
The mystery metric of long-term spatial memory

Can Euclidean metric information from a VR triangle completion task be transferred into long-term memory?

Masterarbeit

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

Vorgelegt von

Melissa R. Munzing

Tübingen, März 2021

Erklärung

Hiermit erkläre ich,

- dass ich diese Arbeit selbst verfasst habe.
- dass ich keine anderen als die angegebenen Quellen benutzt und dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.
- dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.

Tübingen, den

Contents

1	Abstract	iv
1.1	German Abstract	iv
1.2	English Abstract	v
2	List of abbreviations	vi
3	Introduction	1
3.1	Basic concepts in spatial cognition	1
3.2	Idiothetic navigation	1
3.2.1	Path integration and triangle completion	1
3.2.2	Neural mechanisms	3
3.3	Spatial long-term memory and mental maps	5
3.4	Iterated triangle completion	7
3.5	Advantages of virtual reality use	8
4	Materials and methods	10
4.1	Participants	10
4.2	Experimental design	10
4.3	Experimental set-up	14
4.4	Procedure	21
4.5	Analysis methods and calculations	24
5	Results	31
5.1	Calibration	31
5.2	Angle error and distance error	32
5.2.1	Verification analyses	32
5.2.2	Over- and undershooting	35
5.2.3	Angle error	39
5.2.4	Distance error	41
5.2.5	Corrected angle and distance errors	43
5.3	Residual deviation and estimated representations	47
5.4	Precision and accuracy	49
5.5	Sketched target locations	52
6	Discussion	55
6.1	Calibration	55
6.2	Verification analyses	56
6.3	Over- and undershooting	57
6.4	Angle error and distance error	57

6.4.1	Angle error across participants	58
6.4.2	Distance error across participants	59
6.4.3	Individual angle error:	61
6.4.4	Individual distance error:	62
6.4.5	Back condition	62
6.5	Residual deviation and estimated representations	64
6.6	Precision and accuracy	65
6.7	Sketched target locations	66
6.8	Summary	68
7	Possible improvements and future outlook	71
8	Attachments	74
8.1	Target orders	74
8.2	Phase 1 target locations	75
8.3	Angle error	79
8.4	Distance error	83
8.5	Residual deviation and estimated representations	87
8.6	Precision and accuracy	92
8.7	Sketched target locations	96
9	References	98

1 Abstract

1.1 German Abstract

Die Fähigkeit die Umwelt zu navigieren ist essenziell für menschliches Überleben. Komplexe räumliche Aufgaben erfordern das Speichern räumlicher Informationen im Arbeitsgedächtnis (WM) und/oder Langzeitgedächtnis. Menschen können sogar ohne Landmarken navigieren, indem idiothetische Informationen der Eigenbewegung integriert und aktualisiert werden. Wird bei der Navigation die eigene Position kontinuierlich und egozentrisch, in Bezug auf den unsichtbaren Ausgangspunkt des aktuellen Laufweges aktualisiert, spricht man von *Wegintegration*. Diese Fähigkeit kann experimentell anhand einer iterierten Dreiecksvervollständigung untersucht werden. Die Dreiecksvervollständigung kommt in der alltäglichen menschlichen Navigation selten vor, doch Untersuchungen der Wegintegration ermöglichen ein besseres Verständnis der menschlicher Raumkognition, denn sie scheint wichtige Basis und Bestandteil der Fähigkeit zu sein, Euklidische metrische Informationen aus der Umgebung zu extrahieren und in das WM zu integrieren. Hierbei soll der Hippocampus als Brücke des Transfers metrischer Informationen zwischen der Selbst-lokalisierung des WM und dem LTM fungieren. Diese Hypothese wird weitgehend akzeptiert, obwohl es an experimentellen Nachweisen fehlt. Das vorliegende Experiment soll diese Lücke in der Literatur mithilfe einer iterierten Dreiecksvervollständigung in virtueller Realität (VR) schließen.

Unter der Annahme, dass dieser metrische Transfer stattfindet, ist weiter postuliert worden, dass der Raum in LTM in einem kartografischen, globalen Koordinatensystem dargestellt wird. Dies soll das räumliche Gerüst für die euklidische *kognitive Karte* darstellen. Diese Theorie, insbesondere der Aspekt der Euklidischen Metrik, wird bis heute hitzig diskutiert. Alternative Theorien sprechen sich für eine topologische, graphische Darstellung im LTM aus, die nicht in ein Euklidisches Koordinatensystem gebettet ist. Z.B. postuliert die etikettiert Graphenhypothese, dass die metrische Information im LTM wie ein Netzwerk aus topologisch verbundenen Regionen aufgebaut ist, die mit lokalen metrischen Informationen etikettiert sind. LTM unterliegt also immernoch einer weitgehend unbekanntem „rätselhaften Metrik“. Diese wurde in der vorliegenden Arbeit untersucht, indem die interne Darstellung der VR Umgebung der Versuchspersonen mit dem Euklidischen („realen“) Layout der Umgebung verglichen wurde.

Die Ergebnisse sprechen dafür, dass räumliche Informationen einer iterierten Dreiecksvervollständigung in das räumliche LTM übertragen werden können. Es wurde gezeigt, dass die räumliche Einbettung im LTM keiner strikten Euklidischen Metrik folgt. Stattdessen scheint die Metrik der kognitiven Karte einem etikettierten Graphen zu gleichen. Dennoch konnte die Frage nach der „wahren“ Metrik der internen räumlichen Darstellung nicht abschließend geklärt werden.

1.2 English Abstract

One of the most essential survival mechanisms in human behavior is the ability to navigate the environment. Complex spatial tasks require spatial information to be stored and updated in spatial working memory (WM) and/or long-term memory (LTM). Humans are even able to navigate without landmarks, by integrating idiothetic cues generated by self-movement. Navigation, by continuously, egocentrically updating ones position in reference to the invisible starting point of the current path, is called *path integration*. This ability can be experimentally investigated via an iterated triangle completion task. While triangle completion itself is uncommon in everyday human navigation, studying path integration is a very important tool for understanding of human spatial cognition, as it has been shown to be an important basis and integral part of the ability to deduct and integrate Euclidean metric information from the environment into WM. The hippocampus is suggested to form the link of metric transfer between the self-location system of spatial WM and LTM. This idea has been widely accepted, although little experimental evidence for this transfer can be found. This experiment was aimed to fill this gap in literature by utilizing a VR triangle completion task.

Assuming, that metric information can be transferred, researchers have suggested, that within LTM, space is represented in a cartographic, global coordinate system. It is thought to provide the spatial scaffold for creating Euclidean-like *cognitive maps*. Since its postulation, this concept, and especially the suggested Euclidean-like metric, have been heatedly discussed. Alternate theories have advocated for a topological or graph-like representation in LTM, not metrically embedded in a Euclidean coordinate system. E.g. the labeled graph hypothesis postulates that the metric in LTM is like a network of topological connections between places, labeled with local metric information. The spatial LTM therefore underlies a still widely unknown "mystery metric". This was explored in the present study, by comparing the subjects' mental representation of the VR environment to its Euclidean ("actual") layout.

The findings support the idea that spatial information from an iterated triangle completion task can be transferred into spatial LTM. The metric embedding in LTM was shown to not be strictly Euclidean. Instead, the metric of the cognitive map in spatial LTM appears to more closely resemble a labeled graph. Nonetheless, the present study is not sufficient to draw finite conclusions and more research is necessary to unravel the "true" metric embedding underlying humans internal representation of space in LTM.

2 List of abbreviations

Acc	Accuracy
AE	Angle error
DD	Distance difference
DE	Distance error
DE	Distance error
LTM	Long-term memory
lsq fit	Least-squares fit
OS	Overshoot
Ph	Phase
Prec	Precision
US	Undershoot
VR	Virtual reality
WM	Working memory

3 Introduction

3.1 Basic concepts in spatial cognition

One of the most essential survival mechanisms in human behavior is spatial cognition and more specifically, the ability to navigate the environment. Spatial cognition is made possible by various sensory modalities which provide feedback and information, as well as numerous cognitive systems and processes (Montello, 2005). These allow for a multitude of different navigational strategies to be available to humans. These strategies vary vastly in their demands on memory and cognition, dependent on the degree to which they are based on the environment, senses or memory and whether the final goal is visible or not (Mallot, 2012). Simple spatial processes that are solely based on sensory input are possible entirely without memory demand, e.g. by balancing optical flow during translation or by simple interaction and reactions to the environment (Warren, 2006). However, more complex spatial tasks require more than just a temporary acquisition of spatial knowledge from the environment. They also call for the storage of this information in spatial working memory and/or long-term memory and the continuous updating of orientation (knowledge of location and heading) during movement (Mallot, 2012; Montello, 2005). Updating of orientation can be done by use of allothetic cues as beacons or the relational properties of landmarks (Whishaw et al., 2001). Most humans are very urbanely socialized, which means constant surroundings of highly salient, distinct landmarks. Additionally, humans are highly capable in object recognition, using this ability to recognize landmarks and their features as navigational aids. Therefore it comes as no surprise that, when given the option, humans prefer the use of landmarks in navigation tasks (Foo et al., 2005).

3.2 Idiothetic navigation

However, when no landmarks are available humans are still able to navigate space (i.e. to find a way to a non-visible target (Mallot, 2012)). This can be done with the help of idiothetic cues generated by self-movement. These self-movement cues are integrated over time to locate a certain position, or to return to a starting location (Whishaw et al., 2001). These self-motion cues are provided by visual (optical flow), afferent commands to the musculature and afferent proprioceptive feedback (Golledge et al., 1999).

3.2.1 Path integration and triangle completion

If the position, which is continuously, egocentrically updated is the starting point of the current path, then this process is called **path integration**. In this process the self-motion cues are combined in spatial working memory over time in relation to the starting point, i.e. the *home vector* between self-location and starting point is continuously updated

within a larger spatial framework (Mallot, 2012; Golledge et al., 1999). This allows for the return to the unmarked starting point, which is referred to as *homing*. Path integration is also utilized by a number of animals, for example by insects, who achieve an impressive homing accuracy (Wehner et al., 2003). In humans the most classic paradigm to test path integration ability is the **triangle completion task** (Figure 1) (e.g. Foo et al., 2005; Klatzky et al., 1998; Riecke et al., 2002). In this approach a participant is led or guided from a starting position to a point A. They are prompted to turn by a certain angle and guided to a second point B. From here the participant is asked to walk or point back to where they believe the original starting location to be. The homing error is defined as the angular deviation between the true home position and the reported home position. In humans the average homing error has been found to be around 20° (e.g. Klatzky et al., 1998).

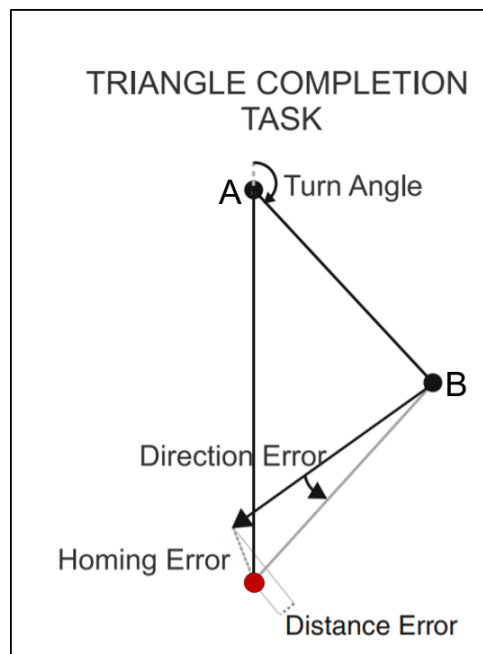


Figure 1: Visualization of a triangle completion task. A participant is guided from a starting point *home* (red dot) to a point A. They are prompted to turn by a certain angle and guided to a second point B. From there the participant is asked to walk or point back to where they believe the home location to be. The homing error is the deviation between the true home position and the reported position (Adapted from Wiener et al., 2011).

Nowadays, most sighted humans rarely encounter situations in which reliance on path integration alone is necessary. But the low average homing error and its independence of previous visual experiences (Loomis et al., 1993) indicate that this navigational strategy has been preserved throughout the course of human evolution. Investigation of path integration tasks has been found to be a "powerful tool" within the investigation of spatial cognition and navigation (Wan et al., 2012, p. 11). During path integration, metric information about the environment is deducted and integrated. This metric information

contains knowledge of Euclidean distances, direction and angles between the navigator and other locations in the environment (Wiener and Mallot, 2006; Warren, 2019). The metric of this information thus is Euclidean, as is indicated by participants low angle and distance errors in homing (Chrastil et al., 2015). Since path integration is a process within spatial working memory and has been shown to follow a Euclidean metric, it can therefore be concluded that human spatial working memory can process and hold Euclidean metric information (Kessels et al., 2001). This process is facilitated by multiple neuron classes and brain regions, which have been found to be associated with Euclidean metric self-location processing.

3.2.2 Neural mechanisms

There is a vast amount of literature concerned with the neural basis of human navigation, self-location and path integration. At this point only a brief overview of the most prominent underlying neuronal mechanisms of self-location and path integration is provided. Four main neuron classes have been found within the hippocampus and associated brain areas, which have been shown to play an important role in the process of self-location (Barry and Burgess, 2014):

Place cells (Figure 2a) were first found in the hippocampus of rodents, but have since been confirmed in humans as well. A place cell will only elicit action potentials when the person is within a certain area of the environment (Barry and Burgess, 2014). Thus it carries information about the current location and is part of encoding the euclidean distances from navigator to the goal. This enables navigation even in the dark and it has been widely suggested that place cell firing is the result of path integration (Redish and Touretzky, 1997).

Head direction cells (Figure 2b) have been found in the thalamus and appear to signal the orientation of the head direction, whereby individual cells respond to a narrow range of head directions (Barry and Burgess, 2014). The firing rates of the cells is directly correlated to the Euclidean angle of the head orientation.

Grid cells (Figure 2c) are most strongly associated with the enthorinal cortex in rats and humans (Doeller et al., 2010). They show a grid-like firing, resulting in lattice-like firing fields. This grid-like firing has been found to be quite stable and increase its scale in discrete steps (Barry and Burgess, 2014).

Boundary vector cells (Figure 2d) located in the subiculum, project to the hippocampus and may shape place cell firing. The firing rates of boundary vector cells are dependent on allocentric direction from an environmental boundary.

Chrastil et al. (2015) and Sherrill et al. (2013) have further described the most prevalent brain regions specifically involved in path integration. They suggest that a homing vector mechanism tracks Euclidean movement and distance from the home location by recruiting

hippocampus, retrosplenial cortex and **parahippocampal cortex**. Of these areas the hippocampus has received the most attention, with multiple studies confirming the importance of the hippocampus in spatial navigation and path integration, where it integrates relative self-motion and tracks Euclidean distance from the starting point/goal (Whishaw et al., 2001; Chrastil and Warren, 2014; Sherrill et al., 2013; Guterstam et al., 2015). The important role of the hippocampus in declarative memory formation has long been shown, as well as the observation that declarative memory is not stored in the hippocampus. It can therefore be seen as a sort of "gate" between working memory and long-term memory (Lisman and Grace, 2005). Similarly Redish and Touretzky (1997) suggested the hippocampus as the link between the self-location system of spatial working memory and spatial declarative memory, both subserved by the hippocampus. This idea is supported by findings of the involvement of the posterior parietal cortex (PPC) in recall of metric information (Guterstam et al., 2015; Brodt et al., 2016). Brodt et al. (2016) suggest that the hippocampus is involved in early spatial learning in novel environments and acts as an encoder for spatial information to be consolidated from spatial working memory to storage in other areas, such as medial temporal lobe or PPC. PPC is then thought to integrate the egocentric input to allow for allocentric recall. In line with this idea, hippocampal lesions cause a decreased ability in navigation of novel environments, as well as the encoding of the environmental spatial layout in LTM (Kessels et al., 2001). This provides a neuronal basis for the idea that egocentric, local metric information from a path integration task could be transferred into LTM (O'keefe and Nadel, 1978).

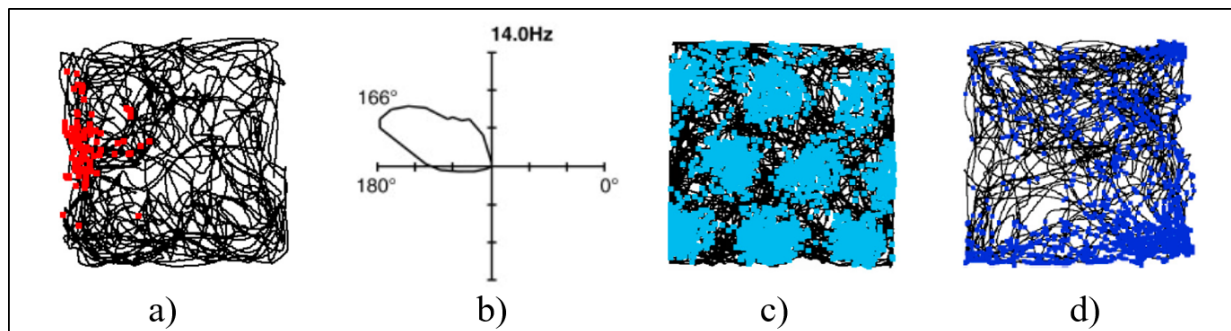


Figure 2: Neural representations of self-location in the hippocampal formation a) Raw data from a place cell (black = the animal's path, red = action potentials of the cells at this location). b) Polar plot of a head direction cell, showing firing rate as a function of head direction. c) Raw data of a grid cell (blue = action potential), firing fields are arranged in a hexagonal lattice. d) Raw data of a boundary vector cell, which fires (blue) whenever there is an environmental boundary in close proximity to the south. (Adapted from Barry and Burgess, 2014, p. R331).

3.3 Spatial long-term memory and mental maps

O'keefe and Nadel (1978) were among the first to suggest such a transfer of metric information from path integration to spatial LTM via the hippocampus. In fact, this notion forms the basis of their hypothesis that the hippocampus provides an a priori and absolute Euclidean spatial framework for navigation (O'keefe and Nadel, 1978; Golledge et al., 1999). The theory contains the idea, that within LTM space is represented with a global metric, i.e. all objects and self in space are regarded as being within a global coordinate system. This cartographic representation is thought to provide the spatial scaffold for creating Euclidean-like *cognitive maps* (Nadel, 2013; O'keefe and Nadel, 1978; Golledge et al., 1999). According to Gallistel (1990) a cognitive map is a mental representation of the environment. He states that geometric relationships between points in the environment are represented on a larger scale by metric position vectors, which are used for planning of movement and navigation. Gallistel (1990) suggests the geometric information of these mental maps to be Euclidean and cartographic. A cognitive map is thought to be held in spatial declarative memory (LTM) and allows for flexible navigation with varying goals and pathways, within the locations included in the map (Mallot, 2012). Cognitive maps may integrate landmark information, but may also be formed and utilized in the absence of landmarks (O'keefe and Nadel, 1978). Since its postulation the concept of cognitive maps has been heatedly discussed. Some researchers disagree with the concept, because simpler mechanisms of navigation may account for what seems to be map-based behavior (Foo et al., 2005; Gibson, 2001). Others have criticized the validity or the overuse of the term and the numerous contradictory definitions that have emerged (Bennett, 1996). The term "mental map" is sometimes used interchangeably or as a form of differentiation. All in all, across literature, it is generally agreed, that humans do build up some form of mental map and that path integration serves as a basis for this map formation (McNaughton et al., 2006; Wang, 2016; Sherrill et al., 2013). However, there has been a lot of contention about the suggested Euclidean-like metric within mental maps and LTM. Alternate theories advocate for a non-Euclidean metric in LTM, with a coarse topological representation (Foo et al., 2005) or a graph-like representation (Warren, 2019; Gillner and Mallot, 1998; Chrastil and Warren, 2014). These hypothesis are aimed to account for the fact, that responses given by participants in spatial navigation tasks have often been found to be geometrically inconsistent and in violation of a Euclidean metric (Moar and Bower, 1983). The labeled graph (or cognitive graph) hypothesis (Figure 3) has emerged as a highly interesting model. It postulates that the metric in LTM is "similar to a labeled graph: a network of topological connections between places, labeled with local metric information" (Chrastil and Warren, 2014, p. 1). A stronger emphasis is put on local metric, which may be Euclidean, modified by Euclidean metric or follow a non-Euclidean metric. This idea of interconnected regions, which are not metrically embedded in a Euclidean coordinate

system has been shown to explain behavioral data, with its spatial inconsistencies, quite well (Warren, 2019; Chrastil and Warren, 2014; Moar and Bower, 1983).

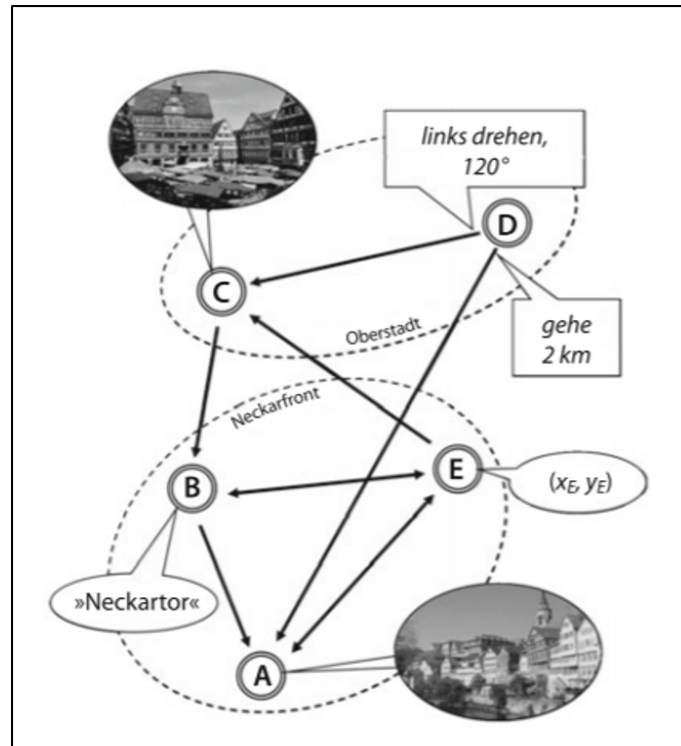


Figure 3: Model of a cognitive map, following the idea of a labeled graph. It consists of places (A-E) and their known junctions/connections. The metric information included in the locations and connections is not metrically consistent. The representation of a place may be characterized by a snapshot (A, C), a coordinate (E) or the name of a location (B). Connections between places can be marked by local metric information (D) or landmarks. Places may be structured according to regions (A,B,E or C,D) (Mallot, 2012, p. 223).

Both notions (Euclidean versus non-Euclidean mental map) have certain aspects and findings speaking for them and can be argued from a theoretical standpoint. Although Experimentally, there has been surprisingly little conclusive evidence to back either idea. Behaviorally, the most common test for metric within spatial LTM is an assessment of participants' ability to form novel shortcuts between previously learned locations (Mallot, 2012). The success of these studies in proving a transfer of Euclidean metric information into LTM is sobering (Foo et al., 2005). Other studies aiming to find evidence of an embedding of mental maps into an absolute Euclidean metric have also lacked in doing so (Mallot, 2012). Therefore the first goal of this study was to analyze whether clear indication could be found, that Euclidean-like metric information from a path integration task could be transferred into spatial LTM. To investigate this question, a different approach was taken than in previous studies. In other modalities it has been shown that human performance in a task improved after information from the task had been learned and transferred into LTM, indicating a beneficial contribution of LTM (Oberauer et al., 2017). Furthermore, it was proposed that "the exchange of information between LTM and

WM appears to be controlled by a gating mechanism that protects the contents of WM from proactive interference but admits LTM information when it is useful" (Oberauer et al., 2017, p. 1). This could also apply to spatial memory, as build-up of a representation of the environment in the LTM (specifically in PPC and MTL (Brodt et al., 2016)) would reduce the need for continuous holding of the home vector in spatial WM. We would therefore expect a backward inference from LTM to path integration after a learning period. In other words, a feedback-cycle is expected to improve the performance of triangle completion, if a global structure is provided from which metric information can be transferred into LTM. If this metric was indeed of stable Euclidean nature between WM and LTM, this should lead to a clear improvement in angle and distance estimations within the task. Lastly, the data was analyzed for aspects such as accuracy, precision and mental representation, to further understand what metric underlies the LTM representation, or mental map.

3.4 Iterated triangle completion

The triangle completion task has been widely used to test for path integration ability. In this experiment the aim was to utilize this task to test for a transfer of metric information into LTM. In order for this to be feasible, multiple criteria has to be fulfilled: Participants need a large number of repetitions, so that additional metric information can be used to improve the home vector through iterations (Hopf, 2008). Multiple targets/locations are necessary to create more than one triangle (> 3); The targets should form a consistent global structure, which can be represented in LTM; In order for the targets to enable the build up of a mental map, each of the targets must be visited from different directions and should be used as home, way point and goal in different rounds (Hopf, 2008) (if the order of the targets was always the same the task would be testing for route-knowledge rather than for a mental map).

It is evident that a simple triangle completion task would not suffice to fulfill these criteria. Therefore the use of an iterated triangle completion task is proposed. In this task participants complete a series of linked triangle completion tasks on a set of five targets (Figure 4). A participant starts at position A and is guided to targets B and C consecutively. At C they are asked to complete a simple triangle completion task, by pointing back at position A. Their current position C then becomes the first step in the next triangle and after being guided to positions D and E they are asked to point back at C. The targets may form a consistent global pattern. Having five targets allows for a variation in targets used to make up a triangle, creating six possible triangle configurations. This method was previously used by Hopf (2008) in a virtual reality (VR) environment to investigate the transfer of metric information into LTM. She found no clear indication of such a transfer. This study aims to adapt and improve the design by offering a modern

VR set-up with higher resolution, a larger visual field, more repetitions and importantly, in a larger experimental space ($A_{Hopf} = 20 \text{ m}^2$, $A_{Munzing} = 92.11 \text{ m}^2$). Additionally, a stage was included to test participants' baseline iterated triangle completion ability in a task without a global structure of targets.

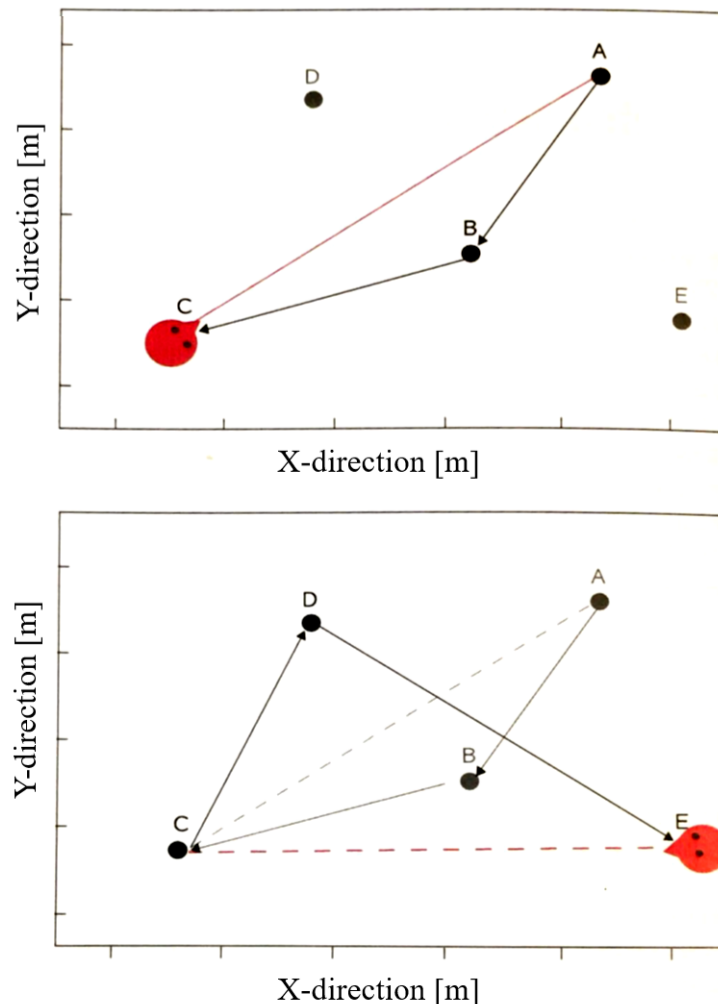


Figure 4: Visualization of an iterated triangle completion task. Top image: A participant (red head) starts at position A and is guided to targets B and then C (black lines). At C they are asked to complete a simple triangle completion task, by pointing (red line) back at position A. Bottom image: The participants position C from the previous triangle (grey lines) becomes the first target in the next triangle (CDE). After walking to positions D and E they are asked to point back at C. (Adapted from Hopf, 2008).

3.5 Advantages of virtual reality use

With the rise of virtual reality the advantages of its use in spatial cognition research has become evident. It has since been used in numerous 3D navigation and orientation paradigms. There seems to be "no significant difference in the accuracy of orientation

judgement in real and virtual environments" (Steck and Mallot, 2000b, p. 3). It has been shown that within a VR environment human visual homing and path integration is possible, even in the absence of landmarks or similar spatial cues (Riecke et al., 2002). In fact, the absence of local landmarks seems to be important in path integration tasks to facilitate mental map formation Foo et al. (2005). Therefore, an advantages of VR use in this study is the possibility of creating an environment without landmarks, which would not be realizable in an experimental room or an urban setting. Another advantage is the possibility of covertly moving target locations. This would be nearly impossible to do in real life. Lastly, the experiment was aimed to investigate global structure, so the availability of a large enough experimental space was of great importance. In VR locomotion can be simulated, which allows for a much larger range of motion and experimental space than is physically available in real world.

The goal of the experiment was to answer the following questions: **1.** Can metric information from an iterated path integration task be transferred into long-term memory? **2.** Does the transferred metric appear to be Euclidean? These are aimed to ultimately aid in answering the following big question. **3.** What is the underlying metric used in human spatial long-term memory?

The main expected outcome of the experiment was a learning curve which lead to lower errors in a task with a set global structure of the targets, than in a similar task without a global target structure. Additional measures were also recorded and analyzed to aid in understanding of the underlying metric.

4 Materials and methods

4.1 Participants

Twelve participants were recruited (6 females, 6 males) with an age range from 21 to 30 years of age. All participants were students of the University of Tübingen and spoke English sufficiently to understand written instructions. They were all right-handed, had naturally good or corrected vision and had no other physical or ocular disabilities.

4.2 Experimental design

The following section is intended to give an overview of the overall experimental design. The exact set-up, procedures and analysis methods will be explained in more detail in the following chapters. To obtain a within-subject comparison, all participants underwent two main experimental phases. Since the two phases were repeated with different participants, a between-subject comparison was also possible. Before entering the two main experimental phases, each participant underwent a calibration and a training, resulting in an experiment constituting four stages (Figure 5). All stages took place within the virtual reality (VR) world, in which participants navigated by means of a VR headset and controller. The VR world offered no local landmarks, however distant global landmarks (skyline cues) were present.

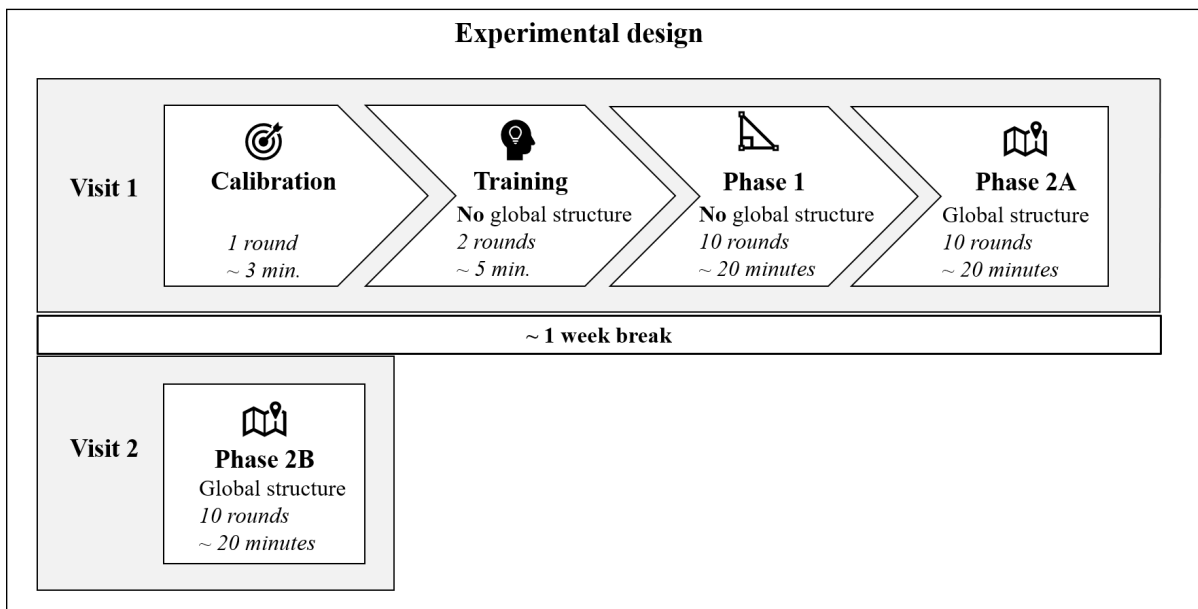


Figure 5: Visualization of the experimental design. During their first visit, participants underwent all four stages of the experiment, including the main experimental Ph1 and Ph2 (2A). Participants returned about a week later for their second visit in which only the Ph2 (2B) was repeated. An estimated average time for completion of each stage is given.

Calibration: During calibration the participants used the controller’s VR laser pointer to point at ten visible targets at varying distances. The distances were matched to the pointing distances later used in Ph2. Any angle or distance deviation from the center of the target was taken as the participants individual baseline pointing error for that specific target distance. This measure was used to offset any bias in varying distance or individual pointing ability in the later measures.

Training: Training, Phase 1 and Phase 2 each contained five targets, which were consecutively visited in a pre-determined order. One round of the experiment was defined as each of the five targets being visited exactly once. During training participants were given the same task, targets and environment as in Ph1, but the training only consisted of two instead of ten rounds. The order in which the targets were visited and pointed to during training was the same for all participants.

Phase 1: (Ph1) the main experimental Ph1 and Ph2 were differentiated by the target locations and the resulting structure. In Ph1 the location of each of the five targets changed with every round. Every round targets assumed a random new location within a radius of $+/- 5$ m from their original location (i.e. the location in round one). Thus the targets did not form a consistent global structure.

Phase 2: (Ph2) In contrast, during Ph2 the target locations were fixed and did not change between rounds, resulting in a consistent global structure. Since only one target was visible at a time, in Ph2 the target locations were additionally marked by distinct images, to aid in recognition of the global structure. After completing all four stages, the first visit of the experiment was completed. At the end of the visit, participants were given a short questionnaire.

Task: All participants were given the same task during training, phase 1 (Ph1) and phase 2 (Ph2), which was to perform a series of triangle completion tasks (iterated triangle completion). Within the VR world they would be presented with a series of wooden pole targets, which they had to walk to, using the controller. When a target was reached it disappeared and the next target became visible, meaning only one target was visible at once. Every three targets the participants were asked to remember and report the location of the target they visited two targets ago (i.e. two targets before their current position, 2-back) (Figure 6). Then they were asked to attempt to report the location of the target they visited before arriving at their current position (1-back). The positions that were presumed to be the locations of the targets 2-back and 1-back were reported by means of the VR controller’s built-in laser pointer. Pointing to both 1-back and 2-back targets was a novel feature, compared to previous experiments. The 1-back pointings only required participants to hold one distance and angle in WM, which does not constitute a triangle completion. The 2-back pointings on the other hand made up the iterated triangle completion task. The back pointing conditions served as within-subject controls. For the 1-back pointings, the angle between the current location and the previously visited target

was always 180°. Thus, no benefit of a set global structure of targets (Ph2) was expected. Beyond a small learning curve within Ph1, due to increased familiarity with the task and environment, no significant changes between the different phases was expected for 1-back pointing performance. The 2-back pointing performance on the other hand were expected to benefit from a set global structure of the targets (Ph2). This back condition is thought to allow for a transfer of metric information into LTM and thus a build up of a mental representation of the target layout. The 2-back pointing performance was expected to improve during Ph2 but not in Ph1. Over the course of the experiment, each participant completed 48 pointings, of which 24 were directed at the target 1-back from their current position, and 24 at the target 2-back. After every pointing, they were given feedback on the position of true target location they were supposed to "hit".

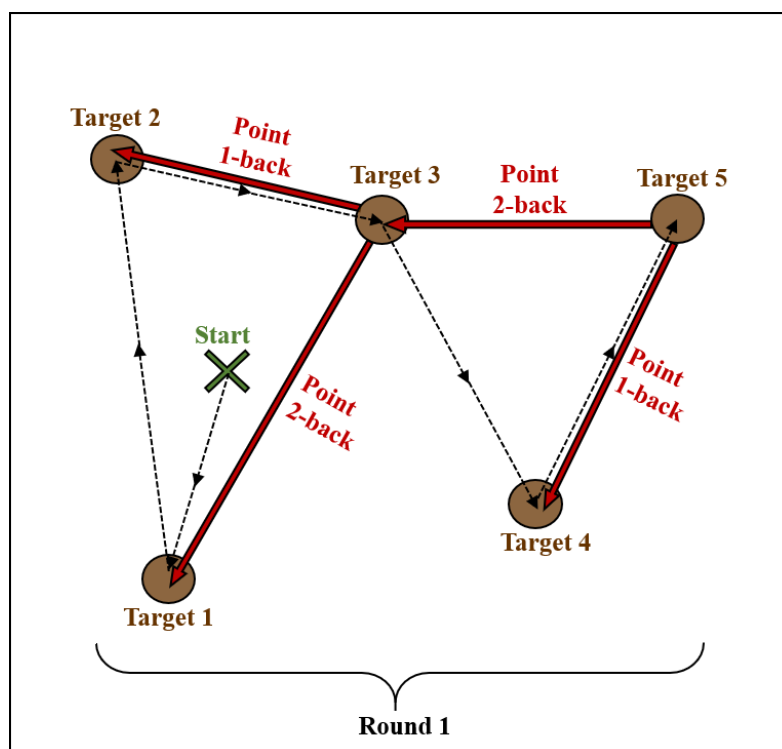


Figure 6: Example course for round one of Ph1 or Ph2. The participant is first immersed in the VR world at the starting point. From this position, the participant was asked to turn until the first target was visible. The VR controller was used to move to the positions of targets 1 to 3. Once target 3 was reached, the participant was asked to use the controller's laser pointer to indicate where they presumed the previously visited position of target 2 and target 1 to be (i.e. the positions, which were located 1-back and 2-back).

Target order: The order in which the five targets were visited and pointed to was semi-randomized and pre-determined. To control for an effect of this order in Ph1 and Ph2, two different target orders were defined (Attachment Table 7). Both orders fulfilled the following criteria: They consisted of ten rounds; in every round each target was visited exactly once; from each target position (numbers one to five) all other targets were pointed

to at least once; ten pointings originated from the positions of target numbers 1 to 4; eight pointings originated from the position of target number 5. The order used in Ph1 was randomly assigned to the participants in an alternating fashion, depending on their participant number (Attachments Table 8). The other target order was then used for Ph2 (i.e. a participant assigned target order 1 for Ph1 was then assigned target order 2 for Ph2).

Second visit: To help assess whether metric information from the VR triangle completion tasks had been transferred into LTM, participants returned to the lab for a second visit approximately one week later (Figure 5). This visit only required a repetition of Ph2 (Ph2B) with the same environment, targets, target positions and order of visited targets as was used during the first visit (Ph2A). After the second visit, participants were given a shortened questionnaire.

The main **independent variables** of the experiment were the phase (Ph1, Ph2A or Ph2B) and the back condition (1-back or 2-back).

The main **dependent variables** were calculated from the collected data: The angle error (AE, in degrees) and distance error (DE, in VR meters) between the actual/true target location and the participant's reported target location (Figure 7).

