How far can we get with just visual information? Path integration and spatial updating studies in Virtual Reality

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften
der Fakultät für Mathematik und Physik

der Eberhard-Karls-Universität zu Tübingen

vorgelegt von

Bernhard E. Riecke

aus Reutlingen

2003

1 Summary

How do we find our way around in everyday life? In real world situations, it typically takes a considerable amount of time to get completely lost. In most Virtual Reality (VR) applications, however, users are quickly lost after only a few simulated turns. This happens even though many recent VR applications are already quite compelling and look convincing at first glance. So what is missing in those simulated spaces? Why is spatial orientation there not as easy as in the real world? In other words, what sensory information is essential for accurate, effortless, and robust spatial orientation? How are the different information sources combined and processed?

In this thesis, these and related questions were approached by performing a series of spatial orientation experiments in various VR setups as well as in the real world. Modeling of the underlying spatial orientation processes finally led to a comprehensive framework based on logical propositions, which was applied to both our experiments and selected experiments from the literature. Using VR allowed us to disentangle the different information sources, sensory modalities, as well as possible spatial orientation processes and strategies. It further offered the precise control, repeatability, and flexibility of stimuli and experimental conditions, which is difficult to achieve in real world experiments.

A first series of experiments (part II) investigated the usability of purely visual cues, with particular focus on optic flow, for basic navigation and spatial orientation tasks. According to the prevailing opinion in the literature, those cues should not be sufficient: Proprioceptive and especially vestibular cues are supposedly prerequisites even for simple navigation and spatial orientation tasks if they involve rotations of the observer. Furthermore, visual cues alone are often considered insufficient for good spatial orientation, especially when useful reference points (landmarks) are missing. To test this notion, we conducted a set of experiments in virtual environments where only visual cues were provided. Participants had to execute simulated turns, reproduce distances or perform triangle completion tasks. Most experiments were performed in a simulated 3D field of blobs, thus restricting navigation strategies to path integration based on optic flow. For our experimental setup (half-cylindrical 180°x50° projection screen), optic flow information alone proved to be sufficient for untrained participants to perform turns and reproduce distances with negligible systematic errors, irrespective of movement velocity. Path integration by optic flow was sufficient for homing by triangle completion, but homing distances were biased towards the mean response. Additional landmarks that were only temporarily available did not improve homing performance. Navigation by stable, reliable landmarks, however, led to almost perfect homing performance. Compared to similar experiments using virtual environments (Kearns et al., 2002; Péruch et al., 1997) or blind locomotion (Loomis et al., 1993; Klatzky et al., 1990), we did not find any distance undershoot or strong regression towards mean turn responses. Using a Virtual Reality setup with a half-cylindrical 180° projection screen allowed us to demonstrate that visual path integration without any vestibular or kinesthetic cues *can* indeed be sufficient for elementary navigation tasks like rotations, translations, and homing via triangle completion.

Nevertheless, we did observe some systematic errors that could not be convincingly explained by the literature or by the experiments themselves. A detailed analysis of participants' behavior suggested that general cognitive abilities and mental spatial reasoning abilities in particular might have been the determining factor. Positive correlations between navigation performance and mental spatial abilities test scores corroborated this hypothesis. In comparable real world situations, however, no higher cognitive processes seem to be needed (even animals as simple as ants can perform comparable homing tasks). Instead, we seem to know automatically and effortlessly where relevant objects in our immediate surround are when moving about without having to think much about it. Hence, we hypothesized that this "automatic spatial updating" of self-to-surround relations during ego-motion was not functioning properly in our and many other VR studies. So what was missing in the simulations? The literature suggests that vestibular cues from physical motions are indispensable for

automatic spatial updating. Furthermore, visual cues alone should be insufficient, especially when ego-rotations are involved.

To test these hypotheses, we established a rapid pointing paradigm and performed a second series of experiments that investigated the influence and interaction of visual and vestibular stimulus parameters for spatial updating in real and virtual environments (part III). After real and/or visually simulated ego-turns, participants were asked to accurately and quickly point towards different previously-learned target objects that were currently not visible. The rapid egocentric response ensured that participants could not solve the task cognitively.

Unpredicted by the literature, visual cues alone proved sufficient for excellent automatic spatial updating performance even without any vestibular motion cues. Furthermore, participants were virtually unable to ignore or suppress the visual stimulus even when explicitly asked to do so. This indicates that the visual cues alone were even sufficient to evoke reflex-like "obligatory spatial updating". Comparing performance in the real environment and a photorealistic virtual replica revealed similar performance as long as the field of view was the same. That is, a simulated view onto a consistent, landmark-rich environment was as powerful in turning our mental spatial representation (even against our own conscious will) as a corresponding view onto the real world. This highlights the power and flexibility of using highly photorealistic virtual environments for investigating human spatial orientation and spatial cognition. It furthermore validates our VR-based experimental paradigm, and suggests the transferability of results obtained in this VR setup to comparable real world tasks. From a number of additional parameters investigated, only the field of view and the availability of landmarks had a consistent influence on spatial updating performance. Unexpectedly, motion parameters did not show any clear influence, which might be interpreted as a dominant influence of static visual (display) information over dynamic (motion) information.

Modeling spatial orientation processes in a comprehensive framework based on logical propositions (part IV) allowed for a deeper understanding of the underlying mechanisms in both our experiments and experiments from the literature. Furthermore, the logical structure of the framework suggests novel ways of quantifying spatial updating and "spatial presence" (which can be seen as the consistent feeling of being in a specific spatial context, and intuitively knowing where one is with respect to the immediate surround). In particular, it allows the disambiguation between two complementary types of automatic spatial updating found in our experiments: On the one hand, the well-known "continuous spatial updating" induced by continuous movement information. On the other hand, a novel type of discontinuous, teleport-like "instantaneous spatial updating" that allowed participants to quickly adopt the reference frame of a new location without any explicit motion cues, just by presenting a novel view from a different viewpoint. Last but not least, the framework suggested novel experiments and experimental paradigms, was used to generate new hypotheses and testable predictions, and already stimulated the scientific discussion in the presence research community.

