actual prepared tumbled room for the experiment. Second, instead of tilting the body forward, it was tilted backwards. Additionally, instead of asking to reach for an object, the authors asked the participants to compare the length of a projected line to the length of a rod held in their unseen hands. In spite of the methodological differences, the results of this experiment fit well in a gravity oriented framework. They found that (real or illusory) backward tilting of the participant caused overestimation of the length of the rod, and hence, the wall seemed presumably closer.

Both experiments raise, however, interesting questions. Even though they attribute the effects to the perception of gravity, it is still unresolved whether the effect is a pure visual illusion or a genuine multisensory phenomenon. In both experiments, only one direction tilting was used; hence, it is inconclusive under which circumstances should we expect underestimation and overestimation. Moreover, since illusory tilting also caused estimation bias (Harris & Mander, 2014), an alternative explanation could be that all unnatural poses/situations biases the estimations and not specifically those that included actual change in the gravity vector. Additionally, both experiments used whole body tilting, which is quite unnatural in everyday scenarios. Normally, the vestibular system perceives that the ground is tilted and, hence, via an interplay between posture control, the body's tilt is adjusted to avoid falling backwards to the ground (Nashner, Shupert, Horak, & Black, 1989). However, unlike full body tilting, tilting of the head is a frequent activity: we look down and up to things when their position in not on the horizon (Gajewski, Wallin, & Philbeck, 2014; Wu, Ooi, & He, 2004). Change in the angle of the head also produces vestibular input; therefore, to address the question whether the vestibular sense modulates visual distance perception, we designed a virtual reality experiment where participants were instructed to tilt their head up and down to judge the distance of an environmental object.

9 EXPERIMENT 4: VESTIBULAR CONTRIBUTION TO VISUAL DISTANCE PERCEPTION[5]

9.1 INTRODUCTION AND HYPOTHESES

Perceiving how far away an object is from one's own body is essential for interacting with the environment. Although spatial localization of distant objects is primarily based on the visual modality (Andre & Rogers, 2006; Loomis & Knapp, 2003), distances often seem shorter or longer if the target objects are not on the level of horizon. For example, a mountain refuge seems farther or closer depending on whether we look up at it from below or down at it from above. In fact, increasing amount of evidence suggests that visual distance perception is affected by other senses (Di Cesare et al., 2014; Harris & Mander, 2014) and by top-down influences, such as fear of height (Jackson & Cormack, 2007) or perceived effort (Proffitt, Stefanucci, Banton, & Epstein, 2003).

Contrasting explanations have been proposed for such *illusory distance biases*. One the one hand, the gravity theory claims that distance perception is based on the estimated motor effort of navigating to the perceived object (Howard & Templeton, 1966; Proffitt et al., 2003). According to this reasoning, upward distances should be overestimated (Bhalla & Proffitt, 1999). On the other hand, the evolved navigation theory posits an evolutionary advantage in overestimating the risk of falling (Jackson & Cormack, 2007; Willey & Jackson, 2014). In this case, contrary to gravity theory, downward distances should be overestimated. Both theories assume that current gaze elevation is combined with internally-*stored* information in order to compute distance. Gravity theories require stored information about previous motor efforts (Howard & Templeton, 1966), while evolved navigation theories require internal information about potential risks of falling (E. J. Gibson & Walk, 1960).

Supporting evidence was found for the role of conscious processes; for example, both reduced fear of falling (Jackson & Cormack, 2010) and reduced anticipated effort

---

(Bhalla & Proffitt, 1999) diminishes the illusion. Nevertheless, it is equally possible that instead of direct top down influence on the visual system, the illusion is based on genuine multisensory interaction and only the direction of the effect is modulated by top down influence of stored information. Such a system would lead to coherent distance percept from early stages of processing, and at the same time it would allow slow recalibration of the system (Bhalla & Proffitt, 1999).

Experimental evidence shows that visual distance estimation is affected by multisensory processes (Di Cesare et al., 2014; Harris & Mander, 2014). In particular, the vestibular system provides a signal, relating the current head orientation to the direction of gravity. Combining a vestibular signal with an eye position signal specifies whether a visual object is located above or below the eye. Although vestibular signals do not directly code the spatial location of external objects, the interaction between vestibular and visual information is essential in providing the organism with space representation (Clément, Fraysse, & Deguine, 2009; Clément et al., 2013; Villard et al., 2005). For instance, vestibular peripheral organs detect the motion of the head, producing experiences of self-motion in three-dimensional space. Cortical vestibular pathways integrate information from other sensory modalities to generate appropriate and accurate responses to self-motion, such as the stabilization of gaze, balance, and postural motor commands (Cullen, 2012) and the perception of the subjective visual vertical (Bömmer & Mast, 1999). Indeed, earlier studies showed that the body-tilt has an effect on perceived distance (Di Cesare et al., 2014; Harris & Mander, 2014). However, the results do not explain the link between multisensory integration and the direction of the illusory distance bias. In these experiments, only either downward or upward angles were tested; therefore, they do not specify under which circumstances one expects underestimation vs. overestimation.

In the present study, we were seeking evidence for a stored information modulated multisensory interaction underlying the illusory distance bias. If this hypothesis is true, then, one expects illusory distance bias even when no anticipated effort or perceived risk is present in the task. Furthermore, if we assume a multisensory link underlying the illusion, then, it should occur when the visual stimulation is constant and only the information provided by the vestibular system changes. And finally, if stored information affects visual distance estimation, then, we should observe both under- *and* overestimation under the right circumstances.

Therefore, in the current experiment, we asked participants to judge the distance of an object presented on a plane at different distances and different gaze elevations (Figure 19). We developed a novel virtual reality environment in which neither risk of falling (Jackson & Cormack, 2010) nor navigational effort were actually present (Proffitt, 2006). The participants' head inclination was systematically varied by asking them to gaze upwards or downwards at the object. Further, we applied event-related galvanic vestibular stimulation (GVS) during each judgement to investigate whether vestibular signals indeed contributed to distance perception biases.

9.2 METHODS

9.2.1 PARTICIPANTS

Sixteen healthy participants volunteered for the study. Data from two participants was discarded because they proved unable to follow the instruction of the experiment (see below). Thus, fourteen participants (5 females, mean age ± standard deviation: 26.64 ± 6.64 years) completed the experiment. All participants were right-handed according to their Edinburgh handedness inventory scores. The sample size was decided *a priori* based on similar psychophysical experiments. Participants gave their written informed consent before the experiments. The experimental protocol was approved by the research ethics committee of University College London. The study adhered to the ethical standards of the Declaration of Helsinki.

9.2.2 GALVANIC VESTIBULAR STIMULATION

Galvanic vestibular stimulation (GVS) was applied using a commercial stimulator (Good Vibrations Engineering Ltd., Nobleton, Ontario, Canada) delivering a boxcar pulse of 1mA for 3s. This low intensity was used to minimise non-specific cueing effects, such as arousal from cutaneous sensations or vertigo. Importantly, several studies confirm that this level of GVS activates the vestibular organs, without effects persisting beyond the period of stimulation (Fitzpatrick & Day, 2004). Carbon rubber electrodes (area 10 cm^2) coated with electrode gel were placed binaurally over the mastoid processes and fixed in place with adhesive tape. The area of application was

first cleaned, and electrode gel was applied to reduce the impedance. Left anodal and right cathodal configuration is named 'L-GVS' (Figure 19). The inverse polarity, namely right anodal and left cathodal configuration, is named 'R-GVS'. GVS is known to increase the firing rate in vestibular afferents on the cathodal side and to decrease the firing rate on the anodal side (Goldberg, Smith, & Fernández, 1984). We also applied a sham stimulation using electrodes placed on the left and right side of the neck, about 5cm below the GVS electrodes (Ferrè, Vagnoni, & Haggard, 2013; Lopez, Lenggenhager, & Blanke, 2010) with left anodal and right cathodal configuration. This sham stimulation evoked similar tingling skin sensations to GVS but not modulation of vestibular afferents and, thus, functions as a control for non-specific alerting effects and for the knowledge that an unusual stimulation is occurring.

9.2.3 VIRTUAL REALITY ENVIRONMENT

The experiment was carried out in the Immersive Virtual Reality Laboratory at University College London, using a facility commonly referred to as a CAVE (Cruz-Neira et al., 1993). This system consists of four stereo-projected surfaces: three back-projected vertical walls, each 3m wide x 2.2m high, and the floor (3m x 3m) form a continuous projection surface. The virtual reality environment was created using Unity3D game engine (www.unity3d.com), rendered using a K5000 graphics card to drive 4 Christie Mirage DLP projectors, each of which projected to one of the 4 screens at 96Hz. The participant wore shutter glasses synchronized with the projectors creating active stereo-projection in each eye at 48Hz. The position of the glasses was tracked by an InterSense™ IS-900 system with high accuracy. The system was calibrated to the participant's own eye height at the beginning of every experiment, and this data was used to accurately compute object distances for the upward, downward, and level inclinations. The virtual scene was a green grass-like plane with blue skies and no visible landmarks. The experimental object was a 2m*2m gift box with purple ribbon. The object rested on the ground, and the same proportion was visible at all inclinations.

9.2.4 PROCEDURE

Verbal and written instructions about the task were given to participants prior to the experiment. Participants were seated in the centre of the CAVE, 1.5m from the front screen. A visual scene was presented on vertical screens and on the floor in order to create a seamless, wide field-of-view immersive display. Participants made absolute judgements about the distance between their own body and an object (a gift box) that appeared in front of them (Klein, Swan, Schmidt, Livingston, & Staadt, 2009; Figure 19). At the beginning of the task, object positions slightly under (1.5m) and slightly over (30m) the experimental range were presented, and the experimenter informed the participant about the actual distance in metres. Participants were encouraged to use these as anchor points to calibrate distance judgements in experimental trials. The positions of the present box were distributed logarithmically between 5m and 25m; thus, the possible distances were 5, 6.9, 9.52, 13.3, 18.2, 25m. These distances were chosen to produce a wide range of distance percepts. Our predictions did not focus on the effects of object distance itself but on the effects of two other experimental factors: head elevation and vestibular stimulation. The object appeared on a smooth plane that was inclining (+20°), flat (0°), or declining (-20°). All inclinations of the plane, including the object on it, were visually the same. The experiment was divided into blocks; head inclination and stimulation (L-GVS, R-GVS, and SHAM) changed only between blocks. Each block consisted of 18 trials; there were three repetitions of the same distance in each block. Distances were presented in random order. Block order followed a Latin square design. Each trial started with the presentation of the grass-like plane in the actual inclination and the blue sky. Participants adjusted their head pitch angle to fixate the object and, therefore, the horizon, while head tracking monitored their posture. This procedure ensured that participants saw the same proportion of grass and sky at all head inclinations. The head tracking system measured the inclination of the head and a sound signalled when the participant's head reached the correct vertical angle. They were told to keep their head at the same position for the duration of the block. Then GVS/SHAM started and lasted for 3s. 1s after GVS onset, the gift box became visible for 1s and then disappeared. The image was then blurred, and the GVS/SHAM pulse ended. Participants made absolute verbal judgements (in metres) of the distance of the object after the screen was blurred. The response was recorded, and the next trial started. This method of reporting distance percepts was preferred to the

method of limits and method of constant stimuli used in psychophysical studies of visual depth perception, for reasons of efficiency. We wanted to sample a range of environmental distances to minimise the number of GVS stimulations (GVS can cause mildly unpleasant sensations) and to minimise duration of the CAVE immersion. Absolute judgements can sometimes be difficult to interpret since different participants may use different subjective standards for one metre, resulting in different reported values. However, our experimental design was based only on within-participant comparisons; therefore, differences between individuals in reported values do not affect our inferences.

9.2.5 DATA ANALYSIS

The dependent variables were defined as the error of estimation relative to the real distance. Thus, negative values denoted underestimation, and positive values denoted overestimation. We ran first a 3 (GVS type) * 3 (Inclination) repeated measures ANOVA. We did not include Position in the ANOVA because the inclusion of different positions were primarily included to add variability to the task. In a second step, we fitted a mixed-effects model on the data to explicitly model intersubject variability. During the model building process we followed the guidelines of Barr and colleagues (Barr et al., 2013; Barr, 2013) and included all critical effects in the random terms, too. In the mixed-effects approach, we entered Inclination (-20°, 0°, 20°) and Visual stimulus position as a scalar variable and scaled it for easier interpretability (D. M. Bates, 2008), while GVS type (L-GVS, R-GVS, Sham) was handled as factorial. Mixed-effects modelling was done in R (Team, 2014)using lme4 (D. M. Bates, 2008); Visualization was done in ggplot2 (Wickham, 2008).
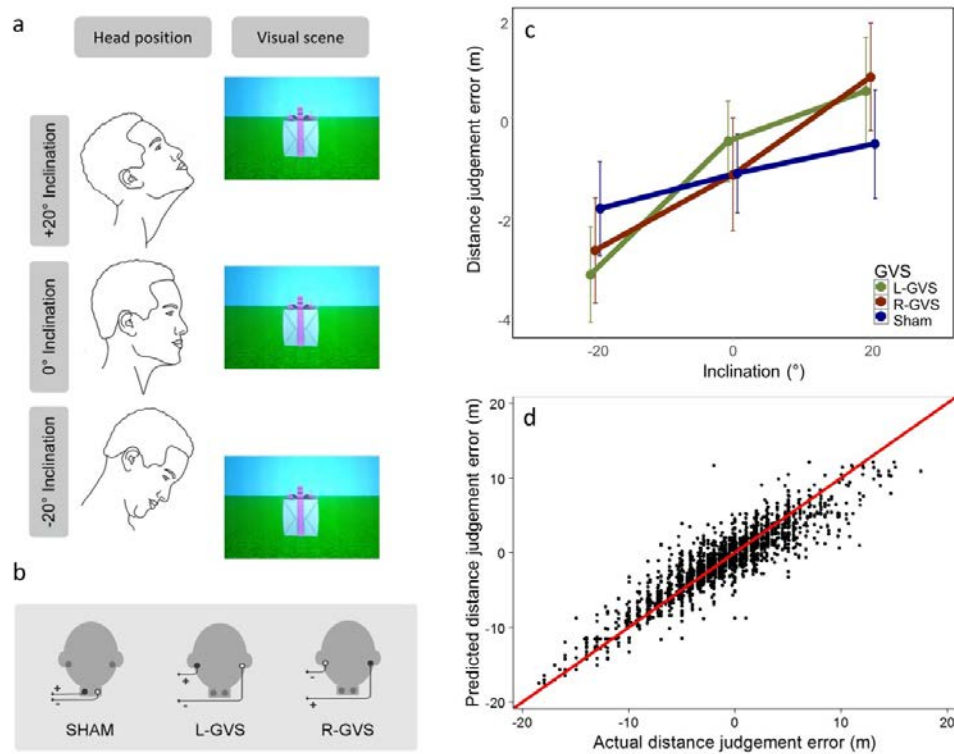
Figure 19 Setup and results of Experiment 4. (a) Participants were seated in the centre of the CAVE. During the experiment, participants made absolute judgements of the distance between their own body and an object (a gift box) appearing in front of them. The positions of the gift box were distributed logarithmically between 5m and 25m. The same distances were presented on the three head inclinations -20°, 0°, and +20°. (b) Left anodal and right cathodal configuration is named 'L-GVS'. The inverse polarity, namely right anodal and left cathodal configuration, is named 'R-GVS'. A sham stimulation was also applied placing the electrodes to the left and right side of the neck about 5cm below the GVS electrodes. GVS and Sham stimulation were applied delivering a boxcar pulse of 1mA for 3s. (c) *Distance Errors* have been calculated by subtracting the actual distance from the judged distance. Distance underestimations are negative; overestimations are positive. Distance perception varied significantly according to environmental inclination. Specifically, downward distances were underestimated, while upward distances were overestimated, relative to ground level. This pattern of distance illusions is in line with the predictions of the gravity theories. Note that GVS enhances this pattern. (d) Predictions based on the linear mixed-effects model. The built model explained 82 % variance in the data, showing that the collected data was well characterized by the estimated fixed effect and random effect parameters. This means that the variables which characterize a given data point strongly define the data point. The amount of variance explained by the fixed effects (23 %) shows the amount of variance one could reliably explain based on only the Inclination, GVS, and Position factors for a new sample with unknown participants.