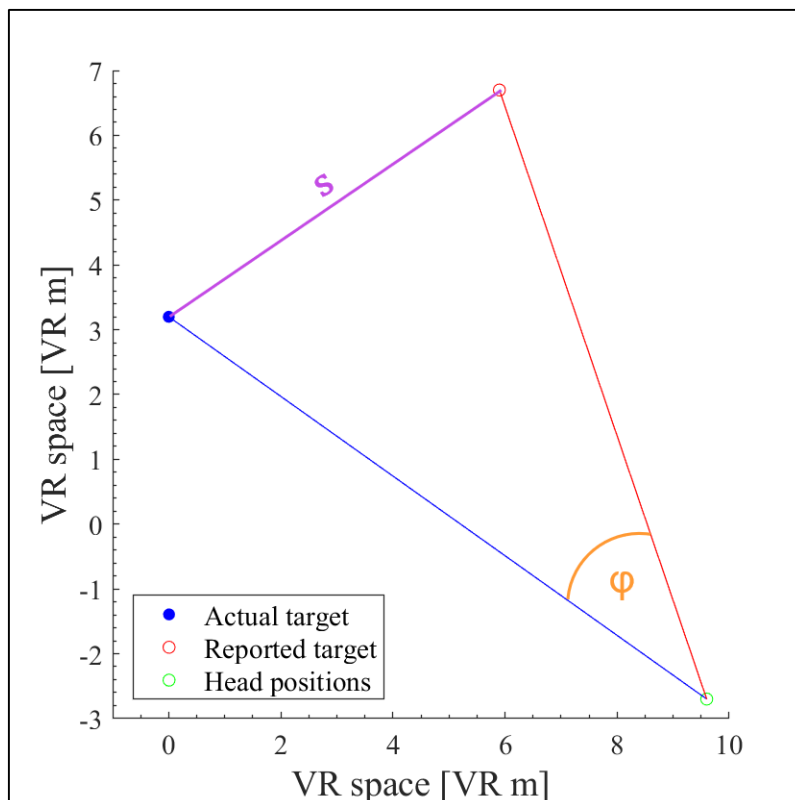
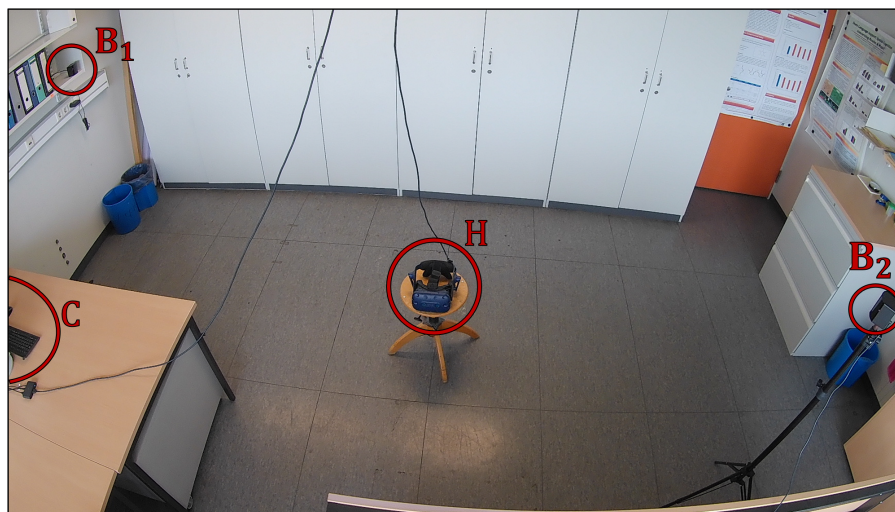


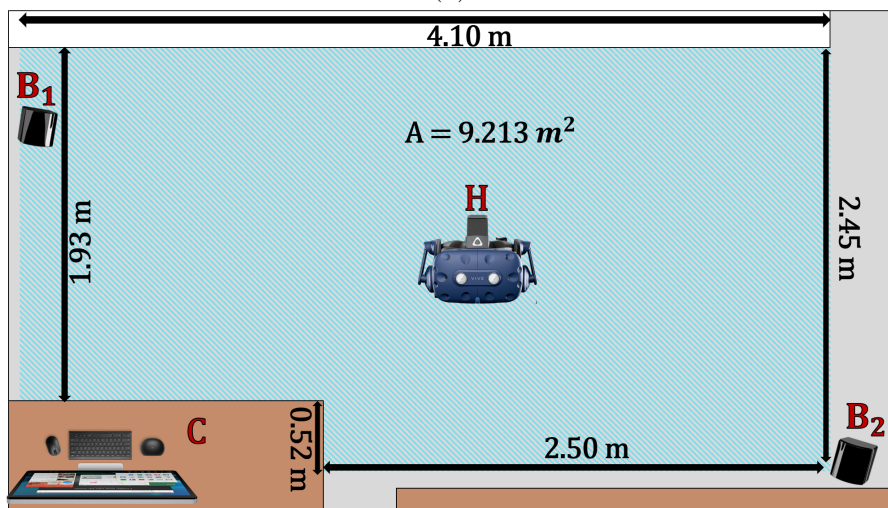
Figure 7: Visualization of the two main dependent experimental variables within the VR space in VR meters. ϕ (orange): angle error in degrees between the actual target location (blue) and the reported target location (red) in reference to the participant's head position (green). s (purple): distance error between the actual and reported target location in VR meters.

4.3 Experimental set-up

Real world set-up: The experiment was conducted in an office room at the university of Tübingen (Figure 8). The size of the experimental space within the room was 9.213 m^2 (Figure 8b). The participant was seated on a swivel chair in the middle of the open space of the room, while wearing a VR head-set and holding a VR controller in their right hand (Figure 9; VIVE Pro Full Kit © 2011-2021 HTC Corporation). They were able to move their head and hands freely and rotate around their own axis by means of the swivel chair, which they were seated on. The headset cable was attached to the ceiling to not limit the range of motion.



(a)



(b)

Figure 8: Photograph (a) and schematic drawing (b) of the experimental room set-up. The participant was seated on a swivel chair at position H (head and headset position). B_1 and B_2 denote the positions of the tracking base stations, and C indicates the fixed position of the computer (and experimenter). (b) The blue shaded area represents the actual size of the tracked space in meters, which was a surface area of 9.213 m^2 .



Figure 9: Photograph of the seating set-up for each participant: Participants were seated on a swivel chair while wearing the VR headset and holding the VR controller in their right hand. Rotation in the VR world was possible via self-rotation of the participant on the swivel chair, as shown. Locomotion in the VR world was possible by pulling the trigger of the VR controller. (Consent for taking and using the images was obtained.)

Virtual reality set-up: The experiment was programmed using the game engine (Real-Time Development Platform) Unity3D Version 2018.2.14f1 (64-bit) (© 2021 Unity Technologies) and the object-based programming language C# (Anders Hejlsberg for © Microsoft), within the integrated development environment Visual Studio version Community 2015 (© 2021 Microsoft). The experiment was run on a LG Tarox Computer (© 2018 TAROK Aktiengesellschaft) with an intel® core™ i7 processor running the digital video game distribution service Steam (© 2021 Valve Corporation).

As previously described, each participant underwent four stages: calibration, training and the two main stages Ph1 and Ph2. These different stages were reflected in the overall VR environment in order to reiterate the switch of task and stage (Figure 10). The calibration task and a brief training period were conducted in a grassy environment (Figure 10a). Ph1 took place in a yellow sand desert and Ph2 in a red sand desert environment (mars-like) (Figure 10b & c). None of the VR environments contained any proximal landmarks, since it's been shown that humans tend to prefer landmark navigation, when available (Foo et al., 2005), which was not of interest in this study. There has also been evidence that local targets may interfere with goal processing in VR path integration (Wan et al., 2012; Foo et al., 2005). The sun was fixed above the participant, and the floor textures were kept minimal, comparable and very repetitive (e.g. the same groove in the sand reappeared over and over), rendering them impossible to be used as landmarks. A global frame of reference was given in the form of global landmarks (i.e. skyline cues; Figure 10). The global landmarks were each split into four distinct areas, offering compass-like information. It has been shown that global landmarks can facilitate navigation in simulated (Basten and Mallot, 2010) and VR environments (Steck and Mallot, 2000b).

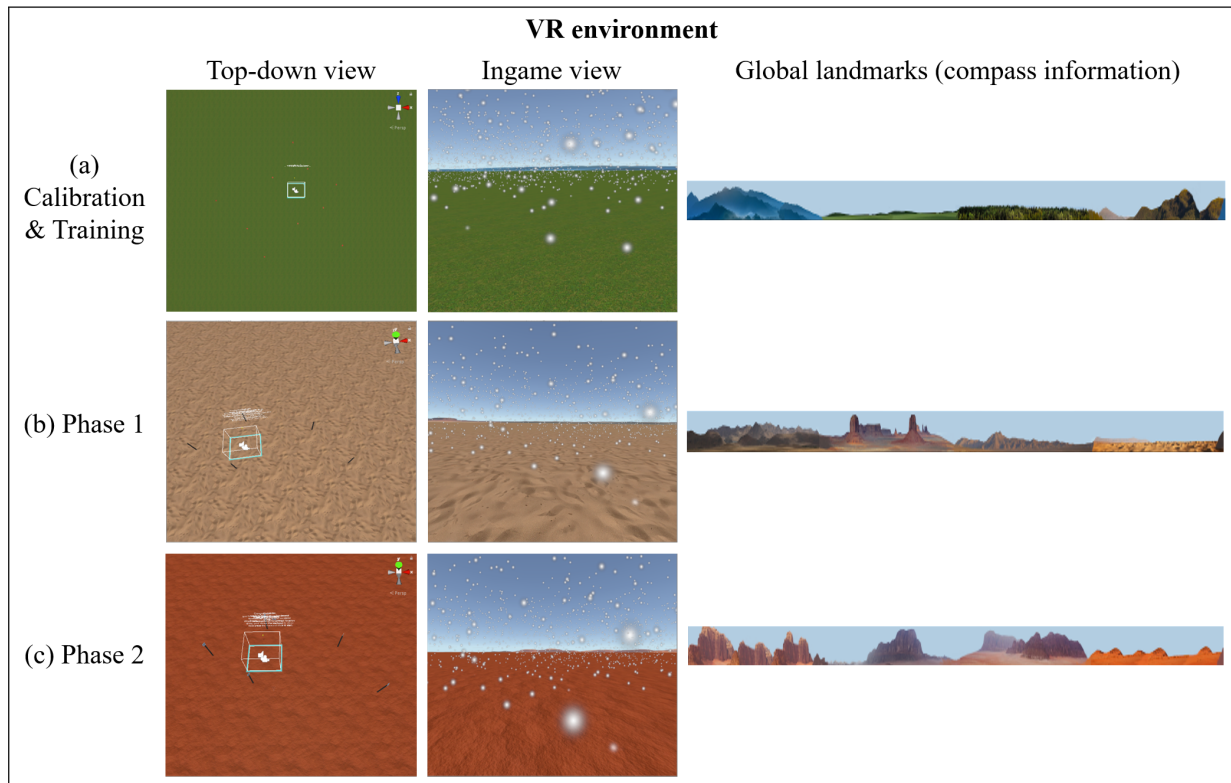


Figure 10: Overview of the three different VR environments used for the four stages of the experiment. In order to allow for navigation without proximal landmarks, very distant global landmarks were used, which could be used as compass information.

This is done by offering a "world-centred framework within which spatial information can be organized" (Ruddle and P eruch, 2004, p.303). The landmarks were presented as an image overlaid onto a large cylinder around the experimental space. The diameter of the cylinder was much larger than the experimental space, making it appear like hills or mountains far off in the distance. The colors were matched for smooth transitions from ground to landmarks to sky. The global landmarks could be used for basic compass navigation (distinguishing heading direction), but could not be used to distinguishing individual points/targets within the experimental area, which means they could not be used as local landmarks. In line with Klatzky et al. (1998), who found that the vestibular feedback of physical rotation was highly relevant in path integration tasks, rotation in the real and VR world was possible via head and body rotation of the participants on the swivel chair. The participant was physically permanently seated in one position, but translation was possible by pulling the trigger of the VR controller (Figure 11). While the trigger was pulled, a forward translation with a speed of 1.5m/s was possible. This walking speed is within the normal walking speed range of a younger pedestrian (Montufar et al., 2007). To ensure a smooth and more realistic "walking" experience, despite the lack of physical walking, a minimal acceleration period was automatically included upon trigger pulling. Proprioceptive and vestibular translation cues were unavailable to the participants, but it has been found that in a VR environment "path integration by optic

flow was sufficient for homing by triangle completion" (Riecke et al., 2002, p. 69). To reinforce optical flow for forward translation and rotation perception, the VR environment contained small white particles, comparable to snowflakes fixed in the air (Figure 10). The second relevant button on the controller was the trackpad (Figure 11), which fired a shot from the controller's laser pointer. A green laser beam appeared out of the tip of the ingame visible controller whenever the participant was required to perform a pointing task. A yellow ball appeared at the contact/puncture point of the laser with the ground plane (Figure 12b). Participants were asked to use this puncture point to report where they believed the targets 1-back and 2-back to have been. The laser and trackpad were only activated during the pointing events.

As visual and auditory feedback of a fired shot, the laser color changed from green to red and a laser shooting sound was played. To ensure that no outside noises or other auditory cues within the real world were used as navigational aids, the participants wore the headphones included in the headset. Both the audio feedback from the laser shots as well as a background noise audio file was played during each stage. The background audio was chosen to fit the respective VR surrounding to add to the immersion, as well as occlude any outside noises. During calibration and training, a background audio of wind whistling in the trees and birds chirping was chosen to match the grassy green environment (Relax Sleep, 2018). For Ph1 and Ph2 a background audio file of wind whistling through desert sand was used to match the desert environments (Pure Relaxing, 2019). Both audio files contained consistent noise and sound levels, were long enough to span the entire stage without interruption and were originally designed as relaxing background noises.

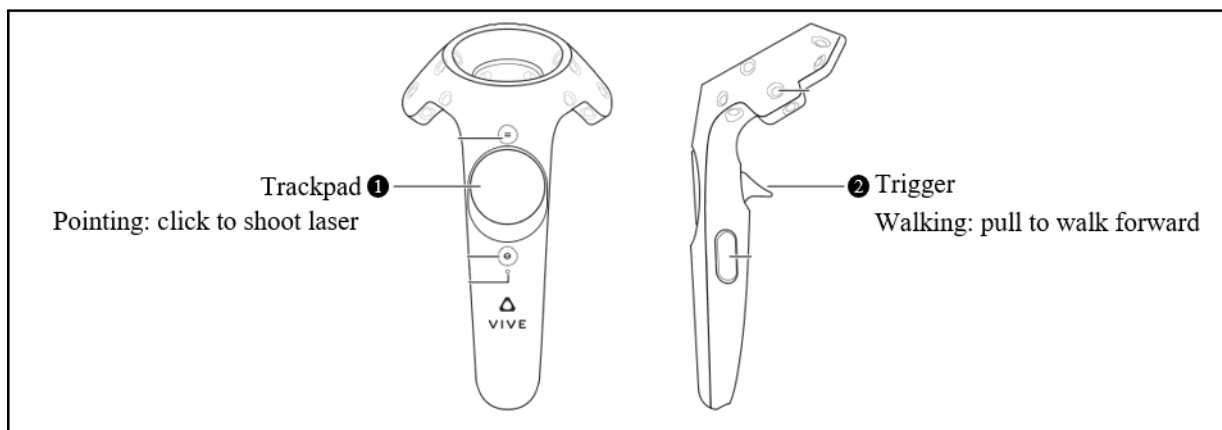


Figure 11: Outline of the HTC VIVE pro VR controller. The numbers indicate the two buttons that were relevant for the experiment. The trackpad (1) was only active during the pointing events, when the laser was visible. Clicking the trackpad resulted in a shooting of the laser (color change and laser sound were given as visual and auditory feedback of a fired shot). Pulling of the trigger (2) led to forward translation with a speed of 1.5 m/s within the VR space. A minimal acceleration time was automatically included when the trigger was pulled to ensure a comfortable, smooth translation).

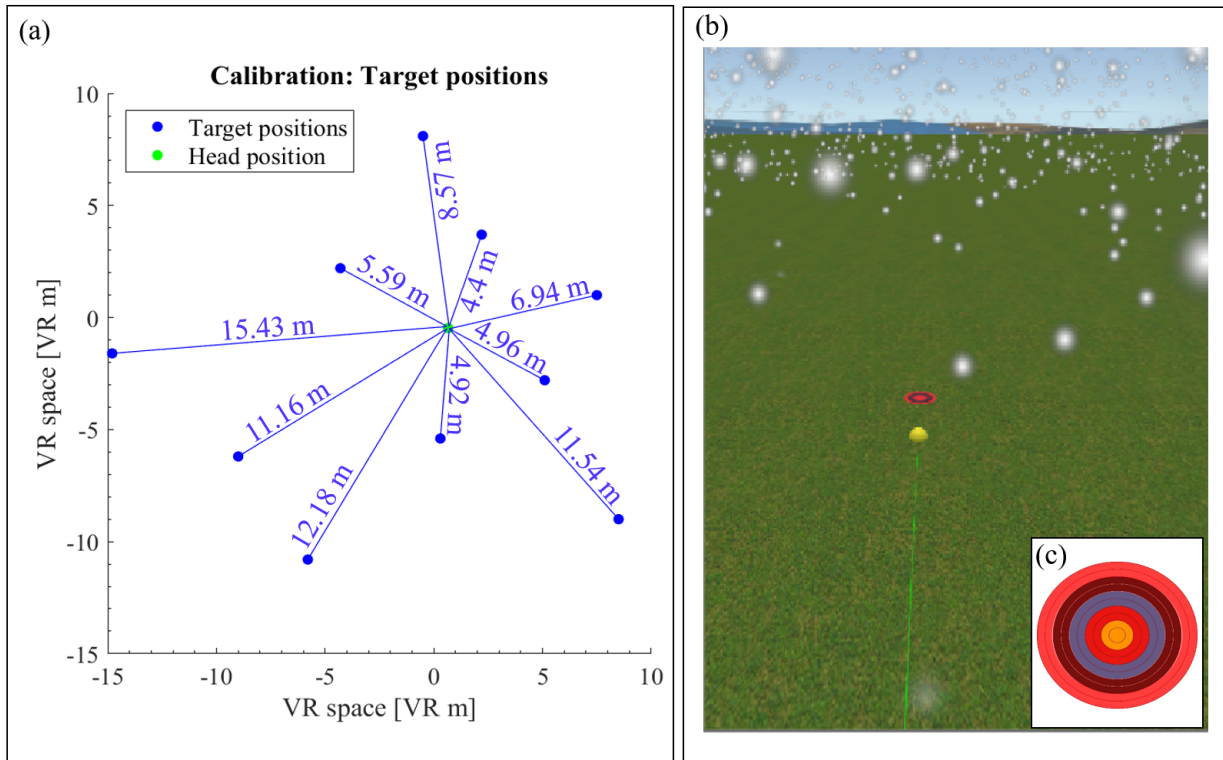


Figure 12: (a) Visualization of the experimental space, the participants head position (green dot) and the target positions (blue dots) in the calibration phase. The distances of the calibration targets were matched to the distances between the targets (i.e. the pointing distances) in Ph2. (b) Ingame frame of the calibration stage. The targets were placed on the ground plane at differing distances to the participant’s fixed position. Participants were instructed to point at the targets with the laser as accurately as possible. (c) Close up of a calibration target which was colored red for better visibility.

The four stages of the experiment shared many similarities in the overall VR set-up, as described above. However, there were some important differences in the VR set-up between the stages: During the calibration stage, no translation was necessary. Red targets were placed circularly on the ground around the participant in varying distances (4.40 m - 15.43 m from head to target position) (Figure 12a). The distances between the participant’s head position and the calibration targets were matched to resemble the pointing distances (i.e. the distances between the targets) used in Ph2. The targets were colored red for better visibility and placed on the ground plane (Figure 12b & c). Although the training and Ph1 took place in different VR environments, the task and targets remained the same between the two. The targets were five wooden poles which were each 2 m high (Figure 13a). Training and Ph1 called for translation via trigger pulling. This allowed for the experimental space within the VR world to be much larger than the real-world space. In these two stages the positions of the five targets changed for every round (one round \equiv all five target positions visited) by a random jitter of between ± 5 meters for each target (note: the configuration of the five targets was the same

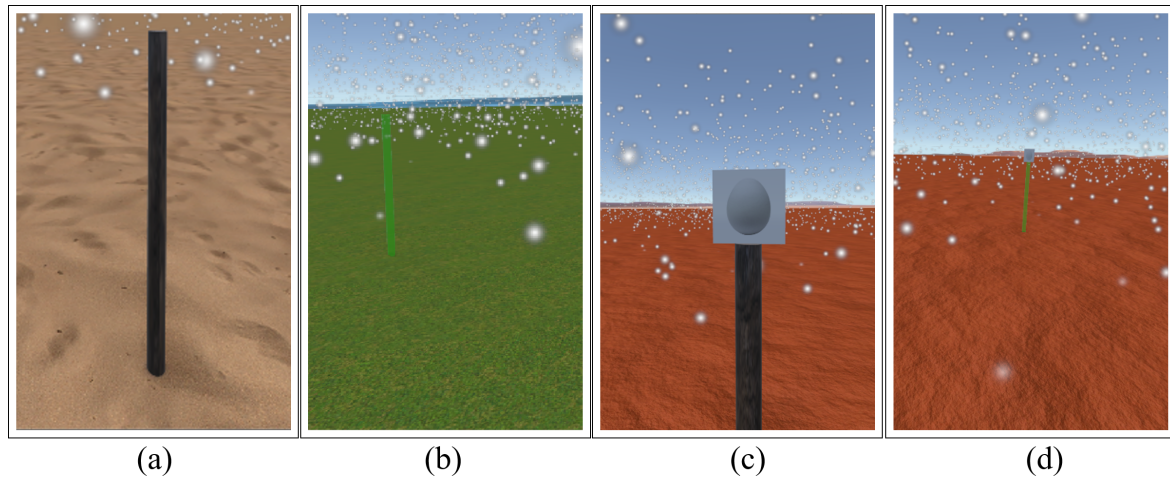


Figure 13: **Training and Ph1:** (a) Ingame view of one of the five 2 m high target poles used during training and experimental Ph1. (b) Green glass feedback pole, used to give the participants feedback about the actual location of a previously visited target after their pointing response. **Ph2:** (c) Example of target pole number 1, which was topped by a billboard image of an egg. (d) Green glass feedback pole and billboard.



Figure 14: The image billboards used atop the target poles in phase 2.

for all participants in round one, which was the same configuration later used in Ph2). Due to the jitter, the exact size of the experimental space during training and Ph1 varied slightly for each round and participant (Attachments, Figures 34, 35, 36 and 37). After each pointing, feedback was given on the true location of the previously visited target through the appearance of a green glass target pole with the same dimensions in the respective target location (Figure 13b). In the experimental Ph2, the same 2 m tall wood posts were used as targets, with two important distinctions: 1. The five targets were

fixed in a set position (Figure 15 and 16) and 2. Each pole was topped with a billboard of a distinct image (Figure 13c and 14). Translation via controller was necessary within the experimental VR space of 92.11 m^2 , with an area of 52.63 m^2 enclosed by the targets (Figure 15). The global structure of the targets was selected to contain a variety of distances, as well as a range from narrow to wide angles. The selected billboard images depicted a single object: an egg, a car, a key, a shoe or beer (Figure 14). All object images had the same art style (photograph, no background) and were presented in grey scale at the same contrast level for comparable salience. The images were presented on a white square billboard, fixed to the top of the targets (Figure 13c). The billboards constantly faced the participant regardless of approach direction (billboard rotation linked to head rotation) to ensure that each image was always visible, and that billboard orientation could not be used as a navigational aid. The words behind the images were selected using the MRC Psycholinguistic Database (Wilson, 1987; Fearnley, 1997) and were matched for: number of syllables, number of letters, number of phonemes, brown verbal frequency (Brown, 1984), familiarity rating, imagability rating and meaningfulness. In English these words contain no capitalization, are all nouns (same comprehensive syntactic category) and are characterized as standard words (contextual status).

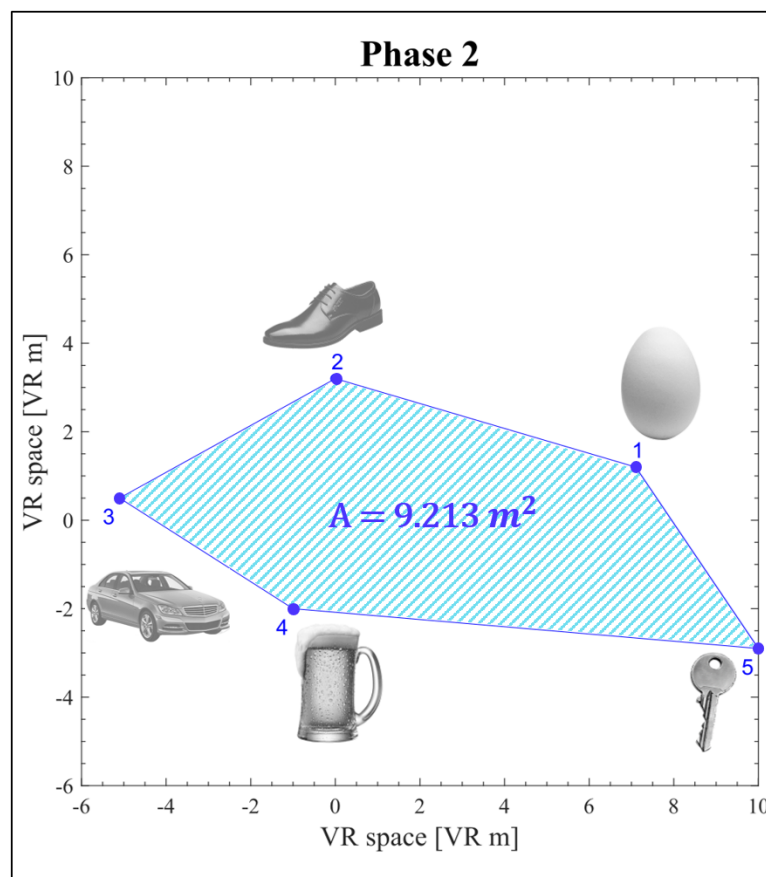


Figure 15: A top-down representation of the five target positions and the experimental VR space. Each target pole was topped by an image billboard of an object, e.g. target 1 is topped by the image of an egg.

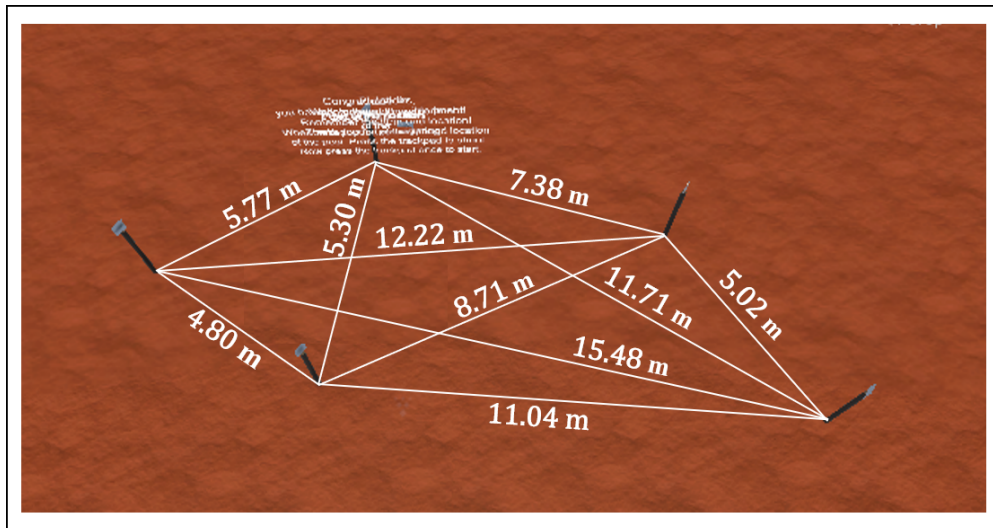


Figure 16: Top-down view of the distances between the targets in Ph2.

4.4 Procedure

In order to comply with the Coronavirus pandemic regulations set forth by the state of Baden-Württemberg and the University, all surface areas and objects the participants came in contact with were disinfected after every visit. Additionally, only one participant was permitted in the experimental room, and the experimenter continuously maintained a distance of at least 1.5 meters to the participants. In line with regulations, personal contact information (phone number or e-mail address) was retained for each participant. This information was destroyed two weeks after the participant's final visit.

Prior to participation, each participant was given standardized written instructions in English. All questions were answered and additional verbal instructions and answers to questions were given in German or English, depending on the participant's preference. An informed consent form was signed. Each participant was required to show up for the experiment on two separate days, which were to be roughly a week apart. All participants adhered to this: three participants came in for the second visit after six days, eight participants after seven days and one participant after eight days. Participants were seated on the swivel chair and adjusted the headset to their satisfaction. The controller was to be held in the right hand and kept straight in front of oneself when turning and pointing.

The **calibration** marked the first stage of the experiment. The VR environment was entered and the background audio of wind in the trees and birds chirping was played over the headset headphones. A brief task description was displayed which could be deactivated by clicking the trackpad once. The green laser and yellow ball, indicating the puncture point, were automatically activated and remained visible for the entire calibration. In order to fulfill the task of hitting all ten of the red targets surrounding the participant,

self-rotation of 360° was necessary. A target could be shot at by clicking the trackpad when the yellow ball had been aligned in accordance with the target. The participants were all given one shot per target to obtain calibration and were aware they should be as accurate as possible. The participants shot the laser at their discretion by clicking the trackpad, when they felt the yellow ball was aligned with and centered on a target. When a target (or its vicinity) was hit, it disappeared. Once all ten targets had disappeared, the laser automatically was shut off and a text appeared informing, that the task had been completed and the next stage was about to begin. participants were requested to close their eyes before transitioning to the next stage.

Training, phase 1 and phase 2 (A & B) each contained five wooden target posts, which were visited and pointed at by the participants in a set order. Differences between training and the Ph1 and Ph2: the training consisted of two rounds while Ph 1 and Ph2 each consisted of 10 rounds (one round \equiv each of the five targets is visited exactly once). For the training, the same grassy environment and background audio were used as during calibration. Ph1 took place in a yellow sand desert and Ph2 in a red sand desert. Ph1 and Ph2 shared the background audio of wind whistling through sand. Differences between Ph1 and Ph2: The target positions in Ph1 were not consistent across rounds. In every round the target positions varied by a jitter with a radius of ± 5 VR meters from their positions in round 1. Figure 17 shows the Ph1 target positions of participant 1 as an example (Note: the target positions in Ph1, round 1 were the same as in Ph2). The target positions in Ph1 for all participants can be found in the attachments (Figures 34, 35, 36, 37).

Participant 1

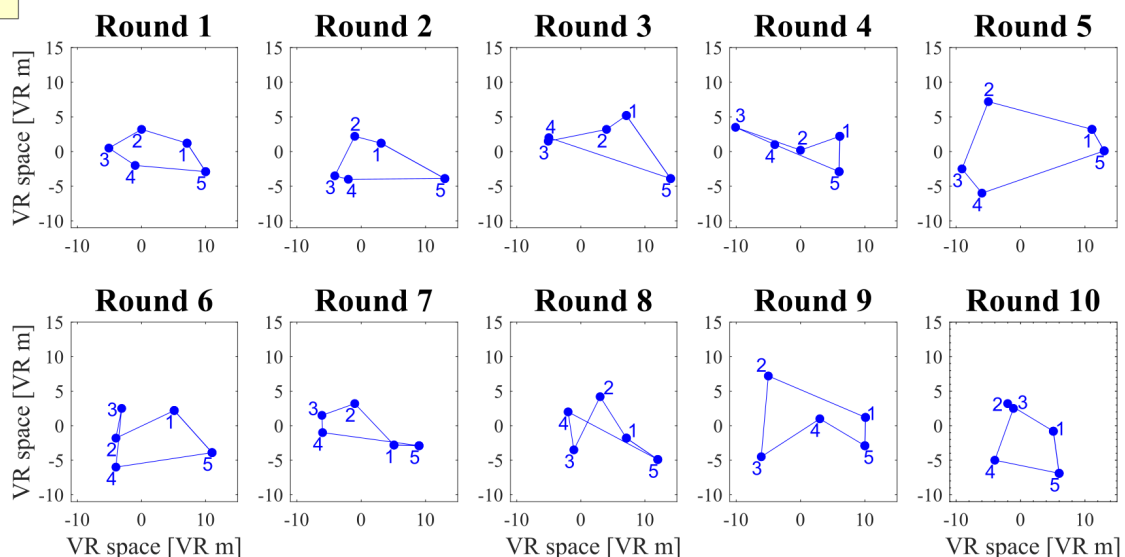


Figure 17: Varying target positions in phase 1 for each round. Example data from participant 1.

In contrast, in the Ph2A and Ph2B, the target positions were set and the same target

number was always associated with the same location. In Ph2B the target posts were additionally topped with one of the following images: an egg, a car, a key, a shoe or beer. These images were strictly associated with one target number and the corresponding target location (Figure 15). Aside from these differences, training, Ph1 and Ph2A and Ph2B all shared the following same basic procedure. After entering each stage, an instruction text appeared, which could be turned off by clicking the trackpad once. This activated the first wooden pole target. Then the trigger of the controller had to be pulled in order to translate to the location of the first visible target. It was necessary to walk up to the target close enough to "collide" with it. Then the first target disappeared, and the next target became visible. Again location, translation and collision with the second target were necessary before the third target appeared. When the third target had been reached a text appeared for two seconds within the participant's line of vision asking them to indicate where the target "2-back" was located. For training and Ph1 this text read "point 2 posts back", for Ph2 the text read "point at the position of the" next to a small picture of the billboard image associated with the 2-back target. When the text had disappeared, the green laser was automatically activated. The yellow ball was used to locate the position, which was assumed to be the location of the first target (i.e. the 2-back target). There was no time limit for a response, but only one shot/answer could be given by pressing the trackpad. A color change of the laser during button pressing and a laser shot sound gave feedback of a carried out shot. The laser disappeared again after one second and a green glass pole appeared at the actual (true) target location to give the participants direct feedback of their response accuracy. The feedback pole was visible for 3.5 seconds before disappearing. Immediately another text appeared in their line of sight for two seconds. This time the text read "point 1 post back" for Ph1, and for Ph2 "point at the position of the" next to the image that had topped off the 1-back target. The pointing and feedback occurred in the same manner as before. After both pointings were completed, the next (fourth) target materialized. The participants proceeded to targets number four and five as described. At the fifth target, the next pointing event took place in the same manner as the previously detailed one, again asking the participants to point 2-back (third target) and 1-back (fourth target). This process was repeated until all 48 pointings had taken place and all ten rounds were completed (for training: 8 pointings and two rounds). After all stages were completed in the first visit, participants were given a questionnaire which contained questions on their strategy, asking whether they noticed differences between the phases and asking them to draw a sketch of what they believed the target locations looked like from a bird's eye view for Ph1 and Ph2 separately. In the second visit, Ph2 was repeated as previously described. After the second visit, participants were asked to complete another drawing of what they thought the locations looked like from a top-down view.

4.5 Analysis methods and calculations

Data analysis was done with Matlab[®] Versions R2017b and R2018b (© 1994-2021 The MathWorks, Inc.). IBM[®] SPSS[®] Statics Version 27 (IBM, 2020) and Excel Microsoft 365 (© Microsoft) were used for the statistical analysis. In all four stages multiple measures were logged and collected for each participant:

- The assigned target Order
- p , pointing number (equivalent to a time point post experiment start; Calibration: $p \in [1, 2, 3, \dots, 10]$, Training: $p \in [1, 2, 3, \dots, 8]$, Ph1 & Ph2: $p \in [1, 2, 3, \dots, 48]$)
- a , target number (Calibration: $a \in [1, 2, 3, \dots, 10]$, Training, Ph1 & Ph2: $a \in [1, 2, 3, 4, 5]$)
- x_a and y_a represent the x and y values of the actual (true) target location which was pointed at during pointing number p in Ph1 and Ph2
- $x_{cal,a}$ and $y_{cal,a}$ represent the x and y values of the actual calibration target location which was pointed at during calibration
- x_r and y_r represent the x and y values of the reported target location (puncture point of the laser on the plane) for pointing number p in Ph1 and Ph2
- $x_{cal,r}$ and $y_{cal,r}$ represent the x and y values of the reported calibration target location which was pointed at during calibration
- x_h and y_h denote the x and y values of the participant's head position during pointing number p

The data collected during all phases was also analyzed, but since the two Ph1 and Ph2 were the main experimental phases only the calculations and analysis methods of the data relevant to those two phases will be described in the following. The collected data specified above was used to calculate the dependent variables of the experiment:

Pointing distance $dist$ in VR meters. This measure was calculated for:

The distance between the participant's head position and the actual calibration target position during calibration

$$dist_{cal,a} = \sqrt{(x_h - x_{cal,a})^2 + (y_h - y_{cal,a})^2} \quad (1)$$

The distance between the participant's head position and the reported calibration target

position during calibration

$$dist_{cal,r} = \sqrt{(x_h - x_{cal,r})^2 + (y_h - y_{cal,r})^2} \quad (2)$$

The distance between the participant's head position and the actual target position in Ph1 and Ph2

$$dist_a = \sqrt{(x_h - x_a)^2 + (y_h - y_a)^2} \quad (3)$$

The distance between the participant's head position and the reported target position in Ph1 and Ph2

$$dist_r = \sqrt{(x_h - x_r)^2 + (y_h - y_r)^2} \quad (4)$$

Calibration distance difference *CalibrDD* in VR meters, denotes the difference between the actual pointing distance $dist_{cal,a}$ and the reported target distance $dist_{cal,r}$ during calibration. For each participant this was calculated individually for each of the ten calibration pointing distances $dist_{cal,a}$. This measure is direction insensitive, it was later used to offset distance biases during the calculation of the participants' over- and undershooting behavior.

$$CalibrDD = |dist_{cal,r} - dist_{cal,a}| \quad (5)$$

with equations (1) and (2):

$$CalibrDD = \left| \sqrt{(x_h - x_{cal,r})^2 + (y_h - y_{cal,r})^2} - \sqrt{(x_h - x_{cal,a})^2 + (y_h - y_{cal,a})^2} \right| \quad (6)$$

Over- and undershooting represented by the signed distance difference *DD* in meters. The participants over- and undershooting behavior when aiming for the targets was analyzed in order to gain insight into their distance estimation ability.

1. Calculate the reported distance $dist_r$ between head and reported target for that pointing p (see equation (4)).
2. Subtract the previously calculated $dist_a$ (equation (3)), which is the true distance between the head and actual target location.
3. Finally, depending on whether the pointing was deemed an over- or undershot, subtract or add the calibration distance difference *CalibrDD* corresponding to the calibration distance which was closest to the pointing distance of this pointing ($dist_{cal,a} \approx dist_a$).

$$DD_p = dist_r - dist_a \begin{cases} \text{undershoot,} & \text{if } DD_p < 0 \Rightarrow OU + CalibrDD_{dist} \\ \text{overshoot,} & \text{if } DD_p > 0 \Rightarrow OU - CalibrDD_{dist} \end{cases} \quad (7)$$

Calibration angle error $Calibr\theta$ in degrees, the angle deviation between actual target location and reported target location during calibration. For each participant this was calculated individually for each of the ten calibration pointing distances $dist_{cal,a}$. This measure was used to offset any individual or distance based bias during the calculation of the angle error.