In addition to assessing spatial cognition, the logical framework proved helpful in tackling the human-computer-interface issue. Several critical simulation and display parameters required for quick and effortless spatial orientation were pinpointed: First of all, any application that does not enable automatic spatial updating is bound to decrease quick and effortless spatial orientation performance and hence unnecessarily increase cognitive load. In addition, most current VR-displays do not allow for effective ego-motion simulation and/or tend to produce rather large artifacts in ego-motion perception. This is especially true for head-mounted displays. Hence, the importance of designing effective VR displays can hardly be overestimated. Furthermore, the simulated objects should be salient enough, non-repetitive, and constitute one coherent scene that can be updated as a whole. Maybe most critical, the physical reference frame of the VR display and the surround should become "transparent", i.e, vanish perceptually or at least be clearly dominated by the simulated (i.e., intended) spatial reference frame. Failure to do so will lead immersion and spatial presence to decrease, resulting in impaired spatial updating, which in turn prevents quick and effortless spatial

orientation. Thus, by gaining a deeper understanding of how the different sensory cues are integrated in the human brain (spatial cognition aspect) we also approach human factors issues. This highlights the truly interdisciplinary nature of this research area and opens up potential applications.

2   Zusammenfassung in deutscher Sprache

Wie finden wir uns tagtäglich in unserer Umgebung zurecht? In realen Umgebungen dauert es relativ lange bis wir komplett die Orientierung verloren haben. In den immer häufiger verwendeten virtuellen Umgebungen hingegen finden sich Benutzer oft schon nach wenigen simulierten Drehungen nicht mehr zurecht. Dies geschieht, obwohl viele neuere Virtual Reality (VR) Anwendungen auf den ersten Blick überzeugend und realistisch aussehen. Was also fehlt diesen simulierten Umgebungen? Warum ermöglichen sie keine genauso gute Raumorientierung wie ihr reales Pendant? Welche Sinnesinformationen sind essentiell für genaue, mühelose, und robuste Raumorientierung? Und wie werden die verschiedenen Informationsquellen im Gehirn zusammengefügt und verarbeitet?

Diese und andere offene Fragen waren Untersuchungsgegenstand dieser Arbeit. Dazu wurden eine Reihe von Raumorientierungsexperimenten in verschiedenen VR-Versuchsaufbauten und verschiedenen virtuellen als auch realen Umgebungen durchgeführt. Die Verwendung virtueller Umgebungen erlaubte dabei die experimentelle Unterscheidung zwischen verschiedenen Informationsquellen, Sinnesmodalitäten, Raumorientierungsprozessen und -strategien. Zudem ergab sich dadurch eine präzise Kontrolle, hohe Flexibilität und problemlose Reproduzierbarkeit der Stimuli und der experimentellen Bedingungen, was in natürlichen Umgebungen so kaum zu erreichen ist. Abschließend wurden die zugrunde liegenden Raumorientierungsprozesse theoretisch modelliert. Dieses umfassende, auf logischen Verknüpfungen aufgebaute Modell wurde sowohl auf unsere eigenen Experimente als auch Befunde aus der Literatur angewendet.

Eine erste Serie von Experimenten (Teil II) untersuchte die Verwendbarkeit rein visueller Stimuli und insbesondere optischen Flusses für grundlegende Navigations- und Raumorientierungsaufgaben. Gemäß der in der Literatur vorherrschenden Meinung dürfte dies nicht zufrieden stellend möglich sein: Angeblich seien propriozeptive und vestibuläre Reize unabdingbar selbst für einfachste Navigations- und Raumorientierungsaufgaben, insbesondere wenn diese Beobachterrotationen beinhalten. Zudem seien visuelle Stimuli allein unzureichend für gute Raumorientierung, vor allem wenn nützliche Orientierungspunkte (Landmarken) fehlen. Wir überprüften diese Behauptung und führten eine experimentelle Reihe durch, in der in verschiedenen virtuellen Umgebungen ausschließlich visuelle Stimuli dargeboten wurden. Versuchspersonen sollten dabei simulierte Drehungen ausführen, zurückgelegte Distanzen reproduzieren und Dreiecksvervollständigungsaufgaben durchführen. Die meisten Experimente wurden in einer simulierten 3D Punktewolke durchgeführt, wodurch mögliche Navigationsstrategien auf Pfadintegration anhand optischen Flusses eingeschränkt wurden. Untrainierte Versuchspersonen konnten in der verwendeten Versuchsumgebung (eine halbzylidrische 180°x50° Projektionsleinwand) allein mit Hilfe optischen Flusses Drehungen ausführen und Distanzen reproduzieren. Ihre systematischen Fehler waren dabei vernachlässigbar und unabhängig von der Bewegungsgeschwindigkeit. Pfadintegration basierend auf optischem Fluss erlaubte ebenfalls Dreiecksvervollständigung, die Distanzantworten waren jedoch zum Mittelwert hin verschoben. Zusätzliche, nur vorübergehend sichtbare Landmarken zeigten keinen Einfluss auf die Heimfindeleistungen. Verlässliche Landmarken hingegen ermöglichten nahezu perfekte Heimfindeleistungen. Wir fanden bei allen Experimenten weder eine Distanzüberschätzung noch eine beträchtliche Tendenz zu mittleren Drehwinkeln, wie sie in ähnlichen Dreiecksvervollständigungsexperimenten in virtuellen Umgebungen (Kearns et al., 2002; Péruch et al., 1997) oder bei Laufexperimenten mit verbundenen Augen (Loomis et al., 1993; Klatzky et al., 1990) beobachtet wurden. Unter Verwendung eines VR-Aufbaus mit einer halbzylindrischen 180° Leinwand konnten wir somit zeigen, dass visuelle Pfadintegration ohne zugehörige vestibuläre oder kinästhetische Reize für elementare Navigationsaufgaben wie Rotationen, Translationen, und Heimfinden nach Dreiecksexkursion prinzipiell ausreichen kann.