9.3 RESULTS

Trials containing either recording errors or multiple responses were eliminated before the analysis. Less than one percent of all participants' data was removed according to this criterion. First, we calculated the Distance judgement errors by subtracting the real distance from the judged distance. Accordingly, distance underestimations are negative and overestimations are positive.

9.3.1 RESULTS OF THE ANOVA ANALYSIS

Distance judgement errors for each participant were averaged for each combination of head inclination and stimulation conditions and analysed using factorial repeated measures ANOVA and planned contrasts. Our theoretical predictions focused on the combination of head elevation and vestibular stimulation factors; therefore, in this analysis we pooled data across the different distances judged. Distance perception varied significantly across inclinations ($F(2,26) = 23.694$; $p < .001$; $\eta_p^2 = 0.65$, see Figure 19). Downward distances were underestimated by 1.65m ($SD = 3.50$), while upward distances were overestimated by 1.19m ($SD = 3.90$), compared to ground level. This pattern of results fits the predictions of gravity theories but opposes the predictions of evolved navigation theories. A planned linear trend contrast confined to the sham condition also showed a trend in the direction predicted by gravity theories (down vs. up inclination $t(1,13) = 1.670$; $p = .059$, *Cohen's d* = 0.45, one-tailed, numerical effect present in 10/14 participants). The corresponding planned contrast for evolved navigation was not supported (flat vs. down inclination: $t(1,13) = -1.274$, n.s.). The main effect of GVS condition was not significant ($F(2,26) = 0.196$; $p = .823$). Most importantly, we found an interaction between GVS condition and inclination ($F(4,52) = 3.318$; $p = .017$; $\eta_p^2 = 0.20$). This interaction occurred because the linear trend predicted by gravity theories was amplified by both polarities of GVS (down vs. up inclination L-GVS $t(1,13) = 4.891$, $p < .001$, *Cohen's d* = 1.31; R-GVS $t(1,13) = 6.585$, $p < .001$, *Cohen's d* = 1.76, numerical effect averaged across GVS polarities present in 14/14 participants). This is consistent with an inclination effect generated online by a vestibular signal that is boosted by artificial vestibular stimulation.

9.3.2 RESULTS OF THE MIXED-EFFECTS MODELLING

In the analysis of repeated measures designs we face two challenges. First, observations from the same participant are usually more correlated than observations between participants. Second, effects are usually slightly different for each participant. Therefore, without explicitly modelling individual differences, generalization of the results must be limited. In mixed-effects models, we explicitly include both fixed (i.e. population general) and random terms (i.e. subject specific) and, hence, are able to model how much of the sample variance is estimable based on the available variables (D. M. Bates, 2008; Quene, Hugo Bergh, 2008). Inspection of the variation of distance judgement errors as a function of object position shows that the mixed-effects model is justified (see Figure 20).
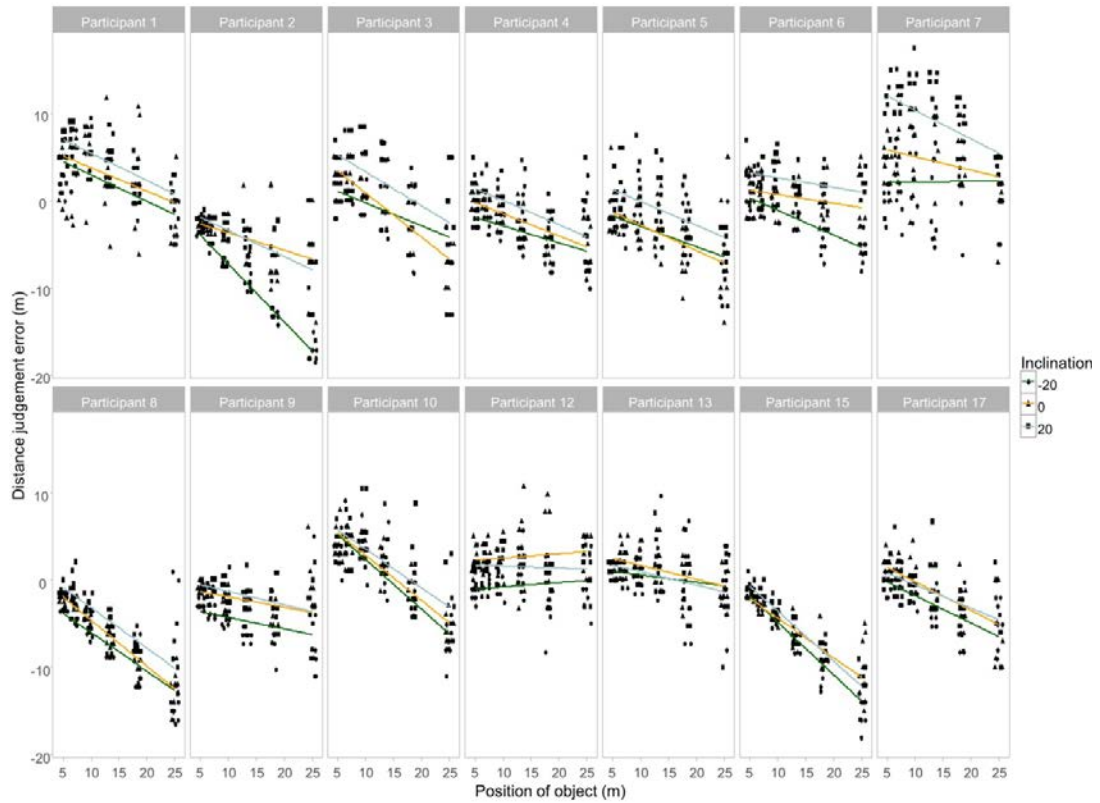

Figure 20 Between subject variability of the effect of inclination and object Position on Distance perception. Farther distances are more underestimated, but the scale of increase different for each participant. It seems that the difference between upward and downward distances also consistently appears for all participants, in different sizes

We estimated random slopes for Position at all combinations of subject level Inclination and GVS type random intercepts (Barr et al., 2013; D. M. Bates, 2010). This model appeared to be the best fit to our data according to an Akaike Information Criteria (AIC) of 3243.747 (df = 31, the baseline model containing no fixed effects, and only the subjects as random effects resulted in AIC 5211.152). The fixed effect significances were tested using F tests, where p values were based on the Kenward-Roger approximation of the degrees of freedom (Kenward & Roger, 1997). We found a main effect of Inclination ($F(2, 26.17) = 23.70$, $p < .001$, $\eta_p^2 = 0.64$) consistent with our hypothesis and a main effect of Position ($F(1, 19.87) = 41.75$, $p < .001$, $\eta_p^2 = 0.68$). The

Table 3 Summary of the Mixed-effects model in Experiment 4

	Dependent variable:
	Difference (M (SE))
Right GVS	0.098 (0.108)
Sham GVS	0.264** (0.108)
Inclination 0	0.533*** (0.120)
Inclination 20	0.735*** (0.120)
Position	-0.545*** (0.076)
Inclination 0:Right GVS	-0.230 (0.152)
Inclination 20:Right GVS	-0.041 (0.152)
Inclination 0:Sham GVS	-0.393*** (0.152)
Inclination 20:Sham GVS	-0.477*** (0.152)
Right GVS GVS:Position	0.087 (0.061)
Sham GVS:Position	0.374*** (0.061)
Inclination 0:Position	0.412*** (0.070)
Inclination 20:Position	0.145** (0.069)
Inclination 0:Right GVS:Position	-0.545*** (0.080)
Inclination 20:Right GVS:Position	0.111 (0.080)
Inclination 0:Sham GVS:Position	-0.631*** (0.080)
Inclination 20:Sham GVS:Position	-0.554*** (0.080)
Constant (Left GVS,Inclination -20)	-0.416** (0.199)
Observations	2,265
Log Likelihood	-1,590.873
Akaike Information Criterion	3,243.747
Bayesian Information Criterion	3,421.232

Note: *p<.05 **p<.01 ***p<.001

main effect of GVS was not significant ($p > .8$). Moreover, we found an interaction between Inclination and GVS ($F(4, 52.92) = 3.68$, $p = .010$, $\eta_p^2 = 0.22$), which is consistent with the hypothesis of a vestibular effect. Least square means post hoc contrast revealed that similarly to the ANOVA analysis, difference was only significant between -20 and 20 degrees in the Sham condition ($p = .035$). Additionally, we found a three way interaction between Inclination, GVS, and Position ($F(4, 67.56) = 39.10$, $p < .001$, $\eta_p^2 = 0.70$). We did not find theoretically motivated interpretation behind this interaction, it was primarily due to different slopes estimated for some Inclination and Condition combinations (see Appendix 10).

The final model (see summary Table 3) explained more than 80 % of total variance in the data (*Conditional* $r^2 = 0.82$) with a compelling contribution of the fixed effects (*Marginal* $r^2 = 0.23$).

9.3.3 SUMMARY OF THE RESULTS

The ANOVA and the mixed-effects model showed converging results. We found a main effect of Inclination: downward distances were underestimated, whereas upward distances were overestimated compared to estimations made on flat surface. Furthermore, we found an interaction between GVS type and Inclination because GVS stimulation (no matter if it was L-GVS or R-GVS) caused significant difference between estimations on different Inclinations. Without artificial vestibular stimulation (Sham), difference was only significant between upward and downward estimations. Moreover, we found that the linear mixed-effects model explained 80 % of the total variance.

9.4 DISCUSSION AND CONCLUSIONS

Participants overestimated distances when looking up and underestimated them when looking down. This result supports gravity theories and opposes evolved navigation theories. More strikingly, the effect increased strongly with event-related GVS. Our results suggest that the gravitational modulation of visual distance perception depends

on on-line vestibular signals. Illusory distance bias is, therefore, not merely a product of learned contextual associations but rather reflects a specific multisensory integration mechanism.

Gravitational signals are coded by vestibular receptors in the inner ear, whose signal depends on the position of the head relative to gravitational vertical (Barany, 1906; Guldin & Grüsser, 1998). The precise mode of action of GVS remains debated, but recent evidence confirms activation of otolithic fibres (Curthoys & MacDougall, 2012). Recent studies suggested that otolithic gravitational inputs in the vestibular system have a direct influence on cognitive tasks involving distance perception (Clément et al., 2009; Di Cesare et al., 2014; Harris & Mander, 2014; Villard et al., 2005). However, neither of these studies was able to consolidate multisensory distance illusions with the results of cognitive and affective factor motivated models so far (Howard & Templeton, 1966; Jackson & Cormack, 2007; Proffitt, 2006). Critically, in all the cited studies vestibular alterations influenced perception only in one direction; for example, either underestimation or overestimation was observed. Our study is the first that shows how vestibular activation leads to illusory distance bias of which direction is modulated by the stored memories of previous climbs and descends. The pattern of our GVS effect suggests that GVS amplifies a vertical, gravity-related component of a head-position signal, thus, increases the effects of gaze elevation on visual distance perception. Interestingly, GVS did not interfere with distance perception when the head's inclination was zero. Presumably, lack of head inclination is processed by the brain as distance across level ground. This represents an intermediate, neutral situation where there is neither cost nor benefit of gravity (cf. Howard & Templeton, 1966; Proffitt et al., 2003; Willey & Jackson, 2014). In this special case, the online vestibular-gravitational signal generated by GVS does not need to be integrated.

Importantly, although the current results supported the predictions of the gravity theory, they do not imply that the distance errors would linearly increase with uphill and decrease with downhill angle. In fact, we are usually not exposed to all angles equally in real environments (Bhalla & Proffitt, 1999; Proffitt, Bhalla, Gossweiler, & Midgett, 1995). Everyday experience of angles is either slight elevations (e.g. the steepness of roads is usually under 20 degrees) or risky, non-navigable surfaces (e.g. a balcony, a visual cliff (E. J. Gibson & Walk, 1960)). Because the direction and size of the effect is possibly modulated by higher cognitive and affective processes, it is reasonable to

assume that for large angles the evolved navigation theory holds. Importantly this assumption is in line with the results of the experiments of both theories (Bhalla & Proffitt, 1999; Di Cesare et al., 2014; Harris & Mander, 2014; Jackson & Cormack, 2007)

A limitation of the current paradigm is the use of verbal reports of absolute judgements; therefore, future research should discover whether the current effects generalize to tactile or reaching judgements (Bhalla & Proffitt, 1999; Di Cesare et al., 2014). However, this method has been used earlier, and although it affects the response accuracy, it does not affect the difference between our experimental conditions (Alexandrova et al., 2010; Loomis & Knapp, 2003; Proffitt et al., 2003; Sugovic & Witt, 2013). The results of those studies were broadly similar to those that used other measures to assess distance estimation (Servos, 2000; Sun, Campos, Young, Chan, & Ellard, 2004; but for differences see Andre & Rogers, 2006).

Previous accounts of visual distance perception identified a gravitational bias but did not investigate the underlying mechanism. We replicated these results and provided evidence for the underlying cause. We showed that a visual-vestibular interaction underlies illusory distance bias and the direction of the effect is consistent with the predictions of gravity theory.

10 APPLIED PERSPECTIVES OF THE COGNITIVE MAP IN

VIRTUAL REALITY

In the previous chapter, we presented a study that showed how the vestibular system affects the basic perceptual process of visual distance estimation. This result has interesting implications for studying human spatial navigation and for the use of virtual reality. The finding that posture has a significant influence on how we process distances emphasizes the possibility that cognitive maps should not be regarded as visual representations. Since the notion of the term cognitive collage by Tversky (1993), ample evidence supports the view that our internal representations of space are not purely visual (G. Chen, King, Burgess, & O'Keefe, 2013; Ravassard et al., 2013; Sharp et al., 1995). One important consequence of handling the cognitive map as a visual representation is that we implicitly apply the regularities of the physical world to the mental representation, too. Although experiments did show that the visual system extracts basic visual properties of a scene to learn real world conditional probabilities (Fiser & Aslin, 2001), these learnt features may easily combine into impossible objects (Schacter et al., 1995). By analogy, we are able to judge distances accurately both in real and in virtual environments (Loomis & Knapp, 2003) Nevertheless, we perceive distances differently, depending on whether the target object is downwards or upwards.

The previous experiments presented in this work have similar implications. In the experiment detailed in Chapter 3, we showed that the reference frame used to define the location of object depends on our viewpoint. While first person and third person avatar following camera views are associated with egocentric reference frame, when our viewpoint is outside of the behavioural area (e.g. an aerial viewpoint) allocentric reference frame is activated. Thus a simple change in the viewpoint could change entirely the way we represent the environment. Then, in Chapter 5 we went further and showed that not even viewpoint change is required, in a simple and well-known environment the introduction of teleportation (and so uncertainty) is enough to invoke a switch from egocentric to allocentric frame of reference. Our spatial representations thus adapt to a change in circumstances. The experiment presented in Chapter 7 showed that the adaptivity of spatial cognition is not only a higher level process, it is present already on the perceptual level. The cognitive system integrates information from

different senses in a near optimal manner (Alais & Burr, 2004), so that the most reliable perception can be guaranteed. In fact, the change in preferred reference frame in the two earlier experiment can be viewed also as a pursuit for optimal space perception. It was shown that disruption of path integration by disorientation makes egocentric reference frame use difficult (Wang & Spelke, 2000), on the other hand allocentric representations take more time to establish and are initially more coarse (Waller & Hodgson, 2006). Thus the cognitive system has to constantly evaluate the circumstances and activate the more reliable reference frame. The cognitive map is therefore an interpreted representation, importantly; our main argument is that at the same time it is not objective. We claim that the cognitive map is not only an impossible figure, but oddly it is an interpreted impossible figure.