1. $u_{cal,0}$, the normalized vector between head position and actual calibration target position was calculated.
2. $v_{cal,0}$, the normalized vector between head position and reported calibration target position was determined.
3. The angle between reported and actual target location in reference to the head position was quantified

$$Calibr\theta = \sin^{-1} \left(|u_{cal,0} \times v_{cal,0}| \right) * 180^\circ / \pi \quad (8)$$

Note: calculated in Matlab using $\text{rad2deg}(\text{atan2}(\text{norm}(\text{cross}(u_{cal,0}, v_{cal,0})), \text{dot}(u_{cal,0}, u_{cal,0})))$

Calibration distance error $CalibrDE$ in VR meters, the distance between actual target location and reported target location during calibration. For each participant this was calculated individually for each of the ten calibration pointing distances $dist_{cal,a}$. This measure was used to offset any individual or distance based bias during the calculation of the distance error.

$$CalibrDE = \sqrt{(x_{cal,r} - x_{cal,a})^2 + (y_{cal,r} - y_{cal,a})^2} \quad (9)$$

Angle error θ_p in degrees, the angular deviation between the actual (true) target position $(x_a|y_a)$ and the reported target position $(x_r|y_r)$ for each pointing number p .

1. u_0 , the normalized vector between head position and actual target position was calculated.
2. v_0 , the normalized vector between head position and reported target position was determined.
3. The angle between reported and actual target location in reference to the head position was quantified (note: calculated in Matlab using $\text{rad2deg}(\text{atan2}(\text{norm}(\text{cross}(u_0, v_0)), \text{dot}(u_0, v_0)))$)

$$\theta_p = \left(\sin^{-1} \left(|u_0 \times v_0| \right) * 180^\circ / \pi \right) - Calibr\theta \quad (10)$$

with equation (8):

$$\theta_p = \left(\sin^{-1} \left(|\vec{u}_0 \times \vec{v}_0| \right) * 180^\circ / \pi \right) - \left(\sin^{-1} \left(|u_{cal,0} \times v_{cal,0}| \right) * 180^\circ / \pi \right) \quad (11)$$

4. To offset any bias due to individual pointing ability or pointing distance *Calibr* θ was subtracted. This was done by determining the pointing distance $dist_a$ for each individual pointing. This was then compared to the calibration distances $dist_{cal,a}$ and the *Calibr* θ was subtracted, which corresponded to the same/similar pointing distance ($dist_{cal,a} \approx dist_a$) (the calibration distances matched the distances used in Ph2, since Ph1 target locations were random, the calibration distance closest to the respective pointing distance was used).

5. The angle error was determined separately for each Ph1, Ph2A and Ph2B. It was also separated and compared for 1-back and 2-back tasks. It was analyzed both within participant and across all participants (by averaging across all errors for each pointing p).

Distance error DE_p in VR meters, distance between the actual (true) target position ($x_a|y_a$) and the reported target position ($x_r|y_r$). This was calculated individually for each pointing p

$$DE_p = \sqrt{(x_r - x_a)^2 + (y_r - y_a)^2} - CalibrDE \quad (12)$$

with equation (9):

$$DE_p = \sqrt{(x_r - x_a)^2 + (y_r - y_a)^2} - \sqrt{(x_{cal,r} - x_{cal,a})^2 + (y_{cal,r} - y_{cal,a})^2} \quad (13)$$

1. For each pointing p the distance between actual and reported target location was calculated

2. To offset any bias due to individual pointing ability or pointing distance *Calibr* DE was subtracted. This was done by determining the pointing distance $dist_a$ for each pointing p and comparing it to the calibration distances $dist_{cal,a}$. The calibration distance error *Calibr* DE of the calibration distance, which most closely matched the distance of the pointing ($dist_{cal,a} \approx dist_a$) was subtracted (the calibration distances matched the distances used in Ph2. Since Ph1 target locations were random, the calibration distance closest to the respective pointing distance was used).

3. The distance error was determined separately for Ph1, Ph2A and Ph2B. It was also separated and compared for 1-back and 2-back tasks. It was analyzed both within participant and across all participants (by averaging across all DE_p for each pointing p).

Residual deviation d , in VR meters is a measure that takes a different approach. It's based on the idea that participants may have formed a non-euclidean representation of

the goal positions in their mind that mirrors the overall structure of the targets, but not the exact euclidean relationships. Hereby d is the remaining deviation between the actual target positions and the estimated represented goal positions in the participant's memory (based on the idea and calculation kindly provided by H. A. Mallot in personal communication, November 18, 2020).

1. $r_p = (x_r|y_r)$ represents each participants measurements (reported target locations) at pointing number p . For measurements taken at pointing $p \geq 3$ the 1-back and 2-back tasks can be denoted as $r_{p,p-1}$ and $r_{p,p-2}$. After eight pointings a participant had pointed to every one of the five targets at least once, these 8 measurements sufficed to uniquely determine an estimated vector of five represented goal positions in the participant's memory: z_1, \dots, z_5 .

For the 1-back and 2-back tasks ($p \geq 3$) it can be presumed:

$$r_{p,p-1} = z_{p-1} - z_p \quad (14)$$

$$r_{p,p-2} = z_{p-2} - z_p \quad (15)$$

2. for $p = 3, \dots, 6$ this can be written in a matrix notation

$$\underbrace{\begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}}_M \underbrace{\begin{pmatrix} z_1^T \\ z_2^T \\ z_3^T \\ z_4^T \\ z_5^T \end{pmatrix}}_Z = \underbrace{\begin{pmatrix} r_{3,1}^T \\ r_{3,2}^T \\ r_{4,2}^T \\ r_{4,3}^T \\ r_{5,3}^T \\ r_{5,4}^T \\ r_{1,4}^T \\ r_{1,5}^T \end{pmatrix}}_R$$

Note: without losing generality, it can be assumed that $z_1 = (0, 0)^T$ and the first column of the matrix can be disregarded.

To obtain the unknown Z -vector:

$$MZ = R \quad (16)$$

$$\Rightarrow Z = (M^T M)^{-1} M^T R \quad (17)$$

3. The complete estimated Z -vector was rotated and shifted to most closely match the corresponding actual target position values, which are stored in a 5×2 -matrix A consisting of the five actual target positions ($(x_a|y_a)$ with $a = 1, \dots, 5$). This was done in Matlab using `[d, U] = procrustes(x,Z,'scaling',false,'reflection',false)`. Hereby U holds

the final estimated coordinates and finally, d expresses the residual deviation between the actual target positions and the final estimate U . Per participant, six d values were obtained, one for every group of eight pointings.

Precision and **Accuracy** in meters, for each target number a . These variables were only calculated for Ph2, since the actual target positions did not change. Of the the 48 pointings nine to ten pointings were directed at each of the five actual target positions. In order to analyse whether the participants' precision ($Prec$) and/or accuracy (Acc) improved over the course of the experiment the pointings were separated by early versus late, where the first 24 pointings ($p = 1, \dots, 24$) were considered "early" and the last 24 pointings ($p = 25, \dots, 48$) were considered "late". $Prec$ and Acc were calculated across these early versus late pointings and compared. Per participant and time-point (early vs. late), separately for Ph2A and Ph2B:

1. A covariance SD error ellipse was calculated around all reported target positions of pointings directed at target number a (Schoeder, 2018), or rather its corresponding position $(x_{a,p}|y_{a,p})$ (i.e. five SD error ellipses, one for each actual target position, for both early and late pointings). \mathcal{X}_a and Υa represent the x and y values of the of the SD ellipse around the reported target positions aimed at the same target number a . n denotes the number of "shots" at target number a (i.e. the number of reported target positions aimed at the same target number a).
2. The precision and accuracy were calculated separately for each target number a .

$$Prec_a = \frac{1}{n} \sum_{i=1}^n \sqrt{(\mathcal{X}_a - x_{r,i})^2 + (\Upsilon a - y_{r,i})^2} \quad (18)$$

$$Acc_a = \sqrt{(\mathcal{X}_a - x_a)^2 + (\Upsilon a - y_a)^2} \quad (19)$$

3. The mean across all five accuracy or precision values was calculated individually for each participant.

$$Prec = \frac{1}{5} \sum_{i=1}^{a=5} Prec_a \quad (20)$$

$$Acc = \frac{1}{5} \sum_{i=1}^{a=5} Acc_a \quad (21)$$

4. A *large* value indicates a higher distance between centroid and real or reported target positions and therefore shows a *low* accuracy or precision! To allow for an intuitive reading of the $Prec$ and Acc values and to make a comparisson of the values easier between participants, the measures are reported as absolute numbers on a scale from zero

to one. Hereby one denotes the highest *Prec* or *Acc* achieved among participants (across all five targets). This was done by dividing the values by *MinPrec* or *MinAcc*, which are the lowest *Prec* and *Acc* values in meters (and therefore the highest achieved precisions and accuracy).

$$Prec = \frac{\frac{1}{Prec}}{\frac{1}{MinPrec}} \Rightarrow \frac{Prec}{MinPrec} \quad (0 \leq Prec \leq 1) \quad (22)$$

$$Acc = \frac{\frac{1}{Acc}}{\frac{1}{MinAcc}} \Rightarrow \frac{Acc}{MinAcc} \quad (0 \leq Acc \leq 1) \quad (23)$$

5 Results

5.1 Calibration

During calibration a baseline for each participant's angle error (AE) and distance error (DE) in a VR setting was obtained by testing each participants ability to point at 10 visible targets at different distances (example of visualized calibration pointings for participant three: Figure 18). The calibration DE and AE was calculated for each participant (Figure 19). Although there were individual differences between participants' baseline pointing abilities, no unexpectedly large errors occurred. The largest AE among all participants was 1.76° and the largest DE was measured at 1.02 VR meters (both by participant three, Figures 18 and 19). Nine participants showed a clear decrease in calibration AE with increasing distance from head to target (pointing distance) (e.g. Figure 19a). Inversely, nine participants showed a clear increase in DE with an increasing target distance (e.g. Figure 19b).

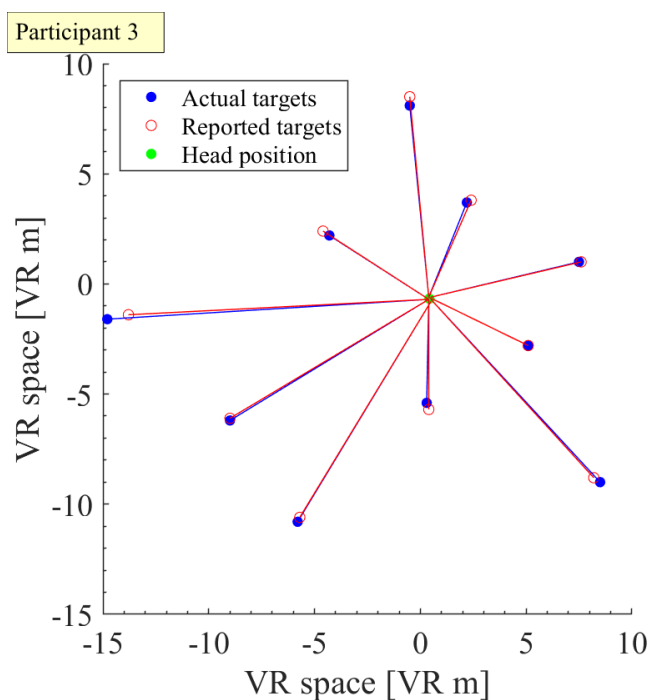


Figure 18: Top view (eagle's eye) of the locations of the participants head (green), the actual (true) calibration targets (blue) and the participants reported targets (red) during calibration. The image shown is the example data from the calibration performance of participant three.

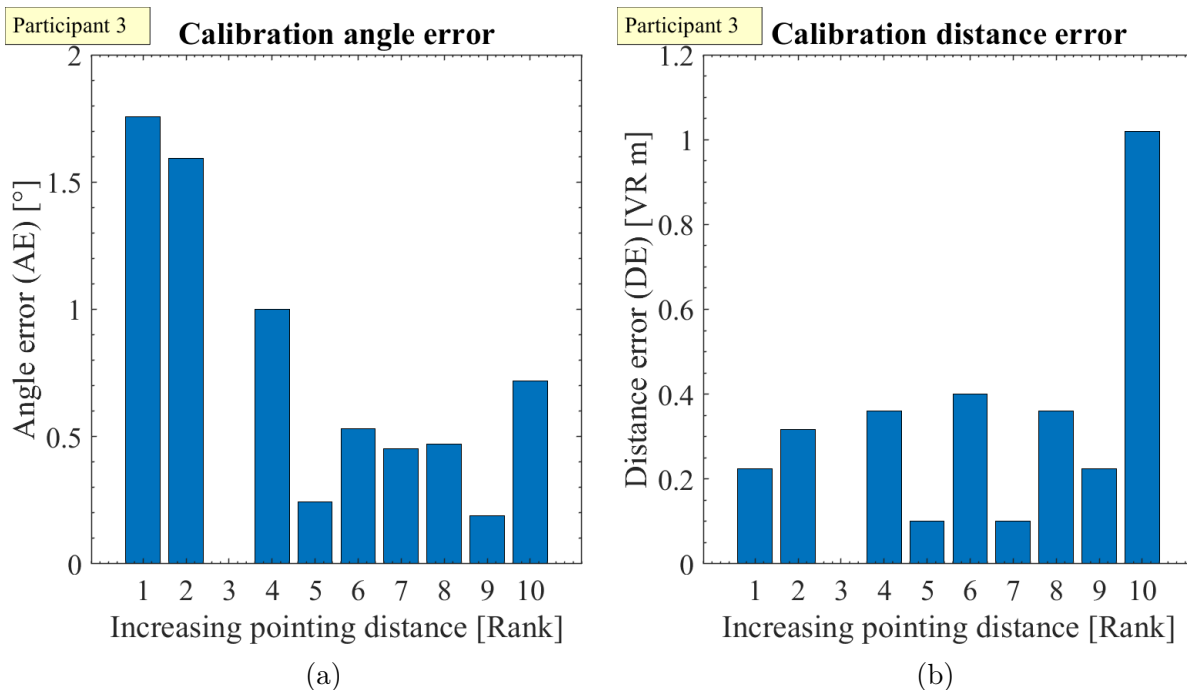


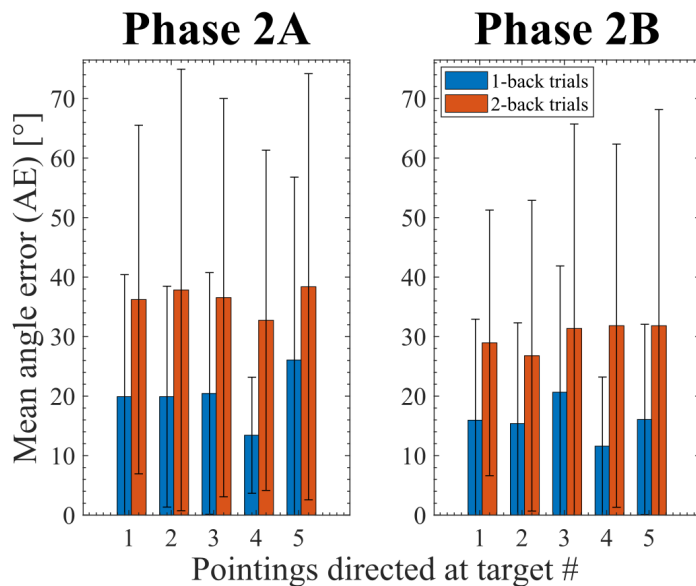
Figure 19: Example data of participant three: The angle error (AE) (a) and distance error (DE) (b) between the actual target positions and the reported target positions, in degrees or VR meters respectively, is shown over increasing target distances (ranked by distance from target to head position). An empty column indicates an error of 0.

5.2 Angle error and distance error

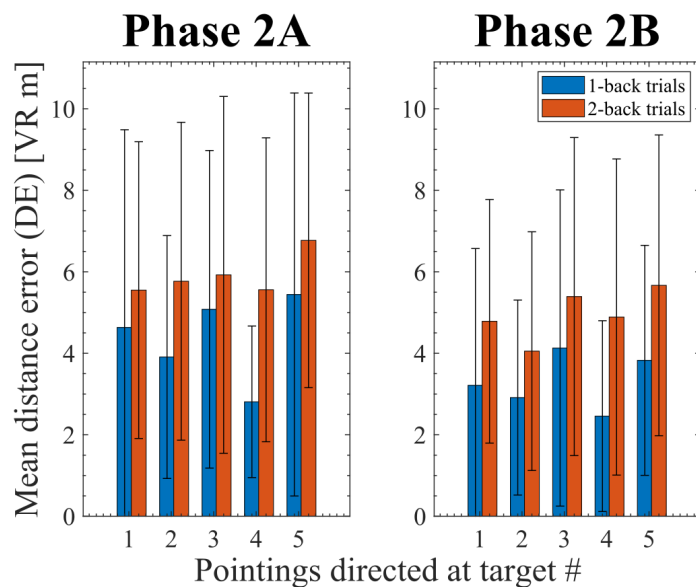
5.2.1 Verification analyses

It was necessary to ensure no undue influence on the data caused by the two different assigned orders in which the participants visited and pointed at the targets (Table 7). A two sample t-test assuming equal variance proved there to be no significant influence on the angle error (AE) caused by the assigned target order in either Ph1 ($t(22) = -1.581$, $p = 0.128$) or Ph2 ($t(46) = 0.281$, $p = 0.780$). The distance error (DE) was also not significantly influenced by the target Order in either Ph1 ($t(22) = -1.124$, $p = 0.273$) or Ph2 ($t(46) = 0.407$, $p = 0.686$). It will therefore henceforth be assumed that the Order had no influence on the variables.

It was also analyzed whether a specific target location elicited significantly higher AEs or DEs than others. The target locations in Ph1 were different for every participant and round, so a statistical comparison could not be done. However, since each target took on ten different positions throughout the phase, it can be assumed that a systematic bias caused by target location can be ruled out. Within Ph2A & Ph2B the AE and DE were investigated separately for each target location (number) by averaging across the errors of all the pointings of all participants which were directed at the same target number (Figure 20a&b). The data was additionally separated for 1-back and 2-back conditions.



(a)

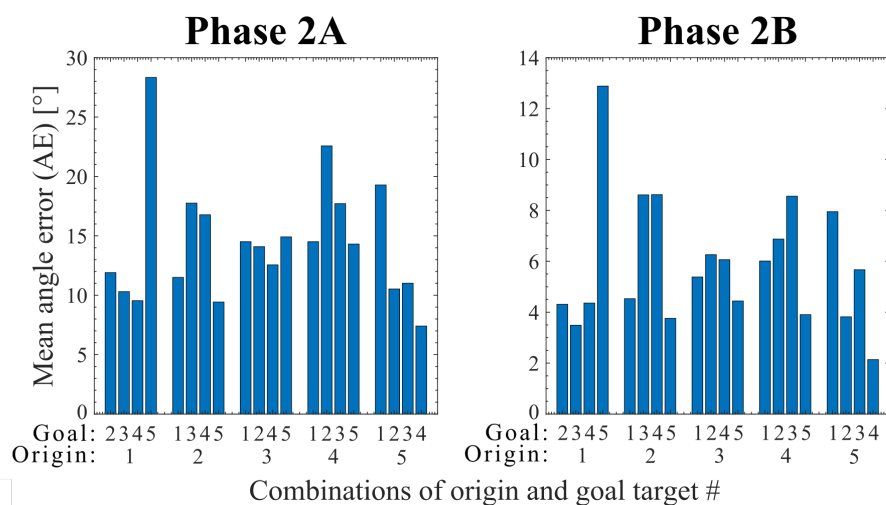


(b)

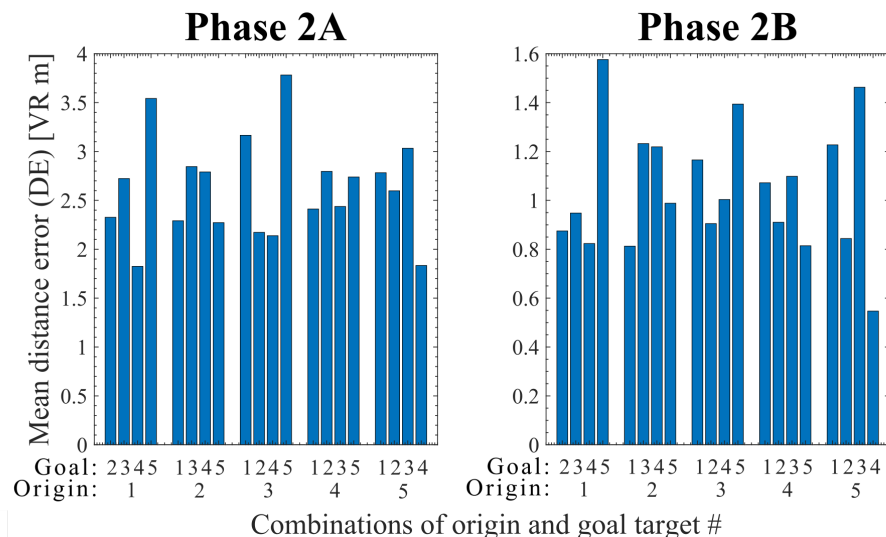
Figure 20: The mean angle error (AE) in degrees (a) and distance error (DE) in VR meters (b) are shown for each individual target number (1-5) by depicting the average error across all pointings directed at this target number. The measures are taken across all participants and separated by 1-back and 2-back trials.

A Bonferroni corrected repeated measure ANOVA analysis with SPSS revealed that the target number, which was pointed to, had no significant effect on the AE in both Ph2A ($F(4, 44) = 2.067$; $p = 0.101$; $\eta_p^2 = 0.158$) and 2B ($F(4, 44) = 0.735$; $p = 0.573$; $\eta_p^2 = 0.063$). In contrast the influence of the back condition (1-back or 2-back) on the AE was highly significant (2A: $F(1, 11) = 50.413$; $p < 0.05$; $\eta_p^2 = 0.821$ and 2B: $F(1, 11) = 38.296$; $p < 0.05$; $\eta_p^2 = 0.777$). There was no significant interaction effect between the back condition and the target number for either phase (2A: $F(4, 44) = 0.195$; $p = 0.939$;

$\eta_p^2 = 0.017$ and 2B: $F(4, 44) = 0.513$; $p = 0.726$; $\eta_p^2 = 0.045$). In contrast, for the DE a Bonferroni corrected repeated measure ANOVA showed a significant effect of the target number on the distance error (2A: $F(4, 44) = 3.731$; $p < 0.05$; $\eta_p^2 = 0.253$ and 2B: $F(4, 44) = 4.624$; $p < 0.05$; $\eta_p^2 = 0.296$). The 1-back or 2-back condition also displayed a significant effect on the distance error (2A: $F(1, 11) = 35.029$; $p < 0.05$; $\eta_p^2 = 0.761$ and 2B: $F(1, 11) = 37.735$; $p < 0.05$; $\eta_p^2 = 0.774$). There was no significant interaction effect between the back condition and the target number for either phase (2A (corrected using Greenhouse-Geisser estimates of sphericity $\varepsilon = 0.55$): $F(2.179, 23.973) = 0.867$; $p = 0.442$; $\eta_p^2 = 0.073$ and 2B: $F(4, 44) = 0.722$; $p = 0.581$; $\eta_p^2 = 0.062$).



(a)



(b)

Figure 21: The mean angle error (AE) in degrees (a) and mean distance error (DE) in VR meters (b) for each combination of origin and goal target numbers. Origin: The target at which the participants head was located during pointing. Goal: The target to which the participant was directed to point. The errors are taken across participants and separated by Ph (2A versus 2B).

Additionally, all combinations of origin and goal target were analyzed for the same measures. The origin target is defined as the target closest to the participant's head position during pointing (the target from which was pointed). The goal target indicates the target number to which the participant was directed to point (the target that they were aiming for) (Figure 21a&b). The pointings with origin at target 1 and goal target 5 (1-5 target combination) evoked a significantly higher mean AE (Ph2A: $t(18) = 16.503$; $p < 0.05$ and Ph2B: $t(18) = 16.729$; $p < 0.05$) and DE (Ph2A: $t(18) = 8.903$; $p < 0.05$ and 2B: $t(18) = 10.734$; $p < 0.05$) than any other target combinations. No other target combinations caused a significantly higher AE value. But the pointings from target 3 to target 5 also led to a higher mean DE than other combinations (Ph2A: $t(18) = 12.042$; $p < 0.05$ and Ph2B: $t(18) = 6.478$; $p < 0.05$). The AE and DE are subsequently first analyzed without any modifications or corrections accounting for the discussed target locations and later including corrections.

5.2.2 Over- and undershooting

To observe each participant's ability to estimate the correct target distance their reported "shots" were analyzed for over- and undershooting. A pointing was deemed an undershoot (US) if the reported target distance fell short of the actual target distance (negative sign). A reported distance that was further away from the participants' head than the actual target distance signaled an overshoot (OS; positive sign). For each participant this signed distance difference (DD) was plotted over pointing numbers (example of participant nine in Figure 22). Across participants there was no clear indication for a systematic reduction of OS or US over the time course of the experiment. Rather, some participants tended toward OS and others towards US. It proved to be more informative to compute the OS and US dependent on the actual target distance (distance from head to actual target), sorted by increasing size of the target distance (example of participant nine in Figure 23). For all participants, phases and back conditions there was a clear negative correlation coefficient R between the DD and the target distances (Table 1). The closer the true target location was to the participant, the larger and more positive the DDs in the reported pointings were, i.e. the more participants tended to overshoot. Inversely, the larger the distance to the true target, the larger and more negative the DDs were, so the more participants tended to undershoot in their responses. The smallest DD errors were generally achieved for targets in a medium distance range to the participants. A Bonferroni corrected repeated measure ANOVA revealed no significant effect of either phase ($F(2, 22) = 0.603$; $p = 0.556$; $\eta_p^2 = 0.052$) or back condition ($F(2, 22) = 0.401$; $p = 0.539$; $\eta_p^2 = 0.035$) on R .

Table 1: The correlation coefficient R between the signed distance difference (height of the over- or undershoot) and the actual target distance, sorted by increasing size. R is shown for each back condition, participant and phase.

Participant no.	Phase 1		Phase 2A		Phase 2B	
	R		R		R	
	1-back	2-back	1-back	2-back	1-back	2-back
1	-0.862	0.450	-0.761	-0.741	-0.961	-0.938
2	-0.494	-0.532	-0.638	-0.664	-0.735	-0.519
3	-0.492	-0.228	-0.326	-0.435	-0.304	0.025
4	-0.436	-0.764	-0.423	-0.645	-0.344	-0.935
5	-0.788	-0.841	-0.786	-0.954	-0.605	-0.924
6	-0.726	-0.210	-0.632	-0.856	-0.653	-0.145
7	-0.834	-0.602	-0.839	-0.853	-0.786	-0.863
8	-0.613	-0.602	-0.737	-0.357	-0.350	-0.875
9	-0.841	-0.768	-0.958	-0.743	-0.847	-0.989
10	-0.878	-0.587	-0.751	-0.713	-0.895	-0.602
11	-0.747	-0.679	0.195	-0.807	-0.959	-0.522
12	-0.739	-0.868	-0.840	-0.694	-0.789	-0.737
Mean	-0.704	-0.519	-0.625	-0.705	-0.686	-0.669

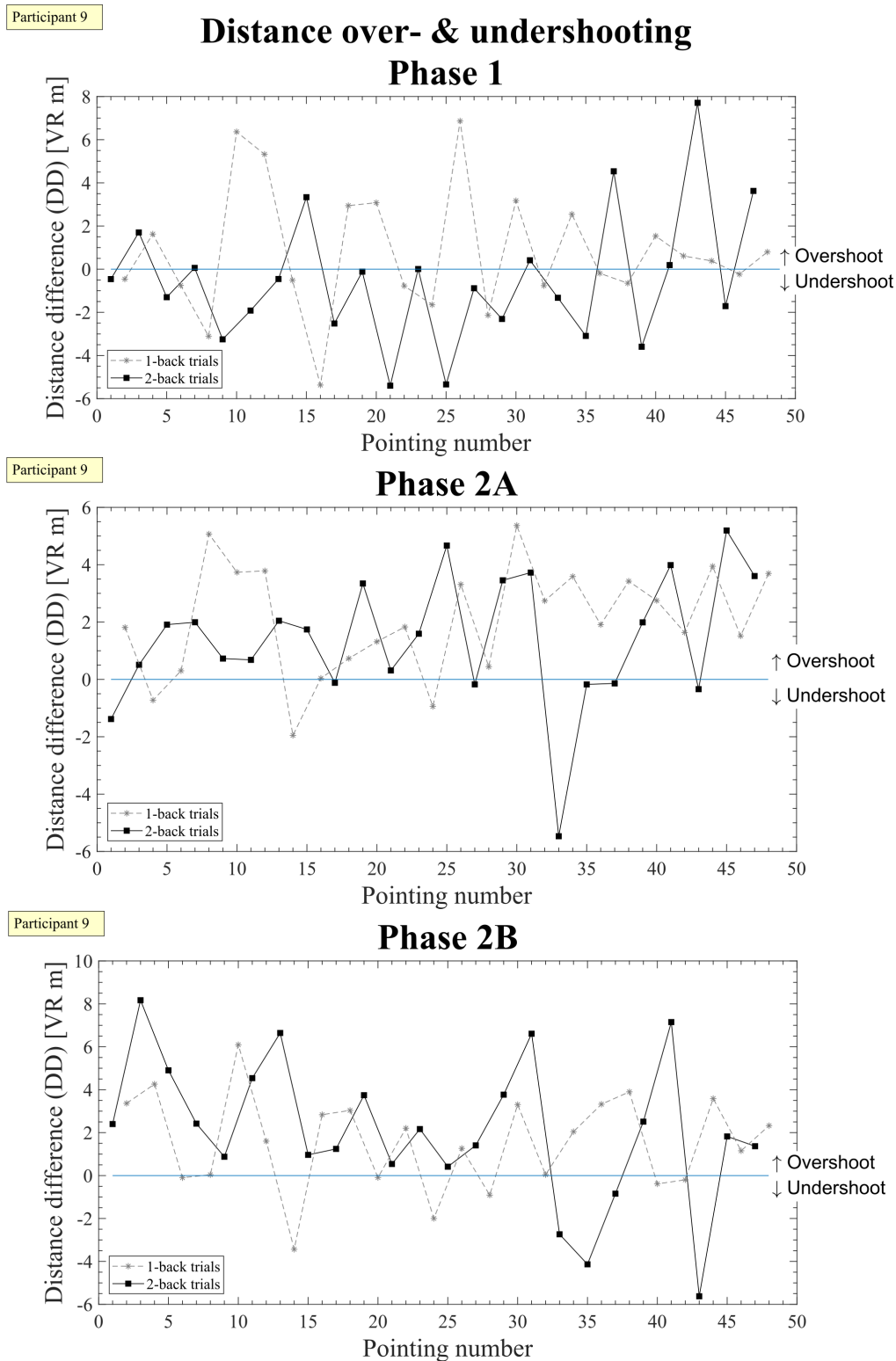


Figure 22: Participant 9: The signed distance difference (DD) between head to reported target and head to actual target is separated by 1-back (grey) and 2-back (black) trials. A value under zero indicates undershooting (underestimated of the target distance). DD above zero indicates overshooting (overestimation of the target distance).

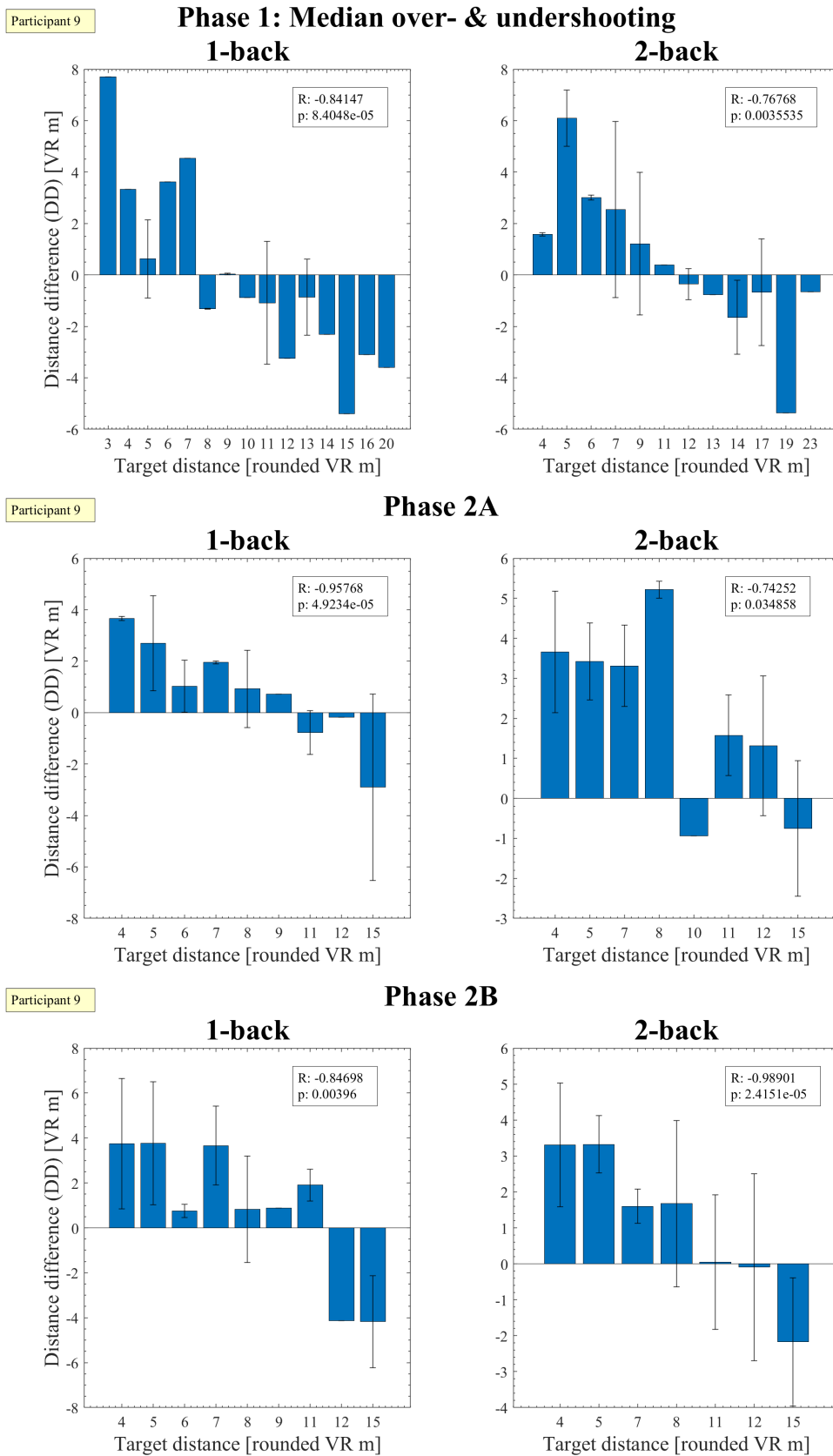


Figure 23: Participant 9: The mean signed distance difference (DD) separated by back condition in each phase, over the rounded target distance (head to actual target) sorted by increasing value in VR meters. If the DD is above zero it indicates overshooting (overestimation of target distance), below zero illustrates undershooting (underestimation of target distance).

5.2.3 Angle error

The angle error (AE) between the actual and reported target locations over the pointing number (i.e. time course of the experiment) was calculated for each individual and across all participants. It was analyzed separately for each phase (combined AE) and additionally separated by 1-back or 2-back condition (separated AE). For each of the three phases a linear least squares (lsq) fit curve was calculated and overlaid onto the data. For the separated AE this was done separately for the 1-back and 2-back tasks.

Combined angle error: The mean combined AE across all participants (Figure 24a) was compared by means of the lsq fit slope of each phase. This showed the strongest reduction in AE over time, i.e. the steepest slope, for Ph1 ($m = -0.133$), followed by Ph2A ($m = -0.059$) and lastly Ph2B ($m = -0.032$). Here the fit of the first pointings in Ph2A started at a lower AE than the in Ph1 ($\sim 30^\circ$ versus $\sim 35^\circ$). But because of Ph2A's flatter learning curve the mean AE for the final pointings was about the same in both phases ($\sim 28^\circ$). The fit for Ph2B was overall lower than for the other two phases (max. $\sim 24^\circ$), but held the flattest learning curve. Examining the individual participant's slopes for the combined AE (Table 2; Attachments Figures 38 and 39) displayed that four participants failed to show a learning curve (negative slope) in Ph1 and Ph2B each, in Ph2A this was the case for five participants. Three participants had a steeper learning curve in Ph2A compared to Ph1.

Separated angle error: The mean separated AE across all participants was overall lower in the 1-back tasks compared to the 2-back tasks in all three phases (Figure 24b). The slope of the linear lsq fit of the 1-back trials decreased over time in Ph1 ($m = -0.108$). Interestingly it showed an increase in AE over time in Ph2A ($m = 0.069$) and Ph2B ($m = 0.114$). Yet the fit for early pointings was still lower in Ph2A than Ph1 ($\sim 28^\circ$ vs. $\sim 19^\circ$). But because of the opposite slopes both phases had an AE of $\sim 22^\circ$ for late pointings. The fit of Ph2B was overall the lowest (max. $\sim 19^\circ$). Looking at the slopes of the separated AE fits for individual participants, it can be seen that four participants did not display a learning curve during Ph1 (Table 2; Attachments Figures 40 and 41). This was also the case for seven participants during Ph2A. The slope of the fit was steeper in Ph2A versus Ph1 in four participants. Within the 2-back tasks of the mean AE across all participants the slope of the fits decreased in all three phases. Ph2A showed the steepest learning curve ($m = -0.144$). Ph1 ($m = -0.122$) and Ph2B ($m = -0.142$) followed closely. The fit for Ph2B early pointings was the same as Ph2A late pointings ($\sim 33^\circ$), but decreased to exhibit the lowest overall AEs during its later pointings ($\sim 26^\circ$). On an individual participant basis, four participants showed no learning curve during Ph1 (Table 2) and three participants in Ph2A. The slope was steeper in Ph2A versus Ph1 four times. A Bonferroni corrected repeated measure ANOVA revealed that both the phase ($F(1.306, 14.368) = 5.133$; $p < 0.05$; $\eta_p^2 = 0.318$) and the back condition

($F(1, 11) = 103.869$; $p < 0.05$; $\eta_p^2 = 0.904$) had a significant effect on the AE (Mauchly's test indicated a violation of the assumptions of sphericity for the main effect of phase $X^2(2) = 7.58$, $p < 0.05$, so the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.65$)) (Field, 2009). There was no significant interaction effect between phase and back condition ($F(2, 22) = 0.453$; $p = 0.642$; $\eta_p^2 = 0.040$).