Dennoch beobachteten wir systematische Fehler, die nicht zufrieden stellend durch die Literatur oder die Experimente selbst erklärt werden konnten. Eine genauere Analyse des Antwortverhaltens

ließ vermuten, dass allgemeine kognitive Fähigkeiten und insbesondere das räumliche Vorstellungs-vermögen die Heimfindeleistungen beeinflussten. Positive Korrelationen zwischen den Navigations-leistungen und zwei Tests zum räumlichen Vorstellungsvermögen bestärkten diese Hypothese. In vergleichbaren Situationen in realen Umgebungen scheinen höhere kognitive Prozesse jedoch nicht nötig zu sein. (Selbst einfache Tiere wie Ameisen können vergleichbare Heimfindeaufgaben pro-blemlos ausführen.) In realen Umgebungen scheinen wir also auch während unserer Eigenbewegun-gen automatisch und mühelos zu wissen, wo relevante Objekte in unserer unmittelbaren Umgebung sind, auch ohne darüber nachdenken zu müssen. Folglich stellten wir die Hypothese auf, dass dieses "automatische spatial updating" der selbst-zur-Umgebung-Relationen während der simulierten Ei-genbewegungen in unseren und vielen anderen VR Untersuchungen nicht richtig funktionierte. Was also fehlte in diesen Simulationen? Die Literatur legt nahe, dass automatisches spatial updating oh-ne die vestibulären Sinnesreize physikalischer Bewegungen nicht ausreichend funktioniert. Zudem seien visuelle Stimuli allein unzureichend, speziell für Beobachterdrehungen.

Um diese Hypothesen zu testen, führten wir schnelle Zeigebewegungen als Messgröße ein und führ-ten damit eine zweite Experimentreihe durch (Teil III). Diese untersuchte Einfluss und Wechsel-wirkung visueller und vestibulärer Reizparameter für spatial updating in realen und virtuellen Um-gebungen. Nach realen und/oder visuell simulierten Eigendrehungen sollten die Versuchspersonen möglichst genau und schnell auf verschiedene zuvor gelernte Zielobjekte zeigen, die gerade nicht sichtbar waren. Durch die schnelle egozentrische Bewegungsantwort wurde vermieden, dass die Auf-gabe abstrakt-kognitiv gelöst werden konnte.

Entgegen den Vorhersagen der Literatur erwiesen sich die visuellen Stimuli als ausreichend für aus-gezeichnetes automatisches spatial updating, selbst ohne jegliche vestibuläre Bewegungsreize. Zu-dem konnten Versuchspersonen den visuellen Stimulus praktisch nicht ignorieren oder unterdrücken, auch wenn sie ausdrücklich dazu instruiert wurden. Dies bedeutet, dass die visuellen Stimuli sogar ausreichten, um reflexhaftes "obligatorisches spatial updating" auszulösen. Vergleiche zwischen spa-tial updating in einer realen Umgebung und ihrem virtuellen Pendant ergaben vergleichbare Leistun-gen, solange das visuelle Gesichtsfeld gleich groß war. Folglich erwies sich die simulierte Ansicht einer konsistenten Umgebung voller Landmarken als genauso einflussreich im Drehen unserer men-talen Raumrepräsentation (selbst gegen unseren ausdrücklichen Willen) wie die entsprechende Sicht auf die reale Umgebung. Dies zeigt das Potential und die Flexibilität der Verwendung photorealisti-scher virtueller Umgebungen bei der Untersuchung der menschlichen Raumorientierung und Raum-kognition. Zudem validiert es unser VR-basiertes experimentelles Paradigma und lässt vermuten, dass sich die in diesem Versuchsaufbau gefundenen Ergebnisse auf entsprechende reale Situationen übertragen lassen. Von den weiteren untersuchten Parametern hatten lediglich das visuelle Gesichts-feld und die Landmarken einen konsistenten Einfluss auf die spatial updating Leistung. Unerwar-teterweise zeigten Bewegungsparameter keinen klaren Einfluss, was als Dominanz der statischen (Display-) Information über die dynamische (Bewegung-) Information interpretiert werden könnte.

In Teil IV wurden die beteiligten Raumorientierungsprozesse theoretisch modelliert. Das resultie-rende Modell basiert auf logischen Verknüpfungen und ermöglicht dadurch ein tieferes Verständnis der zugrunde liegenden Mechanismen in unseren Experimenten als auch Experimenten der Literatur. Zudem ließen sich aus den logischen Verknüpfungen des Modells neuartige Methoden zur Quantifi-zierung von spatial updating und "spatial presence" ableiten. (Räumliche Präsenz bezeichnet hierbei das konsistente Empfinden, in einem bestimmten räumlichen Kontext zu sein und intuitiv zu wis-sen, wo man sich bezüglich der unmittelbaren Umgebung befindet.) Insbesondere konnte so klarer zwischen zwei komplementären Arten von automatischem spatial updating unterschieden werden, die in unseren Experimenten beobachtet wurden: Einerseits das bekannte "kontinuierliche spatial updating", welches durch kontinuierliche Bewegungsstimuli ausgelöst wird; andererseits eine neue Art von diskontinuierlichem, Teleport-artigem, "augenblicklichen (instantaneous) spatial updating", welche es Versuchspersonen ermöglichte, den Referenzrahmen eines neuen Ortes augenblicklich an-

zunehmen. Dies geschah ohne jegliche Bewegungsreize, nur indem eine neue Ansicht aus einem anderen Blickwinkel präsentiert wurde. Nicht zuletzt motivierte das Modell neuartige Experimente und experimentelle Paradigmen, wurde benutzt um Hypothesen und überprüfbare Vorhersagen abzuleiten, und regte bereits die wissenschaftliche Diskussion in der presence community an.

Zusätzlich zu dem untersuchten Raumkognitionsaspekt erwies sich das logische Modell als hilfreich, um das Problem der Mensch-Maschine-Schnittstelle anzugehen. Mehrere für schnelle und mühelose Raumorientierung relevante Simulations- und Displayparameter konnten so herausgearbeitet werden: Vor allem zeigte sich, dass Anwendungen, die kein automatisches spatial updating ermöglichen, die schnelle und mühelose Raumorientierung behindern und so die kognitive Last unnötig erhöhen. Des weiteren ermöglichen die meisten zur Zeit verwendeten VR-Aufbauten keine überzeugende Eigenbewegungssimulation und/oder führen zu beträchtlichen Artefakten in der Eigenbewegungswahrnehmung, insbesondere bei Verwendung von Visualisierungsbrillen (head-mounted displays). Dementsprechend kann die Notwendigkeit der Entwicklung effektiver VR-Aufbauten kaum überschätzt werden. Die simulierten Objekte sollten zudem auffällig (salient) genug und nicht repetitiv sein, sowie eine kohärente Szene bilden, die als Ganzes transformiert (spatially updated) werden kann. Der vielleicht kritischste Punkt ist, dass der reale Referenzrahmen des VR Display und der Umgebung "transparent" werden sollte, also nicht mehr wahrgenommen oder zumindest von dem simulierten (d.h. intendierten) räumlichen Referenzrahmen klar dominiert werden sollte. Ansonsten würden sich Immersion und räumliche Präsenz verringern, was wiederum spatial updating verschlechtert und dadurch letztendlich schnelle und mühelose Raumorientierung verhindert. Indem wir also ein tieferes Verständnis dafür gewinnen, wie die verschiedenen Sinnesreize im Gehirn integriert werden (Raumkognitionsaspekt) nähern wir uns gleichzeitig der Problematik der Mensch-Maschine-Schnittstelle. Die unterstreicht den wahrhaft interdisziplinären Ansatz dieser Forschungsrichtung und eröffnet interessante Anwendungsmöglichkeiten.