Related to this, people easily adapt to an environment where the rules of the Euclidean space does not apply. In one experiment (Schnapp & Warren, 2007), participants had to navigate in a hedge maze that contained two teleportation wormholes. The participants not just easily adapted their wayfinding behaviour, but most of the times they did not even notice that the Euclidean rules were broken. Thus, we build spatial knowledge in a non-Euclidean environment, but whether this representation is different from that of real spaces, is an open question. The microgenesis of spatial knowledge (Siegel & White, 1975) consists of three levels, and while the first two levels (landmark and route knowledge) could easily be a non-Euclidean map, the third level representation is considered a metric representation of spatial surrounding (survey knowledge). In fact, the tessellating hexagonal grid firing pattern of entorhinal cells is thought to give the neural basis of such a metric knowledge (Hafting et al., 2005). It is relatively hard to imagine how grid cells would represent a non-Euclidean space. However, a recent study suggests that non-Euclidean spaces are indeed represented at the level of hippocampal oscillations (Vass et al., 2016). In that study, epileptic patients with implanted electrodes in their hippocampi performed a task that involved teleportation. They found greater oscillatory activity between 1.6 and 8 Hz (delta and theta band) when the distance travelled during the teleportation was longer. Thus, it seems that the Euclidean geometry and the physical space-time continuum do not limit the cognitive map (cf. Dragoi & Tonegawa, 2011; Horner, Bisby, Zotow, Bush, & Burgess, 2016).

On the one hand, real physical teleportation has only been done at the quantum level, and reasonably human teleportation is not going to be a reality for a long time

(Bouwmeester et al., 1997; Janszky, Gábris, Koniorczyk, Vukics, & Asbóth, 2002). On the other hand, teleportation in virtual reality is easy and frequently used (e.g. in games) (Inchamnan, Wyeth, & Johnson, 2013). Virtual reality application utilize not only teleportation but three dimensional navigation (Galambos & Baranyi, 2011; Hámornik, Köles, Komlódi, Hercegfi, & Izsó, 2014) and situations where the actors size change (e.g. zooming in by shrinking). Since virtual reality applications and devices are increasingly popular (Desai, Desai, Ajmera, & Mehta, 2014; Matthies et al., 2014), knowing how to design optimal virtual environments and navigation interfaces are especially up to date.

Changing the size of the observer (e.g. shrinking ourselves to the cell level) is useful when, for example, an educational application is teaching the medical students the structure and function of the organs at multiple levels (McCloy & Stone, 2001). Studies on rodents suggest that spatial specific neurons adapt their firing when the enclosure's size changes, which can be viewed alternatively as a change in the observer size. For example, grid cells stretch or squeeze their grid vertex when the environment is increased or decreased, respectively (Barry et al., 2007). Although studies with humans and with larger scale changes are yet to be done, these results suggest that changing the scale in virtual reality might not be problematic to process for humans.

Unlike scale change and teleportation, three dimensional navigation is proven to be difficult for humans, especially, when it implies rotational movements in three dimensional space (Peters, 1969). This is not surprising since, spatial perception is essentially a multisensory process where the earth-vertical axis remains the most basic spatial knowledge for humans (Angelaki et al., 2009; Clément et al., 2013) even if views can be visually similar in any direction. In fact, representation of three dimensional space has only been verified in bats (Finkelstein et al., 2015; Yartsev & Ulanovsky, 2013). Bats are flying animals, and they use echolocation as their primary distal sensory system. Importantly, the activity of the hippocampal formation in bats does not exhibit oscillatory activity in the theta band, which, in turn, is an essential functional correlate in both rodents and humans (Heys, MacLeod, Moss, & Hasselmo, 2013; Yartsev & Ulanovsky, 2013). Therefore, the spatial representation in bats is different from that in rats (Geva-Sagiv, Las, Yovel, & Ulanovsky, 2015) and presumably from that in humans, too. Thus, although some nervous systems have developed to deal with three dimensional navigation, the human brain has not. Hence

we show impaired performance in tasks that require simultaneous use of three dimensions (cf. Lógó et al., 2014).

These views are in line with the arguments of ecological psychology (J. J. Gibson, 1979) and those of embodied cognition (Haselager, van Dijk, & van Rooij, 2008; M. Wilson, 2002). They both claim that the cognitive system is not separable from the body (Proffitt, 2006) and the environment (E. J. Gibson & Walk, 1960). The natural fit between our cognition and the environment makes us able to cope with the wast amount of information reaching our senses at any given moment (Haselager et al., 2008). This situatedness is crucial for making reactions based on quickly emerging representations. A good example for this is the experiment in Chapter 5 of the current dissertation, where participants reoriented in only 100 msec after teleportation. Interestingly, the current work takes a rather unique way to study the environment embodied mind: virtual reality. These arguments explain the multisensory phenomena of ventriloquism in Chapter 7. Furthermore, the embodied nature of perception is what helping us explain the direction of distance bias in the experiment of Chapter 9. Interestingly, the current work takes a rather unique way to study the environment embodied mind: virtual reality. This raises further the importance of our results of viewpoint and reference frame associations in Chapter 3. Further studies should investigate the how the bird's eye view is grounded to real world experiences, and whether if it is grounded at all without exposure to maps (Barsalou, 2008; H. Zhang et al., 2012).

These results may have important implications for several applied fields and fit well in the synergetic field of cognitive infocommunication (Baranyi & Csapó, 2012). The main objective of cognitive infocommunication is to systematically define how cognitive processes and infocommunication devices can co-evolve. Essentially, the studies summarized in the present chapter collected evidence from the fields of behavioural, system, and human neuroscience and show how the brain copes with the physics of virtual reality. Together with the results of our own experiments the significance of the current work is not limited to exploratory science but has direct implications for the design of future infocommunication devices.

Of course, the current thesis is not the first initiative to study how humans interact with complex three dimensional virtual environments (e.g. Bohil, Alicea, & Biocca, 2011; Keszei et al., 2014; Sziebig & Øritsland, 2012). Nonetheless, integrative view of

neuroscientific, cognitive, and psychological results makes the present work a unique contribution to the field. The presented studies along with others not presented here (Honbolygo et al., 2014; Á. Török, Tóth, et al., 2013; Á. Török, Csépe, et al., 2015; Á. Török, Sulykos, et al., 2014) examined different levels of interaction with different types of virtual environments, thus, provide a good overview on how human spatial cognition works and can be manipulated in virtual reality.

In summary, four studies were presented here. In the first study, using a paradigm involving navigation on tablet pc based virtual environment, we showed that there is an implicit association between viewpoints and reference frames. Ground level viewpoint is associated with egocentric, while bird's eye view is associated with allocentric reference frame use. In the second study, we investigated under which conditions people still use allocentric reference frame in ground level navigation. In a CAVE virtual environment EEG was measured while participants collected rewards in the arms of a cross shaped maze. According to our results, participants quickly reorient themselves at the beginning of each trial, and this reorientation is signalled by a modulation of early visual evoked potentials. Furthermore, we found that when the task implies the possibility of different starting points, reward object locations are represented in an allocentric reference frame. Because both studies relied primarily on the visual modality, in the third study we looked at whether visual perception indeed dominates other senses, specifically hearing, in virtual reality. Although sound generation mode (free field speaker and surround set) has some effects on the localization of sounds, the locations are primarily affected by the position of a concurrent visual stimulus on both the horizontal and the vertical axes. Finally, in the fourth study we examined whether the vestibular system affects visual distance estimations, providing evidence for the notion that even though vision dominates our spatial localization responses, multisensory links can change the visual percept. We demonstrated that the pitch of the head affects the estimated visual distance; namely, distances look shorter downwards and longer upwards.

These results call for further interdisciplinary exploration on how the cognitive map functions in our head. We showed that several factors contribute to how we see space, among others: our viewpoint (Chapter 3), the task specifications (Chapter 5), the reliability of individual senses (Chapter 7), and gravity (Chapter 9). Thus, different aspects of the cognitive map are activated under different circumstances, making the

cognitive map more like a cognitive collage (Tversky, 1993). This map contains the relative positions of places in a hierarchical manner (Tversky, 1981) where route length and intersections are stored with local metric labels (Chrastil & Warren, 2014) in a form that allows the existence of non-Euclidean shortcuts (Schnapp & Warren, 2007; Vass et al., 2016). On the other hand, there are relations and routes that are difficult for our brain to process (Peters, 1969; Shepard & Metzler, 1971; Spelke et al., 2010).

The aim of this chapter was to put the cognitive map in an applied perspective; thus, we would like to close this work with a take home message. Despite the increasing interest in virtual reality tools and applications, there are still unresolved questions in how to develop an optimal virtual reality interface, both from the cognitive and from the affective (i.e. experience) point of views. For example, should we be able to move freely in any direction with any velocity, or should the laws of the virtual navigation somehow reflect the laws of real world navigation? Bearing the objective of cognitive infocommunication research and the results collected in the present work in mind, we suggest a general reframing solution: Instead of freeing us from the bonds of our physical body, virtual reality should utilize these bonds as bridges and extend our mind beyond the laws of physics.

11 REFERENCES

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, *122*(9), 1613–1628.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology : CB*, *14*(3), 257–62. http://doi.org/10.1016/j.cub.2004.01.029

Alexander, A. S., & Nitz, D. (2015). Retrosplenial cortex maps the conjunction of internal and external spaces. *Nature Neuroscience*, *18*(8), 1143–1151. http://doi.org/10.1038/nn.4058

Alexandrova, I. V, Teneva, P. T., de la Rosa, S., Kloos, U., Bülthoff, H. H., & Mohler, B. J. (2010). Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (pp. 57–60). ACM.

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–90. http://doi.org/10.1016/j.tics.2013.06.009

Andersen, N. E., Dahmani, L., Konishi, K., & Bohbot, V. D. (2012). Eye tracking, strategies, and sex differences in virtual navigation. *Neurobiology of Learning and Memory*, *97*(1), 81–9. http://doi.org/10.1016/j.nlm.2011.09.007

Andre, J., & Rogers, S. (2006). Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, *68*(3), 353–361. http://doi.org/10.3758/BF03193682

Angelaki, D. E., Klier, E. M., & Snyder, L. H. (2009). A vestibular sensation: probabilistic approaches to spatial perception. *Neuron*, *64*(4), 448–461.

Angelaki, D. E., Shaikh, A. G., Green, A. M., & Dickman, J. D. (2004). Neurons compute internal models of the physical laws of motion. *Nature*, *430*(6999), 560–564.

Aretz, A. J. (1991). The Design of Electronic Map Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *33*(1), 85–101. http://doi.org/10.1177/001872089103300107

Auger, S. D., Mullally, S. L., & Maguire, E. (2012). Retrosplenial cortex codes for permanent landmarks. *PloS One*, *7*(8), e43620. http://doi.org/10.1371/journal.pone.0043620

Bagrow, L. (2010). *History of cartography*. Transaction Publishers.

Baker, T. E., & Holroyd, C. B. (2009). Which way do I go? Neural activation in response to feedback and spatial processing in a virtual T-maze. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(8), 1708–22. http://doi.org/10.1093/cercor/bhn223

Baker, T. E., & Holroyd, C. B. (2013). The topographical N170: electrophysiological evidence of a neural mechanism for human spatial navigation. *Biological Psychology*, *94*(1), 90–105. http://doi.org/10.1016/j.biopsycho.2013.05.004

Baker, T. E., Umemoto, A., Krawitz, A., & Holroyd, C. B. (2015). Rightward-biased hemodynamic response of the parahippocampal system during virtual navigation.

Scientific Reports, *5*, 9063. http://doi.org/10.1038/srep09063

Balázs, L., Barkaszi, I., Czigler, I., & Takács, E. (2015). Agyműködés Súlytalanságban: Kísérlet A Nemzetközi Űrállomáson. *Magyar Tudomány*, (9), 1045–1052.

Ball, S. (2015). Slow Cities. In *Theme Cities: Solutions for Urban Problems* (pp. 563–585). Springer.

Barany, R. (1906). *Untersuchungen über den vom Vestibularapparat des Ohres reflektorisch ausgelösten rhythmischen Nystagmus und seine Begleiterscheinungen*. Oscar Coblentz.

Baranyi, P., & Csapó, Á. (2012). Definition and Synergies of Cognitive Infocommunications. *Acta Polytechnica Hungarica*, *9*(1), 67–83.

Baranyi, P., Csapó, Á., & Sallai, G. (2015). *Cognitive Infocommunications (CogInfoCom)*. Springer.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Barra, J., Laou, L., Poline, J.-B., Lebihan, D., & Berthoz, A. (2012). Does an oblique/slanted perspective during virtual navigation engage both egocentric and allocentric brain strategies? *PloS One*, *7*(11), e49537. http://doi.org/10.1371/journal.pone.0049537

Barry, C., Hayman, R., Burgess, N., & Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, *10*(6), 682–4. http://doi.org/10.1038/nn1905

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Bates, D. M. (2008). Fitting Mixed-Effects Models Using the lme4 Package in R.

Bates, D. M. (2010). lme4: Mixed-effects modeling with R. *URL http://lme4. R-Forge. R-Project. Org/book*.

Bates, S. L., & Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiology of Aging*, *35*(12), 2761–2769.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489. http://doi.org/10.3758/BF03208826

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: patterns of spatial generalization. *Perception & Psychophysics*, *68*(3), 428–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16900834

Bertini, C., Leo, F., Avenanti, A., & Làdavas, E. (2010). Independent mechanisms for ventriloquism and multisensory integration as revealed by theta-burst stimulation. *The European Journal of Neuroscience*, *31*(10), 1791–9. http://doi.org/10.1111/j.1460-9568.2010.07200.x

Bertrand, O., Perrin, F., & Pernier, J. (1985). A theoretical justification of the average

reference in topographic evoked potential studies. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*(6), 462–464. http://doi.org/10.1016/0168-5597(85)90058-9

Besson, P., Bourdin, C., & Bringoux, L. (2011). A comprehensive model of audiovisual perception: both percept and temporal dynamics. *PloS One*, *6*(8), e23811. http://doi.org/10.1371/journal.pone.0023811

Besson, P., Richiardi, J., Bourdin, C., Bringoux, L., Mestre, D. R., & Vercher, J.-L. (2010). Bayesian networks and information theory for audio-visual perception modeling. *Biological Cybernetics*, *103*(3), 213–26. http://doi.org/10.1007/s00422-010-0392-8

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1076–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10464946

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews. Neuroscience*, *9*(3), 182–194. http://doi.org/10.1038/nrn2335

Bird, C. M., Capponi, C., King, J. A., Doeller, C. F., & Burgess, N. (2010). Establishing the boundaries: the hippocampal contribution to imagining scenes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(35), 11688–95. http://doi.org/10.1523/JNEUROSCI.0723-10.2010

Bjelland, M., Montello, D., Fellmann, J., Getis, A., & Getis, J. (2013). *Human Geography: Landscapes of Human Activities*. McGraw-Hill Science/Engineering/Math. Retrieved from http://www.amazon.com/Human-Geography-Landscapes-Activities/dp/0078021464

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, *13*(8), 556–571.

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Neuropsychology: Stimulating illusory own-body perceptions. *Nature*, *419*(6904), 269–270.

Blodgett, H. C. (1929). The effect of the introduction of reward upon the maze performance of rats. *University of California Publications in Psychology*.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews. Neuroscience*, *12*(12), 752–62. http://doi.org/10.1038/nrn3122

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology : CB*, *17*(19), 1697–703. http://doi.org/10.1016/j.cub.2007.08.050

Botreau, F., & Gisquet-Verrier, P. (2010). Re-thinking the role of the dorsal striatum in egocentric/response strategy. *Frontiers in Behavioral Neuroscience*, *4*, 7. http://doi.org/10.3389/neuro.08.007.2010

Bottini, G., Sterzi, R., Paulesu, E., Vallar, G., Cappa, S. F., Erminio, F., … Frackowiak, R. S. J. (1994). Identification of the central vestibular projections in man: a positron emission tomography activation study. *Experimental Brain Research*, *99*(1), 164–

169.

Botvinick, M., & Cohen, J. (1998). Rubber hands' feel'touch that eyes see. *Nature*, *391*(6669), 756.

Bouwmeester, D., Pan, J.-W., Mattle, K., Eibl, M., Weinfurter, H., & Zeilinger, A. (1997). Experimental quantum teleportation. *Nature*, *390*(6660), 575–579.

Bömmer, A., & Mast, F. (1999). Assessing otolith function by the subjective visual vertical. *Annals of the New York Academy of Sciences*, *871*, 221–231. http://doi.org/10.1111/j.1749-6632.1999.tb09187.x

Brun, V. H., Otnæss, M. K., Molden, S., Steffenach, H.-A., Witter, M. P., Moser, M.-B., & Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science*, *296*(5576), 2243–2246.

Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2012). Going to town: Visualized perspectives and navigation through virtual environments. *Computers in Human Behavior*, *28*(1), 257–266. http://doi.org/10.1016/j.chb.2011.09.008

Brunyé, T. T., Wood, M. D., Houck, L. A., & Taylor, H. A. (2016). The Path More Traveled: Time Pressure Increases Reliance on Familiar Route-based Strategies during Navigation. *The Quarterly Journal of Experimental Psychology*, (just-accepted), 1–38.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. http://doi.org/10.1016/j.tics.2006.10.005

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition*, *94*(2), 149–66. http://doi.org/10.1016/j.cognition.2004.01.001

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, *16*(2), 130–138.

Caplan, J. B., Madsen, J. R., Schulze-Bonhage, A., Aschenbrenner-Scheibe, R., Newman, E. L., & Kahana, M. J. (2003). Human theta oscillations related to sensorimotor integration and spatial learning. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(11), 4726–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12805312

Carr, H., & Watson, J. B. (1908). Orientation in the white rat. *Journal of Comparative Neurology and Psychology*, *18*(1), 27–44.

Chadwick, M. J., Jolly, A. E. J., Amos, D. P., Hassabis, D., & Spiers, H. J. (2015). A goal direction signal in the human entorhinal/subicular region. *Current Biology*, *25*(1), 87–92.

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*(AUG), 304. http://doi.org/10.3389/fpsyg.2012.00304

Chang, Q., & Gold, P. E. (2003). Switching Memory Systems during Learning: Changes in Patterns of Brain Acetylcholine Release in the Hippocampus and Striatum in Rats. *J. Neurosci.*, *23*(7), 3001–3005. Retrieved from http://www.jneurosci.org/content/23/7/3001.full

Chen, C.-H., Chang, W.-C., & Chang, W. (2008). Gender differences in relation to wayfinding strategies , navigational support design , and wayfinding task difficulty. *Journal of Environmental Psychology*, *29*(2), 1–7. http://doi.org/10.1016/j.jenvp.2008.07.003

Chen, G., King, J. A., Burgess, N., & O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 378–83. http://doi.org/10.1073/pnas.1215834110

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*(2), 149–178.

Chiu, T., Gramann, K., Ko, L., Duann, J., Jung, T., & Lin, C.-T. (2012). Alpha modulation in parietal and retrosplenial cortex correlates with navigation performance. *Psychophysiology*, *49*(1), 43–55.

Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P. R., … van den Hove, D. L. A. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiology of Aging*, *34*(9), 2091–2099.

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. *PloS One*, *9*(11), e112544. http://doi.org/10.1371/journal.pone.0112544

Clemens, Z., Borbély, C., Weiss, B., Erőss, L., Szűcs, A., Kelemen, A., … Halász, P. (2013). Increased mesiotemporal delta activity characterizes virtual navigation in humans. *Neuroscience Research*, *76*(1-2), 67–75. http://doi.org/10.1016/j.neures.2013.03.004

Clément, G., Fraysse, M.-J., & Deguine, O. (2009). Mental representation of space in vestibular patients with otolithic or rotatory vertigo. *Neuroreport*, *20*, 457–461. http://doi.org/10.1097/WNR.0b013e328326f815

Clément, G., Lathan, C., & Lockerd, A. (2008). Perception of depth in microgravity during parabolic flight. *Acta Astronautica*, *63*(7-10), 828–832. http://doi.org/10.1016/j.actaastro.2008.01.002

Clément, G., Skinner, A., & Lathan, C. (2013). Distance and Size Perception in Astronauts during Long-Duration Spaceflight. *Life*, *3*(4), 524–537. http://doi.org/10.3390/life3040524

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *113*(4), 507–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11955995

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009.

Colonius, H., & Diederich, A. (2011). The Multisensory Driver: Contributions from the Time-Window-of-Integration Model. In P. C. Cacciabue, M. Hjälmdahl, A. Luedtke, & C. Riccioli (Eds.), *Human Modelling in Assisted Transportation SE - 39* (pp. 363–

371). Springer Milan. http://doi.org/10.1007/978-88-470-1821-1_39

Colonius, H., Diederich, A., & Steenken, R. (2009). Time-window-of-integration (TWIN) model for saccadic reaction time: effect of auditory masker level on visual–auditory spatial interaction in elevation. *Brain Topography*, *21*(3-4), 177–184. http://doi.org/10.1007/s10548-009-0091-8

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-screen projection-based virtual reality. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93* (pp. 135–142). New York, New York, USA: ACM Press. http://doi.org/10.1145/166117.166134

Cullen, K. E. (2012). The vestibular system: Multimodal integration and encoding of self-motion for motor control. *Trends in Neurosciences*, *35*(3), 185–196. http://doi.org/10.1016/j.tins.2011.12.001

Curthoys, I. S., & MacDougall, H. G. (2012). What galvanic vestibular stimulation actually activates. *Frontiers in Neurology*, *JUL*(July), 1–5. http://doi.org/10.3389/fneur.2012.00117

Csapó, Á., & Wersényi, G. (2013). Overview of auditory representations in human-machine interfaces. *ACM Computing Surveys (CSUR)*, *46*(2), 19.

Csibra, G., Davis, G., Spratling, M. W., & Johnson, M. H. (2000). Gamma oscillations and object processing in the infant brain. *Science*, *290*(5496), 1582–1585.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15102499

Desai, P. R. P. N., Desai, P. R. P. N., Ajmera, K. D., & Mehta, K. (2014). A review paper on oculus rift-A virtual reality headset. *arXiv Preprint arXiv:1408.1173*.

Di Cesare, C. S., Sarlegna, F. R., Bourdin, C., Mestre, D. R., & Bringoux, L. (2014). Combined influence of visual scene and body tilt on arm pointing movements: gravity matters! *PloS One*, *9*(6), e99866.

Diamond, M. E., von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., & Ahissar, E. (2008). "Where"and'what'in the whisker sensorimotor system. *Nature Reviews Neuroscience*, *9*(8), 601–612.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*(8), 1388–404. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15813202

Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological Review*, *122*(2), 232.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint Dependence in Scene Recognition. *Psychological Science*, *8*(4), 302. http://doi.org/10.1111/j.1467-9280.1997.tb00442.x

Dix, M. R., & Hallpike, C. S. (1952). The pathology, symptomatology and diagnosis of certain common disorders of the vestibular system. *The Annals of Otology,*

Rhinology, and Laryngology, *61*(4), 987–1016. http://doi.org/10.1136/bmj.2.1821.1289

Dixon, W. J., & Yuen, K. K. (1974). Trimming and winsorization: A review. *Statistische Hefte*, *15*(2-3), 157–170.

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–61. http://doi.org/10.1038/nature08704

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5909–14. http://doi.org/10.1073/pnas.0711433105

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5915–20. http://doi.org/10.1073/pnas.0801489105

Doise, W., Mugny, G., & Perret-Clermont, A.-N. (1982). Pleading for social interaction across the Ocean: A rejoinder to Ames and Murray. In *Neuchâtel et Genève: Universités de Genève et Neuchâte*.

Dokka, K., MacNeilage, P. R., DeAngelis, G. C., & Angelaki, D. E. (2011). Estimating distance during self-motion: a role for visual-vestibular interactions. *Journal of Vision*, *11*(13). http://doi.org/10.1167/11.13.2

Dori, D. (2011). *Object-process methodology: A holistic systems paradigm*. Springer Science & Business Media.

Dragoi, G., & Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, *469*(7330), 397–401. http://doi.org/10.1038/nature09633

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., … Nishida, S. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–8. http://doi.org/10.1038/nature01964

Ekstrom, A. D., Suthana, N., Behnke, E., Salamon, N., Bookheimer, S., & Fried, I. (2008). High-resolution depth electrode localization and imaging in patients with pharmacologically intractable epilepsy. *Journal of Neurosurgery*, *108*(4), 812–5. http://doi.org/10.3171/JNS/2008/108/4/0812

Eley, M. G. (1988). Determining the shapes of landsurfaces from topographical maps. *Ergonomics*, *31*(3), 355–376. http://doi.org/10.1080/00140138808966680

Epstein, J. N., Langberg, J. M., Rosen, P. J., Graham, A., Narad, M. E., Antonini, T. N., … Altaye, M. (2011). Evidence for higher reaction time variability for children with ADHD on a range of cognitive tasks including reward and event rate manipulations. *Neuropsychology*, *25*(4), 427–41. http://doi.org/10.1037/a0022155

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–96. http://doi.org/10.1016/j.tics.2008.07.004

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(23), 6141–9. http://doi.org/10.1523/JNEUROSCI.0799-07.2007

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Farrell, M. J. (1996). Topographical disorientation. *Neurocase*, *2*(6), 509–520. http://doi.org/10.1080/13554799608402427

Ferrè, E. R., Bottini, G., Iannetti, G. D., & Haggard, P. (2013). The balance of feelings: vestibular modulation of bodily sensations. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(3), 748–58. http://doi.org/10.1016/j.cortex.2012.01.012

Ferrè, E. R., Longo, M. R., Fiori, F., & Haggard, P. (2013). Vestibular modulation of spatial perception. *Frontiers in Human Neuroscience*, *7*, 660. http://doi.org/10.3389/fnhum.2013.00660

Ferrè, E. R., Vagnoni, E., & Haggard, P. (2013). Vestibular contributions to bodily awareness. *Neuropsychologia*, *51*(8), 1445–1452.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–54. http://doi.org/10.1038/nn.2983

Finkelstein, A., Derdikman, D., Rubin, A., Foerster, J. N., Las, L., & Ulanovsky, N. (2015). Three-dimensional head-direction coding in the bat brain. *Nature*, *517*(7533), 159–164.

Finney, B. (1995). A Role for Magnetoreception in Human Navigation ? *Current Anthropology*, *36*(3), 500–506.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499–504.

Fitzpatrick, R. C., & Day, B. L. (2004). Probing the human vestibular system with galvanic stimulation. *Journal of Applied Physiology*, *96*(6), 2301–2316.

Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., Walhovd, K. B., & Initiative, A. D. N. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, *117*, 20–40.

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map-versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 195.

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*(12), 2207–2216.

Fouque, F., Bardy, B. G., Stoffregen, T. A., & Bootsma, R. J. (1999). Action and intermodal information influence the perception of orientation. *Ecological Psychology*, *11*(1), 1–43.

Freund, T. F., & Buzsáki, G. (1998). Interneurons of the hippocampus. *Hippocampus*, *6*(4), 347–470. http://doi.org/10.1002/(SICI)1098-1063(1996)6:4<347::AID-HIPO1>3.0.CO;2-I

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science*, *305*(5688), 1258–1264.

Gajewski, D. a, Wallin, C. P., & Philbeck, J. W. (2014). Gaze behavior and the perception of egocentric distance. *J Vis*, *14*(1), 1–19. http://doi.org/10.1167/14.1.20.doi

Galambos, P., & Baranyi, P. (2011). Virca as virtual intelligent space for rt-middleware. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 140–145.

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. http://doi.org/10.1007/s00221-010-2168-8

Geva-Sagiv, M., Las, L., Yovel, Y., & Ulanovsky, N. (2015). Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation. *Nature Reviews Neuroscience*, *16*(2), 94–108.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–85. http://doi.org/10.1016/j.tics.2006.04.008

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10511637

Gibson, E. J., & Walk, R. D. (1960). *The" visual cliff."* WH Freeman Company.

Gibson, J. J. (1979). *The ecological approach to visual perception: classic edition*. Psychology Press.

Goeke, C., Kornpetpanee, S., Köster, M., Fernández-Revelles, A. B., Gramann, K., & König, P. (2015). Cultural background shapes spatial reference frame proclivity. *Scientific Reports*, *5*.

Goldberg, J. M., Smith, C. E., & Fernández, C. (1984). Relation between discharge regularity and responses to externally applied galvanic currents in vestibular nerve afferents of the squirrel monkey. *Journal of Neurophysiology*, *51*(6), 1236–1256.

Goldfarb, E. V., Chun, M. M., & Phelps, E. A. (2016). Memory-Guided Attention: Independent Contributions of the Hippocampus and Striatum. *Neuron*, *89*(2), 317–324. http://doi.org/http://dx.doi.org/10.1016/j.neuron.2015.12.014

Golledge, R. G., Dougherty, V., & Bell, S. (1993). *Survey versus Route-Based Wayfinding in Unfamiliar Environments*. University of California Transportation Center. Retrieved from http://econpapers.repec.org/RePEc:cdl:uctcwp:qt1km115qr

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and

action. *Trends in Neurosciences*, *15*(1), 20–25. http://doi.org/10.1016/0166-2236(92)90344-8

Gottfried, J. A., & Dolan, R. J. (2003). The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron*, *39*(2), 375–386.

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. http://doi.org/10.1016/j.brainres.2006.08.005

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), 2836–49. http://doi.org/10.1162/jocn.2009.21369

Grüsser, O. J., Pause, M., & Schreiter, U. (1990). Localization and responses of neurones in the parieto-insular vestibular cortex of awake monkeys (Macaca fascicularis). *The Journal of Physiology*, *430*, 537.

Guldin, W. O., & Grüsser, O. J. (1998). Is there a vestibular cortex? *Trends in Neurosciences*, *21*(6), 254–259.

Haans, A., & IJsselsteijn, W. A. (2012). Embodiment and telepresence: Toward a comprehensive theoretical framework. *Interacting with Computers*, *24*(4), 211–218. http://doi.org/10.1016/j.intcom.2012.04.010

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–6. http://doi.org/10.1038/nature03721

Hámornik, B. P., Köles, M., Komlódi, A., Hercegfi, K., & Izsó, L. (2014). Features of Collaboration in the VirCA Immersive 3D Environment. *Advances in Cognitive Engineering and Neuroergonomics-Proceedings of AHFE 2014*.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.

Harris, L. R., & Mander, C. (2014). Perceived distance depends on the orientation of both the body and the visual environment. *Journal of Vision*, *14*(12), 17. http://doi.org/10.1167/14.12.17

Hartley, T., Lever, C., Burgess, N., & O'Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1635), 20120510. http://doi.org/10.1098/rstb.2012.0510

Hartnagel, D., Bichot, A., & Roumes, C. (2007). Eye position affects audio-visual fusion in darkness. *Perception*, *36*(10), 1487–1496. Retrieved from http://www.perceptionweb.com/abstract.cgi?id=p5847

Harvey, C. D., Collman, F., Dombeck, D. A., & Tank, D. W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, *461*(7266), 941–946. http://doi.org/10.1038/nature08499

Haselager, P., van Dijk, J., & van Rooij, I. (2008). A lazy brain? Embodied embedded cognition and cognitive neuroscience. *Handbook of Cognitive Science: An Embodied Approach*, *5*, 273–287.

Hein, A., Held, R., & Gower, E. C. (1970). Development and segmentation of visually controlled movement by selective exposure during rearing. *Journal of Comparative and Physiological Psychology*, *73*(2), 181.

Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, *56*(5), 872.

Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments*, *5*(3), 290–301.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*(3), 195–232.

Heys, J. G., MacLeod, K. M., Moss, C. F., & Hasselmo, M. E. (2013). Bat and rat neurons differ in theta-frequency resonance despite similar coding of space. *Science*, *340*(6130), 363–367.

Hitier, M., Besnard, S., & Smith, P. F. (2014). Vestibular pathways involved in cognition. *Frontiers in Integrative Neuroscience*, *8*, 59. http://doi.org/10.3389/fnint.2014.00059

Ho, C., Reed, N., & Spence, C. (2007). Multisensory in-car warning signals for collision avoidance. *Human Factors*, *49*(6), 1107–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18074709

Ho, C., Santangelo, V., & Spence, C. (2009). Multisensory warning signals: when spatial correspondence matters. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *195*(2), 261–72. http://doi.org/10.1007/s00221-009-1778-5

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behavior Research Methods*, *39*(1), 101–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17552476

Holmes, M. C., & Sholl, M. J. (2005). Allocentric coding of object-to-object relations in overlearned and novel environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1069.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679.

Holroyd, C. B., Krigolson, O. E., & Lee, S. (2011). Reward positivity elicited by predictive cues. *Neuroreport*, *22*(5), 249–252.