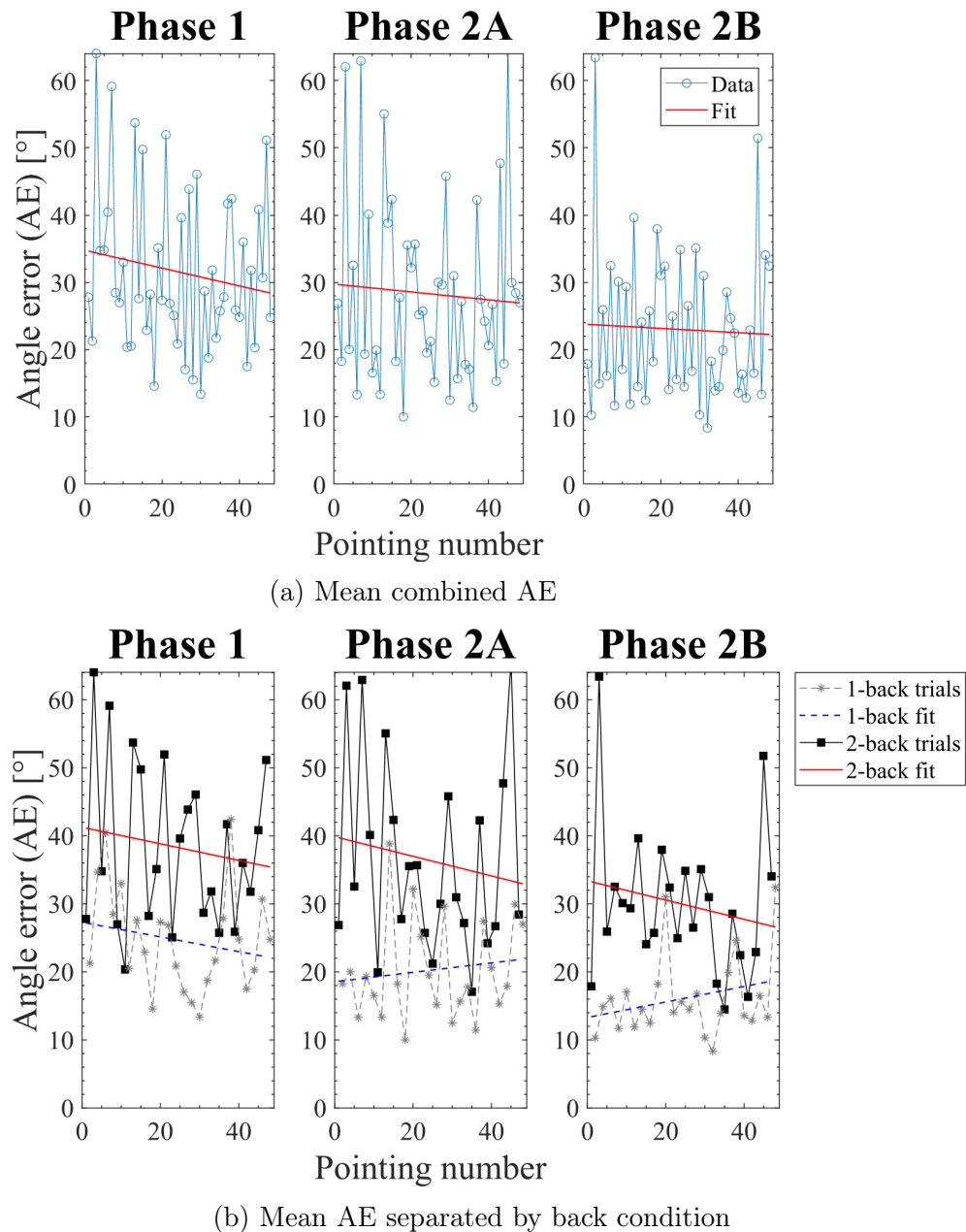


Figure 24: The mean angle error in degrees across all participants, plotted for each of the 48 pointings (i.e. over the time course of the experiment). (a) Combined mean AE (blue) and the corresponding least squares fit (red) for each phase. (b) Separation of 1-back (grey) and 2-back (black) trials with corresponding linear lsq fits (blue and red).

Table 2: The slopes of the linear lsq fits of the angle error (AE) in $^{\circ}$ over pointing number for each participant. The slope is shown for each of the three phases within the combined AE as well as separated for 1-back and 2-back trials.

Participant	Combined angle error			1-back trials			2-back trials		
	Phase 1	2A	2B	Phase 1	2A	2B	Phase 1	2A	2B
1	-0.010	0.084	0.348	0.003	0.325	0.586	0.005	-0.107	0.174
2	-0.290	-0.682	0.101	-0.213	-0.361	-0.049	-0.315	-0.937	0.268
3	-0.320	-0.146	-0.008	-0.147	0.037	-0.009	-0.466	-0.271	0.022
4	-0.585	-0.366	-0.503	-0.332	-0.090	-0.382	-0.809	-0.575	-0.611
5	-0.744	0.379	-0.269	-0.643	1.020	-0.016	-0.824	-0.246	-0.503
6	0.000	0.352	-0.143	-0.110	0.288	0.043	-0.214	0.485	-0.280
7	-0.111	0.044	-0.250	0.002	-0.015	0.014	-0.200	0.152	-0.464
8	0.481	-0.386	-0.151	0.445	-0.191	0.091	0.575	-0.546	-0.353
9	0.387	-0.283	-0.342	0.011	0.012	0.229	0.832	-0.521	-0.840
10	0.353	0.435	0.221	0.501	-0.224	0.130	0.208	1.135	0.339
11	-0.151	-0.069	-0.097	-0.132	0.029	0.145	-0.152	-0.130	-0.315
12	-0.534	-0.053	0.711	-0.752	0.000	0.586	-0.255	-0.110	0.863
Mean	-0.127	-0.058	-0.032	-0.114	0.069	0.114	-0.135	-0.139	-0.142

5.2.4 Distance error

The distance error (DE) between the actual and reported target positions was compared across all participants separately for each of the three phases and a linear lsq fit was applied (Figure 25). Again the combined DE (not separated by back condition) as well as the separated DE were calculated for each participant and averaged across all participants.

Combined distance error: The steepest learning curve (i.e. the biggest DE reduction over time) for the mean DE across all participants was evident in Ph1 ($m = -0.011$) followed by Ph2A ($m = -0.003$) and Ph2B ($m = -0.002$) (Figure 25a). The fit of the mean DE was the highest for Ph1 (max. ~ 6 m, min. ~ 5 m). The fit for the first pointings of Ph2A was just below that of the last pointings of Ph1, but the average DE barely reduced over time. Ph2B displayed the lowest average DE over the entire time course (max. ~ 4.2 m). Figures of each participant's combined DE over pointing number can be found under Attachments Figures 42 and 43. The slopes of the fit for the combined DE for individual participants reveal that five participants failed at achieving a learning curve (negative slope of the data fit) during each phase (Table 3). Five participants exhibited a steeper learning curve in Ph2A compared to Ph1.

Separated distance error: Overall the mean DE of the 1-back tasks was clearly lower than that of the 2-back tasks in all three phases (Figure 25b). The fit of the 1-back tasks had the steepest slope in Ph1 ($m = -0.031$) albeit the slopes in Ph2A ($m = 0.005$) and Ph2B ($m = 0.011$) were slightly positive. Overall the mean DE was reduced from Ph1 to Ph2A, to Ph2B. The fit for the first pointings in Ph2A (~ 4.4 m) was slightly above the fit of the final pointings in Ph1 (~ 4.2 m). Overall the distance error was lowest in Ph2B

(max. DE of ~ 3.6 m). Analysing the fit slopes for the 1-back trials of each participant (Table 3) demonstrated a lack of a learning curve during Ph1 in three participants. The same was true for 50% of the cohort for Ph2A and Ph2B. Yet the slope was steeper in Ph2A compared to Ph1 in five cases. The slopes of the fit of the 2-back tasks for the mean DE across all participants showed the opposite behavior, with a positive slope for Ph1 ($m = 0.013$) but increasingly negative slopes for Ph2A ($m = -0.007$) and Ph2B ($m = -0.010$). Again Ph1 showed the overall highest mean DE (max. ~ 6.8 m). In Ph2A the error of the first pointings was at about the same level as in Ph1, but was subsequently reduced (~ 6.1 to ~ 5.7 m). Ph2B showed the overall steepest slope and lowest DE values of the fit (max. ~ 5.2 m). Depictions of each participant's separated DE can be found under Attachments Figures 44 and 45. Seven individual participants showed a lack of DE reduction over time during 2-back tasks in Ph1 and four participants in Ph2A and Ph2B (Table 3). The slope of the fit was steeper in Ph2A compared to Ph1 five times. A Bonferroni corrected repeated measure ANOVA showed that the phase had a highly significant effect on the DE ($F(2, 22) = 14.079$; $p < 0.05$; $\eta_p^2 = 0.561$), as did the back condition ($F(1, 11) = 111.356$; $p < 0.05$; $\eta_p^2 = 0.910$). There was no significant interaction effect between phase and back condition ($F(2, 22) = 0.035$; $p = 0.966$; $\eta_p^2 = 0.003$).

Table 3: The slope of the lsq fit across the distance error (DE) in VR meters over pointing number for each participant. The slopes are shown individually for each of the three phases within the combined DE as well as for the separated 1-back and 2-back trials.

Participant	Combined distance error			1-back trials			2-back trials		
	Phase 1	2A	2B	Phase 1	2A	2B	Phase 1	2A	2B
1	-0.002	0.002	0.075	-0.015	0.050	0.062	0.012	-0.043	0.094
2	-0.008	-0.047	0.030	-0.016	-0.052	0.008	0.004	-0.037	0.054
3	-0.078	-0.054	-0.047	-0.033	-0.025	-0.085	-0.120	-0.077	-0.006
4	-0.103	-0.020	-0.056	-0.099	-0.002	-0.032	-0.104	-0.030	-0.078
5	-0.052	0.053	-0.022	-0.071	0.105	-0.022	-0.030	0.002	-0.021
6	0.000	0.047	0.012	-0.054	0.038	0.042	-0.004	0.067	-0.011
7	0.028	0.008	-0.020	0.012	-0.017	-0.013	0.047	0.037	-0.024
8	0.071	-0.084	-0.014	0.022	-0.041	0.023	0.124	-0.123	-0.047
9	0.072	-0.002	-0.060	-0.004	0.009	-0.010	0.154	-0.009	-0.101
10	0.025	0.101	0.012	-0.005	0.027	0.020	0.059	0.178	0.009
11	-0.019	-0.029	-0.023	0.015	-0.047	-0.012	-0.051	-0.003	-0.030
12	-0.047	-0.009	0.090	-0.136	0.017	0.148	0.050	-0.035	0.039
Mean	-0.009	-0.003	-0.002	-0.032	0.005	0.011	0.012	-0.006	-0.010

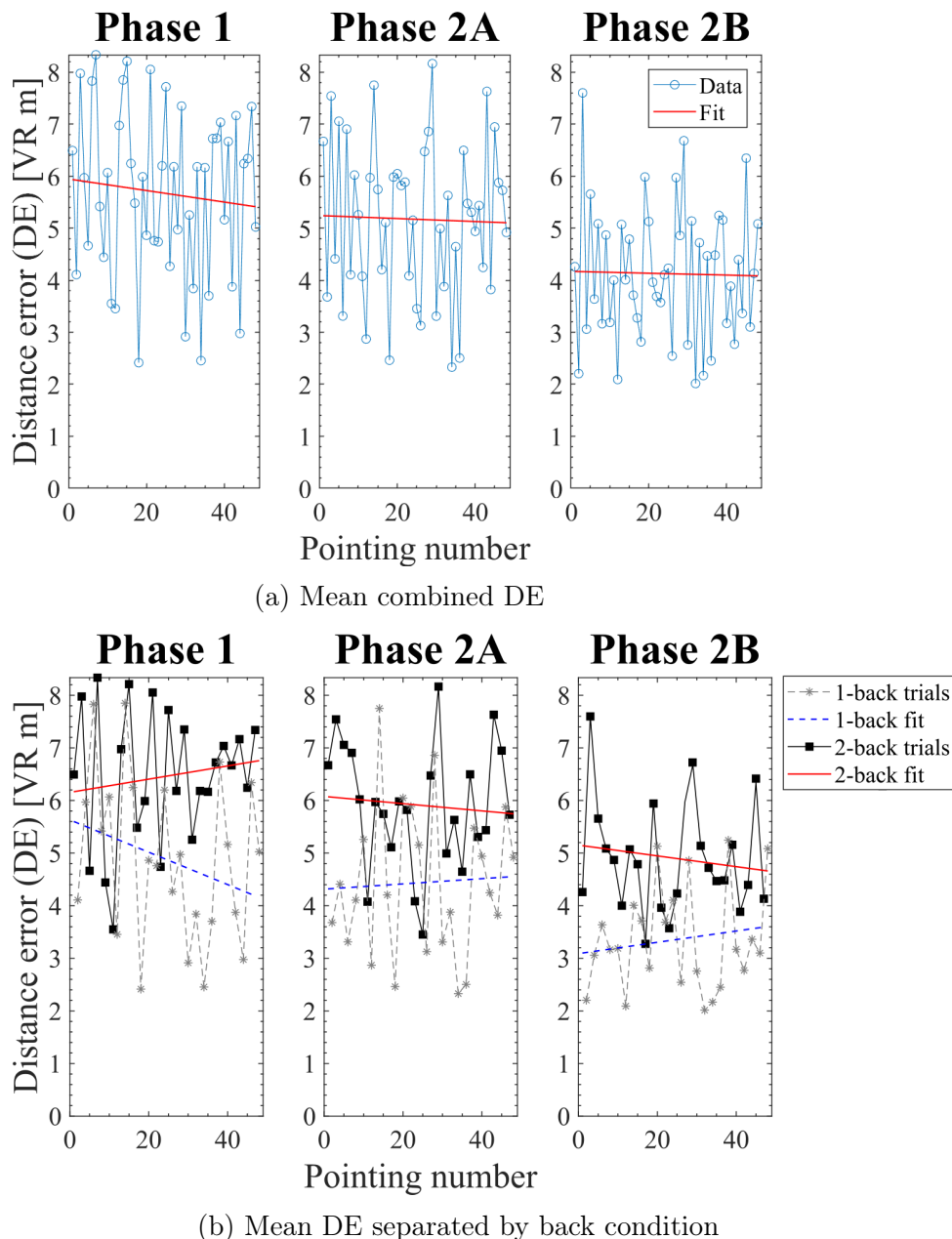


Figure 25: The distance error in VR meters across all participants is plotted for each of the 48 pointings (i.e. over the time course of the experiment). (a) Combined mean DE (blue) and the corresponding least squares fit (red) for each phase across all participants. (b) Separation of 1-back (grey) and 2-back (black) trials with corresponding lsq fits (blue and red).

5.2.5 Corrected angle and distance errors

During the verification analyses it was shown that for Ph2A and Ph2B the origin-goal target combination of 1-5 (i.e. pointing from the location of target 1 at target 5) lead to higher AEs and DEs than other target combinations. The origin-goal target combination of 3-5 also evoked higher DEs than other combinations, although it did not lead to a significantly higher AE. It was investigated, whether said elevations in errors for these

pointings influenced the overall errors. These target combinations affected a total of three pointings for the participants assigned the target order 1 (1-5: pointing no. 43; 3-5: pointing no. 28 and 39). For participants assigned target order 2 this affected six pointings overall (1-5: pointing no. 3, 29 and 35; 3-5: pointing no. 14, 33 and 43). Of these nine pointings three were 1-back tasks, the rest were 2-back tasks. These pointings were excluded and the "corrected" mean AE and DE were calculated across all participants and a linear lsq fit was done for each of the curves (Figures 26 and 27). Table 4 shows the slopes of the fits for the previously discussed, uncorrected data (means across all participants), as well as for the corrected data.

Table 4: Juxtaposition of the slopes of the lsq fits across the mean angle error (top) and distance error (bottom) over pointing number across all participants. The slopes are shown separated for each of the three phases within the combined mean angle error (AE) and distance error (DE) and separated for 1-back and 2-back trials.

	Combined AE			1-back trials			2-back trials		
	Phase 1	2A	2B	Phase 1	2A	2B	Phase 1	2A	2B
Uncorrected	-0.133	-0.059	-0.032	-0.108	0.069	0.114	-0.122	-0.144	-0.142
Corrected	-0.133	-0.113	0.025	-0.108	0.014	0.085	-0.122	-0.176	0.026

	Combined DE			1-back trials			2-back trials		
	Phase 1	2A	2B	Phase 1	2A	2B	Phase 1	2A	2B
Uncorrected	-0.011	-0.003	-0.002	-0.031	0.005	0.011	0.013	-0.007	-0.010
Corrected	-0.011	-0.014	0.002	-0.031	-0.001	0.010	0.013	-0.023	0.000

Combined angle error: (Figure 26a) After correction the slope of the combined mean AEs showed a steeper negative slope for Ph2A ($m = -0.113$), but a shift from a negative to a slightly positive slope in Ph2B ($m = 0.025$). The height relation of the mean angle errors between the phases remained the same, with the fit of Ph1 yielding the highest combined errors (max. $\sim 35^\circ$, min. $\sim 28^\circ$), followed by Ph2A (max. $\sim 30^\circ$, min. $\sim 24^\circ$) and Ph2B, despite its lack of a learn curve (max. $\sim 22.5^\circ$, min. $\sim 21^\circ$).

Separated angle error: (Figure 26b) The fit of the 1-back errors saw Ph2A ($m = 0.014$) and Ph2B ($m = 0.085$) flattening their slopes, which remained positive. The fit of the 2-back trials was affected by the corrections in the same way as the combined error. The fit of the slope for Ph2A turned more negative ($m = -0.176$), whilst Ph2B experienced a shift towards a slightly positive slope ($m = 0.026$). Overall the errors of the 2-back trials were still significantly higher than the 1-back trials in all phases. In both conditions the AEs were overall the highest in Ph1, lower in Ph2A and reached their minimum in Ph2B. A Bonferroni corrected repeated measure ANOVA showed that the mean AE was still significantly affected by both the phase ($\varepsilon = 0.68$, $F(1.365, 15.016) = 5.874$; $p < 0.05$; $\eta_p^2 = 0.348$) and the back condition ($F(1, 11) = 89.135$; $p < 0.05$; $\eta_p^2 = 0.890$).

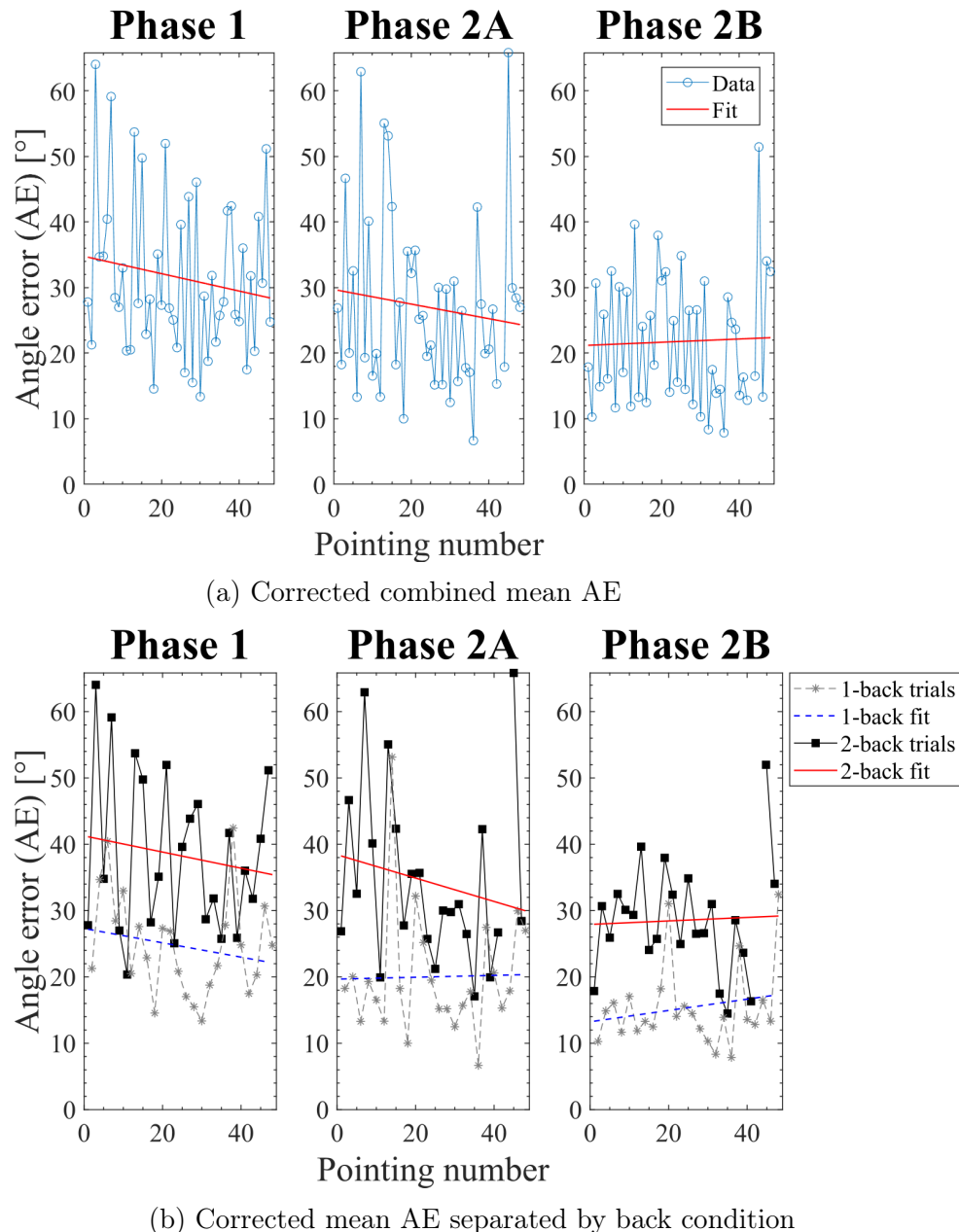


Figure 26: The mean angle error (AE) in degrees across all participants over the pointing number (i.e. time course of the experiment). The data for Ph2A and Ph2B was corrected by excluding the pointings of the two target combinations which had induce significantly higher errors (3 pointings excluded for participants with target Order 1, 6 pointings for order 2). (a) Combined mean AE (blue) and the corresponding least squares fit (red) for each phase. In (b) Separation of 1-back (grey) and 2-back (black) trials with corresponding linear lsq fits (blue and red).

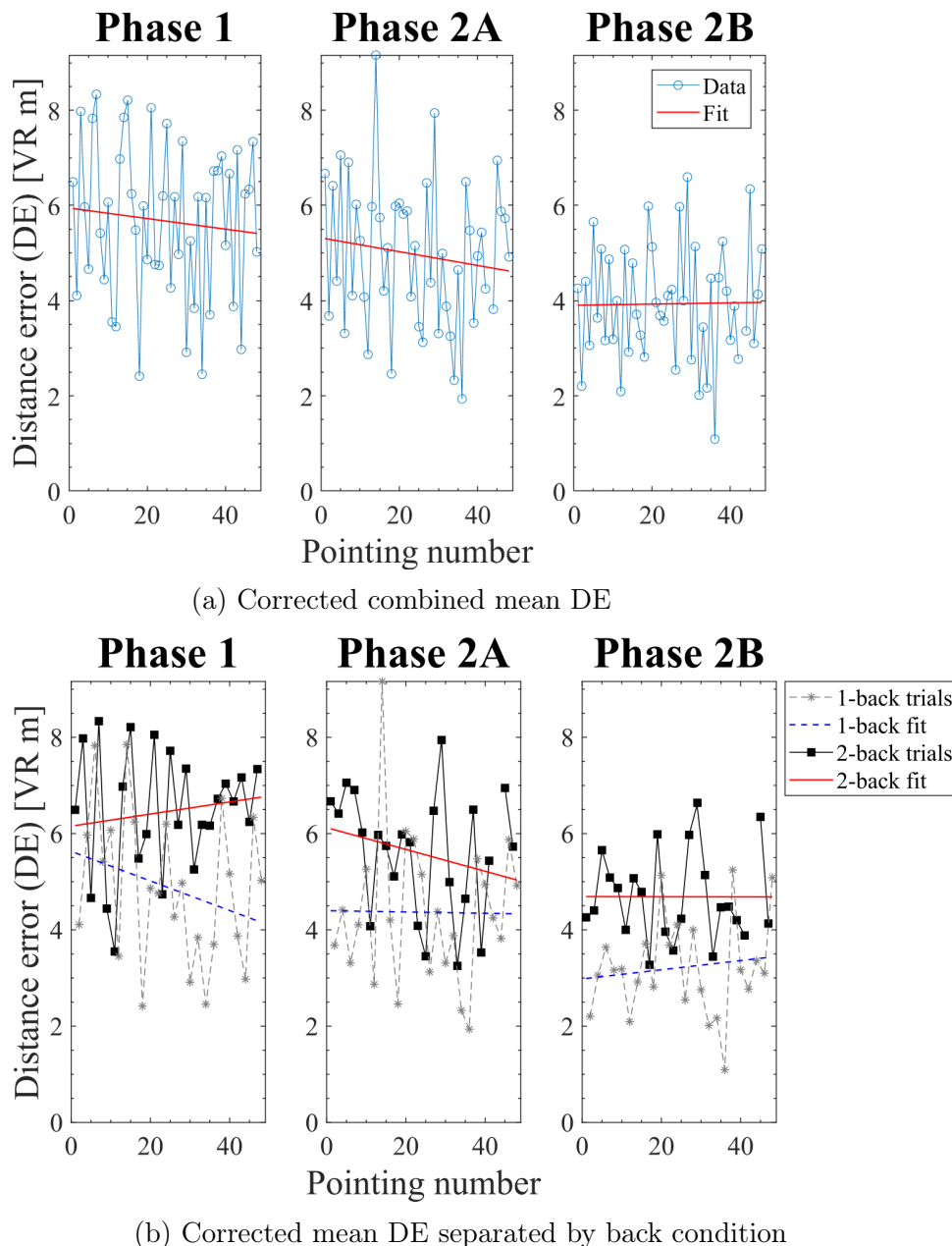


Figure 27: The mean distance error (DE) in VR meters across all participants is plotted for each of the 48 pointings (i.e. over the time course of the experiment). The data for Ph2A and Ph2B was corrected by excluding the pointings of the two target combinations which had evoked significantly higher errors. (a) Combined mean DE (blue) and the corresponding least squares fit (red) for each phase across all participants. (b) Separation of 1-back (grey) and 2-back (black) trials with corresponding lsq fits (blue and red).

Combined distance error: (Figure 27a) Correcting combined mean DE caused the negative slope of the fit in Ph2A to steepen ($m = -0.015$), whereas Ph2B saw a shift from a negative to a slightly positive slope ($m = 0.001$). Once again the highest DE was present in Ph1 (max. ~ 6 m, min. ~ 5.4 m), followed by Ph2A (max. ~ 5.4 m, min. ~ 4.6 m). Ph2B clearly showed the overall lowest mean DEs (fit max. ~ 4 m, min. ~ 3.9 m) despite barely any change in error over the time course of the experiment.

Separated distance error: (Figure 27b) The separation of the corrected DE into 1-back and 2-back trials showed that in Ph2A the slope of the 1-back trial fit had decreased, to a very slightly negative $m = -0.001$. Meanwhile the Ph2B condition only minimally flattened it's still positive slope to $m = 0.010$. The corrected 2-back trials in Ph2A saw a steepening of the learning curve ($m = -0.023$), but in Ph2B the slope increased to form a flat line, indicating no change of DE over the course of the experiment. Again the 2-back DEs were distinctly higher than those of the 1-back trials. The corrected combined and separated DEs both bore the highest mean DEes in Ph1, lower in Ph2A and the lowest DEs in Ph2B. A Bonferroni corrected repeated measure ANOVA confirmed a significant effect of both the phase: $F(2, 22) = 15.403$; $p < 0.05$; $\eta_p^2 = 0.583$) and back condition: $F(1, 11) = 75.078$; $p < 0.05$; $\eta_p^2 = 0.872$) on the mean DE. For both types of error there was no significant interaction effect between any of the phase and back conditions.

5.3 Residual deviation and estimated representations

To test whether the participants had an internal representation of the five targets which deviated from the actual locations, an estimate of the represented target positions U was calculated based on each participants reported target positions. This was calculated for a set of eight pointings, which was necessary for each target to have been pointed at at least once. This made an estimation of the participants represented target positions for all five targets possible. U was compared to the actual target positions and the deviation is reported in the d value in VR meters. This analysis was only done for Ph2, since the actual target locations did not change.

Residual deviation: The average d over each set of eight pointings throughout the experiment was calculated across all participants (Figure 28). During Ph2A d decreased from 0.93 m to 0.67 m over the course of the experiment (Figure 28), except for the value at pointings 17-24. The standard errors were fairly high in relation to the size of the values (0.17 m to 0.37 m). The steepest decrease of d was seen for the last set of eight pointings. In Ph2B overall the trend of a decreasing d with increasing pointing number was visible (0.87 m to 0.74 m). The standard error values here were slightly lower than in Ph2A, but were still fairly high (0.16 m to 0.25 m). The d value was fairly consistent in this phase, with small fluctuations, before decreasing in the last set of eight pointings. d was also calculated for each participants individually (Attachments Figures 46 and 47). During Ph2A d was reduced fairly consistently in six participants, during Ph2B in five people. A paired sample t-test for dependent samples was done to compare the first and last d value for each participant. It revealed a statically significant decrease of d over time for Ph2A ($t(10) = 2.215$; $p = 0.026$; Note: participant six was excluded from the statistical analysis because pointing numbers 47 and 48 were not recorded due to a technical error, this meant a d could not be calculated across pointings 41 - 48). In contrast Ph2B did not see a

significant decrease in d between the first and last eight pointings across all participants ($t(11) = 1.257$; $p = 0.117$).

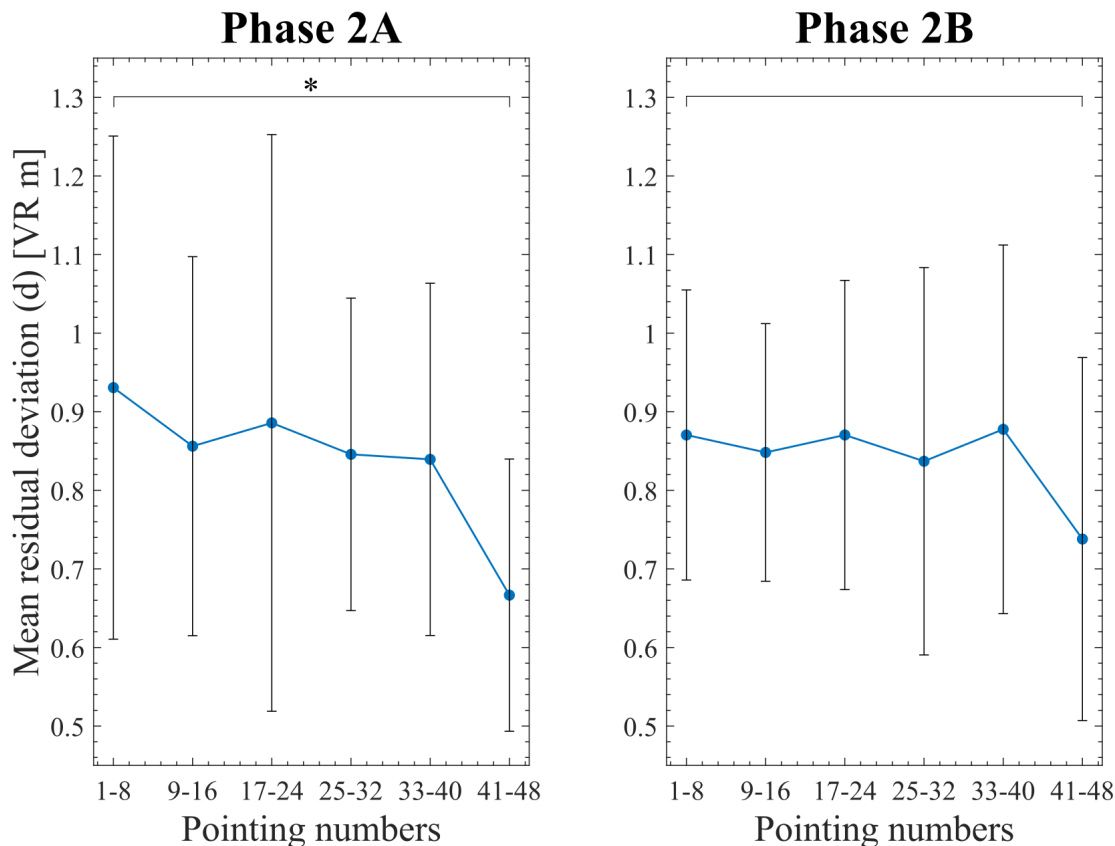


Figure 28: The mean residual deviation (d) between the actual target positions and the estimated represented target positions over the time course of the experiment, across all participants. One d was calculated for each set of eight consecutive pointings throughout the experiment, as eight pointing events uniquely described five estimated represented target positions.

Estimated representations: For a closer look at each participant's final estimated represented target locations, the U estimate, based on the last eight pointings of the experiment, is plotted alongside the true targets (Attachments Figures 48, 49 and 50). At the time of the last eight pointings the participants had spent about 15 (Ph2A) to 35 (Ph2B) minutes within the VR set-up of Ph2 and the deviation d had, on average, sunk to its lowest value. It can therefore be assumed, that if the participants were able to form a mental representation of the global target structure, they would have been able to do so at this point. For many of the participants the represented global structures appeared quite different in both structure and location from the actual global structure. However, some participants displayed a very interesting U . For example in some participants the global structure of their represented targets resembled the global structure and spatial relationships between the actual targets quite closely, albeit at a different magnification or with distortions (e.g. Figure 29a). Interestingly some participants also showed a global

structure which partially coincided with the euclidean structure of the actual targets or equivalent proportions between some of the targets, while other parts were geometrically inconsistent (e.g. Figure 29b). Some participants showed a basic understanding of the overall target structure but certain legs of the pentagon were over- or underrepresented. Analysing the U measure more in-depth was beyond the scope of this thesis, but it could be a very interesting measure to approach in future analyses and experiments.

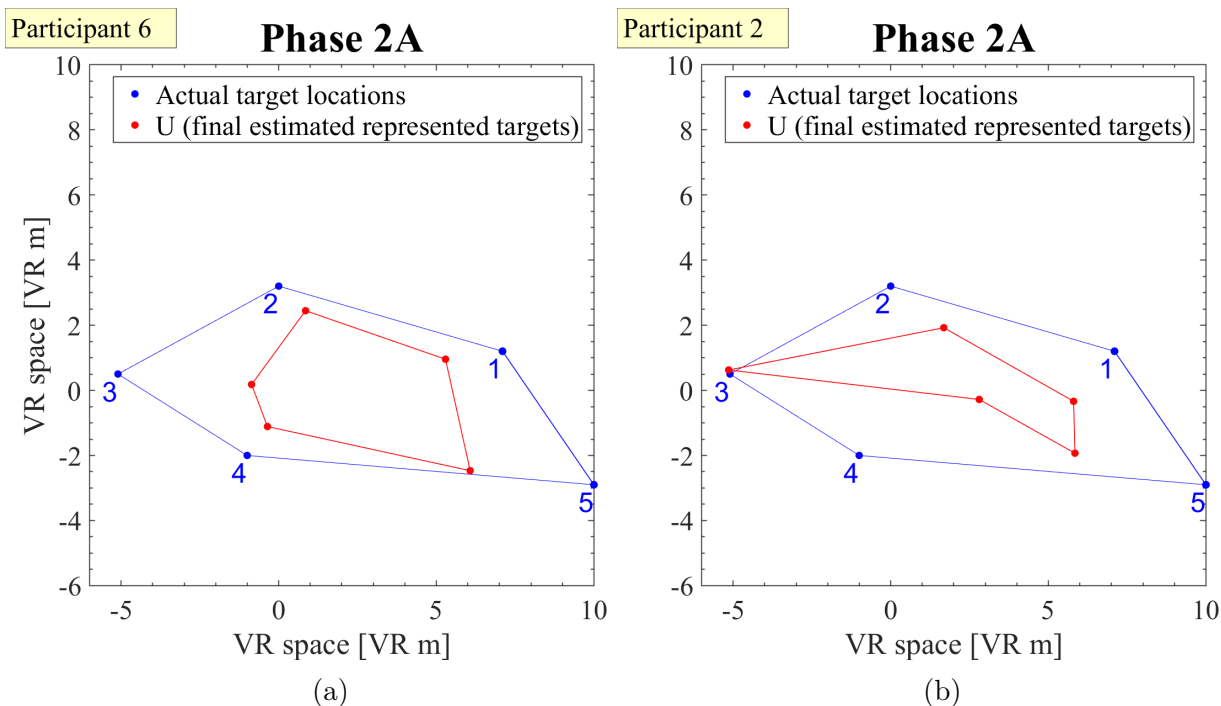


Figure 29: Examples of participant 2 and 6: The participants' final estimates represented target positions U (red) compared to the actual target positions (blue) in Ph2. This U is based on the participant's last eight pointings within the experiment, which uniquely describe the estimated represented five target positions in the participants memory.

5.4 Precision and accuracy

The precision (Prec) and accuracy (Acc) of each participant's reported target locations ("shots") across all five target was compared for the 24 first pointings (early) of the experiment and the 24 last pointings (late) (Table 5). The measure was first calculated in meters and transformed into a normalized value between 0 and 1 for better readability and comparability. A value of 1 indicates the highest achieved Acc or Prec among all participants. The measures were visualized for each participant by plotting a covariance error ellipse (SD error ellipse) around all the pointings directed at the same target number (example data for participant eight in Figure 30; Images for all participant under Attachments Figures 51, 52, 53 and 54).

Precision: On average the precision increased slightly during Ph2A (0.517 to 0.526; Table 5) with six participants showing an increased pointing Prec for the later pointings. In Ph2B this was also the case for seven participants, leading to a small increase of the mean Prec from early to late (0.637 to 0.677). The mean Prec was principally higher in Ph2B than Ph2A. A two-sample t-test for dependent samples comparing the early and late Prec revealed no significant differences for Ph2A ($t(11) = 0.593, p = 0.283$) or Ph2B ($t(11) = 0.141, p = 0.445$). These results were verified by a k related samples Friedman’s ANOVA (exact method). In both phases no significant increase of Prec from early to late pointings was found (Ph2A: $\chi^2(1) = 0.0, p = 1.0$ and Ph2B: $\chi^2(1) = 0.333, p = 0.774$).

Accuracy: Although the accuracy was also increased from early to late pointings in six participants both in Ph2A and Ph2B, the average Acc decreased from early to late pointings (Ph2A: 0.460 to 0.402 and Ph2B: 0.468 to 0.463). The Acc did not increase significantly from early to late pointings in either Ph2A ($t(11) = -0.663, p = 0.260$) or Ph2B ($t(11) = -0.865, p = 0.203$). A k related samples Friedman’s ANOVA (exact method) showed no significant difference in Acc from early to late pointings in either phase (Ph2A: $\chi^2(1) = 0.0, p = 1.0$ and Ph2B: $\chi^2(1) = 0.0, p = 1$). (Note: the statistical analyses were done on the exact Prec and Acc values in meters. The tests were also run on the absolute values, which found the same lack of statistical significance).

Table 5: Accuracy and precision of the reported target positions for each participant, separated by phase and early versus late pointings (the first 24 vs. the last 24 pointings within each phase). The normalized values between 0 and 1 are shown, whereby 1 indicates the highest achieved precision or accuracy across the five targets.