Contents

List of Figures

List of Tables

Part I

General introduction

3   Prologue

Imagine you are at home at night when the main fuse blows. You will have to find your way around in complete darkness until you manage to find candles, the fuse box, or some other light source. Luckily, even though you cannot see anything, you are not completely lost. Instead, you somehow always have some rough idea of where you are with respect to the surround, even while moving around without vision. How does this work? What happens in our brain when locomoting?

Apparently, some largely automated process keeps track of where surrounding objects are with respect to ourselves, and automatically generates an expectancy of where everything is even when we move around with our eyes closed or in the dark. When we re-open our eyes, this expectancy is normally met quite well. That is, we are not surprised at all that we do not see the same view as before, but a completely different view with all self-to-objects relations changed. No cognitive effort seems to be required for this remarkable "spatial updating" of the world inside our head during ego-motions (even though the necessary mental spatial transformations can be quite complex). This geometrical transformation of the egocentric mental spatial reference frame of our immediate surround seems to occur automatically, apparently triggered by some cues about the performed motion, without us having to think about it.

Moreover, unless one is quite talented in visualizing, it is hardly possible to keep the original egocentric spatial reference frame statically in mind during ego-motions. You may want to try this by pointing to familiar objects after blindfolded ego-motions. The underlying spatial updating process thus seems to be obligatory in the sense that it is almost impossible to suppress. Spatial updating can consequently be seen as a reflex-like process that is to a large extent beyond conscious control.

But how can we understand and quantify this process that happens whenever we move? What sensory cues are necessary to trigger this reflex-like process? If we understood the relevant sensory cues and their interaction, could we somehow artificially evoke spatial updating and consequently elicit the illusion of self-motion without any physical motion? And can we gain a deeper understanding of this and related spatial orientation processes by incorporating them into a comprehensive theoretical model? These were some of the questions that motivated this thesis, as will be elaborated in detail below[1]. As an in-depth introduction precedes the individuals part of this thesis, we would like to refrain here from a long general introduction. Instead, we will rather anecdotally present a road map that can guide the reader through this paper.

---

[1]Much of the data presented in this thesis has already been published in part in technical reports, journals, or conference proceedings (Riecke, van Veen, & Bülthoff, 1999; Riecke & van Veen, 1999; Bülthoff, Riecke, & van Veen, 2000; Riecke, van Veen, & Bülthoff, 2000, 2000a, 2000b; Riecke, von der Heyde, & Bülthoff, 2001a, 2001b, 2002a, 2003, 2002b; Riecke, van Veen, & Bülthoff, 2002; von der Heyde & Riecke, 2001; Riecke & von der Heyde, 2002). In this manuscript, we refrain from citing the above-mentioned publications and refer instead to the corresponding parts of this document to make this manuscript more coherent.

4   Introduction and road map

How do we find our way around in everyday life? What sensory cues are needed, and how are the different information sources combined and processed? In real world situations, it typically takes a considerable time to get completely lost. In most Virtual Reality (VR) applications, however, users are quickly lost after only a few simulated turns. So what is missing in most VR applications? The literature typically suggests that proprioceptive and especially vestibular cues are required for navigation and spatial orientation tasks involving rotations of the observer. Furthermore, visual cues alone are often considered not sufficient to allow for good spatial orientation, especially when useful reference points (landmarks) are missing.

Based on earlier experiments, we did not agree with this prevailing opinion, and decided to test this notion by conducting a set of experiments in different virtual environments where only visual cues were provided (see part II). Participants were asked to execute turns, reproduce distances or perform triangle completion tasks. In contrast to predictions from the literature, participants performed rather well, even when only presented with an optic flow pattern. In trying to understand the origin of the remaining systematic errors, we realized that participants had essentially no problem estimating the distances traveled or angles turned. Furthermore, they could execute intended turns and translations quite well. But most of them, however, were nevertheless unable to correctly determine the adequate navigation behavior, even though they had in principle all the information needed. Moreover, participants seemed to think a lot before responding, and produced a number of qualitative errors. We suspected that general cognitive or mental spatial reasoning abilities might be the limiting factor here, and positive correlations between navigation performance and mental spatial abilities test scores corroborated this hypothesis.

Under real world conditions, however, no higher cognitive processes are needed to perform simple navigation or spatial orientation tasks. Rather, even when moving, we seem to know quickly, automatically, and effortlessly where we are and where relevant objects in our immediate surround are, without having to think much about it. So what is the difference between situations where this automatism does and does not work?

We hypothesized that this quick and intuitive spatial orientation during ego-motions exists only if the "world in our head" (i.e., the egocentric mental spatial representation of the immediate surround) is constantly kept in alignment with the outside world. So what is needed for this automatic alignment ("automatic spatial updating") during ego-motions? The literature again suggests that vestibular and proprioceptive cues from physical motions are indispensable, and that visual cues alone should be insufficient, especially when ego-rotations are involved.

To test this claim, we performed a series of experiments and established a rapid pointing paradigm to investigate the influence and interaction of visual and vestibular stimulus parameters for spatial updating in real and virtual environments (part III). After real and/or visually simulated ego-turns, participants were asked to quickly point towards different previously-learned target objects that were currently not visible. The rapid egocentric response ensured that participants could not solve the task cognitively. Unpredicted by the literature, visual cues alone proved sufficient for excellent spatial updating performance even without any vestibular motion cues.

Next, we investigated whether motion cues are at all required for spatial updating. We performed a "teleport" experiment where participants were presented with a new view without any continuous visual or vestibular motion information in between. Unexpectedly, the lack of any motion cues did not impair updating performance at all. To account for that finding, we introduced the distinction between the well-known "continuous spatial updating" and a discontinuous, teleport-like "instantaneous spatial updating" that allows for automatized, reflex-like reorientation without any explicit motion cues.