Honbolygo, F., Babik, A., & Török, Á. (2014). Location learning in virtual environments: The effect of saliency of landmarks and boundaries. In *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 595–598). IEEE. http://doi.org/10.1109/CogInfoCom.2014.7020413

Hori, E., Nishio, Y., Kazui, K., Umeno, K., Tabuchi, E., Sasaki, K., … Nishijo, H. (2005). Place-related neural responses in the monkey hippocampal formation in a virtual space. *Hippocampus*, *15*(8), 991–996.

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like Processing of Imagined Navigation. *Current Biology*.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. John Wiley & Sons Ltd. Retrieved from http://www.amazon.com/Human-Spatial-Orientation-Ian-Howard/dp/0471416622

Hu, H., Zhou, L., Ma, H., & Wu, Z. (2008). HRTF personalization based on artificial neural network in individual virtual auditory space. *Applied Acoustics*, *69*(2), 163–172. http://doi.org/10.1016/j.apacoust.2007.05.007

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*, 574–91. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1363130&tool=pmcentrez&rendertype=abstract

Hughes, K. R. (1965). Dorsal and ventral hippocampus lesions and maze learning: Influence of preoperative environment. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *19*(4), 325.

Huhn, Z., Szirtes, G., Lorincz, A., & Csépe, V. (2009). Perception based method for the investigation of audiovisual integration of speech. *Neuroscience Letters*, *465*(3), 204–9. http://doi.org/10.1016/j.neulet.2009.08.077

Hupbach, A., & Nadel, L. (2005). Reorientation in a rhombic environment: no evidence for an encapsulated geometric module. *Cognitive Development*, *20*(2), 279–302.

Iglói, K., Zaoui, M., Berthoz, A., & Rondi-Reig, L. (2009). Sequential egocentric strategy is acquired as early as allocentric strategy: Parallel acquisition of these two navigation strategies. *Hippocampus*, *19*, 1199–1211. http://doi.org/10.1002/hipo.20595

Inchamnan, W., Wyeth, P., & Johnson, D. (2013). Does activity in computer game play have an impact on creative behaviour? In *Games Innovation Conference (IGIC), 2013 IEEE International* (pp. 77–84). IEEE.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., … Blanke, O. (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron*, *70*(2), 363–374.

Jackson, R. E., & Cormack, L. K. (2007). Evolved navigation theory and the descent illusion. *Perception & Psychophysics*, *69*(3), 353–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17672423

Jackson, R. E., & Cormack, L. K. (2010). Reducing the presence of navigation risk eliminates strong environmental illusions. *Journal of Vision*, *10*(5), 9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20616118

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., … Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience*, *16*(9), 1188–90. http://doi.org/10.1038/nn.3466

Janszky, J., Gábris, A., Koniorczyk, M., Vukics, A., & Asbóth, J. (2002). One-complex-plane representation: a coherent-state description of entanglement and teleportation. *Journal of Optics B: Quantum and Semiclassical Optics*, *4*(3), S213.

Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, *7*(6), 673–7. http://doi.org/10.1038/nn1257

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M.-B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*(7368), 246–9. http://doi.org/10.1038/nature10439

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science*, *145*(3630), 356–362.

Julesz, B. (1971). *Foundations of cyclopean perception.* Oxford, England: U. Chicago Press.

Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, *14*(12), 7347–7356.

Kardos, L., da Pos, O., Dellantonio, A., & Saviolo, N. (1978). Discrimination learning—with continuously changing goal objects. *Magyar Pszichologiai Szemle*.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997.

Keszei, B., Dúll, A., Hámornik, B. P., Köles, M., Tóvölgyi, S., Hercegfi, K., … Lógó, E. (2014). Visual attention and spatial behavior in VR environment: an environmental psychology approach. In *Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on* (pp. 247–250). IEEE. http://doi.org/10.1109/CogInfoCom.2014.7020455

Killian, N. J., Jutras, M. J., & Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex. *Nature*, *491*(7426), 761–764. http://doi.org/10.1038/nature11587

Klatzky, R. L. (1998). Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections. In *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (pp. 1–18). London, UK: Springer-Verlag. http://doi.org/10.1007/3-540-69342-4_1

Klausberger, T., Magill, P. J., Márton, L. F., Roberts, J. D. B., Cobden, P. M., Buzsáki, G., & Somogyi, P. (2003). Brain-state-and cell-type-specific firing of hippocampal interneurons in vivo. *Nature*, *421*(6925), 844–848.

Klein, E., Swan, J. E., Schmidt, G. S., Livingston, M., & Staadt, O. G. (2009). Measurement protocols for medium-field distance perception in large-screen immersive displays. In *Virtual Reality Conference, 2009. VR 2009. IEEE* (pp. 107–113). IEEE.

Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *15*(3 Pt 1), 1648–59. Retrieved from http://www.jneurosci.org/content/15/3/1648.abstract

Knierim, J. J., Neunuebel, J. P., & Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: objects, path integration and local-global reference frames. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1635), 20130369. http://doi.org/10.1098/rstb.2013.0369

Kober, S. E., Kurzmann, J., & Neuper, C. (2012). Cortical correlate of spatial presence in

2D and 3D interactive virtual reality: an EEG study. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *83*(3), 365–74. http://doi.org/10.1016/j.ijpsycho.2011.12.003

Kóbor, I., Füredi, L., Kovács, G., Spence, C., & Vidnyánszky, Z. (2006). Back-to-front: Improved tactile discrimination performance in the space you cannot see. *Neuroscience Letters*, *400*(1), 163–167.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–84. http://doi.org/10.1016/j.actpsy.2010.03.010

Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience*, *2011*, 938925. http://doi.org/10.1155/2011/938925

Koenig, T., & Melie-Garcia, L. (2009). Statistical analysis of multichannel scalp field data. *Electrical Neuroimaging*, 169–189.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography*, *23*(3), 233–242.

Koffka, K. (1922). Perception: An introduction to the Gestalt-theorie. *Psychological Bulletin*, *19*(10), 531–585.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, *88*(1), 46.

Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, *90*(16), 7495–7497. http://doi.org/10.1073/pnas.90.16.7495

Kovács, I., Julesz, B., Kovacs, I., Julesz, B., Kovács, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, *370*(6491), 644–6. http://doi.org/10.1038/370644a0

Krechevsky, I. (1932). " Hypotheses" in rats. *Psychological Review*, *39*(6), 516.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1), 37–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13035466

Kunz, L., Schröder, T. N., Lee, H., Montag, C., Lachmann, B., Sariyska, R., … Messing-Floeter, P. C. (2015). Reduced grid-cell–like representations in adults at genetic risk for Alzheimer's disease. *Science*, *350*(6259), 430–433.

Lackner, J. R., & DiZio, P. (1993). Multisensory, cognitive, and motor influences on human spatial orientation in weightlessness. *Journal of Vestibular Research: Equilibrium & Orientation*.

Lashley, K. S. (1943). Studies of cerebral function in learning XII. Loss of the maze habit after occipital lesions in blind rats. *Journal of Comparative Neurology*, *79*(3), 431–462.

Lashley, K. S. (1950). In search of the engram. *Experimental Biology Symposium No. 4:*

Physiological Mechanisms in Animal Behaviour. http://doi.org/10.1097/00008877-199204001-00015

Lashley, K. S., & Ball, J. (1929). Spinal conduction and kinesthetic sensitivity in the maze habit. *Journal of Comparative Psychology*, *9*(1), 71.

Lashley, K. S., & McCarthy, D. A. (1926). The survival of the maze habit after cerebellar injuries. *Journal of Comparative Psychology*, *6*(6), 423.

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children's use of landmarks: Implications for modularity theory. *Psychological Science*, *13*(4), 337–341.

Lederbogen, F., Kirsch, P., Haddad, L., Streit, F., Tost, H., Schuch, P., … Deuschle, M. (2011). City living and urban upbringing affect neural social stress processing in humans. *Nature*, *474*(7352), 498–501.

Lee, J.-H., & Spence, C. (2009). Feeling what you hear: task-irrelevant sounds modulate tactile perception delivered via a touch screen. *Journal on Multimodal User Interfaces*, *2*(3-4), 145–156. http://doi.org/10.1007/s12193-009-0014-8

Lee, P. U., & Tversky, B. (2001). Costs of Switching Perspectives in Route and Survey Descriptions. In *Proceedings of the twenty-third Annual Conference of the Cognitive Science Society, Edinburgh, Scotland*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.196.6837

Lee, P. U., & Tversky, B. (2005). Interplay Between Visual and Spatial: The Effect of Landmark Descriptions on Comprehension of Route/Survey Spatial Descriptions. *Spatial Cognition & Computation*, *5*(2-3), 163–185. http://doi.org/10.1080/13875868.2005.9683802

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, *48*(6), 609–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6155251

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science (New York, N.Y.)*, *317*(5841), 1096–9. http://doi.org/10.1126/science.1143439

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*(5688), 1295–1298.

Lin, C.-T., Chiu, T.-C., & Gramann, K. (2015). EEG correlates of spatial orientation in the human retrosplenial complex. *NeuroImage*, *120*, 123–132.

Lin, Q., Xie, X., Erdemir, A., Narasimham, G., McNamara, T. P., Rieser, J., & Bodenheimer, B. (2011). Egocentric distance perception in real and HMD-based virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization - APGV '11* (p. 75). New York, New York, USA: ACM Press. http://doi.org/10.1145/2077451.2077465

Linde, C., & Labov, W. (1975). Spatial Networks as a Site for the Study of Language and Thought. *Language*, *51*(4), 924–939. http://doi.org/10.2307/412701

Lithfous, S., Dufour, A., Blanc, F., & Després, O. (2014). Allocentric but not egocentric orientation is impaired during normal aging: An ERP study. *Neuropsychology*,

28(5), 761–771. Retrieved from http://psycnet.apa.orgjournals/neu/28/5/761

Lithfous, S., Dufour, A., & Després, O. (2013). Spatial navigation in normal aging and the prodromal stage of Alzheimer's disease: insights from imaging and behavioral studies. *Ageing Research Reviews*, *12*(1), 201–13. http://doi.org/10.1016/j.arr.2012.04.007

Lloyd, D. (2014). In Touch with the Future: The Sense of Touch from Cognitive Neuroscience to Virtual Reality. *Presence: Teleoperators and Virtual Environments*, *23*(2), 226–227. http://doi.org/10.1162/PRES_r_00182

Lógó, E., Hamornik, B. P., Koles, M., Hercegfi, K., Tovolgyi, S., & Komlodi, A. (2014). Usability related human errors in a collaborative immersive VR environment. In *Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on* (pp. 243–246). IEEE.

Loomis, J., & Knapp, J. (2003). Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, (11), 21–46. http://doi.org/10.1201/9781410608888

Lopez, C., Lenggenhager, B., & Blanke, O. (2010). How vestibular stimulation interacts with illusory hand ownership. *Consciousness and Cognition*, *19*(1), 33–47.

Lynch, K. (1960). *The image of the city* (Vol. 11). MIT press.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, *97*(8), 4398–4403.

Maguire, E., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science*, *280*(5365), 921–924.

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.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

Márkus, A. (2006). *Neurológia. Pszichológia szakos hallgatók számára*. Akadémiai Kiadó, Budapest.

Marton, M. L. (1970). Learning visual-postural body-model and the evolution of consciousness. *Magyar Pszichologiai Szemle*.

Matthies, D. J. C., Manke, F. M., Müller, F., Makri, C., Anthes, C., & Kranzlmüller, D. (2014). VR-Stepper: A Do-It-Yourself Game Interface For Locomotion In Virtual Environments. *arXiv Preprint arXiv:1407.3948*.

McCloy, R., & Stone, R. (2001). Virtual reality in surgery. *British Medical Journal*, *323*(7318), 912.

McCormick, E. P., Wickens, C. D., Banks, R., & Yeh, M. (1998). Frame of Reference Effects on Scientific Visualization Subtasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(3), 443–451.

http://doi.org/10.1518/001872098779591403

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. http://doi.org/10.1038/264746a0

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), 589–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14620351

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the "cognitive map". *Nature Reviews Neuroscience*, *7*(8), 663–678.

Ménière, P. (1861). Memoire sur des lesions de l'oreille interne donnant lieu a des symptomes de congestion cerebrale apoplectiforme. *Gazette Medicale de Paris*, 597–601.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science (New York, N.Y.)*, *221*(4608), 389–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6867718

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*, *56*(3), 640–662. Retrieved from http://jn.physiology.org/content/56/3/640.short

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, *42*, 135–59. http://doi.org/10.1146/annurev.ps.42.020191.001031

Miller, A. M. P., Vedder, L. C., Law, L. M., & Smith, D. M. (2014). Cues, context, and long-term memory: the role of the retrosplenial cortex in spatial cognition. *Frontiers in Human Neuroscience*, *8*, 586. http://doi.org/10.3389/fnhum.2014.00586

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798.

Mizuseki, K., Royer, S., Diba, K., & Buzsáki, G. (2012). Activity dynamics and behavioral correlates of CA3 and CA1 hippocampal pyramidal neurons. *Hippocampus*, *22*(8), 1659–1680.

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment. *Presence: Teleoperators and Virtual Environments*, *19*(3), 230–242. http://doi.org/10.1162/pres.19.3.230

Mollison, M. V. (2005). Event-related potentials in humans during spatial navigation. Brandeis University.

Monmonier, M. (2010). *Rhumb lines and map wars: A social history of the Mercator projection*. University of Chicago Press.

Montello, D. R. (2005). Navigation. In P. Shah & A. Miyake (Eds.), *The Cambridge Handbook of visuospatial thinking* (pp. 257–294). New York: Cambridge University Press.

Moore, D. R., & King, A. J. (1999). Auditory perception: The near and far of sound localization. *Current Biology*, *9*(10), R361 – R363. http://doi.org/10.1016/S0960-9822(99)80227-9

Morgan, M. L., Deangelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–73. http://doi.org/10.1016/j.neuron.2008.06.024

Mou, W., Fan, Y., McNamara, T. P., & Owen, C. B. (2008). Intrinsic frames of reference and egocentric viewpoints in scene recognition. *Cognition*, *106*, 750–769. http://doi.org/10.1016/j.cognition.2007.04.009

Mou, W., & McNamara, T. P. (2002). Intrinsic Frames of Reference in Spatial Memory. *Cognition*, *28*(1), 162–170. http://doi.org/10.1037//0278-7393.28.1.162

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topography*, *20*(4), 249–64. http://doi.org/10.1007/s10548-008-0054-5

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, *22*(5), 336–343. http://doi.org/10.1016/j.intcom.2010.04.005

Nadasdy, Z. (2009). Information encoding and reconstruction from the phase of action potentials. *Frontiers in Systems Neuroscience*, *3*, 6.

Nadasdy, Z. (2010). Binding by asynchrony: the neuronal phase code. *Frontiers in Neuroscience*, *4*, 51.

Nadasdy, Z., Nguyen, P., Török, Á., Shen, J., Briggs, D., & Buchanan, R. J. (2015). *Spatially periodic activity in the human entorhinal cortex scales with virtual environments.* Manuscript in preparation.

Nadel, L., & Hardt, O. (2011). Update on memory systems and processes. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(1), 251–73. http://doi.org/10.1038/npp.2010.169

Nadel, L., Hoscheidt, S., & Ryan, L. R. (2013). Spatial cognition and the hippocampus: the anterior-posterior axis. *Journal of Cognitive Neuroscience*, *25*(1), 22–8. http://doi.org/10.1162/jocn_a_00313

Nashner, L. M., Shupert, C. L., Horak, F. B., & Black, F. O. (1989). Organization of posture controls: an analysis of sensory and mechanical constraints. *Progress in Brain Research*, *80*, 411–418.

Nelson, A. J. D., Powell, A. L., Holmes, J. D., Vann, S. D., & Aggleton, J. P. (2015). What does spatial alternation tell us about retrosplenial cortex function? *Frontiers in Behavioral Neuroscience*, *9*, 126. http://doi.org/10.3389/fnbeh.2015.00126

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*(7), 741–747.