Participant no.	Precision				Accuracy			
	Phase 2A		Phase 2B		Phase 2A		Phase 2B	
	Early	Late	Early	Late	Early	Late	Early	Late
1	0.225	0.289	0.341	0.299	0.423	0.356	0.269	0.199
2	0.382	0.521	0.738	0.769	0.304	0.609	0.591	0.402
3	0.469	0.514	0.609	0.711	0.327	0.507	0.327	0.448
4	0.538	0.516	0.706	1.000	0.460	0.391	0.473	0.570
5	0.737	0.723	0.746	0.908	0.484	0.284	0.651	0.624
6	0.634	0.430	0.538	0.471	0.431	0.285	0.410	0.311
7	0.719	0.650	0.874	0.752	1.000	0.382	0.381	0.440
8	0.412	0.825	0.640	0.756	0.508	0.652	0.458	0.623
9	0.551	0.562	0.408	0.587	0.454	0.454	0.431	0.365
10	0.627	0.456	0.796	0.743	0.567	0.237	0.642	0.793
11	0.547	0.453	0.735	0.766	0.360	0.407	0.440	0.513
12	0.364	0.378	0.509	0.357	0.202	0.264	0.546	0.272
Mean	0.517	0.526	0.637	0.677	0.460	0.402	0.468	0.463

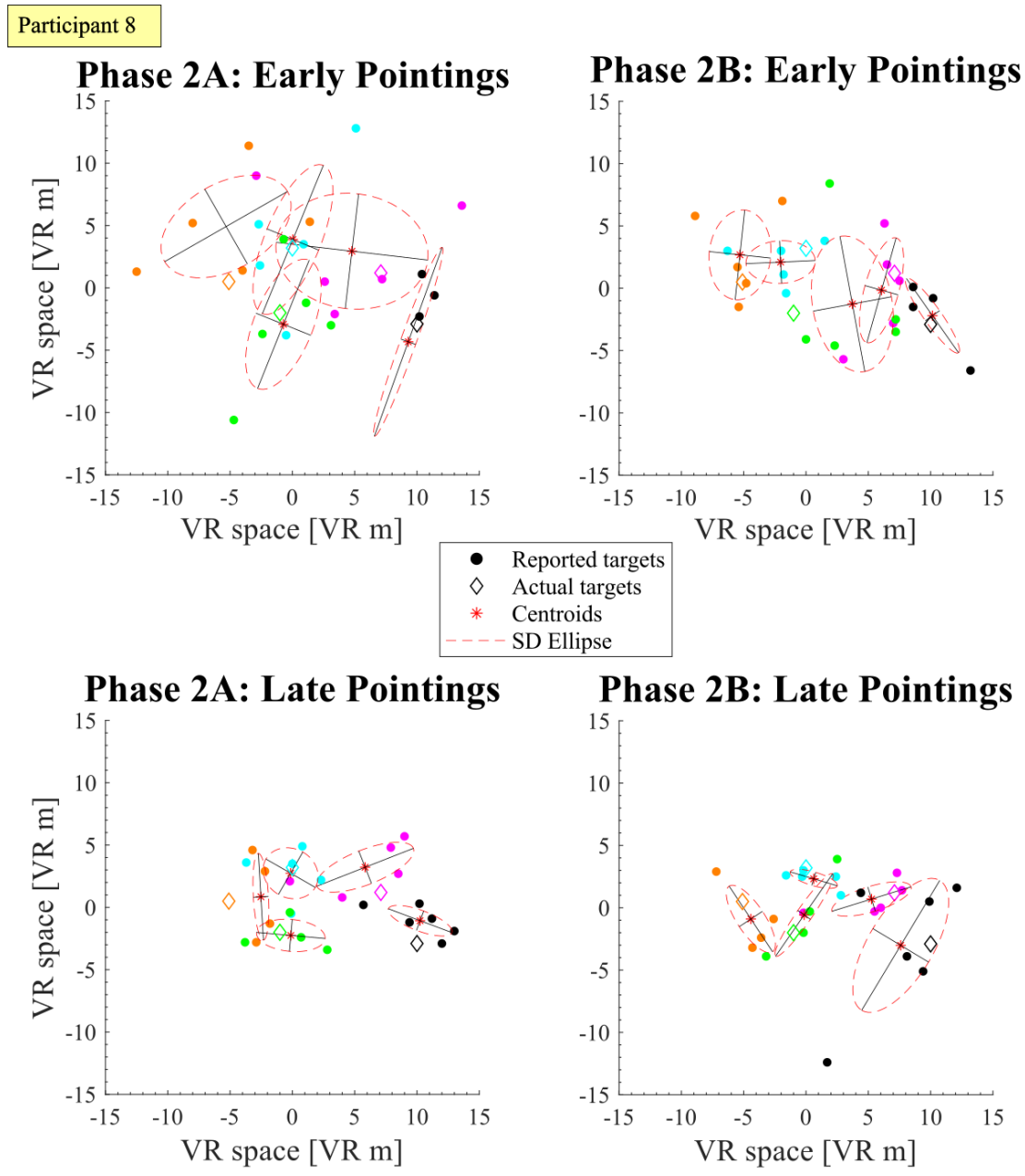


Figure 30: Example of participant eight: Visualization of the accuracy and precision of the first 24 (early) pointings versus the last 24 pointings (late) in Ph2A (left) and Ph2B (right). The covariance (SD) error ellipse is shown for all the pointings (colored dots) directed at the same true target location (colored diamond). "Shots" directed at the same target are indicated by the same color, as is the corresponding actual target location. The diameters of the error ellipse indicate the precision of the pointings. The distance of the error ellipse centroid (red star) to the corresponding actual target location shows the accuracy.

5.5 Sketched target locations

After both visit 1 and 2 participants were given a final questionnaire, which contained the request for them to draw what they believed the top-down view of the target locations looked like. They were asked to draw two separate images for Ph1 and Ph2 after the first visit and only an image for Ph2 after visit 2. To understand what was being asked of them, they were prompted by a small image of five randomly arranged dots. This prompt was the same for all participants and phases. Almost all participants expressed difficulties with the task and the results show mixed success in deduction of the global structure of the target locations (all sketches can be found under Attachments Figures 55 and 56).

In **phase 1** two participants responded with a drawing of multiple triangular target arrangements. Another participant also drew a triangle, symbolizing the connections between three targets. One participant simply wrote "completely random" as a response. Three responses of \geq seven targets were given and five of exactly five targets in varying formations.

For **phase 2A** five participants failed to recognize that there were only five targets presented, despite the set locations and billboards. Of these, the three participants that had reported triangles in Ph1 also did so in Ph2A. Two participants still reported an excess amount of targets, while one person reduced their response from seven to five targets. Overall a constellation of five targets was reported seven times, of which two were by people who had given a different response for Ph1.

After **phase 2B** interestingly five participants still did not report a constellation of five targets, despite the set locations, billboards, repetition of the same task and having completed the first drawings a week prior. Of these, three were done by participants who had never reported an arrangement of five targets before, but two were done by participants who had reported five targets in Ph2A. The remaining seven participants drew five targets in various arrangements. Two of these participants had not responded with a five target constellation in any drawing before. Two participants labeled their targets with the images presented on the billboards. Overall four participants drew five targets in both Ph2A and Ph2B, three only drew five targets in Ph2A, two only in Ph2B and three participants never completed a sketch containing exactly five targets.

Correlation with measures: The participants were sorted by their drawing responses, based on how often they sketched a constellation of five targets in Ph2A or Ph2B. Each individual's drawing response was correlated with the slopes of their learning curve for AEs, DEs in all phases (Figure 32). This was also done with their pointing precisions in Ph2A and Ph2B (Figure 31). The participants, that never understood the five target concept had less of a learning curve (higher positive slope) for DE and AE in Ph1 (Figure 32). Unexpectedly, their learning curve slopes in Ph2A and Ph2B were on average the same or even slightly lower than those of the participants that drew five targets in both

Ph2A and Ph2B. People, who had drawn five targets in only one of the phases, showed the highest, most positive slopes in both Ph2A and Ph2B. There was no recognizable, clear trend or correlation between the participants sketched target locations and their normalized precision (early and late separately) in Ph2A (Figure 31). But in Ph2B participants who had created sketches of five targets for both phases had an overall higher precision than the other categories. Yet, a low R^2 indicated that the difference in understanding (and reporting) five targets only partially accounts for the variability in the data.

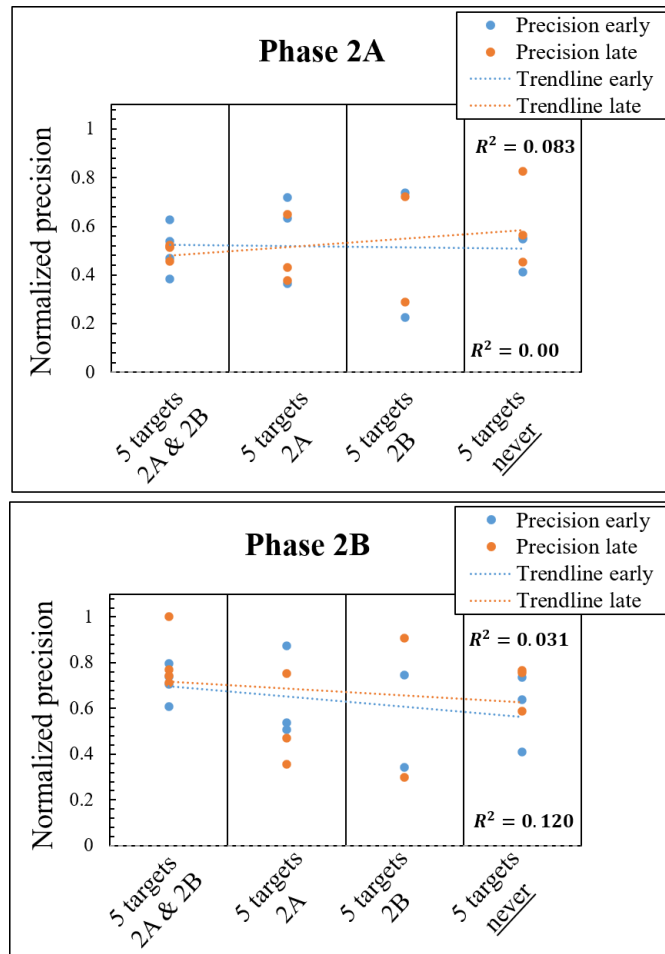


Figure 31: Each participant was sorted into one of four categories, dependent on how often they sketched a constellation of exactly five targets for Ph2A and Ph2B in their post-visit questionnaires. Every individual's normalized precision (Table 5) was plotted within the appropriate category for each of the phases. This was done for the first 24 pointings (early; blue) and for the last 24 pointings (late; orange) of the experiment.

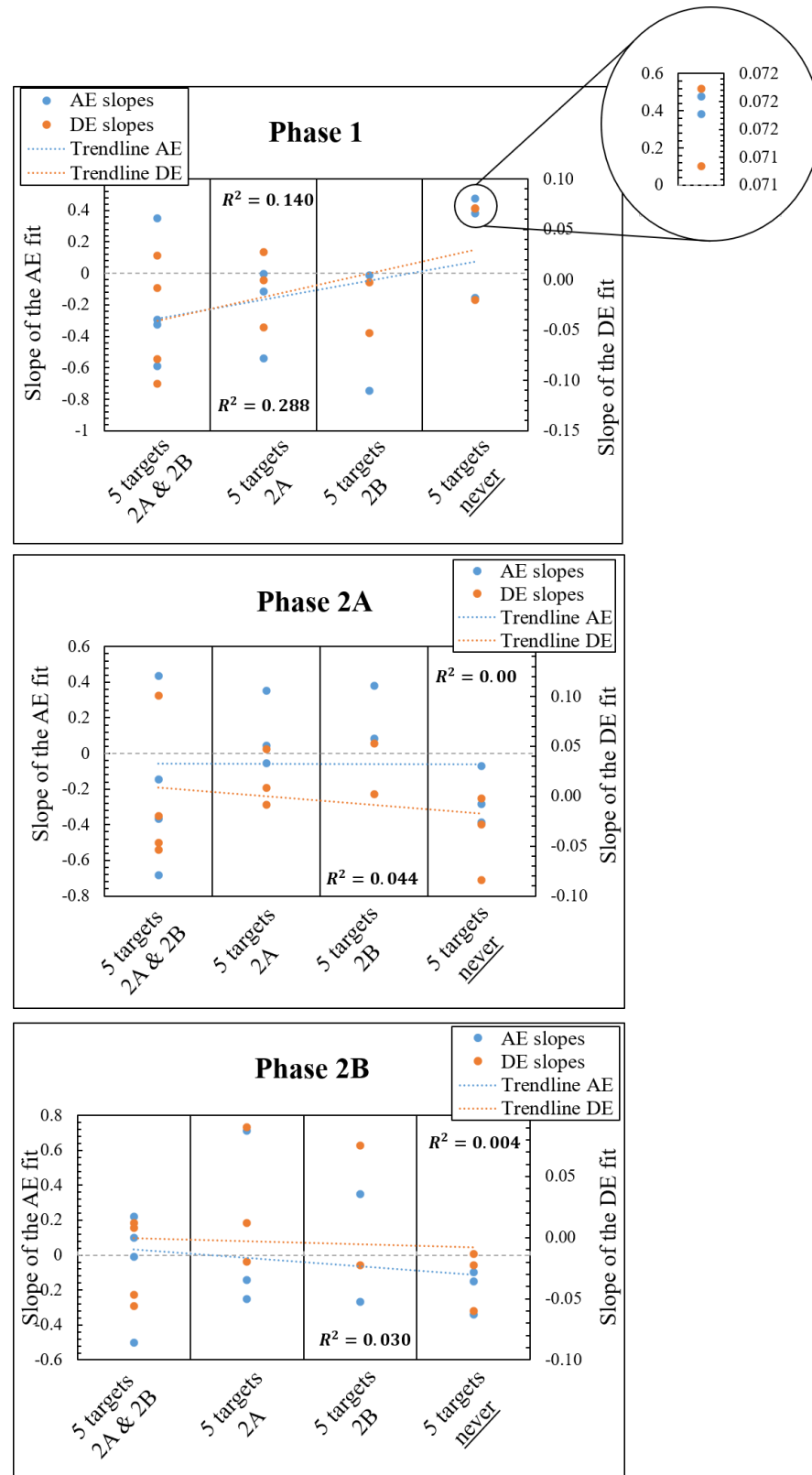


Figure 32: Each participant was sorted into one of four categories, dependent on how often they sketched a constellation of exactly five targets for Ph2A and Ph2B in their post-visit questionnaires. The slope of every participant's combine angle error lsq fit (blue) and combined distance error lsq fit (orange) are plotted over the sketch response. This was done separately for Ph1, Ph2A and Ph2B.

6 Discussion

6.1 Calibration

During calibration a strong correlation was found for decrease in angle error (AE) with increasing distance to the target (Figure 19a). This observation was to be expected, since the same deviation (y_1) from the target leads to a larger angle error (θ_2) at closer distance (x_1) than at a farther distance (x_2) (Figure 33b):

$$\tan(\theta_1) = \frac{y_1}{x_1} \text{ and } \tan(\theta_2) = \frac{y_1}{x_2} \text{ lead to } \tan(\theta_2) = x_1 \tan(\theta_1) \frac{1}{x_2} \Rightarrow \frac{x_1}{x_2} = \frac{\tan(\theta_2)}{\tan(\theta_1)}$$

The opposite is true for the distance error (DE) (Figure 19b). Here an increase of DE was found to correlate with increasing pointing distance. This was to be expected, because for the same angular deviation (θ), a larger distance (x_2) leads to a larger deviation (y) than at a smaller distance (x_1) (Figure 33a):

$$\tan(\theta) = \frac{y_1}{x_1} = \frac{y_2}{x_2} \Rightarrow \frac{x_1}{x_2} = \frac{y_1}{y_2}$$

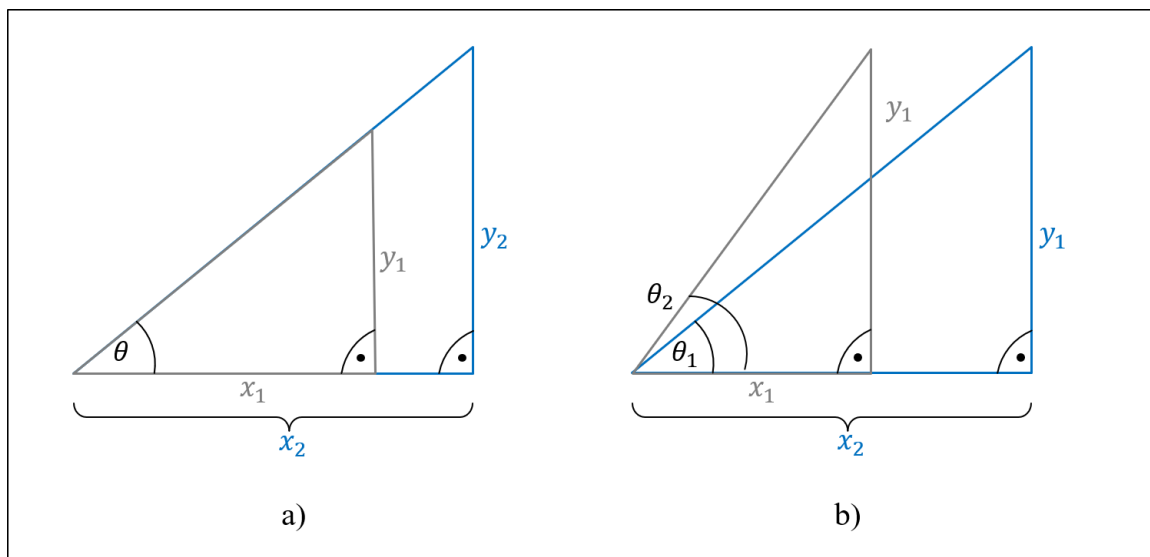


Figure 33: Visualizations of the calculations for a) distance (x) dependent distance error (y), and b) distance (x) dependent angle error (θ).

These factors are nearly impossible to avoid when working with varying target distances. The calibration task was a useful tool in diminishing these biases in the later phases of the experiment. The distances in the calibration task were matched to the distances used during Ph2. This allowed for the derivation of an individual calibration error, not only for each person, but for each of the pointing distances. In turn this allowed for AE and DE in Ph1 and Ph2 to be more independent of distance bias, because a direct removal of the corresponding calibration error was possible. In Ph1 this idea met its limitations, because the target locations varied with a random jitter of ± 5 VR meters. Although a large range of distances were covered during calibration, the distances $<$ than the shortest target distance in Ph2 (4.8 m), and $>$ than the largest target distance (15.4 m) were not

represented during calibration. In Ph1 the calibration errors of the calibration distances were used, which were closest to the actual distances. Because of this approximate correction and the fact, that the measured calibration errors were overall very small, there is no indication of an underlying distance bias of the data in Ph1. Nevertheless, for future studies it is advised to create a wider range of calibration target distances, which could account for more of the various target distances in Ph1. Overall participants were very accurate in their calibration pointings, displaying only small errors ($DE_{max} = 1.02\ m$ and $AE_{max} = 1.76^\circ$). The actual and reported target locations and the head positions were all tracked and reported from within the VR space, eliminating a disconnect between real-world and VR coordinates. The calibration data thus conclusively shows that there appear to be no underlying issues with pointing, measurements or tracking.

6.2 Verification analyses

No influence of target order on either AE or DE was found. Likewise, AE was not affected by which target was being pointed to (Figure 20a). The DE was significantly influenced by the goal target (Figure 20b) however. This result was surprising, because the standard errors between the mean DEs for each target number overlapped by a large margin. The largest difference between highest and lowest mean DE of the targets within the same phase and back condition was $2.634\ m$ (Ph2A, 1-back). Therefore it is believed that this effect was not large enough to significantly bias or impact the DE data. Moreover, when comparing the DE for each combination of origin and goal target (Figure 21b), the only two target combinations that elicited a significantly higher error than others were 1-5 and 3-5. These pointings most likely lead to the previously discussed statistical significance of target number on DE, since both are directed at target number 5. The reason for the increase of DE in the two combinations can be explained by the configuration of the targets in Ph2 (Figure 15 and 16). The distance between the targets 3 and 5 was the largest possible pointing distance and was therefore systematically either underestimated or over-corrected for. Interestingly, the opposite pointings, from target 5 to target 3, did not elicit the same height of DE, although they did also evoke a higher DE than most other combinations. The other origin-goal target combination, which had lead to a higher mean DE, was pointing from target 1 to target 5. This combination also caused a higher AE, which can be explained by the fact, that pointing from 1 to 5 required the largest turning radius and thus pointing angle. This means participants most likely underestimated the turn angle. They may have also over-corrected as a response to feedback in earlier pointings. Target 5 was not farther away from target 1 in distance, but it was located at a more obtuse angle than it would have been, if the pentagon was an equilateral pentagon. Assuming participants built an inner representation of the target positions, this representation may have been biased towards an equilateral pentagon. In

that case a smaller turning angle would have been sufficient. The high angle deviation for 1-5 pointings also explains why it elicited a high distance error, as a misjudged turning angle also causes a larger distance deviation between the reported actual target location. The higher AE and DE caused by the two target combinations may have had an effect on the principal analysis of the DE and AE. Therefore, the data was subsequently analyzed both without this consideration, as well as with. The difference between these analyses will be discussed in a later chapter.

Whether a target was pointed to as part of a 1-back or 2-back task had a significant influence on both AE and DE. This was not surprising and will be re-visited and discussed in more detail in the follow chapter.

6.3 Over- and undershooting

Philbeck and Loomis (1997) presented participants with targets that were only visible for a short period of time, participants were then asked to report the target locations (in their case, verbally or by walking to them). They found, that participants tended to overestimate the distance for near targets, and underestimate the distance of far targets. This effect is called the compressed average effect. This effect was also found within the data of this experiment for all participants, phases and back condition (only exception was participant one, 2-back trials in Ph1) (Figure 23). This can be seen in the negative correlation coefficient R between the true target distance and the signed distance difference (DD) (distance from head to reported target minus distance from head to actual target) (Table 1). This effect has been found to be amplified in VR studies, compared to real-world paradigms (Loomis et al., 1992; Grechkin et al., 2010; Finnegan et al., 2016). The present study did not make changes to the overall configuration of the targets based on this effect. Future studies may opt for a correction of this compression effect to improve participants' distance estimations, by intentionally rendering the visual components of objects incongruently to one another, as proposed by Finnegan et al. (2016). This was shown to improve VR distance estimations, at least for long-distance targets. However, in this experiment a statistical analysis showed, that there was no influence of either phase or back condition on R , proving it a stable effect across all phases. Since the effect had the same influence on the distance error in all phases a comparison between the stages was possible and no undue influence of the effect was to be expected.

6.4 Angle error and distance error

The first goal of the experiment was to establish whether metric information from the iterated triangle completion task could verifiably be transferred into LTM. If no metric information was transferred into LTM during the task, we would expect no difference (lack

of learning curve) in AE or DE between Ph1, Ph2A and Ph2B, beyond a learning curve for the early pointings at the beginning of Ph1 (It is to be noted, that participants had already undergone a short training before Ph1. All participants showed full understanding of the task and opted to forgo additional training). In contrast, it was suggested that the global structure of the targets in Ph2 could allow for the transfer of metric information into LTM, which could lead to the build-up of a mental map or representation of the target structure. If the metric embedding of this mental map was Euclidean, it would in turn be expected to increase performance (i.e. a lower AE and DE) in Ph2A and Ph2B compared to Ph1.

6.4.1 Angle error across participants

Uncorrected: The combined and separated AE of the 2-back tasks, averaged across all participants (Figure 24), showed a consistent decrease of errors over time in each phase. Besides this learning curve in all three phases, the mean AE also significantly decreased from Ph1 to Ph2B in phases, which is particularly interesting. The influence of both the phases and back conditions on AE were found to be statistically significant. These findings are in line with the expected outcome, of the assumption, that metric information is transferred into LTM during Ph2. This was particularly visible in the lsq fit slope of the AEs in the 2-back trials, where early Ph1 and Ph2A showed roughly the same degree of AE, but the steeper learning curve in Ph2A lead to a lower AE in late Ph2A than in Ph1. This lower AE and steeper learning curve in Ph2A are in line with the postulated results of a metric information transfer. Measurements of Ph2B took place one week later. If there was no transfer of metric information into LTM, there would be no reason for the triangle completion performance to yield any lower errors in Ph2B than in Ph1 or Ph2A. One might expect less of a starting error than in Ph1, since the participants would be more used to the VR environment and task. But this would not account for the significantly lower mean AE in Ph2B that was measured, compared to Ph2A, both in the combined and separated mean AE. These finding seem to speak for the idea, that some form of metric information is transferred into LTM. On the other side, it could be argued that the decrease in AE from one phase to the next could merely be based on increasingly better understanding of the task and environment. This explanation would require the participants to still be in the learning phase after up to 40 minutes post experiment start, since the decrease of AE continued throughout Ph2A. participants had completed the first visit after about 1 to maximally 1.5 hours. This time span is very representative of the average time used in previous VR triangle completion tasks (e.g. 1 h in Riecke et al. (2002) and 1.5 h in Loomis et al. (1993)). However, it is possible that the task or the environment were particularly difficult for the participants to adjust to, which could account for the slopes seen in Ph1 and Ph2A. The overall lower slope in Ph2B does pose

a challenge to this line of reason though, as the significantly lower mean AE indicates a lasting effect of the learned information on performance. The lower mean AE in Ph2B, even seen at the beginning of Ph2B, suggests that information about the metric of the set-up was accessible to the participants, even one week after the first visit. The lsq fit slope of the 1-back trial AEs also showed a learning curve in Ph1 and a decrease of mean AE from Ph1 to Ph2B, which is in line with what was just discussed. Not so in Ph2A and Ph2B. The reason for this different behavior between the 1-back and 2-back conditions will be discussed later on.

Corrected: The corrected AE across all participants (Figure 26) showed the same principle behavior as the uncorrected mean AE. The effect of both phase, and back condition on AE were significant. The slopes of the lsq fits of the combined AE, and of both back conditions in Ph2A were steeper after correction, indicating even more of a learning curve, than previously described. Since the significantly higher errors were removed in this measure, it can be assumed that this curve is closer to the unbiased learning curve across participants. Again, the lower AE and steeper decrease in Ph2A compared to Ph1, are in line with the postulated results of the assumption, that metric information about the target structure is transferred into LTM. The correction had less of an effect on mean AEs of the 1-back trials, mainly because only $\frac{1}{3}$ of the corrected pointings were 1-back trials. For Ph2B the correction surprisingly revealed a slight increase of mean AE over time, for the combined and both back conditions. Yet, in relation to the mean AE of Ph1 and Ph2A the AE of Ph2B is still overall significantly lower.

6.4.2 Distance error across participants

Uncorrected: The mean combined and separated DEs across participants (Figure 25) showed the same relationship between phases as the AE. Overall the DE was highest in Ph1, slightly lower in Ph2A and lowest in Ph2B. The effect of the phase on DE was statistically significant, and especially the significantly lower mean DE in Ph2B is notable. This observation hints at a possible recall of metric information about the set-up from LTM in Ph2B, and matches the postulation, that transfer of metric information took place. This would explain the slight increase in performance (i.e. decrease in DE), especially from Ph1 to Ph2B. This idea is not as strongly supported by the slopes of the lsq fits of the DEs. A learning curve was measured in the fit of Ph1 in the combined DE and 1-back trials. But in the 2-back trials, an increase of the DE over time was calculated. This is a highly unexpected finding, and can not be conclusively explained, other than by increasing tiredness and fatigue of the participants throughout the phase. In Ph2A all DE fit slopes are quite close to a zero line. This actually speaks against the idea that metric information about the target layout is integrated and accessed during the task. It speaks more for a classic triangle completion task in which the presence of a set and

global target structure was irrelevant to the task performance, which had leveled off at maximal triangle completion ability after Ph1. The slope in Ph2B is only slightly steeper than in Ph2A, but it is striking, that the overall DE in Ph2B was significantly lower than in Ph1 and Ph2A. These two examinations appear to be contradictory, but correction of the data helped alleviate this discrepancy.

Corrected: Exclusion of the pointings, that had elicited a significantly higher DE yielded an interestingly altered mean DE across participants for Ph2A and Ph2B (Figure 27). Similarly to the AE, the correction revealed a steeper slope of all three data fits for Ph2A. For Ph2B the correction did not cause much difference in either phase. These findings may be due to continued learning of the task and environment from Ph1 to Ph2A and a reaching of maximal familiarity in Ph2B. But, as discussed, this would require participants to have had a very long learning time. Also, especially the combined mean DE coincided strongly with the postulated results for the case, that metric information was transferred into LTM. In this vein, the results can also be interpreted as participants increasing their performance (i.e. reducing their DE) in Ph2A, because of a feedback-loop of the learned metric information about the target location between LTM and hippocampus. Upon returning to the lab, participants could then have used the metric information stored in their LTM, to maintain a lower DE over time, which would also explain the overall lower DE in Ph2B. The slopes of the 1-back fits were barely affected by the fit. This was expected, since within 1-back trials, only 1 or 2 pointings per person and phase were affected by the corrections (depending on the assigned target order).

Summarizing the implications of the findings on the mean AE and DE, especially the AE data and the corrected DE data direct towards the idea, that egocentric information from within the path integration task is transferred into LTM, where it is conformed into a mental map, or mental representation of the targets. This would explain why there was a significant effect of the phase on performance in both measures. This also accounts for the significantly lower errors in Ph2B in both measures, which provides indication for metric information retention. This idea is in line with the current opinion in research, which sees path integration as an important input system into the formation of mental maps (Warren, 2019; Wang, 2016; Chrastil and Warren, 2014). It must be mentioned that participants were asked to draw the global structure of the targets after their first visit. It is possible that this drawing task may have influenced performance and strategy of some participants in Ph2B. This possibility will be discussed in more detail later on. It is noteworthy though, that neither of the measures show a large error at the beginning of Ph2B, which would speak against the argument that participants switched strategies/focus from Ph2A to Ph2B, since that would most likely require a certain additional learning/adjustment phase. At this point a strict differentiation must be made. While the data does lend itself to the idea of building up a mental map in LTM, it does

not offer enough indication to conclusively conclude that this representation is embedded in a Euclidean metric or reference frame. This is particularly evident when comparing the average AE of this experiment to those of studies with a simple triangle completion task. Klatzky et al. (1998) also presented participants with a triangle completion paradigm in VR, in which participants were presented with only virtual optical flow, but underwent real turning. He measured an average angle error of 20° . This angle is roughly in line with the average AEs achieved in Ph2B for the combined mean AE 23° , and for the 1-back trials in Ph2A and Ph2B. Higher errors were measured in the other phases and condition. If there was a transfer and feedback of strictly Euclidean metric information about the target structure into and from LTM, we would have expected errors to be lower than in a classic triangle completion paradigm, which only relies on WM.

6.4.3 Individual angle error:

Switching gears from looking at the errors across all participants to looking at each individual participant (Table 2, Figures 38, 39, 40 and 41) unveiled a surprising discrepancy between the cumulative data and the individual's performance. In each back condition of Ph1 four participants failed to show a learning curve. This is particularly glaring, as Ph1 was intended to function as a means of assessing participants baseline triangle completion ability. The lack of learning curve here would indicate a lack of triangle completion ability and thus path integration ability. It is also possible that these participants struggled with the unfamiliar VR setting or the lack of proprioceptive feedback. Intriguingly, two of the participants that lacked a learning curve in Ph1 did display learning curves for Ph2A and Ph2B. It is difficult to conclusively state causation for this observation, but it is possible, that the participants in question heavily rely on mental maps or landmarks for navigation in the real-world and struggled to navigate without either in Ph1. These participants would therefore particularly benefit from the global structure present in Ph2, which would explain the learning curves. In combined and separated AE only three to four participants showed a steeper learning curve in Ph2A than in Ph1. This could be explained by a lack of metric information transfer. In this, each participant's performance variations was due to learning of the task and VR environment, individual WM capacity and engagement and in- and decrease of concentration, or other individual factors. The data of the individual participants quite clearly disputed the hypothesis, that a mental map was formed which was embedded in a Euclidean global metric. Alternatively, as was indicated in the AE across participants, metric information was transferred to LTM to form a mental representation of the target structure. However, the metric embedding of this map was not Euclidean in nature, and therefore did not lead to a uniform reduction of AE and DE in participant.

6.4.4 Individual distance error:

The individual participants' DEs were similarly high in variation as the AE (Table 3, Figures 42, 43, 44 and 45). In the combined DE five participants failed at achieving a learning curve in each phase, the same was true for up to six participants in Ph1 and Ph2A in the separated DE. The participants, that did not decrease DE within Ph1 showed mixed fit behavior in Ph2A and Ph2B, with most participants sporting a positive slope in one of the phases and a negative one in the other. In each phase only five participants displayed a steeper learning curve in Ph2A compared to Ph1. This highly variable data clearly shows that participants did not use one unified, global reference frame to complete the task. It is possible that some participants were not able to extract the necessary information about the global structure from the target locations in Ph2. This would mean that no local metric was transferred from WM into LTM. Again the alternative is possible as well, in which participants represented the targets in a metrically inconsistent way, in which case the learning of a mental map would not be strictly associated with a decrease in DE.

The data on the level of individual participants can be interpreted to indicate, that no systematic transfer of metric information of the triangle completion task into LTM has taken place. In this case individual performance would be strictly tied to personal triangle completion ability, WM capacity etc. However, this explanation does not account for the statistically significant effect of phase on the Errors, as well as for the lower errors in Ph2B compared to Ph2A. Here the alternate explanation gains traction, as it can explain both findings. The labeled graph hypothesis suggests, that a transfer of Euclidean metric information via hippocampus into LTM does take place, where it is saved as a non-Euclidean mental map/representation of the target layout. This non-Euclidean spatial information could still feedback into the triangle completion task. This would account for the statistically significant decrease in AE and DE across participants as an overall increase in navigation ability (i.e. performance) across participants, but a lack of metrically consistent improvement in the individual. Additionally, Steck and Mallot (2000a) have found, that knowledge of a global configuration did not necessarily improve metric performance in VR navigation. The measures of AE and DE are quite strictly tied to testing for the presence of a Euclidean metric in LTM. The data from this experiment, as well as previous research in the same paradigm indicate that there is no transfer of strictly Euclidean metric into LTM (Hopf, 2008; Warren, 2019). Additional measures were investigated to further analyze the presence of such a mental map and the metric of this map.

6.4.5 Back condition

Another interesting observation was made in the separation of 1-back and 2-back trials. **AE and DE across participants:** Especially in the AE and DE across participants

(Figures 24b and 25b), it was evident that 1-back trial errors did not follow the learning curves, which were seen in the 2-back tasks. In Ph1 both back conditions' fit slopes behaved the same, as the AE and DE across participants saw a decrease in 1-back and 2-back errors. This is not surprising during the phase, in which the participants are adjusting to the task and environment. Both in AE and DE in Ph2A and Ph2B, the mean errors are lower compared to Ph1, but the slopes are slightly positive. The increase in error within the phases is unexpected, but it is most likely attributable to tiring and loss of concentration, which was reported by some participants. The lower errors in 1-back trials were to be predictable and can be explained by the smaller demand on WM (one distance and angle for 1-back, versus 2 angles and three distances in 2-back). During 2-back tasks updating of the homing vector required multiple additions, resulting in a higher error rate. To ensure 1-back pointings did not always rely on a turning angle of 180° , the 2-back pointings were queried first. Nevertheless, during 2-back trials participants were actually fulfilling the triangle completion task, which entails learning the appropriate angles and distances to complete a triangle. Pertaining back to the idea of metric information transfer, 2-back is the back condition in which this transfer is thought to take place and thus the condition in which actual learning about the spatial layout is happening. For the 1-back condition, it seems that participants reached a maximum in their ability to remember their approach direction in Ph1, which was then held for Ph2A and Ph2B, with a small increase of error due to fatigue. If a Euclidean metric map of the targets has been formed on the basis of the metric information from the 2-back trials, it would be expected that the underlying mechanism in 1-back trials may switch to draw from the map in LTM as well. This appears to not be the case, which could either mean, that there was no transfer of Euclidean or any type of metric into LTM. Alternatively, the 1-back tasks, with a much lower demand on WM, may underlie a different mechanism that does not draw metric information from LTM but instead relies solely on egocentric information from WM.

Individual AE and DE: In all participants, AE and DE in 1-back trials were generally lower than in 2-back trials, in all phases (Tables 2 and 3) (Although in ten participants the 1-back and 2-back fit slopes intersected in at least one of the phases). The difference in learning curve between 1-back and 2-back trials falls in line with the previously described complicated and highly variable AE and DE for each participant. There are some participants who show relatively similar behavior of the 1-back and 2-back AE and DE fits. But most participants show the described different slopes of the conditions. Despite the variability in the lsq fit slopes of the individual's 1-back and 2-back trials, across all participants the effect of the back condition on AE and DE was statistically significant.

6.5 Residual deviation and estimated representations

In the previous chapter it was postulated, that participants may have formed an internal representation of the five targets in Ph2, which deviated from the actual target locations, as it did not underlie the Euclidean metric. To put this theory to test, an estimate of represented target positions was calculated, based on each participant's reported target positions. It was calculated for every set of eight pointings within the experiment, which uniquely described the five represented target positions (i.e. made an estimation of the participant's represented positions for all five targets possible). The comparison of this estimate U with actual target positions was reported as deviation value d in VR meters. Under the assumption, that the metric within LTM is Euclidean, we would expect d to reduce over time, since the represented target locations would come closer to the actual target locations. The absence of a decrease in d would indicate either of two things: 1. The participants did not represent the target locations in a mental map. 2. The metric embedding of the mental representation of the targets in LTM is non-Euclidean.

Residual deviation: The d across participants was shown to decrease over time (Figure 28). Within Ph2A the decrease of d between the first and last eight pointings of each individual was statistically significant. Separate analysis of each participant's d over time did not reveal this clear reduction mirrored in the individual participant (Figure 46 and 47). 50% of the participants reduced their d over time in Ph2A and 41.67% in Ph2B. Although the mean d across participants in Ph2A hinted at the presence of a Euclidean-like presentation, the data of the individual participants does not confirm this. The theory of a Euclidean embedding of mental maps is inadequate in explaining the inconsistencies and deviations in d among participants. These mixed findings could be seen as evidence, that no metric information was transferred to form a LTM representation. In this, the reduction of d in Ph2A may be explained by continued learning of the task and new environment, resulting in better performance in triangle completion for later pointings. This explanation once again relies on the assumption, that participants had still not fully learned the task and set-up after 25 to 30 minutes into the experiment. The statistically significant decrease of d in Ph2A is not fully explained by this interpretation. Instead, the reduction speaks for the idea, that metric information was transferred, however this information does not appear to be strictly Euclidean. The alternate assumption, of a local, non-Euclidean metric in LTM, would account for the overall decrease in d , as Ph2A is the phase in which this mental representation is built-up. Meaning, in the first pointings the mental representation is still far from the real target formation, since each target has only been pointed to once. Throughout Ph2A the representation is approximated to represent the structure. At the same time, possible metric inconsistencies within representation account for the variability within participants. This would also explain the relatively flat d curve in Ph2B. At this time the mental map and representation had been formed

and responses were given in line with this representation, resulting in relatively small changes in d . Nonetheless, d is an indirect measure of a Euclidean metric based mental representation. It can therefore be used to strengthen doubts about a strictly Euclidean-like metric in LTM, but can not conclusively demonstrate an alternate metric.

Estimated representation: After review of the data gave interesting indication of an underlying mental map of alternate metric, each individual participant's estimated mental representation of the target locations (U) was plotted (Figures 29, 48, 49, 50). However, a deeper analysis of the measure was beyond the scope of this experiment. It does pose a very interesting measure to be investigated more in-depth in future research though. To allow for enough time for each participant to have formed a mental map, only the last eight pointings within Ph2A and Ph2B were investigated. In multiple participants U formed a global structure that was very different from the actual global structure, both in structure and location. This could speak for the notion that at least some of the participants were not able to form an adequate mental representation. On the other hand, some participants' U stood in very interesting relation to the real structure. None of the U s consistently matched the true targets in a Euclidean sense, but some followed the global shape with certain distortions or magnifications. There were also cases of a combination of Euclidean-like and non-Euclidean location or distance representations within the same U . These findings support the assumption, that at least some of the participants were able to form a mental representation of the targets, and thus, that a form of metric information was transferred into LTM from the triangle completion task. They do not display a Euclidean embedding of the LTM representation. The results could speak for the labelled graph hypothesis, since there seems to be no embedding in a global coordinate system. Rather, a network of paths between places, augmented with local metric information, allowing for geometrical inconsistencies (Warren, 2019; Chrastil and Warren, 2014). Altogether, despite the indications, no unambiguous conclusion on the "true" underlying metric of LTM could be made based on the measure.

6.6 Precision and accuracy

Precision (Prec) and accuracy (Acc) of each participant's early pointings (number 1 to 24) and late pointings (number 25 to 48) were compared for Ph2. The Acc measure allows for an estimation, whether the participants were able to better match their pointings to the Euclidean positions of the targets with time (\equiv increased Acc over time). Assessing Prec gives indication of whether participants "narrowed in" on a specific location over time (\equiv increased Prec over time).