Nevertheless, the results puzzled us, and we decided to model the underlying spatial orientation processes in a theoretical framework based on logical propositions (part IV). Coherently representing our findings as well as findings from the literature in one unifying framework enabled a deeper understanding of the underlying (dis)similar processes and logical interrelations. It further allowed us to pinpoint critical factors for good spatial orientation. Last but not least, it suggests novel experiments and experimental paradigms, allows for testable predictions, and already stimulated the scientific discussion in the presence research community.

Part II

How far can we get with just visual path integration?

5 Introduction

Successful spatial orientation and navigation involve a number of different processes, including sensing the environment, building up a mental spatial representation, and using it (e.g., to plan the next steps). During navigation, one needs to update one's mental representation of the current position and orientation in the environment. The available spatial cues can be classified by the type of information used: Position ("position"- or "recognition-based navigation") or velocity and acceleration ("path integration" or "dead reckoning" Loomis, Klatzky, Golledge, Cicinelli, Pellegrino, & Fry (1993)).

Position- or recognition-based navigation (also called **piloting**) uses exteroceptive information to determine one's current position and orientation. Such information sources include visible, audible or otherwise perceivable reference points, so-called "landmarks" (i.e., distinct, stationary, and salient objects or cues). Many studies have demonstrated the usage and usability of different types of landmarks for navigation purposes. (See Golledge (1999) and Hunt & Waller (1999) for an extensive review.) Only piloting allows for correction of errors in perceived position and orientation through reference points (position fixing) and is thus more suited for large-scale navigation.

Path integration, on the other hand, is based on integrating the perceived velocity or acceleration over time to determine the current position and orientation with respect to some starting point. More generally speaking, path integration is navigation based on means other than position fixing (landmarks), and is thus complementary to piloting (Loomis, Klatzky, Golledge, & Philbeck, 1999). Path integration is based on the perception of time, velocity, and acceleration, and is therefore susceptible to accumulation errors due to the integration process. It is well suited for small-scale navigation and connecting neighboring landmarks, but uncertainty and error increase exponentially with traveled distance. See Loomis et al. (1999) and Klatzky, Loomis, & Golledge (1997) for an overview on human and animal path integration.

For navigation experiments, one might wish to distinguish between the contributions of piloting and path integration. This can be done by excluding one of the two spatial updating cues at a time: Path integration can be rather easily excluded by eliminating all velocity and acceleration information, e.g., through a slide-show type presentation. The elimination of recognition-based spatial updating is more difficult and, perhaps, more critical, as landmarks play a dominant role in normal navigation. The difficulty of navigating in heavy fog or snowfall illustrates this dominance.

Kinesthetic and vestibular cues typically reveal no information about external landmarks, and are as such well suited for path integration studies. *Visual cues* provide information about the location of the objects seen, which can consequently be used for recognition-based navigation. Apart from blindfolding people, the only way to circumvent this navigation-by-landmarks is through displaying optic flow only, (i.e., removing the landmark-character from the visible objects). Methodically, this can be achieved through presenting an abundance of indistinguishable objects that can only be tracked over a short distance. This can be easily implemented using a Virtual Reality setup. The effect is similar to moving through heavy snowfall or flying through clouds that block the vision for all distant landmarks. (See Figure 1.) Warren, Kay, Zosh, Duchon, & Sahuc (2001) have shown that optic flow information can indeed be used for goal-directed walking.

As recognition-based strategies are known to provide sufficient information for accurate homing performance in simple navigation tasks (see section 6 and 7), we focus here on navigation tasks based solely on path integration, without the aid of external reference points (landmarks).

5.1   Outline and motivation

Vestibular and kinesthetic cues are typically thought to be indispensable for navigation and spatial tasks involving ego-rotations (see subsection 5.4). The goal of this study is to test this claim and investigate human navigation and spatial orientation abilities based *solely* on visual path integration. In short: Is visual homing without landmarks possible? More precisely, can the lack of useful vestibular and kinesthetic cues in visually based navigation be compensated for by the external reference frame and broad visual field of view of a curved 180° projection screen?

In the **first Experiment** ("**TURN&GO**", section 6), we investigated how well untrained participants can perform simple rotations and translations, given optic flow information only. If optic flow information is sufficient for performing elementary turns and translations, errors in the subsequent triangle completion tasks can be ascribed to problems in encoding the path traveled and/or in mentally computing the homeward trajectory.

The **second Experiment** ("**LANDMARKS**", section 7) constitutes a baseline for the later experiments: Given an abundance of salient landmarks in a natural-looking virtual environment, how good is visually based homing? If visual cues are indeed sufficient, we expect accurate and precise homing performance.

In the **third Experiment** ("**TOWN&BLOBS**", section 8), we compared homing by optic flow with homing by naturalistic landmarks that were only temporarily available (town with "scene swap"). The primary issues addressed in this experiment are: Is optic flow information alone sufficient for accurate homing? If piloting is the main source for visual navigation, then the elimination of all stable landmarks ("scene swap") should reduce performance to the level in the optic flow condition. If naturalism is important for navigation, optic flow performance should be inferior to "scene swap" performance.

The **fourth Experiment** ("**RANDOM TRIANGLES**", section 9) was designed to investigate the influence of the simplicity of the triangle geometry: How does the homing performance change when each triangle geometry is novel (randomized) instead of isosceles (as in Experiment TOWN&BLOBS)? To our knowledge, so far nobody investigated triangle completion for completely randomized lengths of the first and second segment and the enclosed angle.

Finally, we conducted two standard **mental spatial abilities tests** to investigate whether mental spatial ability might be a determining factor for this type of navigation performance (see section 10).

5.2   Virtual Reality

5.2.1   Definition and applications in spatial cognition

Using Virtual Reality (VR) for experiments on orientation and navigation has several advantages over the classic approach: The **real-time interactivity** of VR makes a closed-loop paradigm possible that is important for studying natural behavior. **Data collection and analysis** can be performed easily and on the fly, allowing for immediate feedback if required. The **experimental design is flexible** and could be changed even during the experiment, depending, for example, on the participant's performance. Most importantly, the **experimental conditions are well-defined** and can easily be **reproduced** (Bülthoff & van Veen, 2001; Loomis, Blascovich, & Beall, 1999).