Nitz, D. (2006). Tracking route progression in the posterior parietal cortex. *Neuron*, *49*(5), 747–56. http://doi.org/10.1016/j.neuron.2006.01.037

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5124915

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. *Hippocampus*. Oxford: Clarendon Press. Retrieved from http://journals.lww.com/jonmd/Abstract/1980/03000/The_Hippocampus_as_a_Cognitive_Map.18.aspx

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, *3*(3), 317–330.

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, *14*(6), 775–82. http://doi.org/10.1038/nn.2815

Ono, T., Nakamura, K., Fukuda, M., & Tamura, R. (1991). Place recognition responses of neurons in monkey hippocampus. *Neuroscience Letters*, *121*(1-2), 194–198. http://doi.org/10.1016/0304-3940(91)90683-K

Packard, M. G., & Goodman, J. (2013). Factors that influence the relative use of multiple memory systems. *Hippocampus*, *23*(11), 1044–1052.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, *65*(1), 65–72. http://doi.org/10.1006/nlme.1996.0007

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, *47*(4), 1747–56. http://doi.org/10.1016/j.neuroimage.2009.04.058

Parron, C., Poucet, B., & Save, E. (2004). Entorhinal cortex lesions impair the use of distal but not proximal landmarks during place navigation in the rat. *Behavioural Brain Research*, *154*(2), 345–352.

Patterson, R. D. (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *327*(1241), 485–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1970894

Pecchia, T., & Vallortigara, G. (2012). Spatial reorientation by geometry with freestanding objects and extended surfaces: a unifying view. *Proceedings. Biological Sciences / The Royal Society*, *279*(1736), 2228–36. http://doi.org/10.1098/rspb.2011.2522

Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., … Hsu, M.-C. (2004). Mining sequential patterns by pattern-growth: The prefixspan approach. *Knowledge and Data Engineering, IEEE Transactions on*, *16*(11), 1424–1440.

Pennartz, C. M. A., Berke, J. D., Graybiel, A. M., Ito, R., Lansink, C. S., van der Meer, M., … Voorn, P. (2009). Corticostriatal Interactions during Learning, Memory Processing, and Decision Making. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(41), 12831–8. http://doi.org/10.1523/JNEUROSCI.3177-09.2009

Persa, G., Török, Á., Galambos, P., Sulykos, I., Kecskés-Kovács, K., Czigler, I., … Csépe, V. (2014). Experimental Framework for Spatial Cognition Research in Immersive Virtual Space. In *Cognitive Infocommunications (CogInfoCom), 2014 IEEE 5th International Conference on* (pp. 587–593). IEEE.

Peters, R. A. (1969). Dynamics of the vestibular system and their relation to motion perception, spatial disorientation, and illusions.

Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 4067–4076). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2556988

Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*.

Proffitt, D. R. (2006). Embodied Perception and the Economy of Action. *Perspectives on Psychological Science*, *1*(2), 110–122. http://doi.org/10.1111/j.1745-6916.2006.00008.x

Proffitt, D. R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review*, *2*(4), 409–428.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, *14*(2), 106–112. http://doi.org/10.1111/1467-9280.t01-1-01427

Quene, Hugo Bergh, H. Van Den. (2008). Memory and Language Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413–425. http://doi.org/10.1016/j.jml.2008.02.002

Quinn, P. C., Doran, M. M., Reiss, J. E., & Hoffman, J. E. (2009). Time course of visual attention in infant categorization of cats versus dogs: Evidence for a head bias as revealed through eye tracking. *Child Development*, *80*(1), 151–161.

Quirk, G. J., Muller, R. U., Kubie, J. L., & Ranck, J. B. (1992). The positional firing properties of medial entorhinal neurons: description and comparison with hippocampal place cells. *The Journal of Neuroscience*, *12*(5), 1945–1963.

Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs. The DO Hebb lecture. *Brain*, *121*(9), 1603–1630.

Ramachandran, V. S., & Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society of London B: Biological Sciences*, *263*(1369), 377–386.

Ratner, M. (2016). Imagining the Unimaginable: Interview with László Nemes on Son of Saul. *FILM QUART*, *69*(3), 58–66.

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D., Cushman, J., … Mehta, M. R.

(2013). Multisensory control of hippocampal spatiotemporal selectivity. *Science (New York, N.Y.)*, *340*(6138), 1342–6. http://doi.org/10.1126/science.1232655

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*(5695), 443–447.

Roland, P. E., & Gulyás, B. (1995). Visual Memory, Visual Imagery, and Visual Recognition of Large Field Patterns by the Human Brain: Functional Anatomy by Positron Emission Tomography. *Cerebral Cortex*, *5*(1), 79–93. http://doi.org/10.1093/cercor/5.1.79

Rolls, E. T., Robertson, R. G., & Georges-François, P. (1997). Spatial view cells in the primate hippocampus. *European Journal of Neuroscience*, *9*(8), 1789–1794.

Rossano, M. J., & Reardon, W. P. (1999). Goal Specificity and the Acquisition of Survey Knowledge. *Environment and Behavior*, *31*(3), 395–412. http://doi.org/10.1177/00139169921972164

Rousselet, G. A. (2012). Does Filtering Preclude Us from Studying ERP Time-Courses? *Frontiers in Psychology*, *3*, 131. http://doi.org/10.3389/fpsyg.2012.00131

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 of the United States of America*, *108*(35), 14654–8. http://doi.org/10.1073/pnas.1105445108

Schacter, D. L., Reiman, E., Uecker, A., Roister, M. R., Yun, L. S., & Cooper, L. A. (1995). Brain regions associated with retrieval of structurally coherent visual information.

Schinazi, V. R., & Epstein, R. A. (2010). Neural correlates of real-world route learning. *NeuroImage*, *53*(2), 725–35. http://doi.org/10.1016/j.neuroimage.2010.06.065

Schnapp, B., & Warren, W. (2007). Wormholes in Virtual Reality: What spatial knowledge is learned for navigation? *Journal of Vision*, *7*(9), 758.

Seeber, B. U., Kerber, S., & Hafter, E. R. (2010). A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hearing Research*, *260*(1-2), 1–10. http://doi.org/10.1016/j.heares.2009.11.004

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–8. http://doi.org/10.1016/j.neuroimage.2011.03.075

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401–409.

Serafin, S., & Serafin, G. (2004). Sound Design to Enhance Presence in Photorealistic Virtual Reality. In *ICAD*.

Servos, P. (2000). Distance estimation in the visual and visuomotor systems. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *130*(1), 35–47. http://doi.org/10.1007/s002210050004

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*(6814), 788. http://doi.org/10.1038/35048669

Sharp, P., Blair, H., Etkin, D., & Tzanetos, D. (1995). Influences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. *J. Neurosci.*, *15*(1), 173–189. Retrieved from http://www.jneurosci.org/content/15/1/173.short

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, *4*, 102–106. Retrieved from http://scholar.google.com/scholar?cluster=2116295900628071159&hl=hu#0

Shelton, A. L., & Pippitt, H. A. (2007). Fixed versus dynamic orientations in environmental learning from ground-level and aerial perspectives. *Psychological Research*, *71*(3), 333–46. http://doi.org/10.1007/s00426-006-0088-9

Shemyakin, F. N. (1962). Orientation in space. *Psychological Science in the USSR*, *1*, 186–255.

Shepard, R. N. (1978). The mental image. *American Psychologist*, *33*(2), 125.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior*, *10*, 9–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1101663

Simon-Thomas, E. R., Brodsky, K., Willing, C., Sinha, R., & Knight, R. T. (2003). Distributed neural activity during object, spatial and integrated processing in humans. *Cognitive Brain Research*, *16*(3), 457–467.

Skalski, P., & Whitbred, R. (2010). Image versus Sound: A Comparison of Formal Feature Effects on Presence and Video Game Enjoyment. *PsychNology Journal*, *8*(1), 67–84.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130–144.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11201094

Small, W. S. (1901). Experimental study of the mental processes of the rat. II. *The American Journal of Psychology*, 206–239.

Snider, J., Plank, M., Lynch, G., Halgren, E., & Poizner, H. (2013). Human cortical θ during free exploration encodes space and predicts subsequent memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(38), 15056–68. http://doi.org/10.1523/JNEUROSCI.0268-13.2013

Snyder, J. P. (1997). *Flattening the Earth: Two Thousand Years of Map Projections*. University Of Chicago Press. http://doi.org/10.1002/9780470979587.ch23

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science (New York, N.Y.)*, *322*(5909), 1865–8. http://doi.org/10.1126/science.1166466

Somekh, J., Choder, M., & Dori, D. (2012). Conceptual Model-Based Systems Biology: Mapping Knowledge and Discovering Gaps in the mRNA Transcription Cycle. *PLoS*

ONE, *7*(12), e51430. Retrieved from http://dx.doi.org/10.1371%2Fjournal.pone.0051430

Spelke, E., Lee, S. A., & Izard, V. (2010). Beyond core knowledge: Natural geometry. *Cognitive Science*, *34*(5), 863–884.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hearing Research*, *258*(1-2), 134–42. http://doi.org/10.1016/j.heares.2009.04.015

Spencer, J., Quinn, P. C., Johnson, M. H., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose: Evidence for young infants categorizing mammals by head and facial attributes. *Early Development and Parenting*, *6*(3-4), 113–126.

Sprague, J. M. (1966). Interaction of Cortex and Superior Colliculus in Mediation of Visually Guided Behavior in the Cat. *Science*, *153*(3743), 1544–1547. http://doi.org/10.1126/science.153.3743.1544

Srikant, R., & Agrawal, R. (1996). *Mining sequential patterns: Generalizations and performance improvements*. Springer.

Steenken, R., Weber, L., Colonius, H., & Diederich, A. (2014). Designing Driver Assistance Systems with Crossmodal Signals: Multisensory Integration Rules for Saccadic Reaction Times Apply. *PLoS ONE*, *9*(5), e92666. Retrieved from http://dx.doi.org/10.1371%2Fjournal.pone.0092666

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses. Book* (Vol. onpp). MIT Press. Retrieved from http://mitpress.mit.edu/catalog/item/default.asp?tid=3184&ttype=2

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, *357*(3), 163–6. http://doi.org/10.1016/j.neulet.2003.12.085

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78. http://doi.org/10.1038/nature11649

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, *10*(4), 422–437.

Sugovic, M., & Witt, J. K. (2013). An older view on distance perception: Older adults perceive walkable extents as farther. *Experimental Brain Research*, *226*(3), 383–391. http://doi.org/10.1007/s00221-013-3447-y

Sulpizio, V., Committeri, G., & Galati, G. (2014). Distributed cognitive maps reflecting real distances between places and views in the human brain. *Frontiers in Human Neuroscience*, *8*.

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. http://doi.org/10.1016/j.bbr.2012.12.031

Sun, H., Campos, J. L., Young, M., Chan, G. S. W., & Ellard, C. G. (2004). The contributions of static visual cues, nonvisual cues, and optic flow in distance

estimation. *Perception*, *33*(1), 49–65. http://doi.org/10.1068/p5145

Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 757–764). ACM.

Sziebig, G., & Øritsland, T. A. (2012). Navigating in 3D Immersive Environments: a VirCA usability study. *Elsevier IFAC Publications/IFAC Proceedings Series*, 380–384.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *The Journal of Neuroscience*, *10*(2), 420–435.

Taylor, H. A., & Tversky, B. (1996). Perspective in Spatial Descriptions. *Journal of Memory and Language*, *35*(3), 371–391. http://doi.org/10.1006/jmla.1996.0021

Team, R. C. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0.

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills*, *36*(3), 1171–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4711968

Tolman, E. (1948). Cognitive maps in rats & in men. *Psychological Review*, *55*(4), 189–208. http://doi.org/10.1037/h0061626

Tolman, E., Ritchie, B. F., & Kalish, D. (1946a). Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, *36*(1), 13.

Tolman, E., Ritchie, B. F., & Kalish, D. (1946b). Studies in spatial learning. II. Place learning versus response learning. *Journal of Experimental Psychology*, *36*(3), 221.

Tootell, R. B. H., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends in Cognitive Sciences*, *2*(5), 174–183.

Török, Á., Csépe, V., Weingrovitz, N., Katsev, G., Heimann, N., & Dori, D. (2015). A téri észlelés konceptuális modellje. In *MPT Nagygyűlés*. Eger.

Török, Á., Kolozsvári, O., Virágh, T., Honbolygó, F., & Csépe, V. (2014). Effect of stimulus intensity on response time distribution in multisensory integration. *Journal on Multimodal User Interfaces*, *8*(2), 209–216. http://doi.org/10.1007/s12193-013-0135-y

Török, Á., Mestre, D., Honbolygó, F., Mallet, P., Pergandi, J.-M. M., & Csépe, V. (2015). It sounds real when you see it. Realistic sound source simulation in multimodal virtual environments. *Journal on Multimodal User Interfaces*, *9*(4), 323–331. http://doi.org/10.1007/s12193-015-0185-4

Török, Á., Nguyen, P., Kolozsvári, O., Buchanan, R. J., & Nadasdy, Z. (2013). Implementation of spatial navigation tasks on tablet PC virtual-reality environment for human studies. In *Workshop on Memory Processes and the Brain*. Budapest.

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*. http://doi.org/10.3389/fnhum.2014.00646

Török, Á., Sulykos, I., Kecskés-Kovács, K., Persa, G., Galambos, P., Kóbor, A., … Honbolygó, F. (2014). Comparison between wireless and wired EEG recordings in a virtual reality lab: Case report. In *Cognitive Infocommunications (CogInfoCom), 2014 5th IEEE Conference on* (pp. 599–603). IEEE.

Török, Á., Tóth, Z., Honbolygó, F., & Csépe, V. (2013). Integration of warning signals and signaled objects to a multimodal object: A pilot study. In *2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 653–658). IEEE. http://doi.org/10.1109/CogInfoCom.2013.6719183

Török, Z. (1993). Social context: Five selected main theoretical issues facing cartography. An ICA report. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *30*(2), 9–11 pp.

Török, Z. (2007). Renaissance Cartography in East-Central Europe, ca. 1450-1650. In W. D.(ed.) (Ed.), *CARTOGRAPHY IN THE EUROPEAN RENAISSANCE*. Chicago: The University of Chicago Press.

Tranter, P. J. (2010). Speed kills: the complex links between transport, lack of time and urban health. *Journal of Urban Health*, *87*(2), 155–166.

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 80.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, *13*(3), 407–433. http://doi.org/10.1016/0010-0285(81)90016-5

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. Frank & I. Campari (Eds.), *Spatial Information Theory A Theoretical Basis for GIS SE - 2* (Vol. 716, pp. 14–24). Springer Berlin Heidelberg. http://doi.org/10.1007/3-540-57207-4_2

van Hoogmoed, A. H., van den Brink, D., & Janzen, G. (2012). Electrophysiological Correlates of Object Location and Object Identity Processing in Spatial Scenes. *PLoS ONE*, *7*(7), e41180. http://doi.org/10.1371/journal.pone.0041180

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*, 1–7. http://doi.org/10.1016/j.neuron.2016.01.045

Villard, E., Garcia-Moreno, F. T., Peter, N., & Clément, G. (2005). Geometric visual illusions in microgravity during parabolic flight. *Neuroreport*, *16*(12), 1395–8. http://doi.org/10.1097/01.wnr.0000174060.34274.3e

Vroomen, J., & Gelder, B. De. (2004). Perceptual Effects of Cross-modal Stimulation : Ventriloquism and the Freezing Phenomenon. *The Handbook of Multisensory Processes*, *3*(1), 1–23. Retrieved from http://books.google.com/books?hl=es&lr=&id=CZS_yDoFV7AC&pgis=1

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(4), 867–82. http://doi.org/10.1037/0278-7393.32.4.867

Waller, D., & Lippa, Y. (2007). Landmarks as beacons and associative cues: their role in route learning. *Memory & Cognition*, *35*(5), 910–924.

Wang, R. F., Hermer, L., & Spelke, E. S. (1999). Mechanisms of reorientation and object localization by children: a comparison with rats. *Behavioral Neuroscience*, *113*(3), 475.

Wang, R. F., & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, *70*(2), 191–210. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10349763

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, *77*(3), 215–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11018510

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in Cognitive Sciences*, *6*(9), 376–382.