On average the precision of late versus early pointings increased slightly for Ph2A and Ph2B (Table 5). This increase was quite small and not statistically significant in either phase. Overall, nine participants increased their Prec from early to late pointings in at

least one of the phases. The accuracy values followed the opposite trend, with a decrease in both phases across all participants from early to late pointings. The decrease was not statistically significant, with only a small average decrease. These findings are interesting, because they indicate that on average participants did not increase the proximity of their reported targets to the actual targets. The overall statistical insignificance of the changes in both cases could indicate a lack of metric transfer from the triangle completion task and therefore lack of mental representation. On the other hand, individual participants sported an increase in Prec of up to 41.3% and of Acc by up to 30.5%. This could mean that at least some of the participants were able to transfer metric information into LTM. In the measures discussed in previous chapters, there was evidence that such a transfer appears to have taken place. So, assuming metric information had been transferred into LTM, the lack of increase in Acc certainly speaks against the underlying metric of this information being Euclidean. Prec is independent of the actual target locations and Acc. Instead, it evaluates the diameter of the error ellipse around the pointings for each targets. The increase in precision, although small, is an interesting discovery, especially in combination with the lack of increase in Acc. An increase in both measures could have been explained by a basic increase in triangle completion ability, due to familiarity with the experiment or VR set-up. But with an increase in Prec, in combination with a decrease in Acc it could be argued, that over time the participants were able to build-up a mental representation of the target locations, to which they adjusted their aim. This would mean an increased precision of shots to the participants individual represented target locations, leading to increase in Prec. Since there was no increase in Acc, the represented target locations metrically deviate from the true locations. It is to be noted, that in this analysis 24 pointings were bundled for each early and late pointings. In this, the values of the "middle" pointings were also divided into either of the categories. This might have lead to an inaccurate or at least rough representation of the development of Prec and Acc over time.

6.7 Sketched target locations

In the post-visit questionnaire participants were asked to draw what they believed the target locations looked like from a birds-eye view. Herein participants were asked to draw what they thought the pole positions looked like from an overhead (bird's-eye) view in the phases they had just completed. The images for Ph1 and Ph2A were drawn at the same time, after the first visit. Their responses were sorted into four categories, depending on, in which phases they had drawn a constellation consisting of exactly five targets (Ph2A & Ph2B, only in Ph2A, only in Ph2B, never). To give an idea of what kind of response was expected a small example image was provided which contained a random constellation of five dots. It could be seen that after Ph1 five participants responded

with a drawing of five targets, although the random nature of the target locations should have rendered the recognition of a five-target structure impossible (Attachments Figures 55 and 56). It is therefore possible, that the example image influenced some of the participants to opt for a drawing of five targets. This makes it harder to determine whether these participants later understood the concept, that five targets were present in Ph2, or whether they just followed the example image throughout. Nevertheless, a correlation of each participants AE, DE and precision was correlated with the participants' drawing response. In all three measures, there generally was a wide spread of the data, describing a large individual differences between the participants. In Ph1, the correlation demonstrated that the participants, who never sketched exactly five targets had a higher average AE and DE slope, indicating less of a learning curve than the other three groups (Figure 32). This is in contrast to Ph2A and Ph2B, where the same participants either had the same or a slightly more negative AE and DE fit slope. However, with an R^2 s of between 0.00 and 0.044 the difference in sketch response only minimally accounted for the variability between the categories. The lower AE and DE slope for participants, that did not appear to even understand the number of targets, is a surprising finding. This could indicate, the participants were generally unable to deduct the global target structure from the task, in turn causing no metric to be transferred to LTM. This would mean, that any trends seen in the correlations were random, which would explain the small R^2 values. That being said, two participants, that had reported a constellation of five images in Ph2B, additionally labeled the targets with the billboard image objects. This strongly speaks for them having understood, that there were five targets, which were strictly associated to one locations and one billboard image. Additionally, several of the sketches, especially for Ph2B, resembled the relationships between the targets of the real global structure fairly well. Considering these findings, it is fair to say, that at least some of the participants were able to deduct the global structure and a metric between the targets. The question, whether this applies to all participants stands though, especially because three people never sketched a structure consisting of five targets. This however, in turn supports the idea, that the transferred metric was not Euclidean, since learning of the global structure did not correlate with a lower AE and DE, which are indicative for the learning of a Euclidean metric. An increased precision measure on the other hand implies the presence of a stable metric, irrespective of the embedding of the mental map. In Ph2A the normalized precision was on average about the same between all four sketching groups. In Ph2B however, the precision was slightly higher for the participants, that had sketched five targets in both phases 2A and 2B. This could mean the participants managed to form a non-Euclidean mental representation of the target, and were able to be fairly precise in pointing at said representation. It must be said, that the R^2 is very small for the late pointings. For early pointings the variability is only explained by the different target sketches by 12%. There are some additional observation

from the task that remain unexplained. For example, why three participants reported five targets in Ph2A, but reported a different number in Ph2B. All things considered, the sketched maps gave strong indication that at least some of the participants were able to deduct the global structure of the targets, even if not in a strictly Euclidean sense. It also showed that some participants appeared to not have understood the global structure. Correlation of the sketches with other measures was inconclusive. Furthermore, it has been found that individual drawing ability may confound participants' ability to convey spatial information about the environment in a sketch map (Bell and Archibald, 2011). And it can generally be seen across spatial map sketching tasks, that the correlation of sketched map and internally represented map is often low, since map drawings are more than just an externalization of map knowledge (Mallot, 2012). Additionally, the question can be raised, whether Ph2B (visit 2) performance was influenced by the sketching task at the end of the first visit. participants may have been made aware of the possibility or presence of a global target structure, which they might not have (consciously) noticed before. This might be illustrated by an increased number of sketches containing five targets after Ph2B. This begs the question, whether asking the participants' to draw might not pose more confounding factors than actual information about the metric of their cognitive map. For similar future studies utilizing a sketching task, it might be useful to consider offering multiple examples of different target numbers, not offering an example at all, only having sketches done after the second visit or only asking a portion of the participants to draw sketches.

6.8 Summary

Multiple measures were employed to answer the main experimental **questions:**

1. Can metric information from an iterated path integration task be transferred into long-term memory? And, if yes, does it help improve task performance?
2. Does the transferred metric appear to be Euclidean?
3. What is the underlying metric used in human spatial long-term memory?

The following Table 6 is intended to give a short summary of the previously discussed possible answers to these questions. The various results of the experiment can be explained by three **explanations:**

A - There was no transfer of metric information from the iterated path integration task transferred into LTM. This results in the absence of a mental map/representation of the target locations. participants simply completed a series of independent triangle completions, utilizing WM.

B - Metric information was transferred into LTM, leading to the formation of a mental representation of the targets. This mental map is embedded in a Euclidean metric.

C - Metric information was transferred into LTM, leading to the formation of a mental

map of the target layout. However, the metric embedding is not global, but local. This means, the metric of the feedback back into the task is not metrically consistent, but does indicate an understanding of the general layout and relationships between the targets, augmented with local metric information.

Table 6: Overview of the measures assessed and analyzed in this experiment. The measures are grouped, dependant on similarity of the aspect they test for. For each measure it was evaluated how they pertain to the three possible explanations/hypothesis: The results of the measure can be fully (+), partially (+-) or not at all (-) explained by this hypothesis. This is done for each measure individually, as well as for the members of a measure group combined. This was done, because in some cases, the combination of measure outcomes was more indicative of which hypothesis was a likely explanation. (* indicates statistically significant results)

Measure	A		B		C	
	No metric transfer		Euclidean metric		Non-euclidean metric	
	Individually	Combined	Individually	Combined	Individually	Combined
AE across participants*	-	+ -	+	+ -	+	+
DE across participants*	-	+ -	+	+ -	+	+
Individual AE	+	+ -	-	+ -	+	+
Individual DE	+	+ -	-	+ -	+	+
d across participants*	-	+ -	+	+ -	+	+
Individual d	+	+ -	-	+ -	+	+
U	+ -	+ -	-	-	+ -	+ -
Sketched targets	+ -	+ -	-	-	+ -	+ -
Precision	-	+ -	+	-	+	+
Accuracy	+	+ -	-	-	+	+

Each of the measures has been previously discussed in-depth and the possible explanations of the observations were assessed. Therefore they shall not be individually reiterated at this point, instead, a summary of the overall implications of the results on the three above question will be given:

1. & A: Table 6 shows that several of the measures could, at least in part, be explained by the simple answer, that participants did not achieve to learn the layout of the targets in Ph2A and Ph2B. Since there was no direct measure for whether a global metric had been learned, the absence of such must be also be inferred. Looking at the individual measures, some participants performances and reports suggest that they may have failed in transferring the metric of the global target structure into LTM, despite the aids (global landmarks and billboards). They may thus have failed to form a mental representation of

the targets. On the other hand, the data of other participants gave strong indication, that they had, in fact, understood the global structure and formed a mental representation of the target goals. Not all participants were able to, other participants showed indication of having transferred metric information from the iterated path integration task into LTM. Therefore, it can be concluded that the transfer is in principal possible.

2. & B: As can be seen in the summary, only few of the measures were comprehensively explained by the idea of a Euclidean metric. Most of the tested measures were geared specifically towards detection of a Euclidean embedding of the mental representation. Therefore, the lack in findings supporting the idea, is telling. In this experiment, the hypothesis that Euclidean metric information is transferred into LTM from an iterated path integration task must be rejected. This can't conclusively determine, whether LTM entails a Euclidean metric at all, but it gives clear indication that this might not be the case.

3. & C: Several of the tested measures could be explained by the idea of a non-Euclidean map being constructed in LTM. Particularly the labeled graph hypothesis was explored (Warren, 2019) in this context. This explanation accounts for most of the measures, since it allows for a metrically inconsistent mental map, that is not embedded in a global, Euclidean-like reference frame. However, due to the loose definition of this labeled graph, further research is necessary, to make sure this hypothesis is not just a "cop out" explanation for inexplicable findings, but instead, actually statistically explains the data better than the idea that no metric information was transferred into LTM. Especially, because very few of the measures tested directly for the presence of a mental representation (but U and the sketched targets were not analyzed in-depth), and non of the measures tested for the metric of a non-Euclidean representation. All in all, there is a fair amount of evidence pointing towards the presence of a non-Euclidean metric in LTM, but the design, task analysis and results are not sufficient to conclude on what the "true" metric within LTM is.

In conclusion, this experiment supports the idea that spatial information from an iterated triangle completion task can be transferred into spatial long-term memory. However, this information appears to not follow a strictly Euclidean metric. In line with Warren (2019) and Chrastil and Warren (2014), there is indication, that the metric of the cognitive map established in human spatial LTM may more closely mirror a labeled graph, a network of connections between locations, that are developed as local metric information, allowing for metric inconsistencies across the global structure. Nonetheless, the present study is not sufficient to draw clear conclusions on the "true" metric within LTM. Therefore more freedom in the design may be necessary for participants to form their own internal representation and a more in depth measuring and analysis of said representation are necessary.

7 Possible improvements and future outlook

Although the study at hand was designed with consideration of many factors, nevertheless certain caveats were not accounted for within the experimental design or analysis. The following section is intended to give an overview of some of these possible factors and offer suggested improvements for future experiments, where applicable.

Individual differences: As the large variation of individual participant data showed, humans have vastly varying triangle completion task abilities, and navigational abilities in general. In this experiment, no participants were excluded from the analysis and no separation of participants was done. The large differences opened the avenue for a broader range of data interpretation. It is very possible, even likely, that not all participants managed to achieve the expected recognition of a five-target structure in Ph2. For future studies it may be useful to pre-define a criterion by which participants are excluded or analyzed separately. Alternatively, a pre-test could be run, to determine which participants are suitable to study the hypothesis and question of interest.

Gender differences: Participants were each 50% male and females, and no distinction was made between the genders for any part of the experiment. The sexes were also not analyzed separately for any of the measures. The possibility of gender differences in spatial navigation and orientation is a mixed bag. Certain spatial tasks and design seem to elicit better performance in males (e.g. Kim et al., 2007), others in females (e.g. Voyer et al., 2007). Interestingly, gender differences have also been found to translate into large scale VR tasks (Castelli et al., 2008) and males may be better at navigating without landmarks, while female performance may also vary by menstrual cycle time point and the use of hormonal contraceptives (Bernal et al., 2020). On the other hand such a dichotomous view may also be too simple and a more multi-faceted approach may be necessary (Coluccia and Louse, 2004; Montello et al., 1999). Be it as it may, for future studies it would be interesting to analyze the two sexes separately and test to what degree gender may account for individual variation.

Landmark preference: This experiment offered no local landmarks but very distant global landmarks in the form of mountains, hills and planes were present. Steck and Mallot (2000b) investigates the role of global versus local landmarks within virtual reality navigation. They found strong individual differences in participants as some participants most heavily relied on local landmarks while others only used global landmarks or even varied between the two for different navigational decisions. This could indicate that participants who naturally preferred global landmark use may perform better than their counter parts. However, it was also shown that the removal of one type of target led to a switch to the other target type for navigation in most participants. It must also be noted, that the experiment did not present local landmarks at any time, meaning there was no possibility for back and forth switching of target strategies within the experiment.

But it is feasible that participants with a local landmark preference might have require more time to adjust to a navigational task without their naturally preferred navigational strategy.

Continuous feedback: After each pointing participants were given feedback about the actual target location by a glass cylinder in the "true" position. The benefit of this constant feedback was a reduction of an accumulation in errors over time, which would otherwise have to be accounted for. It also gave participants the best possible chance to learn the Euclidean global structure and relationships of the targets. On the down side, especially in Ph2, this constant feedback may have facilitated reliance on the feedback poles as navigational aid rather than the formation of ones own, internal representation of the target positions. The feedback was given right away after each pointing, so right after each 2-back pointing, which was performed first. This could have caused a correction or change in orientation before the 1-back responses and thus influenced the pointings. In a future study it may be interesting to have the participants learn the layout in Ph2 with feedback for some time, but also allow for a period without feedback. Alternatively the feedback could appear after both 1-back and 2-back responses had been given at each location, to reduce influence. This might allow for an increase in formation of an internal representation.

Collective pointing: Without a task that specifically tested for the participants' represented target positions, irrespective of metric, many of the described measures rely on the assumption and secondary evidence, that the participants in fact were able to build up an internal representation. Therefore future studies should add a task testing for a general understanding of the global structure. Such an additional measure could be included to ensure the iterated path integration really was more than just classic triangle completion and lead to the formation of a global structure in Ph2. For example before ending the experiment participants could be placed at a novel position and asked to point at all the targets from that positions. This task would defy strict working memory involvement which is necessary for triangle completion.

Vestibular feedback: Lastly, participants were not able to physically walk within the VR space, but instead used the controller for translation. The rotations within the experiment were physically performed by the participants, allowing for the natural and familiar feedback of both a visual and vestibular system. Translation on the other hand only resulted in visual feedback (optical flow). This may have resulted in a conflict of incoming cues and made the estimation of distances versus angles unbalanced, due to the difference of incoming modality cues. Ideally the experiment would be conducted in a large experimental room, in which both walking and turning were physically possible, allowing for a more natural cueing and reducing the cue conflict.

This study presented some interesting findings, suggesting, that the metric in LTM is

not a Euclidean metric. However, no final conclusion could be made on what the "actual" metric is, although data may support the idea of a labeled graph hypothesis. In any case, more future research is necessary to fully understand the true underlying metric in LTM. For this purpose, the presented topic and methodological approach can easily be expanded and improved upon in future studies. The use of increasingly accessible, affordable and life-like virtual reality technology is a valuable and versatile method in studying human navigation and its underlying mechanisms and "mystery metric".

8 Attachments

8.1 Target orders

Table 7: The two target number orders. Each round contained a semi randomized order of the five targets in which they were visited and subsequently pointed at by the participant. The orders were assigned to the participants alternating upon arrival. The red numbers indicate the targets at which a pointing event took place i.e., the head position of the participant was equivalent to the red target number, 1-back and 2-Back target numbers.

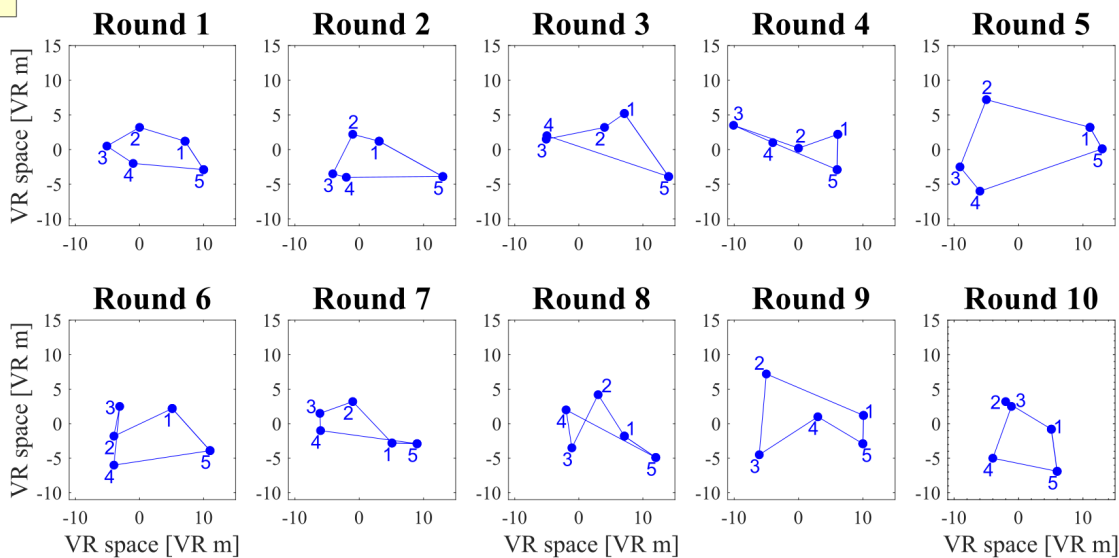
Target Order 1					Target Order 2						
Round no.	Target numbers					Round no.	Target numbers				
1	5	1	2	3	4	1	2	4	5	3	1
2	1	2	5	4	3	2	5	2	4	1	3
3	1	2	3	5	4	3	4	1	2	5	3
4	5	2	4	1	3	4	5	2	3	4	1
5	4	2	1	3	5	5	3	1	2	5	4
6	1	4	5	3	2	6	1	2	3	5	4
7	5	1	2	4	3	7	1	2	5	4	3
8	4	1	3	5	2	8	5	1	2	4	3
9	3	4	5	2	1	9	1	2	5	4	3
10	5	2	1	3	4	10	5	4	2	3	1

Table 8: Overview of the assignment of target orders 1 and 2 to the participants and phases. This determined in which order the participant visited and subsequently pointed at the five different target locations.

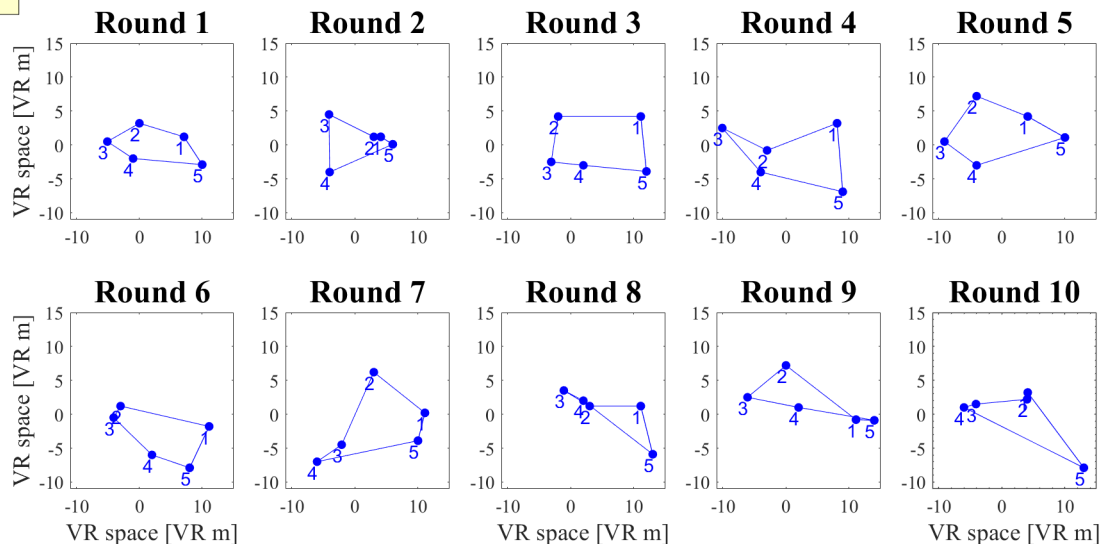
Participant no.	Phase 1	Phase 2
1	Order 1	Order 2
2	Order 2	Order 1
3	Order 1	Order 2
4	Order 2	Order 1
5	Order 1	Order 2
6	Order 2	Order 1
7	Order 1	Order 2
8	Order 2	Order 1
9	Order 1	Order 2
10	Order 2	Order 1
11	Order 1	Order 2
12	Order 2	Order 1

8.2 Phase 1 target locations

Participant 1



Participant 2



Participant 3

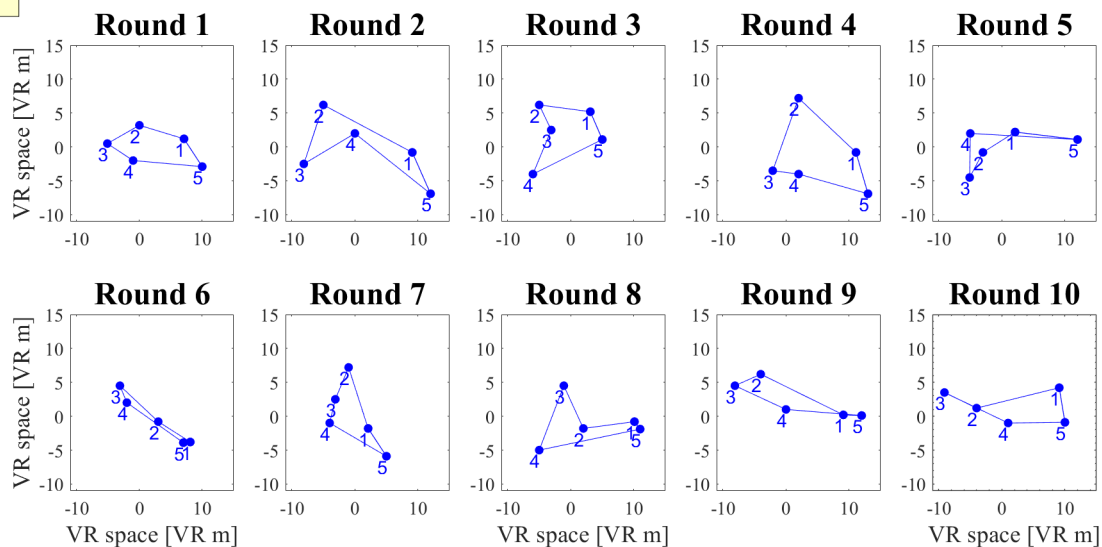
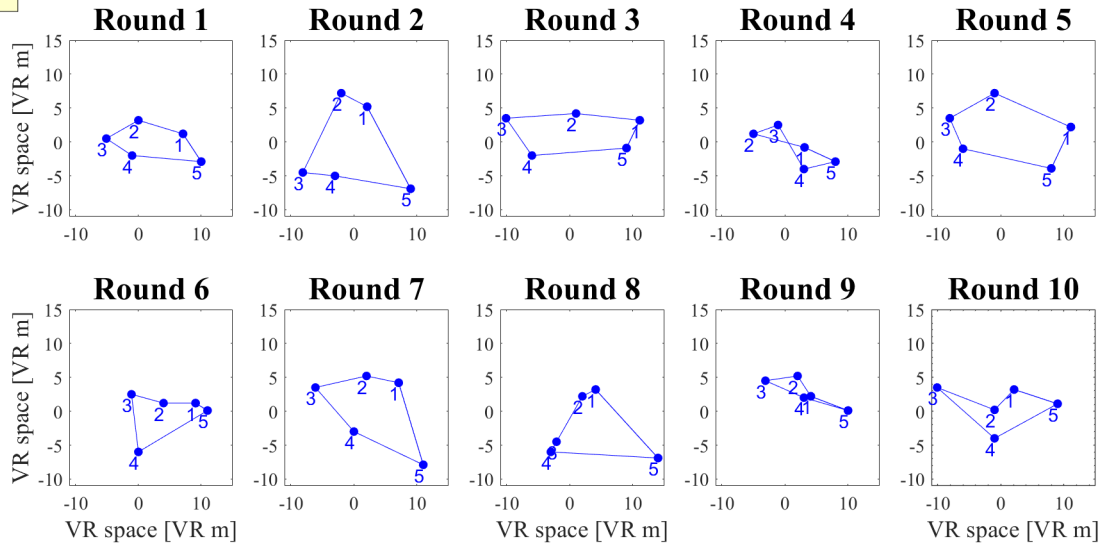
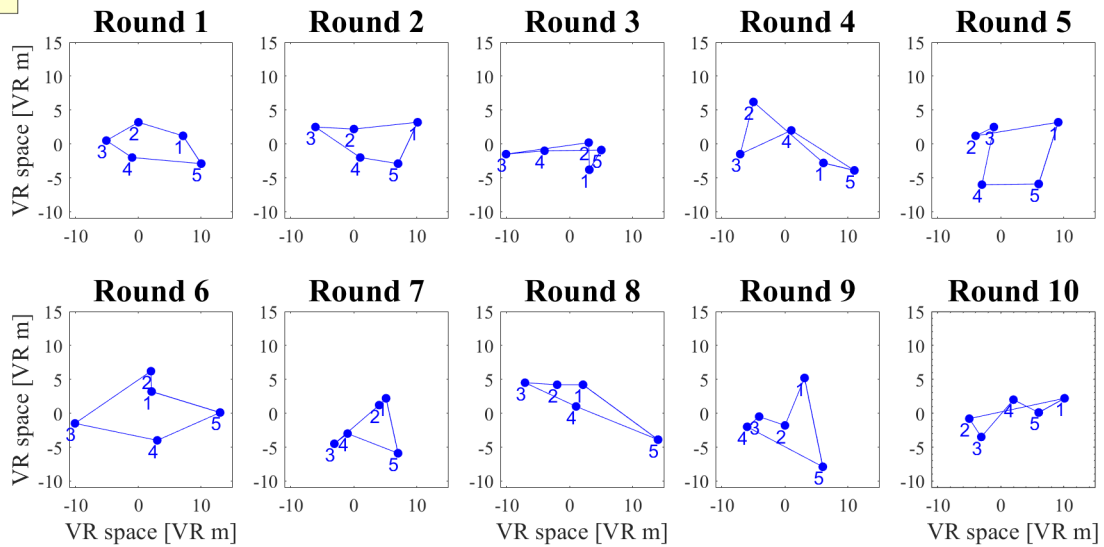


Figure 34: The positions of the five target posts in VR space in each round of phase 1, for participants 1 - 3.

Participant 4



Participant 5



Participant 6

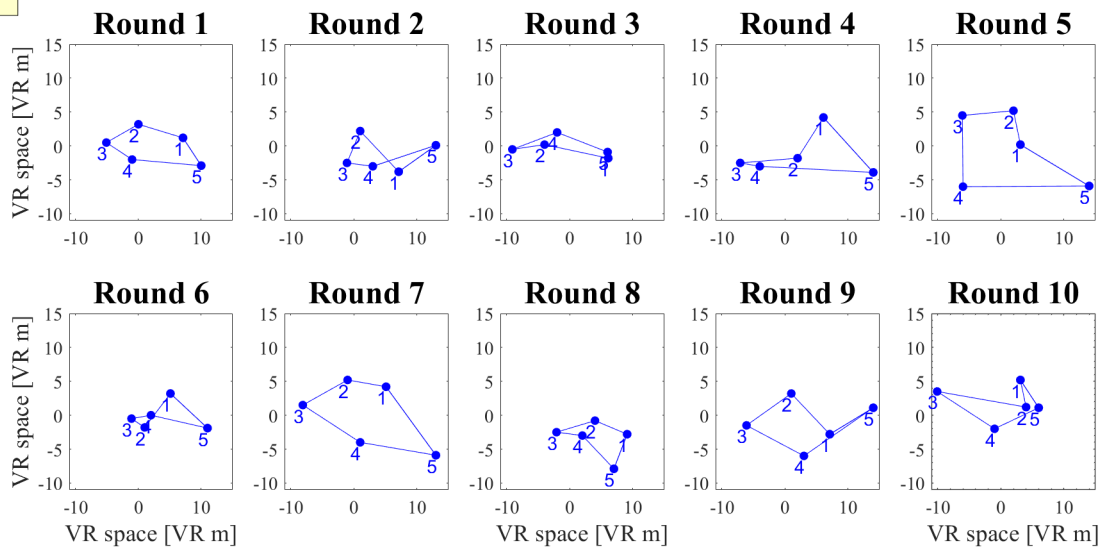
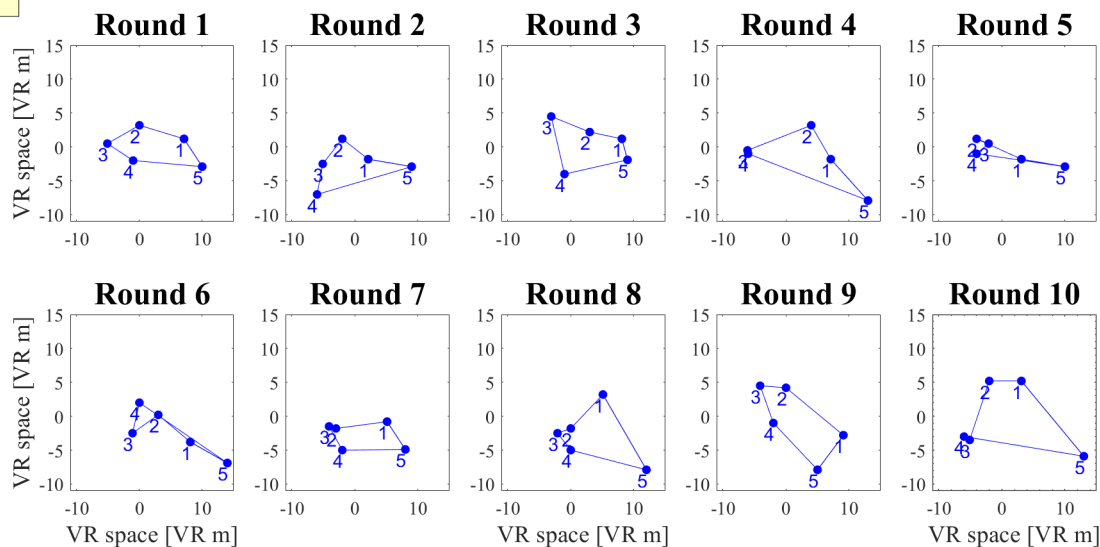
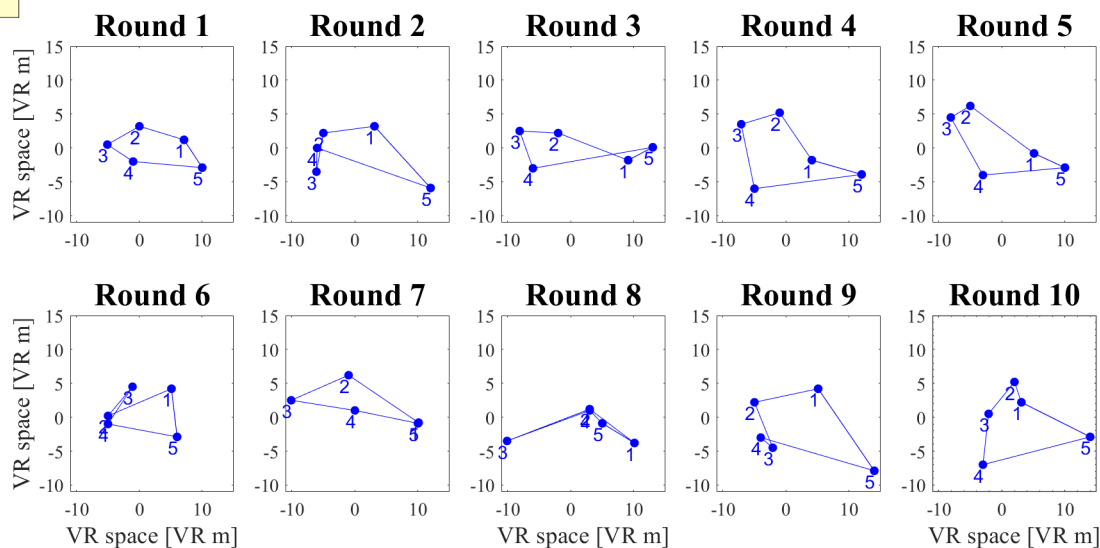


Figure 35: The positions of the five target posts in VR space in each round of phase 1, for participants 4 - 6.

Participant 7



Participant 8



Participant 9

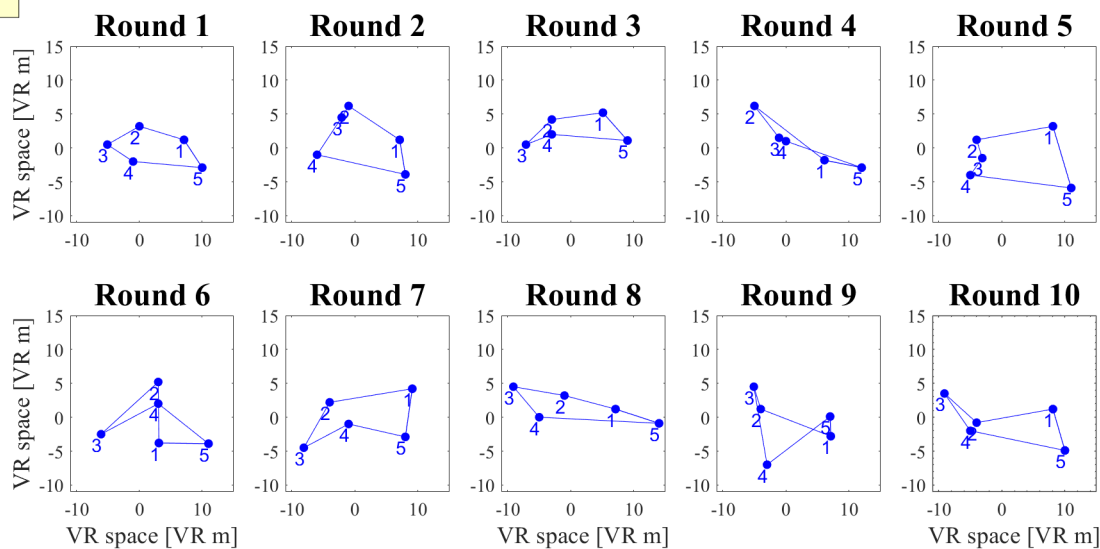
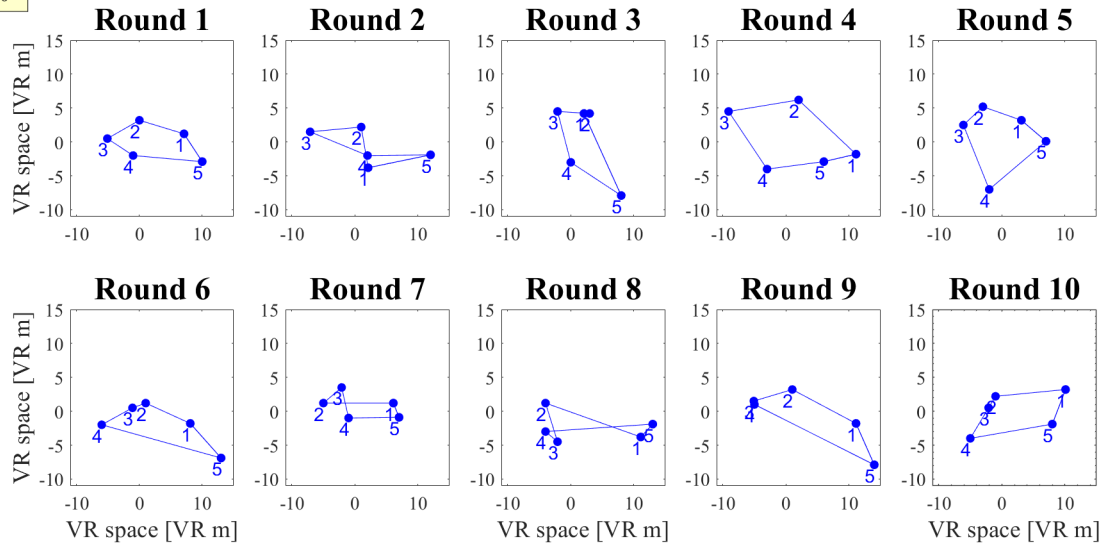
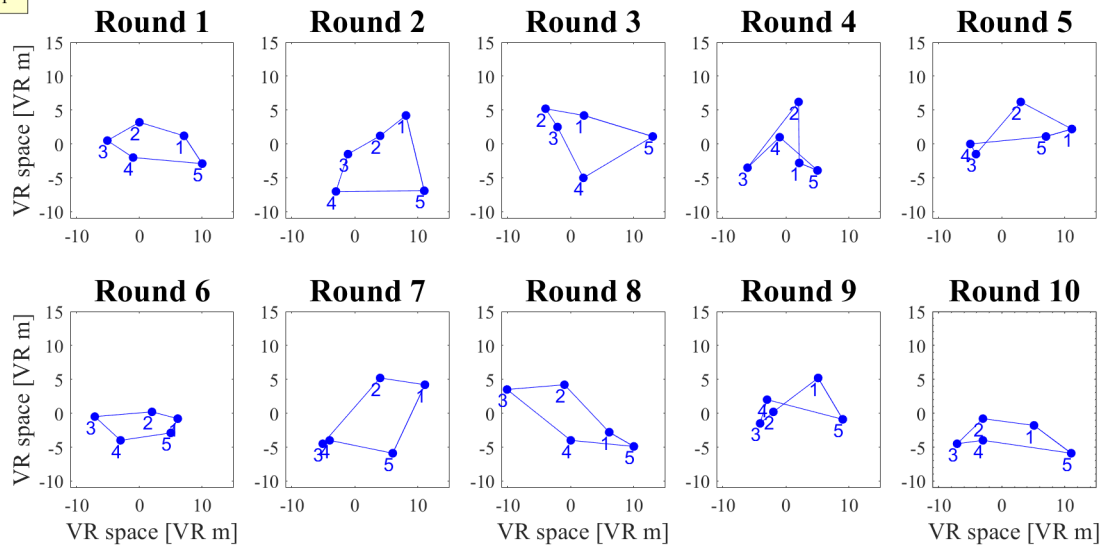


Figure 36: The positions of the five target posts in VR space in each round of phase 1, for participants 7 - 9.

Participant 10



Participant 11



Participant 12

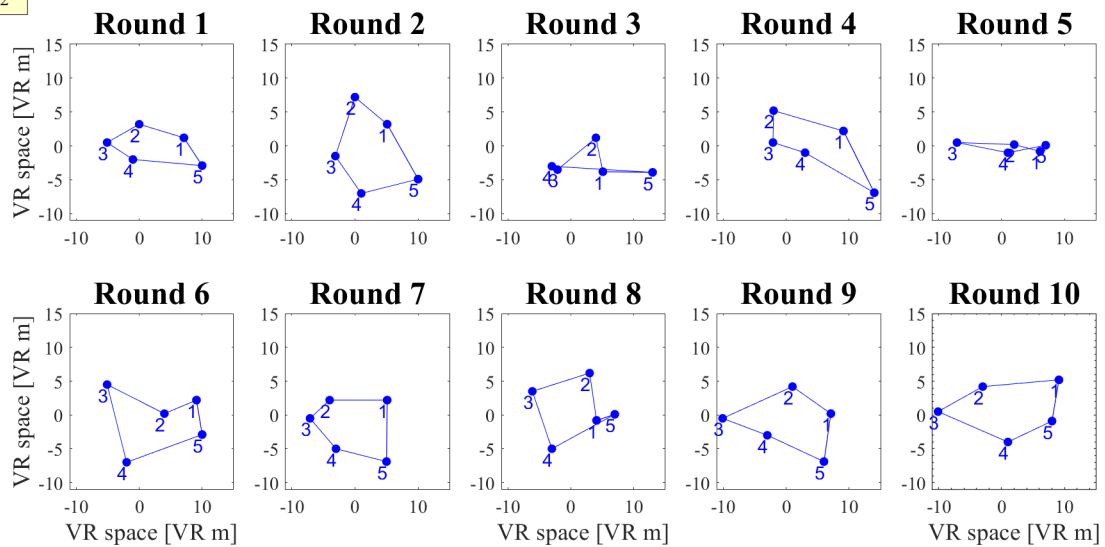


Figure 37: The positions of the five target posts in VR space in each round of phase 1, for participants 10 - 12.