This is often an advantage over navigation experiments performed in real environments, where it is very difficult to control a number of experimental factors. Among them are weather conditions (e.g., sun position, clouds, visibility of landmarks), existence, location and persistence of landmarks (e.g., parked cars, construction work, people walking around, sound sources) and previous knowledge of the environment. To circumvent these issues, experiments on spatial cognition have often used slide shows, film sequences or models/maps of the environment traveled (Goldin & Thorndyke, 1982). All those experiments had in common that they were either highly unrealistic (models and maps) or not interactive (slide shows and film sequences), thus lacking a (possibly important) component of natural navigation (Flach, 1990). The recent evolution of virtual environments technology provides the opportunity to tackle these issues. The number of studies on human spatial cognition and navigation using VR has rapidly increased over the last years and has given rise to a number of interesting results (see Péruch & Gaunet, 1998; Darken, Allard, & Achille, 1998; Christou & Bülthoff, 1998, for extensive reviews).

5.2.2 Virtual Reality as a tool to disentangle different sensory modalities and render piloting impossible

In addition to the above mentioned properties of VR, we used virtual environments in this study for two specific purposes: To disentangle the different sensory modalities and to render piloting impossible. The virtual environment was presented only visually, thus excluding all spatial cues from other sensory modalities, especially kinesthetic[2] and vestibular cues from physical motion. To reduce proprioceptive cues from motion control to a minimum (and consequently restrain motor learning), participants pressed buttons to control their self-motion, instead of using more sophisticated input devices like data gloves or joysticks. However, in previous experiments we have shown that adding proprioceptive cues through the use of a bicycle as a locomotion device only marginally affected homing performance (Riecke, 1998; van Veen, Riecke, & Bülthoff, 1999). To ensure that participants rely on path integration only, piloting was rendered impossible through presenting optic flow information only (in a 3D field of blobs, see subsection 6.1.4) or through making landmarks only temporarily visible (through "scene swap", see subsection 8.1.2).

5.3 Triangle completion studies

In most of the experiments described in part II of this thesis, we used "triangle completion", a paradigm commonly used for navigation tasks without landmarks: Participants are led along two sides of a given triangle and have to find the shortest way back to the starting position by themselves (see Klatzky et al. (1997), Loomis et al. (1999) for a review). Triangle completion uses the simplest non-trivial combination of translations and rotations. A simple experimental paradigm for path integration studies is blind locomotion with ears muffled. Kearns et al. (2002, exp. 3), Klatzky et al. (1990), Loomis et al. (1993), Marlinsky (1999b), and Sauvé (1989) showed in triangle completion studies that kinesthetic and vestibular cues from blind walking allow for homing, but lead to strong systematic errors. In all five studies, participants showed a considerable regression towards stereotyped responses, for example similar turning angles for different triangle geometries. Qualitatively similar results were found for purely visual triangle completion without salient landmarks. Presentation via head-mounted display (HMD) (Kearns et al., 2002; Duchon, Bud, Warren, & Tarr, 1999) as well as via flat projection screen (Péruch et al., 1997; Wartenberg, May, & Péruch, 1998) led to considerably larger systematic errors than in the blind walking studies. Our results showed in contrast *smaller* systematic errors than the blind walking studies. The above studies will be discussed

---

[2]feedback from muscles, joints, and tendons and motor efferent commands.

in more detail in subsection 11.1, where they will be compared to the experiments presented in this part.

Triangle completion tasks without reliable landmarks can be modeled by three distinct, consecutive processes (Fujita, Klatzky, Loomis, & Golledge, 1993):

1. The **encoding phase** refers to the set of processes leading to an internal representation of the navigated area.

2. **Mental spatial reasoning** is used to compute the desired homing trajectory.

3. In the **execution phase**, the intended trajectory (rotations and translations) is executed.

Errors can potentially occur in all three phases. Several studies attributed all systematic errors to the encoding phase (Fujita et al., 1993; Klatzky et al., 1997; Klatzky, 1999; May & Klatzky, 2000; Péruch et al., 1997; Wartenberg et al., 1998), following the main idea of the "encoding error model" by Fujita et al. (1993).

5.4   Differences between updating translations and rotations

The difficulty in updating rotations from visual cues alone is consistent with observed fundamental differences between the updating of rotations and translations: For example, studies by May, Péruch, & Savoyant (1995) and Chance, Gaunet, Beall, & Loomis (1998) revealed that vestibular and kinesthetic cues are more important for updating rotations than for translations. Simulated turns presented only visually resulted in a reduced spatial orientation ability compared to physical rotations with the same visual input. Chance et al. (1998) suggest *"the advisability of having subjects explore virtual environments using real rotations and translations in tasks involving spatial orientation"* (p. 168). However, simply adding physical movements does not necessarily guarantee better spatial orientation performance, as was demonstrated by Kearns et al. (2002): Response variability decreased, but participants were still insensitive to angles turned.

Rieser (1989) and Presson & Montello (1994) found a similar difference between rotations and translations for *imagined* movements: Updating the location of several landmarks during imagined self-rotations (without translations) proved more difficult and error-prone than during translations (without rotations). Klatzky, Loomis, Beall, Chance, & Golledge (1998) proposed that this difficulty in updating rotations is due to the lack of proprioceptive cues accompanying the self-rotation. Comparing visually presented locomotion with and without physical rotations, Klatzky et al. (1998) conclude that *"optic flow without proprioception, at least for the limited field of view of our virtual-display system, appears not to be effective for the updating of heading"* (p. 297). The first experiment of this thesis demonstrates, however, that optic flow without proprioception *can* indeed be sufficient for correct updating of heading, at least if a wide field of view and a curved projection screen is used (see section 6).

5.5   Influence of field of view and external reference frame

The studies on triangle completion by Péruch et al. (1997) and Kearns et al. (2002) and turning studies by Bakker, Werkhoven, & Passenier (1999, 2001) all used a physical visual field of view

(FOV[3]) that was well below the natural FOV of the human eye. Locomotion was visually presented via projection screen or HMD with a horizontal field of view of 45°, 60°, 24°, and 48° respectively, compared to more than 180° for humans. These studies demonstrated that humans can not use visual information for accurate path integration. Might this be due to the unnaturally limited FOV and/or the missing visibility of one's own body and the physical environment, which might serve as a helpful reference frame? To address these questions, we conducted navigation experiments similar to those by Péruch et al. (1997), but using a half-cylindrical 180° projection screen. Furthermore, three different environments were used, providing different types of spatial information: reliable and salient landmarks, temporarily available landmarks, and no landmarks at all (i.e., optic flow only).