Watson, J. B. (1907). Kinæsthetic and organic sensations: Their role in the reactions of the white rat to the maze. *The Psychological Review: Monograph Supplements*, *8*(2), 1–101.

Weidemann, C. T., Mollison, M. V, & Kahana, M. J. (2009). Electrophysiological correlates of high-level perception during spatial navigation. *Psychonomic Bulletin & Review*, *16*(2), 313–319.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, *94*(1), 111–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8354753

Werner, S., Liebetrau, J., & Sporer, T. (2013). Vertical Sound Source Localization Influenced by Visual Stimuli. *Signal Processing Research*, *2*(2). Retrieved from http://www.seipub.org/spr/paperInfo.aspx?ID=2701

Wexler, M., & Van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences*, *9*(9), 431–438.

Wickens, C. D., Liang, C. C., Prevett, T., & Olmos, O. (1996). Electronic maps for terminal area navigation: Effects of frame of reference and dimensionality. *International Journal of Aviation Psychology*, *6*(3), 241–271. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0030310438&partnerID=tZOtx3y1

Wickens, C. D., Liang, C.-C., Prevett, T., & Olmos, O. (1994). Egocentric and Exocentric Displays for Terminal Area Navigation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *38*(1), 16–20. http://doi.org/10.1177/154193129403800105

Wickham, H. (2008). ggplot2: an Implementation of the Grammar of Graphics. R package version 0.7. … *0.7, URL: http://CRAN. R-Project. Org/package= ggplot2*. Retrieved from http://scholar.google.de/scholar?hl=de&q=ggplot2&btnG=&lr=#1\nhttp://scholar.google.de/scholar?hl=de&q=ggplot2&btnG=&lr=#2

Willey, C. R., & Jackson, R. E. (2014). Visual field dependence as a navigational strategy.

Attention, Perception & Psychophysics, *76*(4), 1036–44. http://doi.org/10.3758/s13414-014-0639-x

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science*, *261*(5124), 1055–1058.

Wiltschko, W., & Wiltschko, R. (2005). Magnetic orientation and magnetoreception in birds and other animals. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, *191*(8), 675–93. http://doi.org/10.1007/s00359-005-0627-7

Wirth, M., Horn, H., Koenig, T., Razafimandimby, A., Stein, M., Mueller, T., … Strik, W. (2008). The early context effect reflects activity in the temporo-prefrontal semantic system: evidence from electrical neuroimaging of abstract and concrete word reading. *NeuroImage*, *42*(1), 423–436.

Wozny, D. R., & Shams, L. (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(12), 4607–12. http://doi.org/10.1523/JNEUROSCI.6079-10.2011

Wu, B., Ooi, T. L., & He, Z. J. (2004). Perceiving distance accurately by a directional process of integrating ground information. *Nature*, *428*(6978), 73–77. http://doi.org/10.1038/nature02350

Xin, D., Han, J., Yan, X., & Cheng, H. (2005). Mining compressed frequent-pattern sets. In *Proceedings of the 31st international conference on Very large data bases* (pp. 709–720). VLDB Endowment.

Yartsev, M. M., & Ulanovsky, N. (2013). Representation of three-dimensional space in the hippocampus of flying bats. *Science (New York, N.Y.)*, *340*, 367–72. http://doi.org/10.1126/science.1235338

Yoder, R. M., Peck, J. R., & Taube, J. S. (2015). Visual landmark information gains control of the head direction signal at the lateral mammillary nuclei. *The Journal of Neuroscience*, *35*(4), 1354–1367.

Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nature Reviews Neuroscience*, *17*(3), 173–182. http://doi.org/10.1038/nrn.2015.24

Zhang, H., Copara, M., & Ekstrom, A. D. (2012). Differential Recruitment of Brain Networks following Route and Cartographic Map Learning of Spatial Environments. *PLoS ONE*, *7*(9), e44886. Retrieved from http://dx.doi.org/10.1371%2Fjournal.pone.0044886

Zhang, S., & Manahan-Vaughan, D. (2015). Spatial olfactory learning contributes to place field formation in the hippocampus. *Cerebral Cortex*, *25*(2), 423–432.

Zhou, Z., Cheok, A. D., Yang, X., & Qiu, Y. (2004). An experimental study on the role of 3D sound in augmented reality environment. *Interacting with Computers*, *16*(6), 1043–1068. http://doi.org/10.1016/j.intcom.2004.06.016

Ziemer, C. J., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2009). Estimating distance

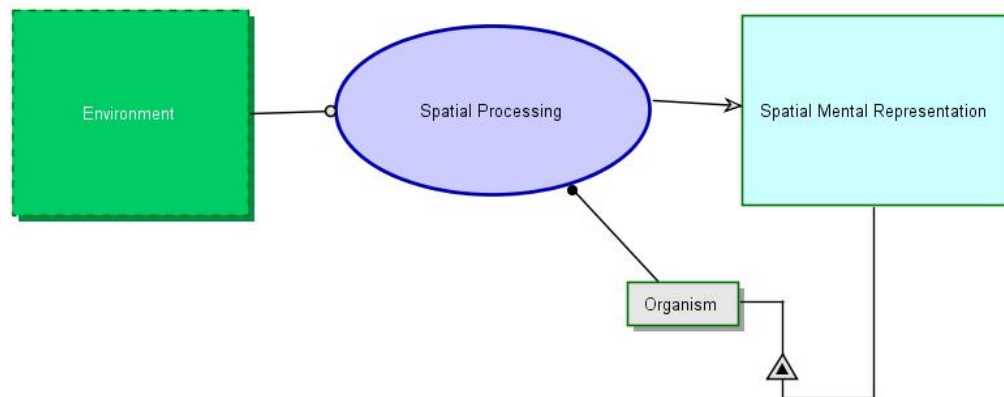in real and virtual environments: Does order make a difference? *Attention, Perception & Psychophysics*, *71*(5), 1095–106. http://doi.org/10.3758/APP.71.5.1096

12 APPENDICES

LIST OF APPENDICES

Environment is environmental and physical.
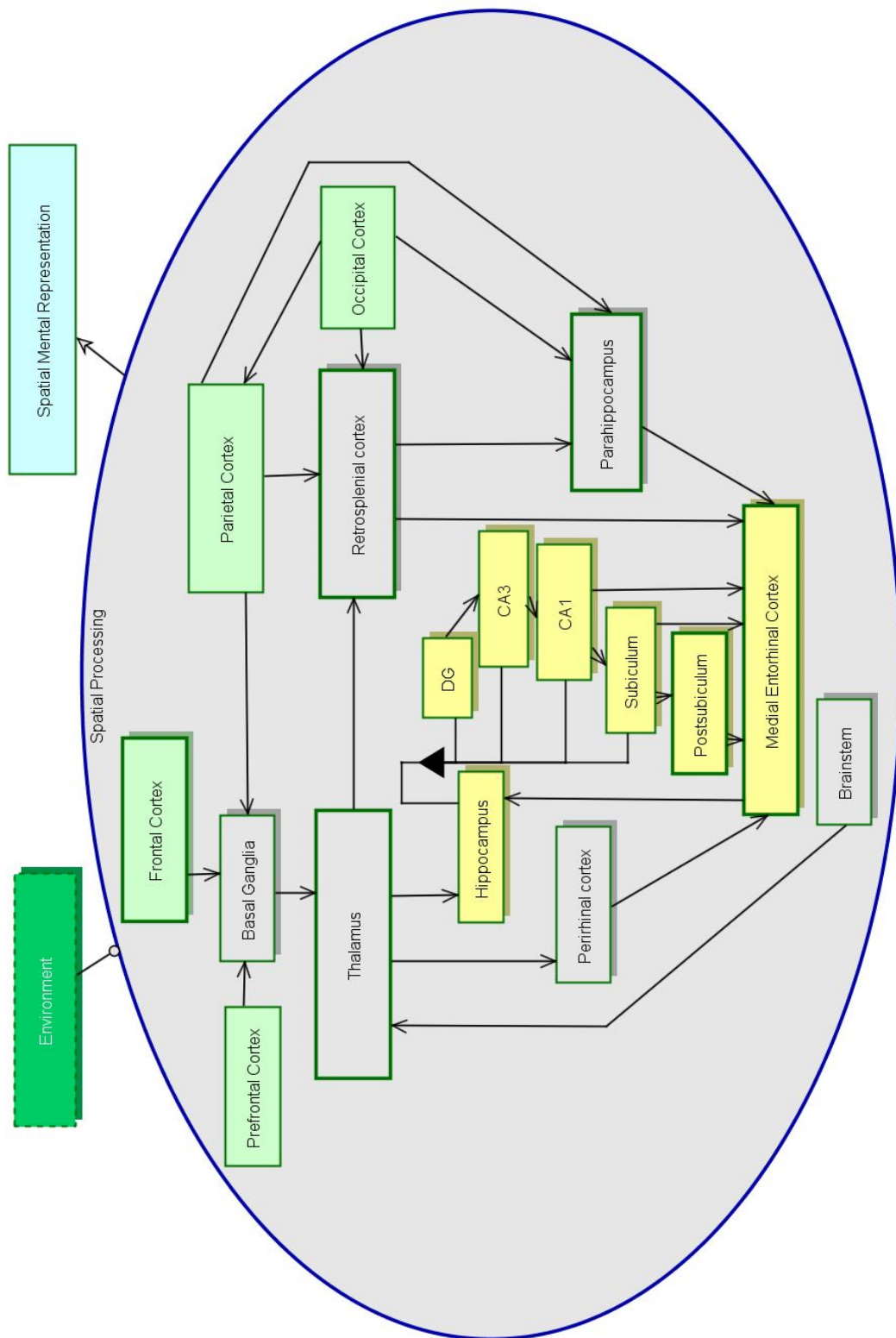Organism is physical.
Organism exhibits Spatial Mental Representation.
Organism handles Spatial Processing.
Spatial Processing requires Environment.
Spatial Processing yields Spatial Mental Representation.

Appendix 2 The first level of the OPM model in diagram (OPD) and in language form (OPL)

Environment is environmental and physical.
Organism is physical.
Spatial Processing requires Environment.
Spatial Processing yields Spatial Mental Representation.

Spatial Processing zooms into Brainstem, Prefrontal Cortex, Frontal Cortex, Parietal Cortex, Thalamus, Subiculum, DG, CA1, CA3, Basal Ganglia, Occipital Cortex, Perirhinal cortex, Postsubiculum, Medial Entorhinal Cortex, Retrosplenial cortex, Parahippocampus, and Hippocampus.

Brainstem is physical.
Brainstem relates to Thalamus.
Prefrontal Cortex relates to Basal Ganglia.
Frontal Cortex is physical.
Frontal Cortex relates to Basal Ganglia.
Parietal Cortex relates to Basal Ganglia.
Parietal Cortex relates to Parahippocampus.
Parietal Cortex relates to Retrosplenial cortex.
Thalamus relates to Hippocampus.
Thalamus relates to Perirhinal cortex.
Thalamus relates to Retrosplenial cortex.
Subiculum is physical.
Subiculum relates to Medial Entorhinal Cortex.
Subiculum relates to Postsubiculum.
DG is physical.
DG relates to CA3.
CA1 is physical.
CA1 relates to Medial Entorhinal Cortex.
CA1 relates to Subiculum.
CA3 is physical.
CA3 relates to CA1.
Basal Ganglia is physical.
Basal Ganglia relates to Thalamus.
Occipital Cortex relates to Parahippocampus.
Occipital Cortex relates to Parietal Cortex.
Occipital Cortex relates to Retrosplenial cortex.
Perirhinal cortex is physical.
Perirhinal cortex relates to Medial Entorhinal Cortex.
Postsubiculum is physical.
Postsubiculum relates to Medial Entorhinal Cortex.
Medial Entorhinal Cortex is physical.
Medial Entorhinal Cortex relates to Hippocampus.
Retrosplenial cortex is physical.
Retrosplenial cortex relates to Medial Entorhinal Cortex.
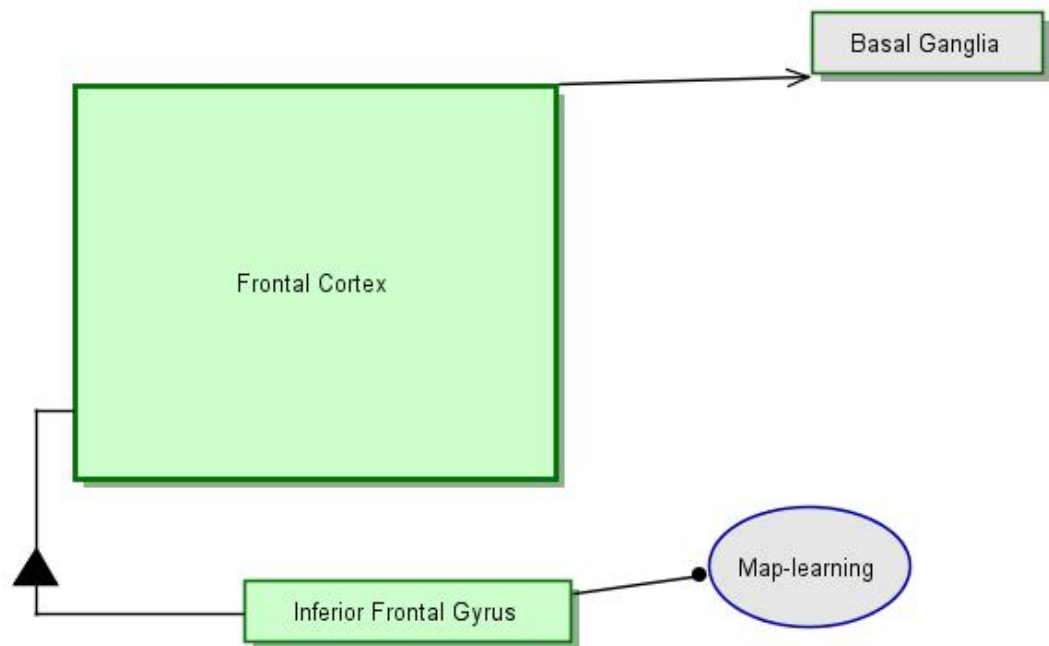Retrosplenial cortex relates to Parahippocampus.
Parahippocampus is physical.
Parahippocampus relates to Medial Entorhinal Cortex.
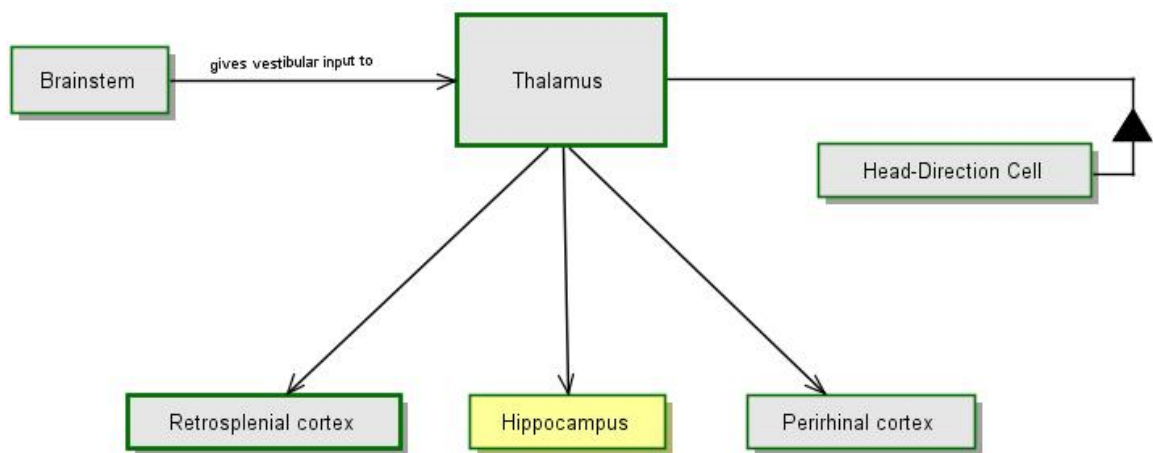Hippocampus is physical.
Hippocampus consists of CA3, CA1, DG, and Subiculum.

Basal Ganglia is physical.
Frontal Cortex is physical.
Frontal Cortex consists of Inferior Frontal Gyrus.
 Inferior Frontal Gyrus is physical.
 Inferior Frontal Gyrus handles Map-learning.
Frontal Cortex relates to Basal Ganglia.