8.3 Angle error

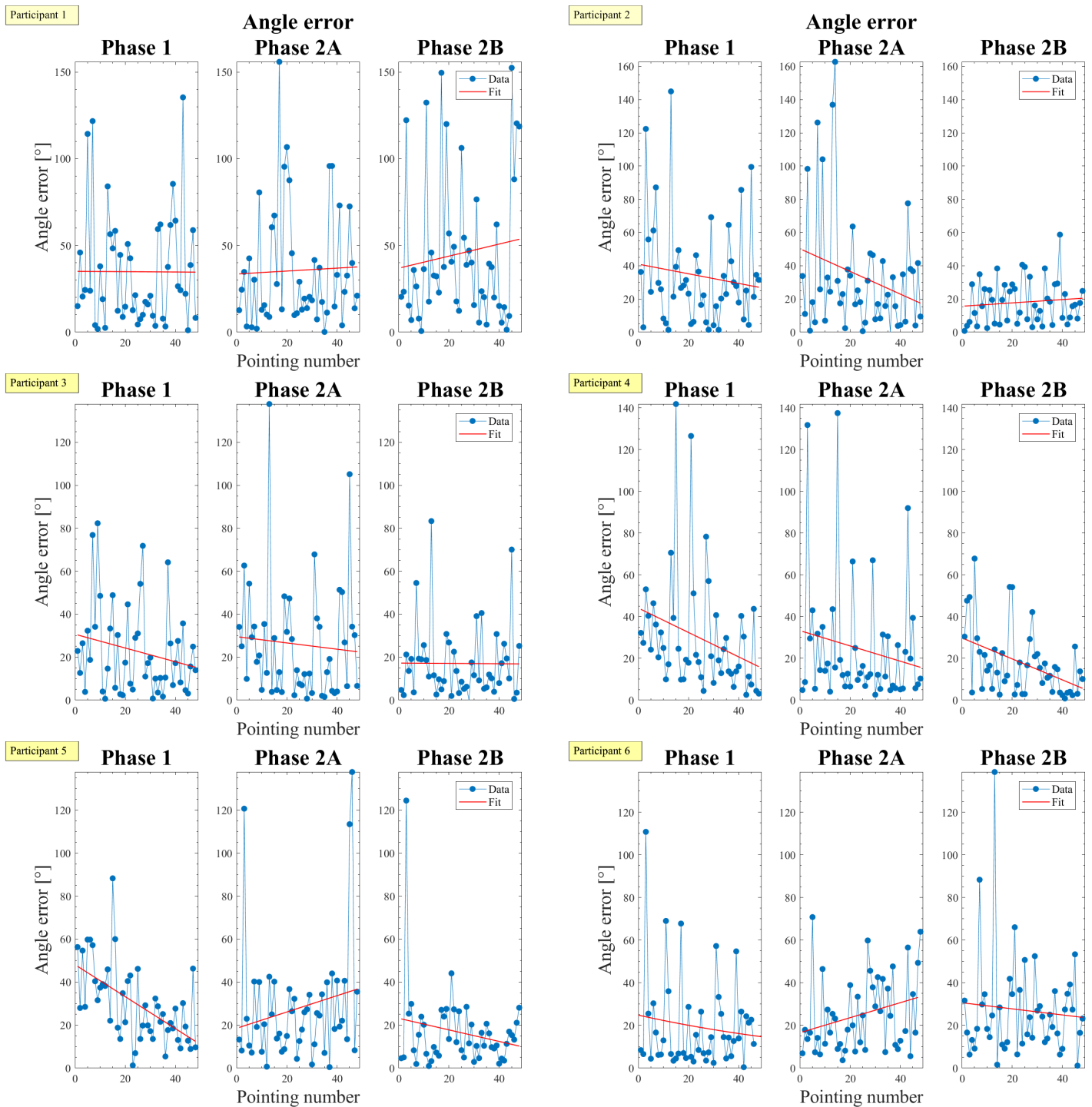


Figure 38: For each of the participants 1 - 6: The combined angle error (AE) in degrees (blue) over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (red) for all three phases is also included.

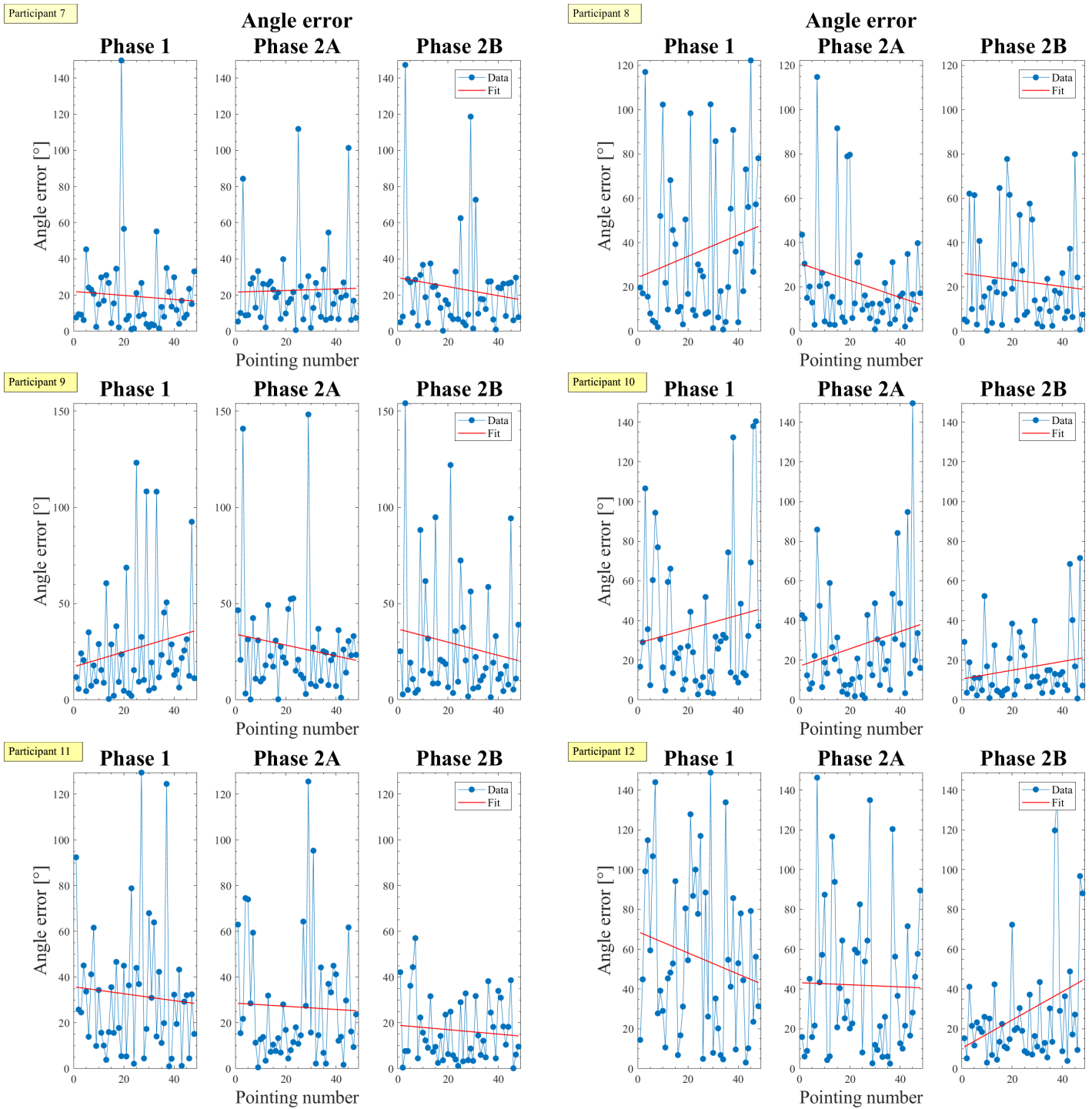


Figure 39: For each of the participants 7 - 12: The combined angle error (AE) in degrees (blue) over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (red) for all three phases is also included.

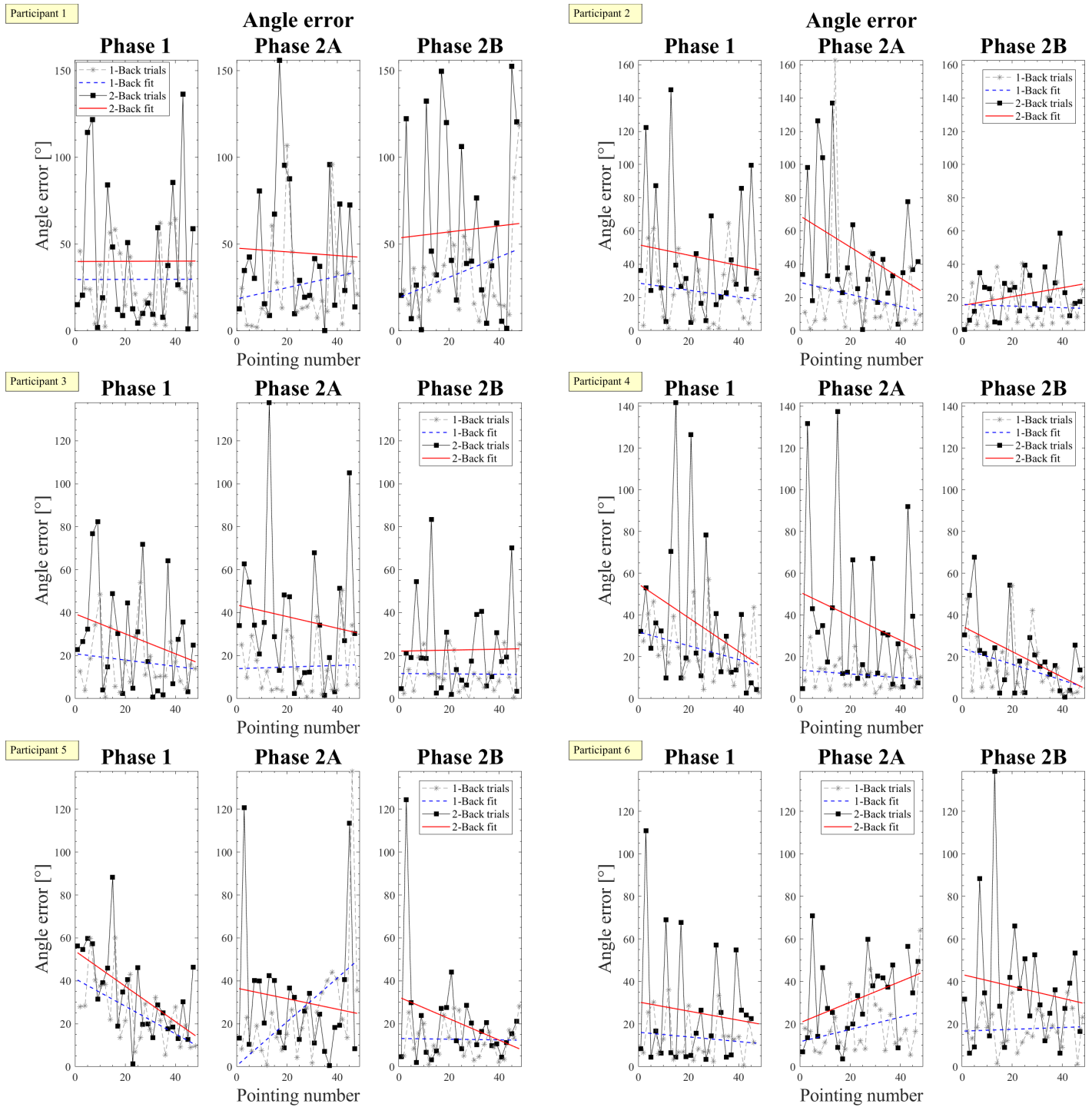


Figure 40: For each of the participants 1 - 6: The angle error (AE) in degrees separated by 1-back (grey) and 2-back (black) tasks over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (blue and red respectively) for all three phases is also included.

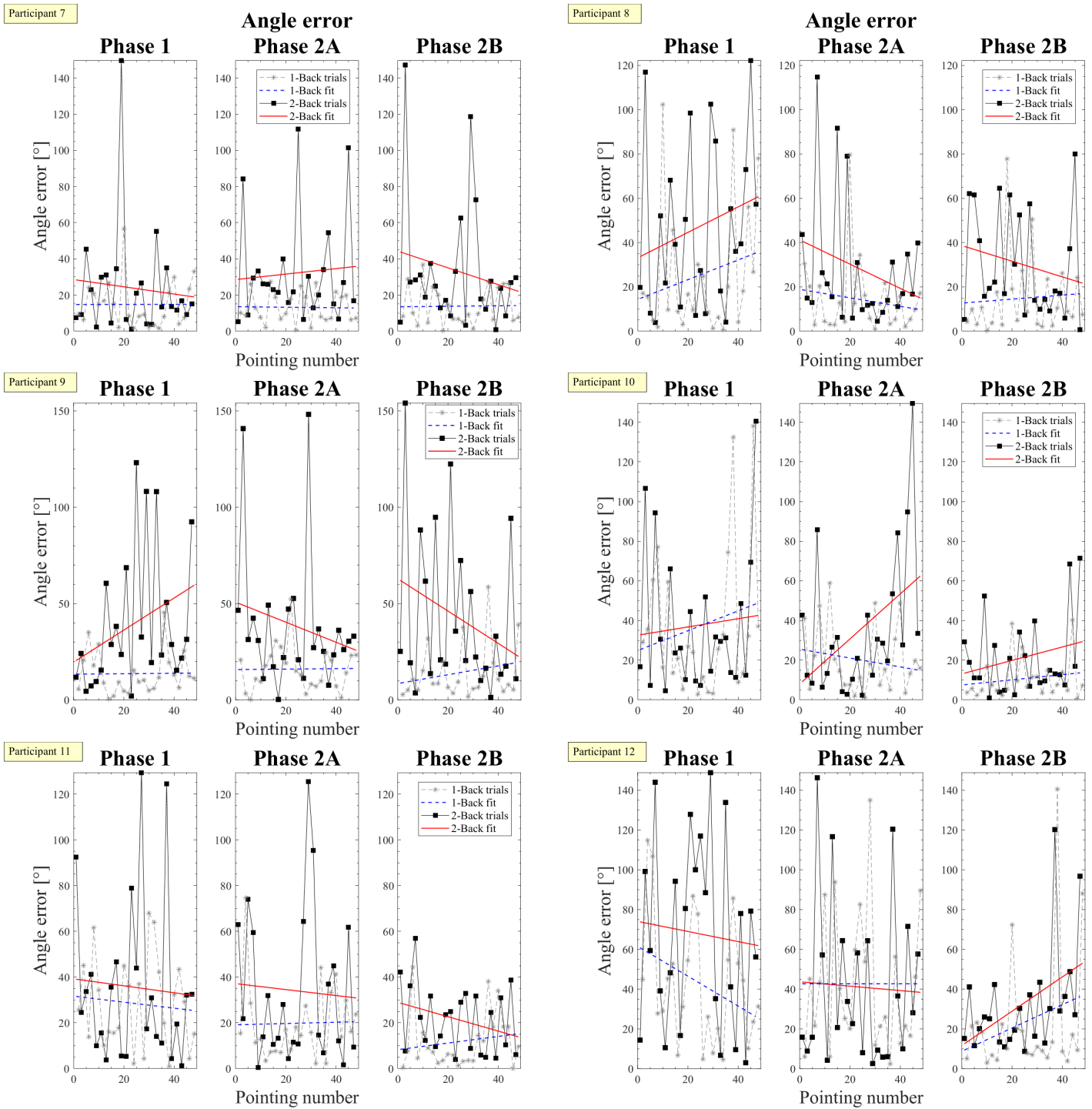


Figure 41: For each of the participants 7 - 12: The angle error (AE) in degrees separated by 1-back (grey) and 2-back (black) tasks over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (blue and red respectively) for all three phases is also included.

8.4 Distance error

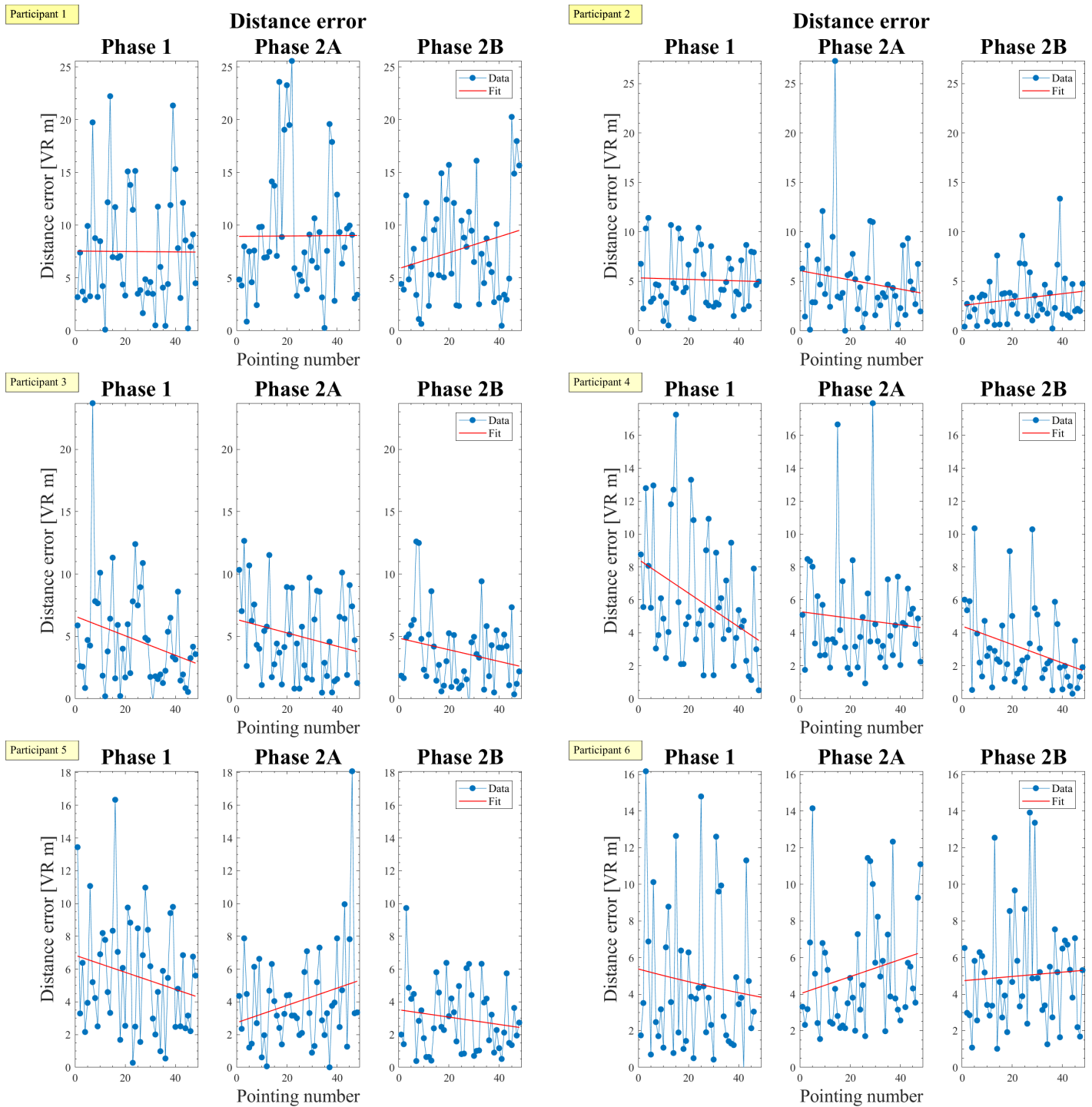


Figure 42: For each of the participants 1 - 6: The combined distance error (DE) in VR meters (blue) over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (red) for all three phases is also included.

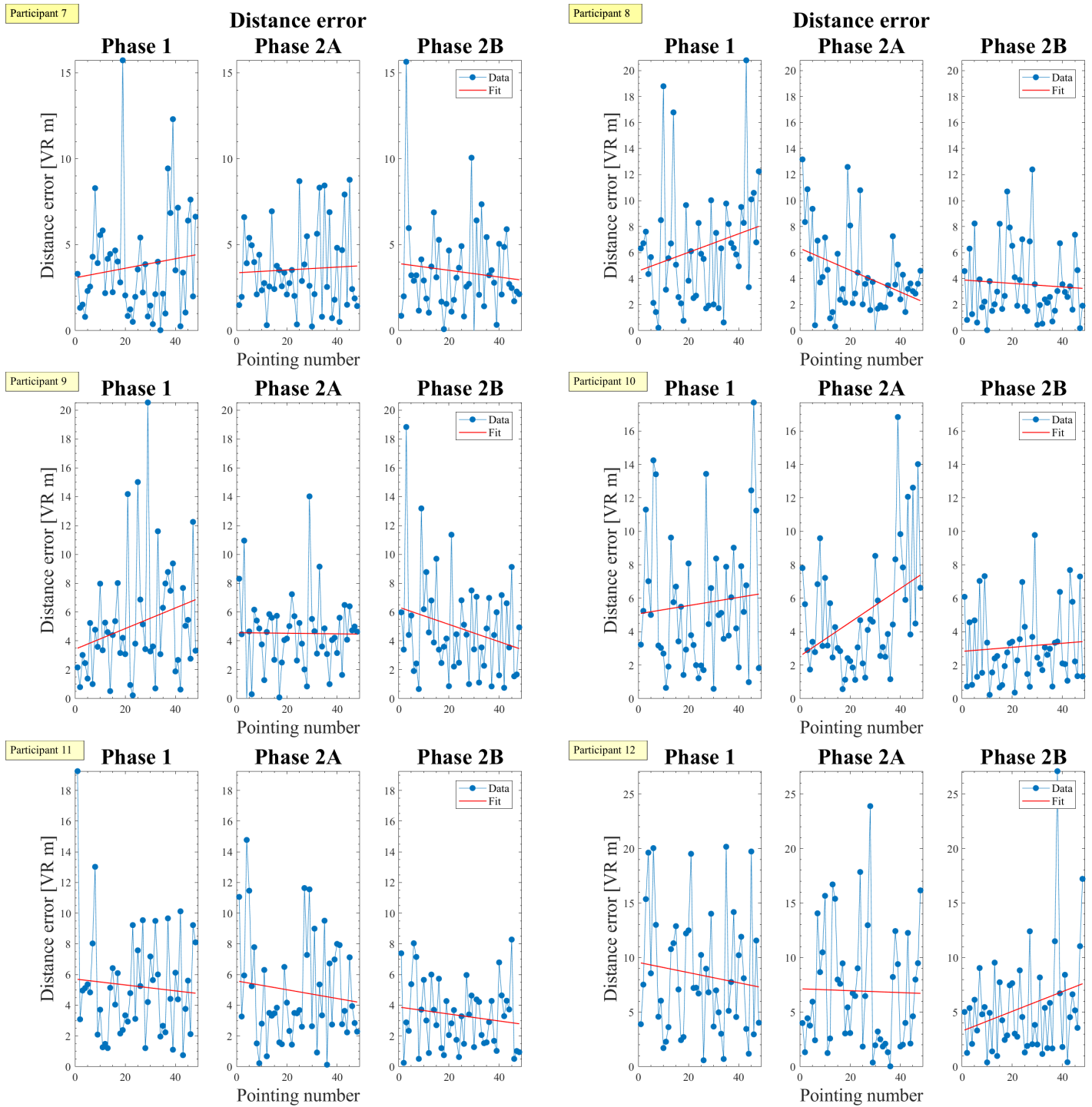


Figure 43: For each of the participants 7 - 12: The combined distance error (DE) in VR meters (blue) over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (red) for all three phases is also included.

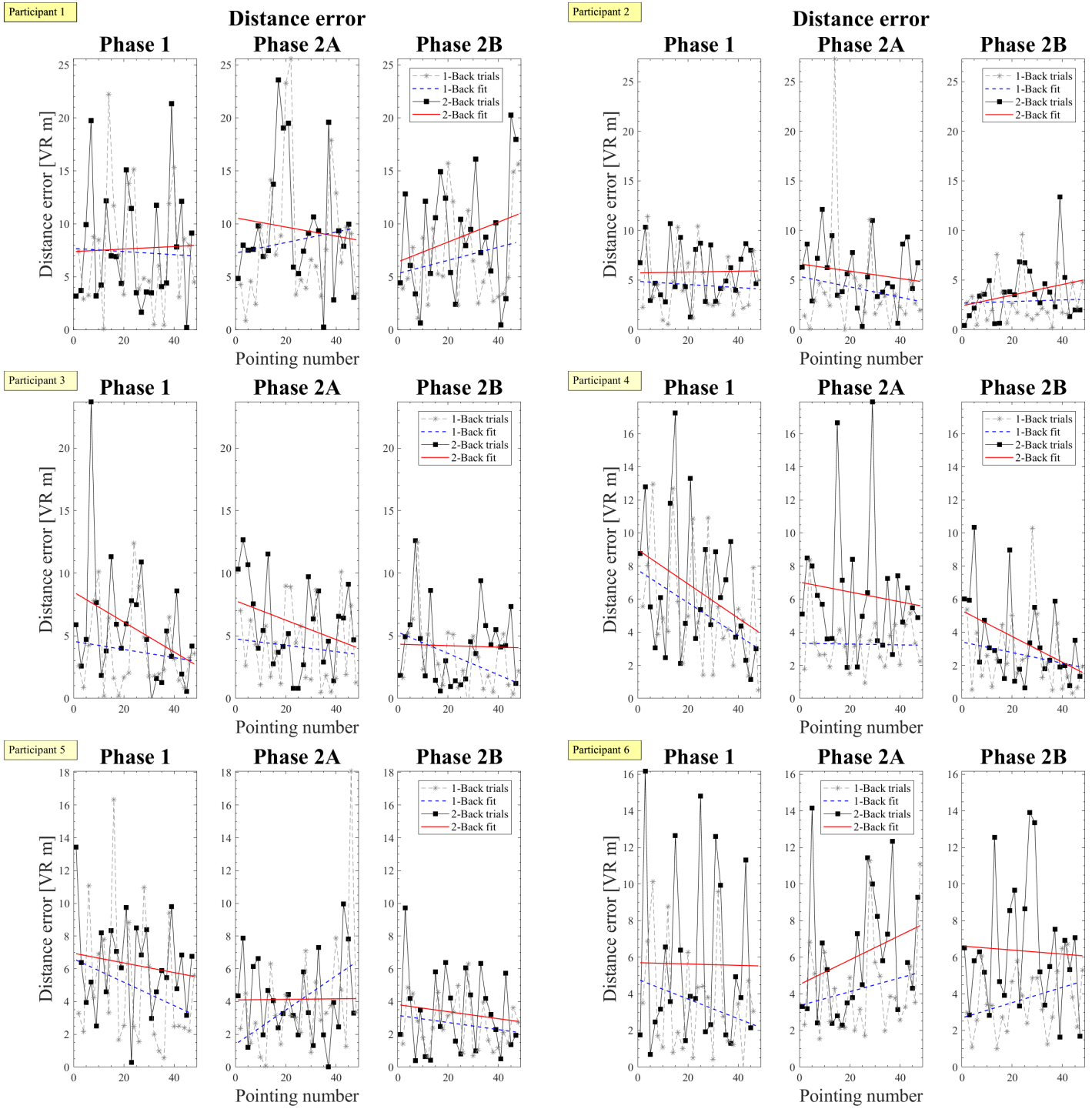


Figure 44: For each of the participants 1 - 6: The distance error (DE) in VR meters separated by 1-back (grey) and 2-back (black) tasks over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (blue and red respectively) for all three phases is also included.

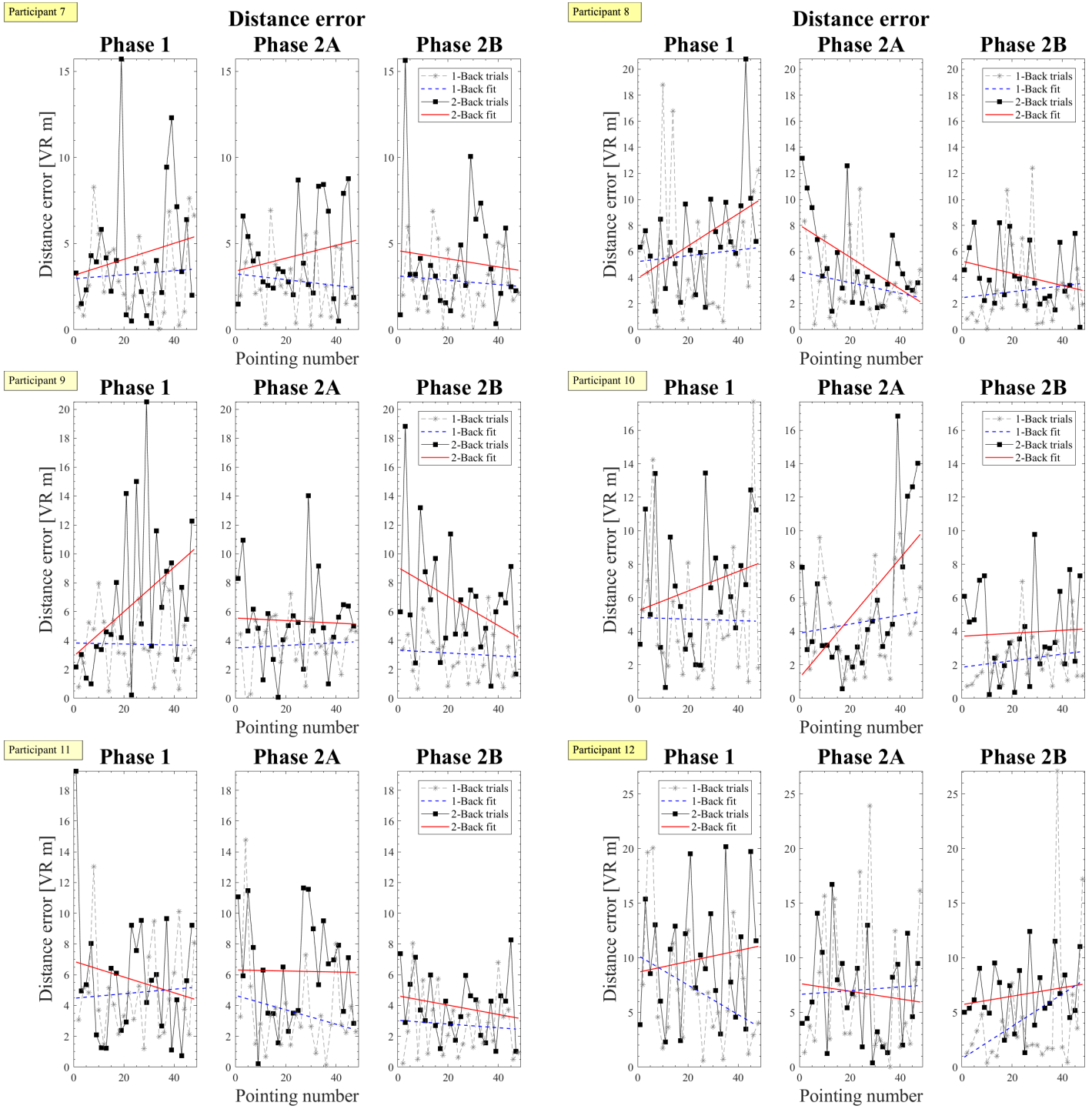


Figure 45: For each of the participants 7 - 12: The distance error (DE) in VR meters separated by 1-back (grey) and 2-back (black) tasks over increasing target number (i.e. over the time course of the experiment). The linear lsq fit (blue and red respectively) for all three phases is also included.

8.5 Residual deviation and estimated representations

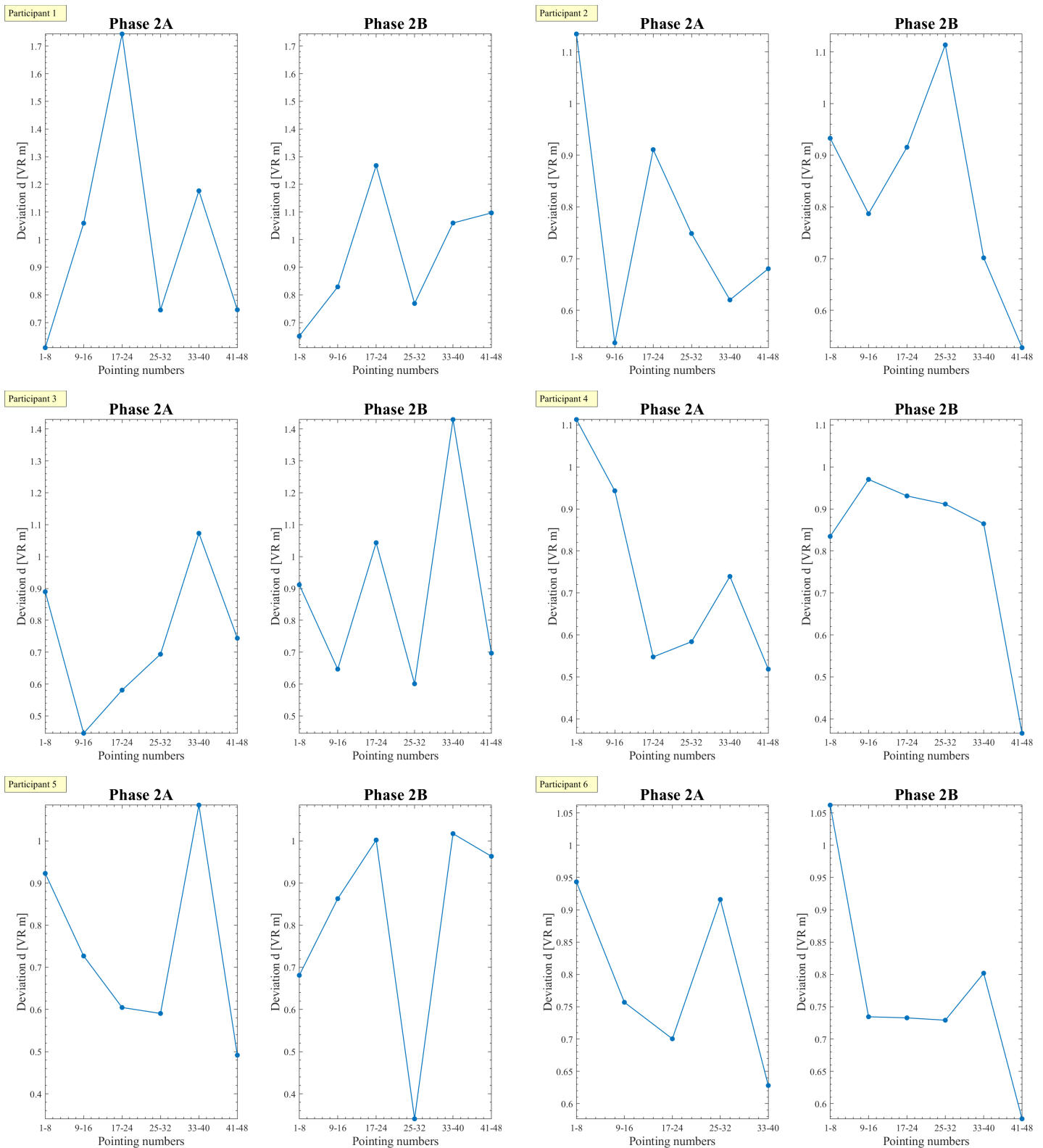


Figure 46: The residual deviation (d) between the actual (true) target positions and the estimated target positions over the time course of the experiment for participants 1 - 6. d was calculated from a set of eight consecutive pointing events and plotted as such (i.e., one d value was calculated for every eight pointing events, as eight pointing events uniquely described five estimated target positions)

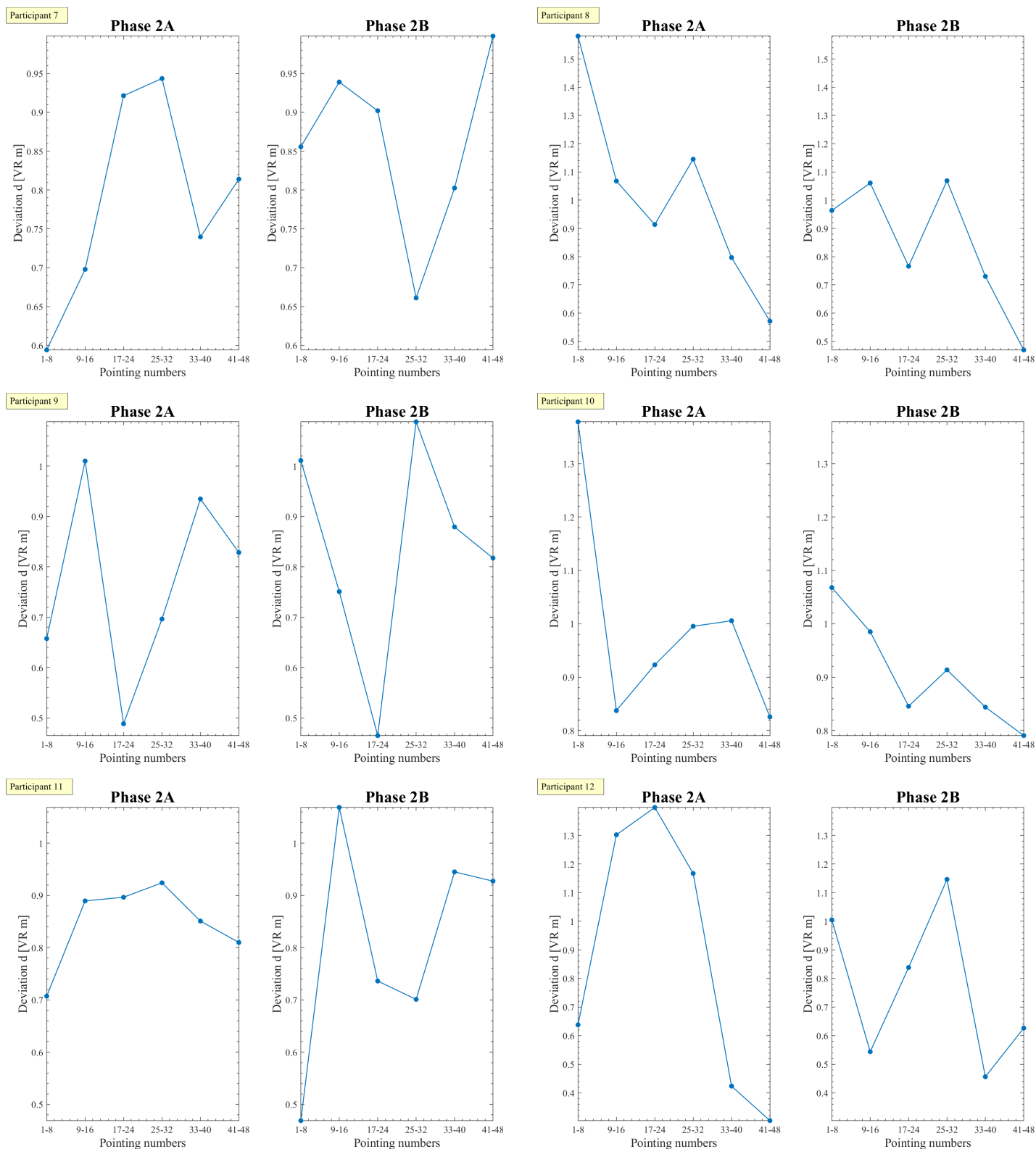


Figure 47: The residual deviation (d) between the actual (true) target positions and the estimated target positions over the time course of the experiment for participants 7 - 12. d was calculated from a set of eight consecutive pointing events and plotted as such (i.e., one d value was calculated for every eight pointing events, as eight pointing events uniquely described five estimated target positions)

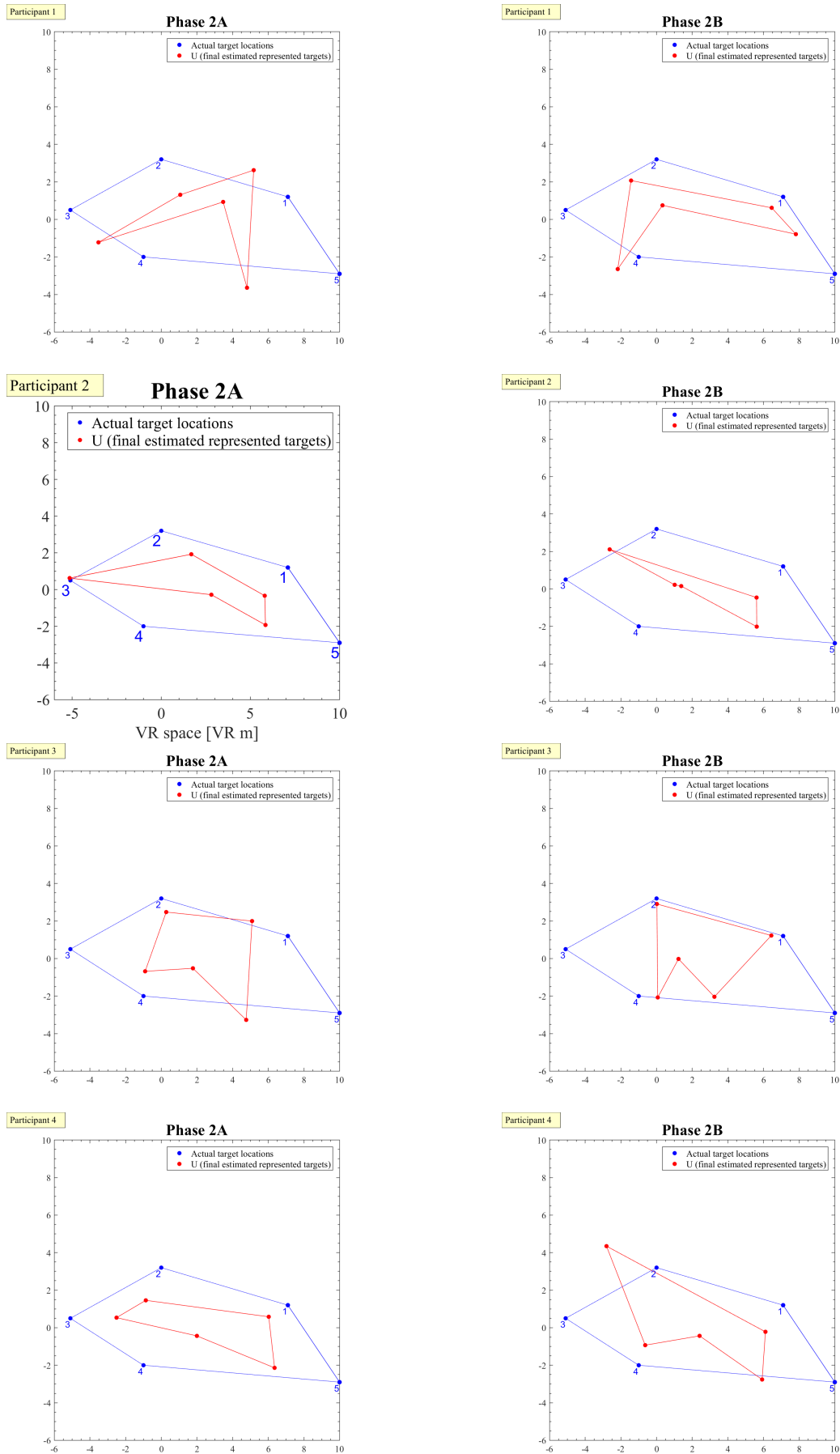


Figure 48: Participant 1 to 4: The participants' final estimated represented target positions (U ; red) compared to the actual target positions (blue) in Ph2A and Ph2B. U is based on the participant's last eight pointings within the phase, which uniquely described the estimated represented five target positions in the participants memory.