It is known that enlarging the FOV results in a more realistic spatial perception and has a positive influence on motion perception, sense of presence, visual recognition, lane-keeping performance, spatial orientation, spatial updating, navigation, spatial perception and visuomotor activities (Alfano & Michel, 1990; Arthur, 2000; Hendrix & Barfield, 1996a; Kappe, van Erp, & Korteling, 1999; Loomis, Klatzky, & Lederman, 1991; Riecke et al., 2001b; Rieser, Hill, Talor, Bradfield, & Rosen., 1992; Ruddle & Jones, 2001). On the other hand, most displays currently have a rather limited FOV (usually below 60° horizontally). This is especially true for HMDs. Arthur (2000) provides an extensive review on past work as well as several experiments on the influence of FOV in HMDs on task performance. Using a custom-built HMD, he found a significant performance benefit in walking tasks even for enlarging the horizontal FOV from 112° to 176°, which is much wider than the FOV of commercially available HMDs.

Comparisons of HMDs and curved projection systems revealed for HMDs an increased workload, fatigue ratings, and reduced visual target detection performance (Hettinger, Nelson, & Haas, 1996; Nelson, Hettinger, Cunningham, Brickman, Haas, & McKinley, 1998). Moreover, HMDs exclude vision of the physical surround and oneself, which might provide an important reference frame: In visual triangle completion experiments by Riecke (1998, chap. 5.4), participants used the physical reference frame of a half-cylindrical projection screen as an external reference frame to better estimate visual turning angles.

---

[3]The physical field of view (FOV, sometimes referred to as absolute FOV) is a property of the physical setup; it is defined by the angle (horizontal and vertical) under which the observer sees the simulation window. The simulated field of view (sFOV) generated by the computer (also referred to as geometric FOV) in contrast is a property of the simulation. It is defined by the geometry of the viewing frustrum, i.e. by the angle (horizontal and vertical) under which the virtual (simulated) eye point sees the virtual environment. For the experiments presented in this thesis and for most immersive simulations the physical and simulated FOV are kept identical. sFOV > FOV corresponds to a wide angle effect, sFOV < FOV corresponds to a telescope-like view.

6   Experiment 1: "Turn&Go"

Recent evidence suggests that optic flow is sufficient for accurate distance reproduction (Bremmer & Lappe, 1999), but insufficient for ego-rotations, where training is needed to correct for systematic errors (Bakker et al., 1999, 2001; Péruch et al., 1997; Sadalla & Montello, 1989). Typically, a considerable variability and compression towards stereotyped turn responses is found. The first Experiment ("Turn&Go") was designed to test these claims and to investigate how well *untrained* participants are able to perform simple visual turns and translations, given only optic flow information. Rotations and translations constitute the basis for all navigation behavior, as all movements can be decomposed into a combination of those elementary operations. Participants were asked to turn by specific angles and reproduce distances traveled using randomized velocities and a simple button-based motion model.

If participants are able to execute intended turns with relatively small systematic errors and variance, we could argue that turn execution errors play only a minor role in the subsequent triangle completion experiments, too. Hence, observed turning angles would reflect the intended turns and give insight in the spatial representation of the participants. Consequently, we could argue that systematic turn errors in the triangle completion experiments should be ascribed to systematic errors in encoding or mental "computation" of the homeward trajectory (encoding phase or mental spatial reasoning phase, respectively).

If participants are able to reproduce traveled distances with relatively small systematic errors and variance, we could argue that encoding and execution errors are either negligible or cancel each other out. That would suggest that systematic distance errors in the subsequent triangle completion experiments have to be attributed to errors in the mental spatial reasoning phase. If participants are able to properly use path integration by optic flow to derive angles turned and distances traveled, we would expect no correlation between movement velocity and turns executed or distances traveled. A significant correlation on the other hand would suggest the usage of a timing strategy (like counting seconds to estimate distances) or general problems with path integration by optic flow.

6.1   Methods

6.1.1   Participants

For all experiments described in this thesis, participants had normal or corrected-to-normal vision. Participation was always voluntary and paid at standard rates. A group of six female and three male naive participants participated in Experiment Turn&Go and later also in Experiment Random Triangles. Ages ranged from 20 to 36 years (mean: 26.6 years, SD: 4.4 years). A tenth participant had to be excluded from the analysis, as she misunderstood the instructions.

6.1.2   Visualization

Experiments were performed on a SGI Onyx2 3-pipe Infinite Reality2 Engine. The experiment took place in a completely darkened room. Participants were seated in the center of a half-cylindrical projection screen (7m diameter and 3.15m height, see Figure 1), with their eyes at a height of 1.25m. Three neighboring color images of the virtual environment were rendered at an update rate of 36 Hz and projected non-stereoscopically side by side, with a small overlap of 7.5° smoothed by Panomaker Softedge Blending. The resulting image had a resolution of about 3500×1000 pixels and subtended a physical field of view of 180° horizontally by 50° vertically. Physical and simulated field of view (used for the image rendering) were always identical. A detailed description of the setup can be found in van Veen, Distler, Braun, & Bülthoff (1998).

6.1.3   Interaction

Participants used the three mouse buttons as an input device to move through the virtual environment. Pressing the middle button produced forward translations that lasted as long as the button was being pressed. Releasing the button ended the motion. Similarly, the left or right button produced left or right rotations, respectively. Pressing or releasing a button resulted in a short acceleration or deceleration phase, respectively, with a constant maximum velocity in between. The button-based motion model was chosen to reduce proprioceptive cues about the motion to the absolute minimum and hence avoid motor learning.

6.1.4   Scenery

The experiment was performed in a 3D field of blobs that consists of a ground plane and four semi-transparent upper horizontal planes, all textured with randomized blob patterns (see Figure 1). The blob environment was designed to create a compelling feeling of self-motion (vection) using optic flow. The individual, similar looking blobs became blurred for simulated viewing distances larger than about 10m, thus providing no salient landmarks that could be used for position-based navigation strategies. Consequently, participants had to rely on path integration.

Figure 1: Virtual environments lab with 180 degree projection screen displaying the 3D field of blobs. The participant is seated behind the table in the center of the half-cylindrical screen. On the table are mouse and keyboard as input devices.