Appendix 4 The Thalamus level of the OPM model in diagram (OPD) and in language form (OPL)

Hippocampus is physical.
Retrosplenial cortex is physical.
Perirhinal cortex is physical.
Brainstem is physical.
Brainstem gives vestibular input to Thalamus.
Thalamus consists of Head-Direction Cell.
 Head-Direction Cell is physical.
Thalamus relates to Hippocampus.
Thalamus relates to Perirhinal cortex.
Thalamus relates to Retrosplenial cortex.
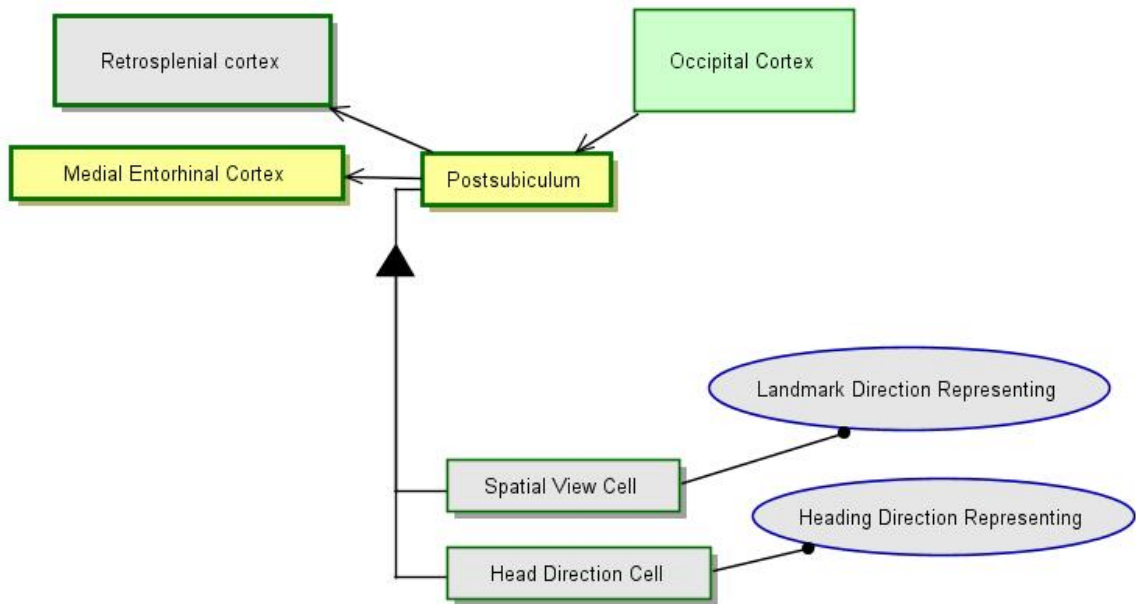
Retrosplenial cortex is physical.
Medial Entorhinal Cortex is physical.
Occipital Cortex relates to Postsubiculum.
Postsubiculum is physical.
Postsubiculum consists of many Spatial View Cells and many Head Direction Cells.
 Spatial View Cell is physical.
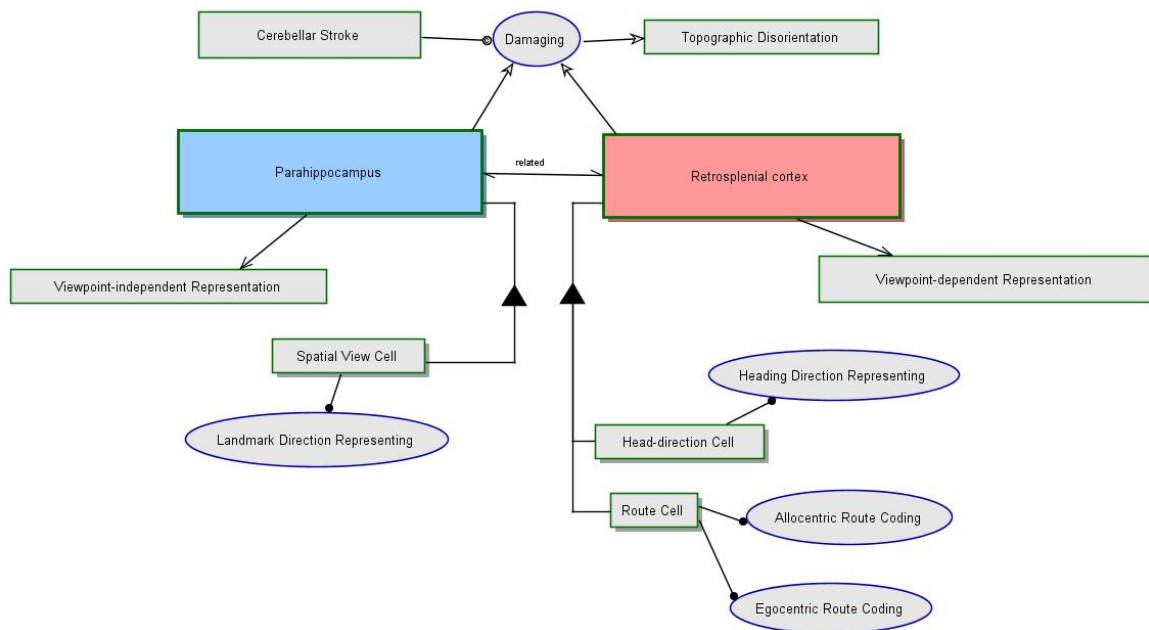 Spatial View Cell handles Landmark Direction Representing.
 Head Direction Cell is physical.
 Head Direction Cell handles Heading Direction Representing.
Postsubiculum relates to Medial Entorhinal Cortex.
Postsubiculum relates to Retrosplenial cortex.

Retrosplenial cortex is physical.
Retrosplenial cortex consists of Head-direction Cell and Route Cell.
 Head-direction Cell is physical.
 Head-direction Cell handles Heading Direction Representing.
 Route Cell is physical.
 Route Cell handles Egocentric Route Coding and Allocentric Route Coding.
Retrosplenial cortex relates to Viewpoint-dependent Representation.
Retrosplenial cortex and Parahippocampus are related.
Parahippocampus is physical.
Parahippocampus consists of many Spatial View Cells.
 Spatial View Cell is physical.
 Spatial View Cell handles Landmark Direction Representing.
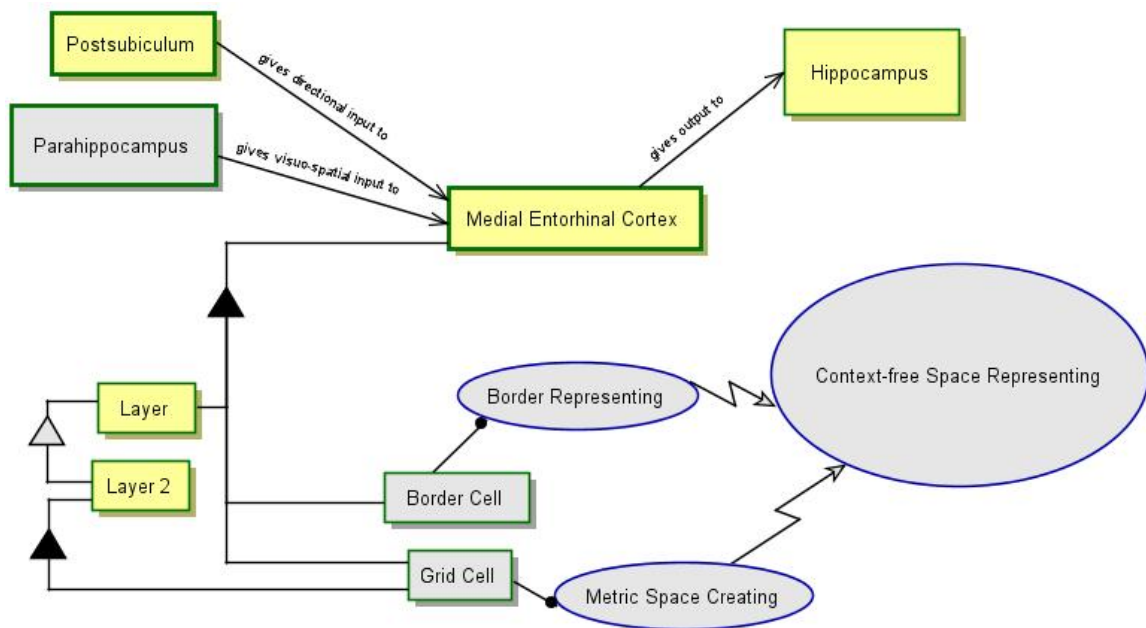Parahippocampus relates to Viewpoint-independent Representation.
Damaging occurs if Cerebellar Stroke is in existent.
Damaging consumes Retrosplenial cortex and Parahippocampus.
Damaging yields Topographic Disorientation.

Hippocampus is physical.
Parahippocampus is physical.
Parahippocampus gives visuo-spatial input to Medial Entorhinal Cortex.
Postsubiculum is physical.
Postsubiculum gives directional input to Medial Entorhinal Cortex.
Medial Entorhinal Cortex is physical.
Medial Entorhinal Cortex consists of 4 Layers, many Grid Cells, and many Border Cells.
 Layer is physical.
 Grid Cell is physical.
 Grid Cell handles Metric Space Creating.
 Border Cell is physical.
 Border Cell handles Border Representing.
Medial Entorhinal Cortex gives output to Hippocampus.
Layer 2 is physical.
Layer 2 is a Layer.
Layer 2 consists of many Grid Cells.
Border Representing invokes Context-free Space Representing.
Metric Space Creating invokes Context-free Space Representing.
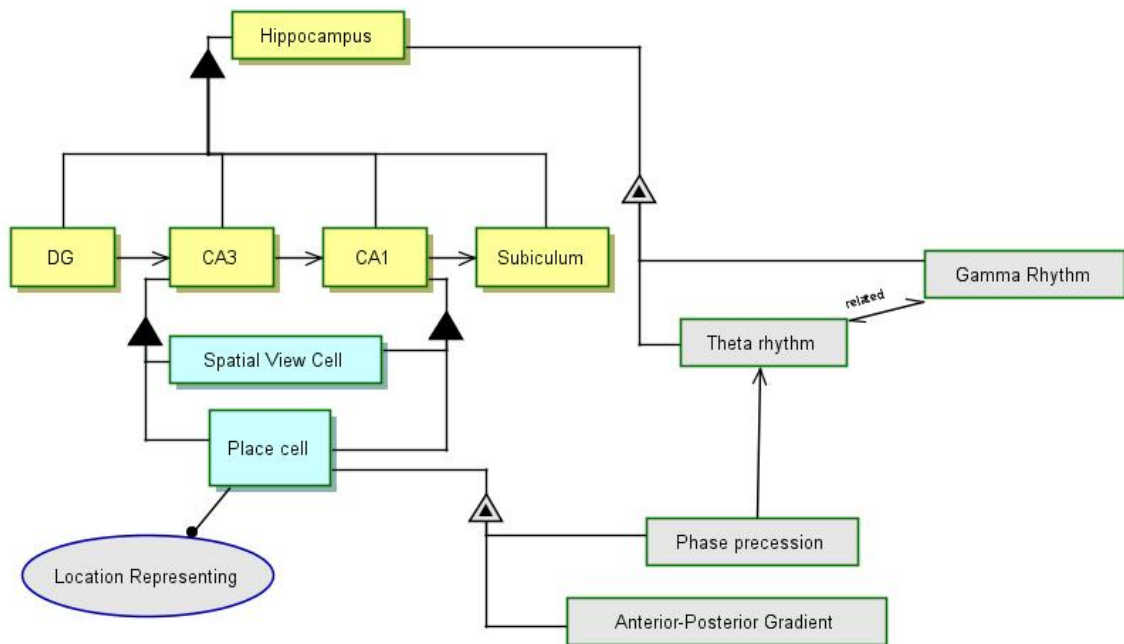
Hippocampus is physical.
Hippocampus exhibits Theta rhythm and Gamma Rhythm.
 Theta rhythm and Gamma Rhythm are related.
Hippocampus consists of CA3, CA1, DG, and Subiculum.
 CA3 is physical.
 CA3 consists of many Place cells and many Spatial View Cells.
 Place cell is physical.
 Place cell exhibits Anterior-Posterior Gradient and Phase precession.
 Phase precession relates to Theta rhythm.
 Place cell handles Location Representing.
 Spatial View Cell is physical.
 CA3 relates to CA1.
 CA1 is physical.
 CA1 consists of many Place cells and many Spatial View Cells.
 CA1 relates to Subiculum.
 DG is physical.
 DG relates to CA3.
 Subiculum is physical.

Appendix 9 Topographic and global field power difference between the two textures irrespective of the location of the alley. We did not apply the duration threshold in the figure to show that there is not even a tendency for difference in processing between the two textures. To note, this does not mean that there is no ERP correlate of texture processing. There is, but in the current paradigm the visual scene was complex, containing at every start almost equal area of each of the four available patterns.
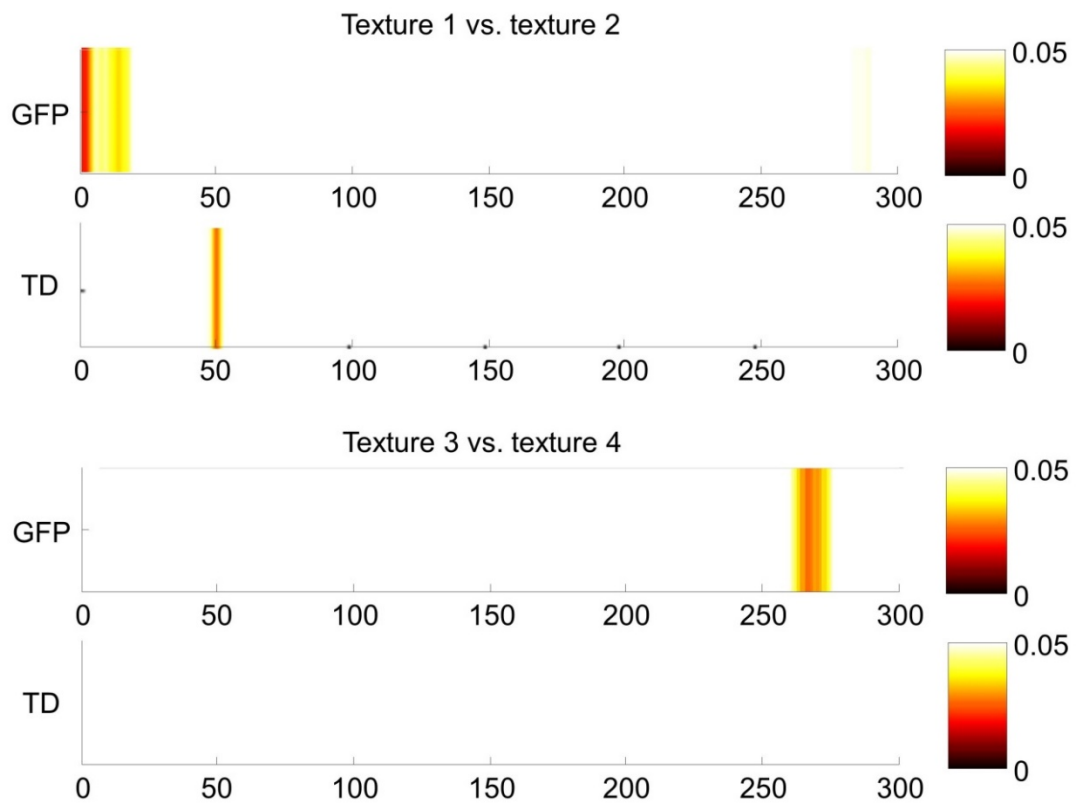
Appendix 10 three way interaction between Inclination, GVS, and Position ($F_{(4, 67.56)} = 39.10$, $p < .001$). We did not find theoretically motivated interpretation behind this interaction, it was primarily due to different slopes estimated for some Inclination and Condition combinations. There is a tendency for opposite effect for farther distances in the sham condition. Alternatively this could indicate that the GVS effect rather extends the radius of the gravity illusion to farther distances.