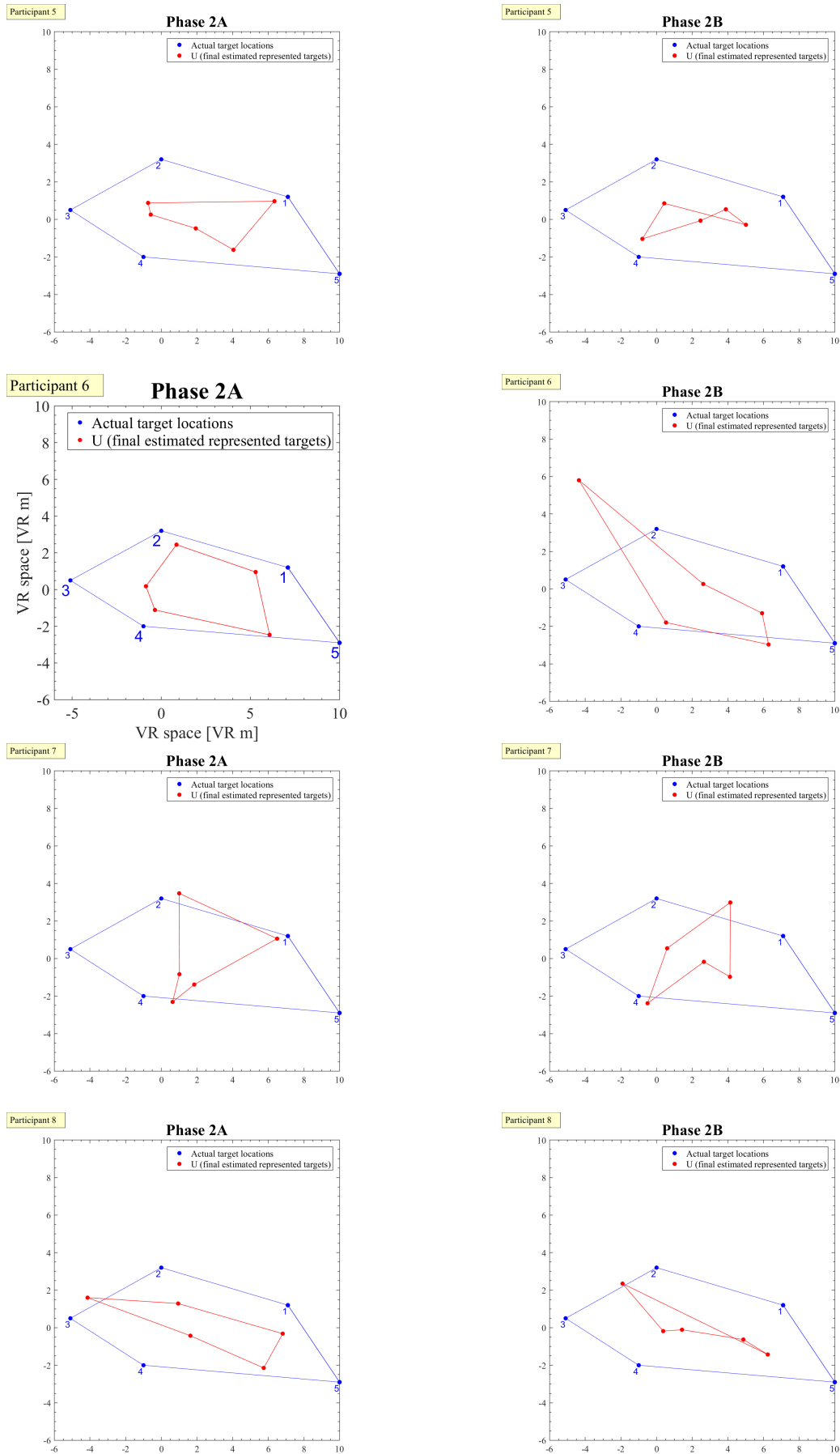


Figure 49: Participant 5 to 8: The participants’ final estimated represented target positions (U ; red) compared to the actual target positions (blue) in Ph2A and Ph2B. U is based on the participant’s last eight pointings within the phase, which uniquely described the estimated represented five target positions in the participants memory.

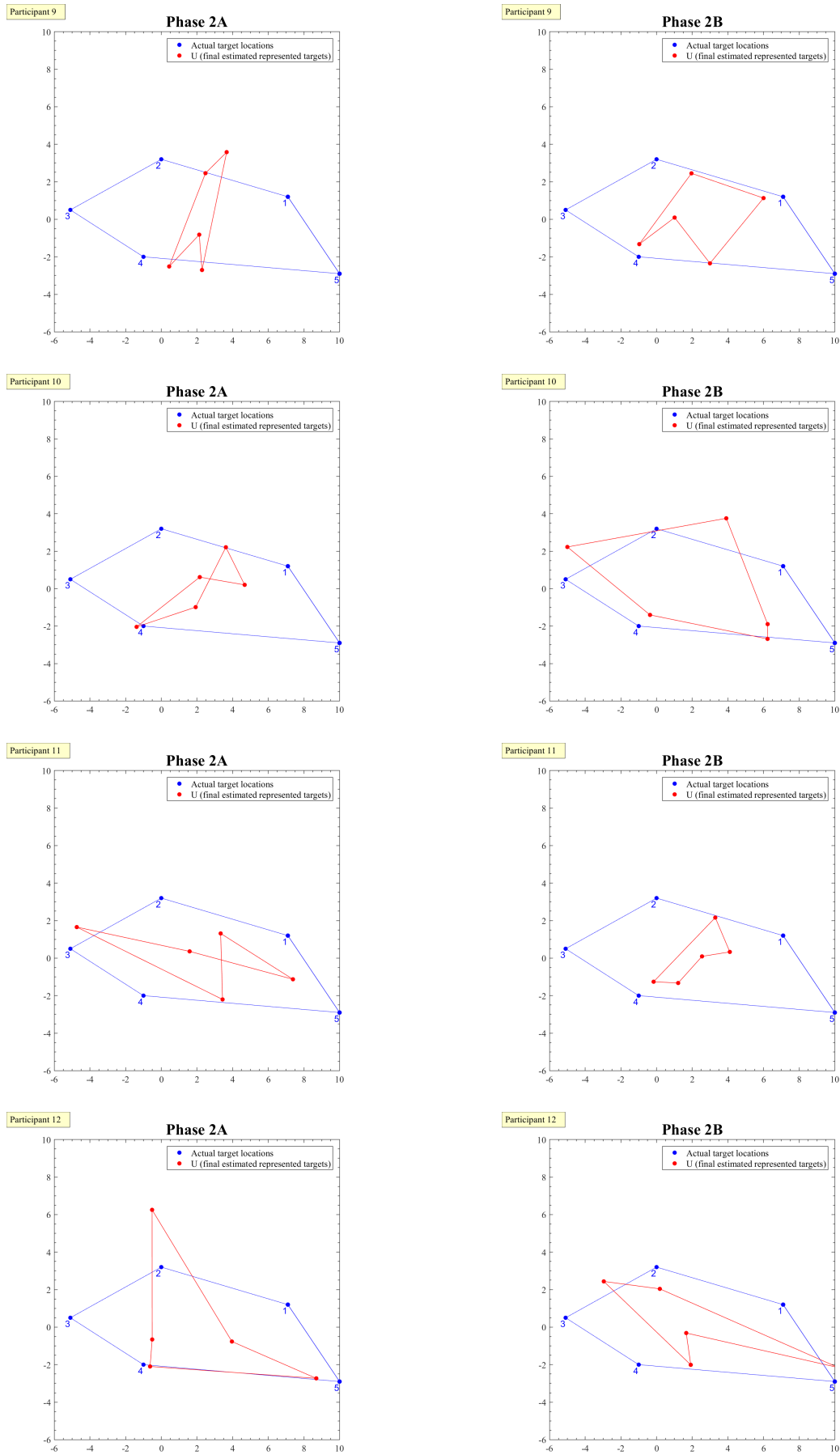


Figure 50: Participant 9 to 12: The participants' final estimated represented target positions (U ; red) compared to the actual target positions (blue) in Ph2A and Ph2B. U is based on the participant's last eight pointings within the phase, which uniquely described the estimated represented five target positions in the participants memory.

8.6 Precision and accuracy

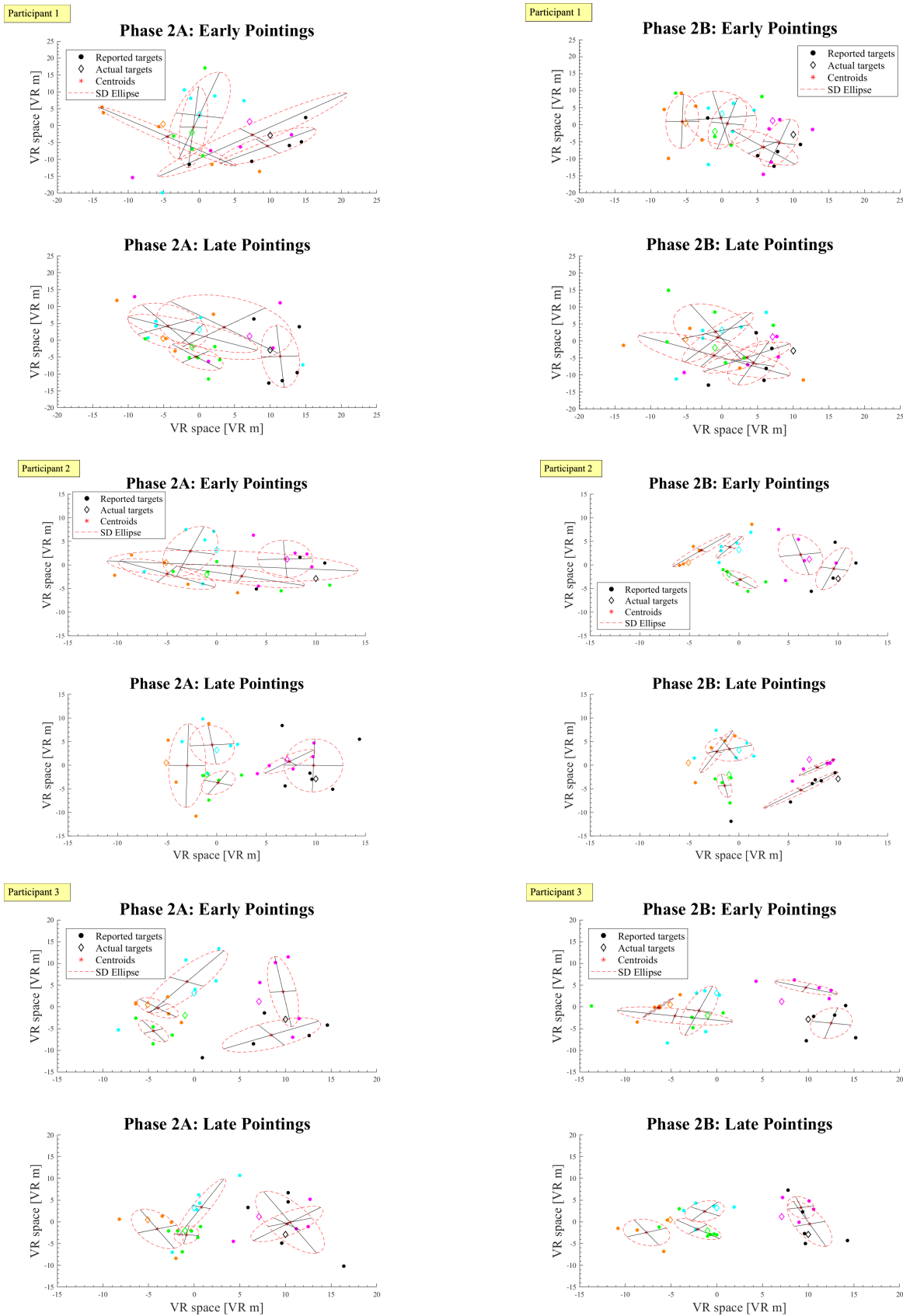


Figure 51: For participant 1 - 3: Visualization of accuracy and precision in early (1-24) versus late pointings (25-48) in Ph2A (a) and Ph2B (b). The covariance (SD) error ellipse surrounds all the pointings (colored dots) directed at the same true target location (colored diamond). "Shots" directed at the same target are indicated by the same color, as is the corresponding actual target location.

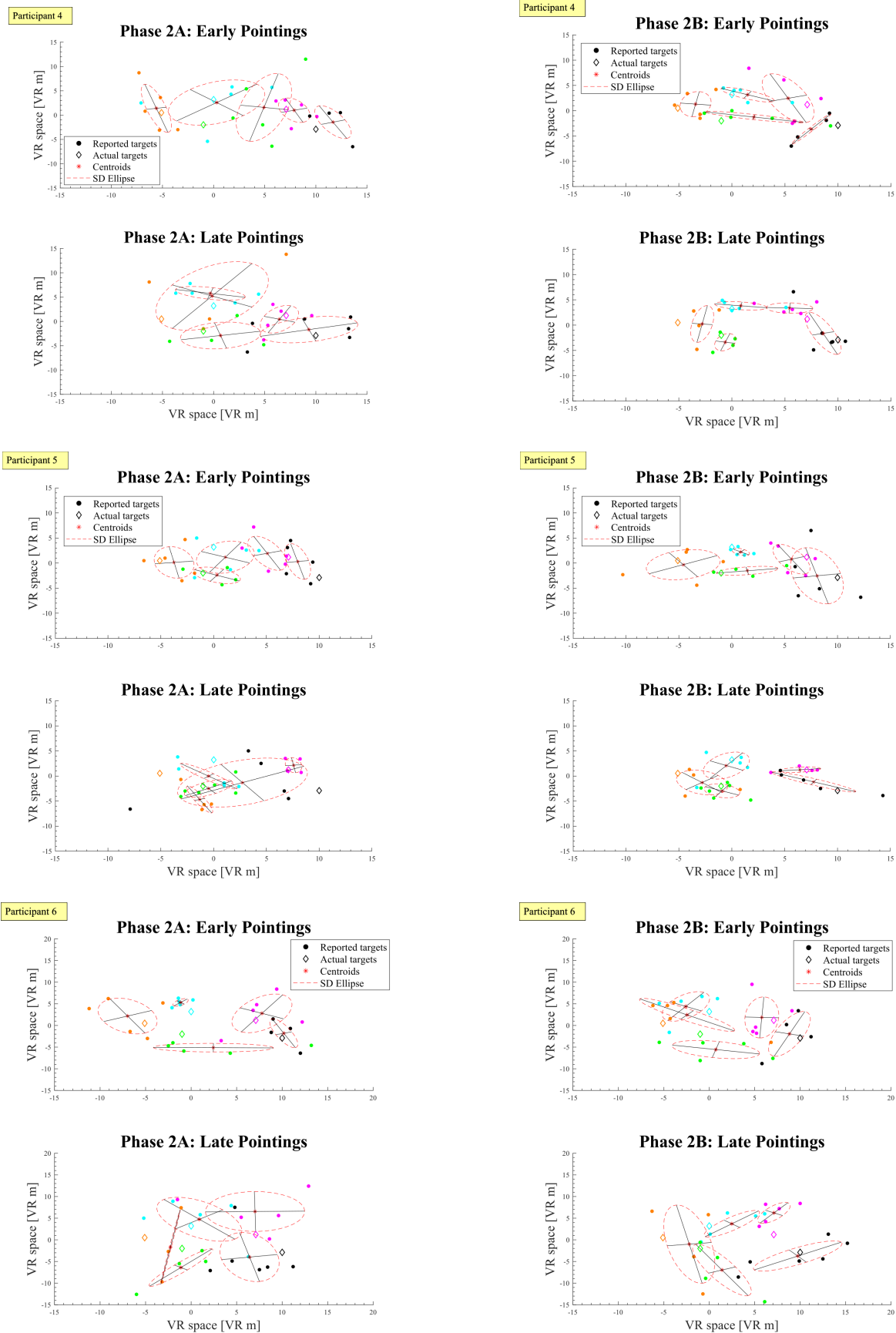


Figure 52: For participant 4 - 6: Visualization of accuracy and precision in early (1-24) versus late pointings (25-48) in Ph2A (a) and Ph2B (b). The covariance (SD) error ellipse surrounds all the pointings (colored dots) directed at the same true target location (colored diamond). "Shots" directed at the same target are indicated by the same color, as is the corresponding actual target location..

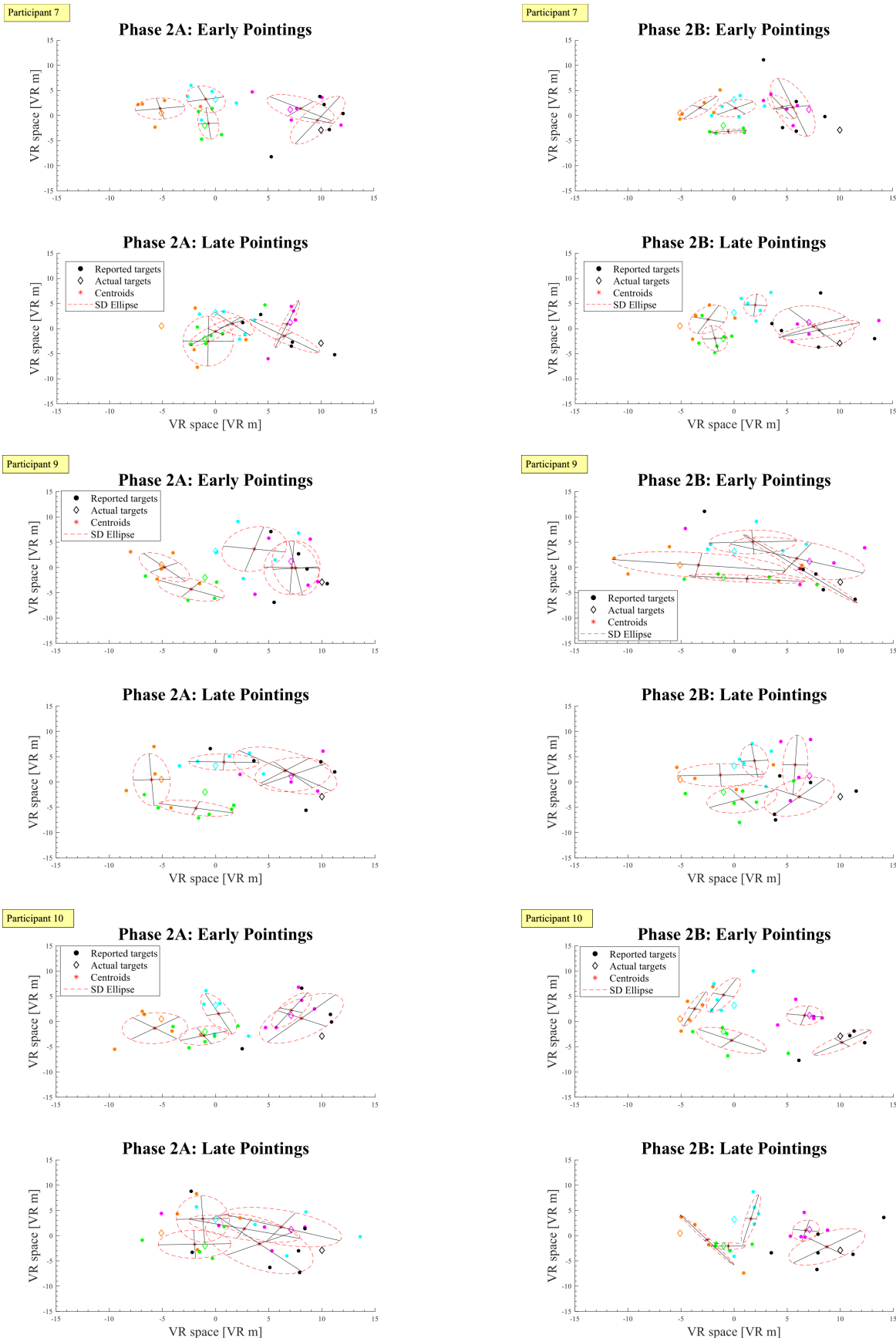


Figure 53: For participant 7,9 and 10: Visualization of accuracy and precision in early (1-24) versus late pointings (25-48) in Ph2A (a) and Ph2B (b). The covariance (SD) error ellipse surrounds all the pointings (colored dots) directed at the same true target location (colored diamond). "Shots" directed at the same target are indicated by the same color, as is the corresponding actual target location.

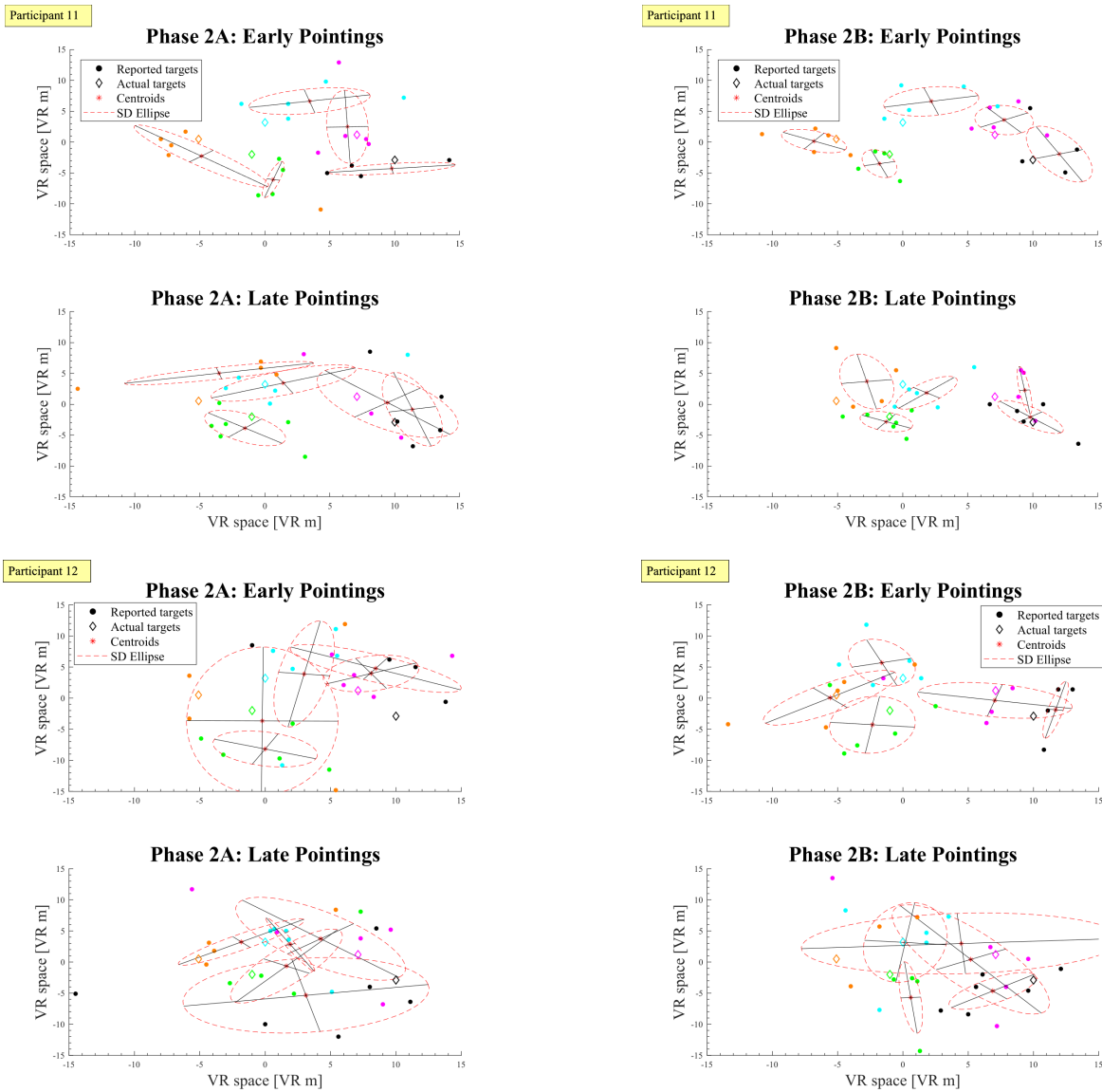


Figure 54: For participant 11 and 12: Visualization of accuracy and precision in early (1-24) versus late pointings (25-48) in Ph2A (a) and Ph2B (b). The covariance (SD) error ellipse surrounds all the pointings (colored dots) directed at the same true target location (colored diamond). "Shots" directed at the same target are indicated by the same color, as is the corresponding actual target location.

8.7 Sketched target locations

	Phase 1	Phase 2A	Phase 2B	
Participant number	1			
	2			
	3			
	4			
	5			
	6			
	7			
	8			

Figure 55: For participant number 1 to 8: Photograph of each participant’s sketched target locations for each of the three phases. The sketches of Ph1 and Ph2A were done as part of the questionnaire at the end of visit 1, Ph2B at the end of visit 2.

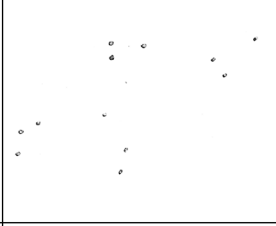
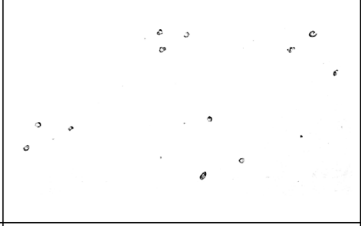
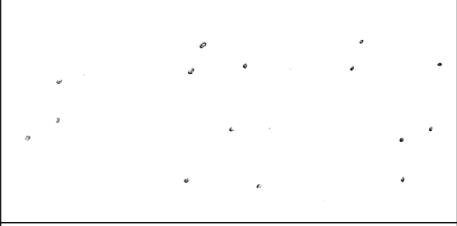
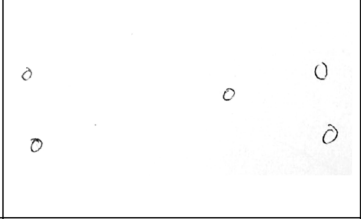

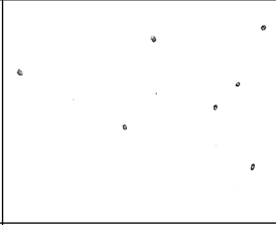
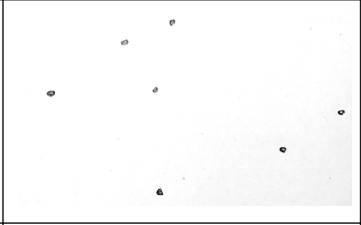
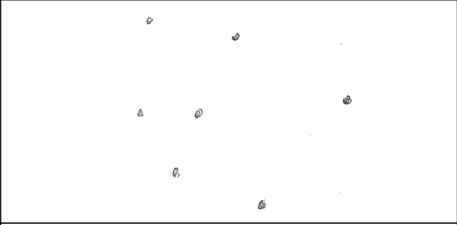



		Phase 1	Phase 2A	Phase 2B
Participant number	9			
	10	gan 7 random „completely random“		
	11			
	12			

Figure 56: For participant number 9 to 12: Photograph of each participant’s sketched target locations for each of the three phases.

9 References

- Barry, C. and Burgess, N. (2014). Neural mechanisms of self-location. *Current Biology*, 24(8):R330–R339.
- Basten, K. and Mallot, H. A. (2010). Simulated visual homing in desert ant natural environments: efficiency of skyline cues. *Biological cybernetics*, 102(5):413–425.
- Bell, S. and Archibald, J. (2011). Sketch mapping and geographic knowledge: What role for drawing ability? In *Understanding and Processing Sketch Maps@ COSIT*, pages 1–10.
- Bennett, A. T. (1996). Do animals have cognitive maps? *Journal of Experimental Biology*, 199(1):219–224.
- Bernal, A., Mateo-Martínez, R., and Paolieri, D. (2020). Influence of sex, menstrual cycle, and hormonal contraceptives on egocentric navigation with or without landmarks. *Psychoneuroendocrinology*, 120:104768.
- Brodth, S., Pöhlchen, D., Flanagan, V. L., Glasauer, S., Gais, S., and Schönauer, M. (2016). Rapid and independent memory formation in the parietal cortex. *Proceedings of the National Academy of Sciences*, 113(46):13251–13256.
- Brown, G. D. (1984). A frequency count of 190,000 words in the london-lund corpus of english conversation. *Behavior research methods, instruments, & computers*, 16(6):502–532.
- Castelli, L., Corazzini, L. L., and Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human behavior*, 24(4):1643–1667.
- Chrastil, E. R., Sherrill, K. R., Hasselmo, M. E., and Stern, C. E. (2015). There and back again: hippocampus and retrosplenial cortex track homing distance during human path integration. *Journal of Neuroscience*, 35(46):15442–15452.
- Chrastil, E. R. and Warren, W. H. (2014). From cognitive maps to cognitive graphs. *PloS one*, 9(11):e112544.
- Coluccia, E. and Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of environmental psychology*, 24(3):329–340.
- Doeller, C. F., Barry, C., and Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, 463(7281):657–661.

- Fearnley, S. (1997). Mrc psycholinguistic database search program. *Behavior Research Methods, Instruments, & Computers*, 29(2):291–295.
- Field, A. (2009). *Discovering Statistics Using SPSS, Third Edition*. SAGE publications.
- Finnegan, D. J., O’Neill, E., and Proulx, M. J. (2016). Compensating for distance compression in audiovisual virtual environments using incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 200–212.
- Foo, P., Warren, W. H., Duchon, A., and 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.
- Gallistel, C. R. (1990). *The organization of learning*. The MIT Press.
- Gibson, B. M. (2001). Cognitive maps not used by humans (homo sapiens) during a dynamic navigational task. *Journal of Comparative Psychology*, 115(4):397.
- Gillner, S. and Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of cognitive neuroscience*, 10(4):445–463.
- Golledge, R. G. et al. (1999). *Wayfinding behavior: Cognitive mapping and other spatial processes*. JHU press.
- Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F., and Kearney, J. K. (2010). How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception (TAP)*, 7(4):1–18.
- Guterstam, A., Björnsdotter, M., Gentile, G., and Ehrsson, H. H. (2015). Posterior cingulate cortex integrates the senses of self-location and body ownership. *Current Biology*, 25(11):1416–1425.
- Hopf, G. (2008). *Iterierte Dreiecksvervollständigung in virtueller Realität - Können metrische Informationen in das räumliche Langzeitgedächtnis übernommen werden?* Unpublished Thesis, University of Tübingen.
- IBM (Released 2020). *IBM SPSS Statistics for Windows, Version 27.0*. Armonk, NY: IBM Corp.
- Kessels, R. P., de Haan, E. H., Kappelle, L. J., and Postma, A. (2001). Varieties of human spatial memory: a meta-analysis on the effects of hippocampal lesions. *Brain Research Reviews*, 35(3):295–303.

- Kim, B., Lee, S., and Lee, J. (2007). Gender differences in spatial navigation. *World Academy of Science, Engineering and Technology*, 31:297–300.
- Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., and Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological science*, 9(4):293–298.
- Lisman, J. E. and Grace, A. A. (2005). The hippocampal-vta loop: controlling the entry of information into long-term memory. *Neuron*, 46(5):703–713.
- Loomis, J. M., Da Silva, J. A., Fujita, N., and Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):906.
- Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., and Fry, P. A. (1993). Nonvisual navigation by blind and sighted: assessment of path integration ability. *Journal of Experimental Psychology: General*, 122(1):73.
- Mallot, H. A. (2012). Raumkognition. In *Kognitive Neurowissenschaften*, pages 217–224. Springer.
- McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006). Path integration and the neural basis of the ‘cognitive map’. *Nature Reviews Neuroscience*, 7(8):663–678.
- Moar, I. and Bower, G. H. (1983). Inconsistency in spatial knowledge. *Memory & Cognition*, 11(2):107–113.
- Montello, D. R. (2005). *Navigation*. Cambridge University Press.
- Montello, D. R., Lovelace, K. L., Golledge, R. G., and Self, C. M. (1999). Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American geographers*, 89(3):515–534.
- Montufar, J., Arango, J., Porter, M., and Nakagawa, S. (2007). Pedestrians’ normal walking speed and speed when crossing a street. *Transportation Research Record*, 2002(1):90–97.
- Nadel, L. (2013). Cognitive maps.
- Oberauer, K., Awh, E., and Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(1):1.

- O'keefe, J. and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Philbeck, J. W. and Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):72.
- Pure Relaxing, V. (2019). Desert Wind Sound - 10 Hours - Stress Relief | Meditate - Sleep - Study | Desert Winds (Video). <https://www.youtube.com/watch?v=uKkJ0etAO5st=3066s>. (Last accessed 14.01.2021).
- Redish, A. D. and Touretzky, D. S. (1997). Cognitive maps beyond the hippocampus. *Hippocampus*, 7(1):15–35.
- Relax Sleep, A. (2018). A Quiet Day with Birds Singing and Light Wind in the Green Forest - Sounds for Relaxation and Sleep (Video). <https://www.youtube.com/watch?v=cvQLjfLw644>. (Last accessed 14.01.2021).
- Riecke, B. E., Veen, H. A. H. C. v., and Bülthoff, H. H. (2002). Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence: Teleoperators and Virtual Environments*, 11(5):443–473.
- Ruddle, R. A. and Péruch, P. (2004). Effects of proprioceptive feedback and environmental characteristics on spatial learning in virtual environments. *International Journal of Human-Computer Studies*, 60(3):299–326.
- Schoeder, J. (2018). *calculation based on and adapted from the author's code*. Scientific work done at the University of Tübingen, provided by mutual supervisor Hardiess, G.
- Sherrill, K. R., Erdem, U. M., Ross, R. S., Brown, T. I., Hasselmo, M. E., and Stern, C. E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. *Journal of Neuroscience*, 33(49):19304–19313.
- Steck, S. and Mallot, H. (2000a). Knowledge of landmark configuration does not improve metric performance in virtual environment navigation. In *Annual Meeting of the Association for Research in Vision and Ophthalmology (ARVO 2000)*, page S224. Association for Research in Vision and Ophthalmology, etc.
- Steck, S. D. and Mallot, H. A. (2000b). The role of global and local landmarks in virtual environment navigation. *Presence: Teleoperators & Virtual Environments*, 9(1):69–83.
- Voyer, D., Postma, A., Brake, B., and Imperato-McGinley, J. (2007). Gender differences in object location memory: A meta-analysis. *Psychonomic bulletin & review*, 14(1):23–38.

- Wan, X., Wang, R. F., and Crowell, J. A. (2012). The effect of landmarks in human path integration. *Acta psychologica*, 140(1):7–12.
- Wang, R. F. (2016). Building a cognitive map by assembling multiple path integration systems. *Psychonomic Bulletin & Review*, 23(3):692–702.
- Warren, W. H. (2006). The dynamics of perception and action. *Psychological review*, 113(2):358.
- Warren, W. H. (2019). Non-euclidean navigation. *Journal of Experimental Biology*, 222(Suppl 1).
- Wehner, R., Srinivasan, M. V., et al. (2003). Path integration in insects. *The neurobiology of spatial behaviour*, pages 9–30.
- Whishaw, I. Q., Hines, D. J., and Wallace, D. G. (2001). Dead reckoning (path integration) requires the hippocampal formation: evidence from spontaneous exploration and spatial learning tasks in light (allothetic) and dark (idiothetic) tests. *Behavioural brain research*, 127(1-2):49–69.
- Wiener, J. M., Berthoz, A., and Wolbers, T. (2011). Dissociable cognitive mechanisms underlying human path integration. *Experimental brain research*, 208(1):61–71.
- Wiener, J. M. and Mallot, H. A. (2006). Path complexity does not impair visual path integration. *Spatial cognition and computation*, 6(4):333–346.
- Wilson, M. (1987). Mrc psycholinguistic database: Machine usable dictionary. version 2.00. <https://websites.psychology.uwa.edu.au/school/mrcdatabase/mrc2.html>. (Last accessed 16.01.2021).