6.1.5 Procedure

The experimental design is summarized in Table 1. Each participant completed 96 trials, corresponding to a factorial combination of 8 distances × 6 turning angles × 2 turning directions. The range of distances corresponds to the range of homing distances s_3 in the subsequent triangle completion experiments. The range of turning angles was considerably larger than that used in subsequent experiments of part II.
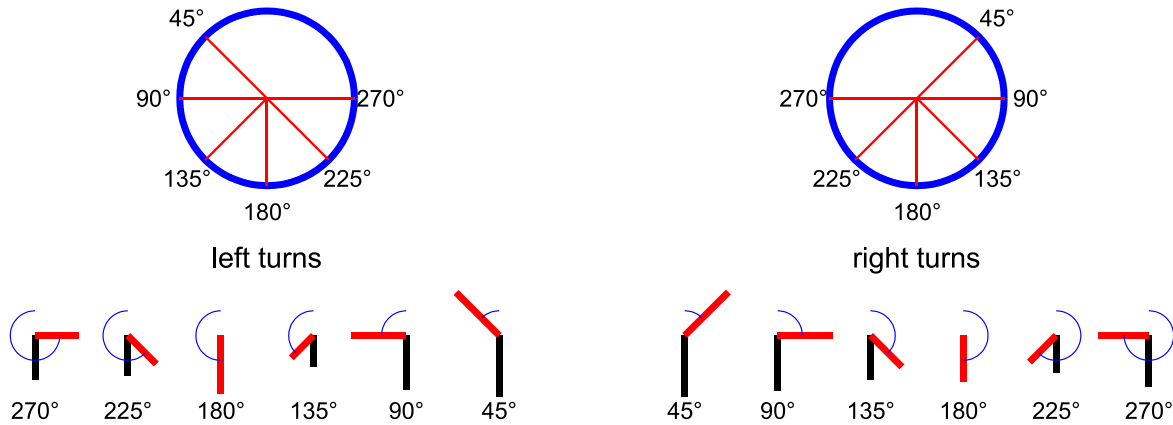
Figure 2: Graphical display used to illustrate the turning angles in the training phase. The angles α to be turned in the turn execution phase are displayed in blue. Examples for the distances s_1 to be reproduced are displayed in black (distance encoding phase), the reproduced distances s_2 in red (distance reproduction phase).

To test the influence of velocity, translational and rotational velocities were randomized independently for each trial and each segment, within an interval centered around the velocity used in the subsequent experiments (see Table 1).

Before the actual experiment, a handout with a graphical representation of the turning angles was shown to the participants (see Figure 2). To ensure that they understood the turning instruction properly, participants were asked to turn physically by angles indicated by the experimenter. Each trial consisted of three phases:

1. **Distance encoding phase**
   Participants were positioned randomly within the 3D field of blobs, facing a yellow "light beam" at a given distance s_1. By pressing the middle mouse button, they moved to the light beam where they stopped automatically upon contact. Turning was disabled during phase 1 and 3.

2. **Turn execution phase**
   Participants were requested to turn using mouse buttons by an angle α_c and in the direction specified by written instructions displayed in the lower part of the screen (e.g., "turn left by 225°"). Translation was disabled during this phase.

3. **Distance reproduction phase**
   Participants were asked to reproduce the distance s_1 from the first phase by traveling that distance in the current direction.

Before the actual experiment, participants performed six practice trials to get accustomed with the interface and the task requirements. Participants were never given any feedback about their performance or accuracy. Just as for the other experiments in part II, there was no time limit for fulfilling the task. The experiment generally lasted about one hour.

	Independent variable	Levels	Values
translations	distance s_1	8 (equally spaced)	$s_1 \in \{20, 28.29, \ldots, 78\}$
	velocity $v_{s_1} = gain_{s_1} \cdot v_0$	randomly selected from a continuous range	$0.75 \leq gain_{s_1} \leq 1.5$
	velocity $v_{s_2} = gain_{s_2} \cdot v_0$	randomly selected from a continuous range	$0.75 \leq gain_{s_2} \leq 1.5$
rotations	turning angle α_c	6 (equally spaced in $45°$ steps)	$\alpha_c \in \{45°, 90°, \ldots, 270°\}$
	turning direction	2	left and right
	rotational vel. $\dot{\alpha} = gain_\alpha \cdot \dot{\alpha}_0$	randomly selected from a continuous range	$0.5 \leq gain_\alpha \leq 2$

Table 1: Experimental design for the TURN&GO experiment. $v_0 = 5m/s$ and $\dot{\alpha}_0 = 40°/s$ are the movement velocities used in the subsequent experiments. Further explanations in the text.

6.1.6   Elimination of outliers

On a few trials, participants accidentally pressed the confirm button before completing the trial or turned in the wrong direction. To reliably eliminate those outliers for all participants, we used the following criterion: A trial was removed if the participant either didn't turn at all or if the turning error was larger than 4 standard deviations. A total of 15 trials or 1.7% of the trials were eliminated due to this criterion.

6.2   Results

There are several ways to look at the data. As the participants' observed responses were clearly linearly correlated to the desired responses, the data were analyzed in terms of signed error and slope (gain factor) of the linear regression for rotations as well as translations. For comparison with the literature, and to give an estimate of the error on a trial-to-trial basis, the absolute error was additionally computed. All of these dependent variables were also used to analyze the subsequent triangle completion data, and are juxtaposed in Figures 13 and 14, allowing for direct comparison among the different experiments. A correlation analysis further reveals the influence of individual dependent variables.

6.2.1   Errors and gain factors

The typical distance reproduction and turn execution performance is displayed in Figure 3 for one representative participant. The individual data for all participants are summarized in the appendix in Figures 54 and 55. The general results are summarized in Figures 13 and 14 for comparison with the other experiments. As for all participants, a linear regression line fits well to the data and captures its main aspects: The slope ("gain factor") for distances (Figure 3 (a)) is slightly less than 1, implying that the range of observed mean reproduced distances is smaller than for the distances to be reproduced. The distance gain factor in this example is 0.9 (0.91 ± 0.05 for all participants), indicating a slight compression of the response range, whereas perfect performance (no compression) would result in a gain factor of 1. The y-intercept above zero indicates a regression (compression) towards distances larger than zero, and not just an overall scaling between stimulus and response.

The angular gain factor (Figure 3 (b)) is 0.99 for this participant and 0.97 ± 0.01 for all participants, indicating negligible systematic errors. There was no significant undershoot or overshoot for distances or turns (see Figure 13). The absolute error for turns and distances is displayed in Figure 14

(a) distance reproduction performance   (b) turning performance
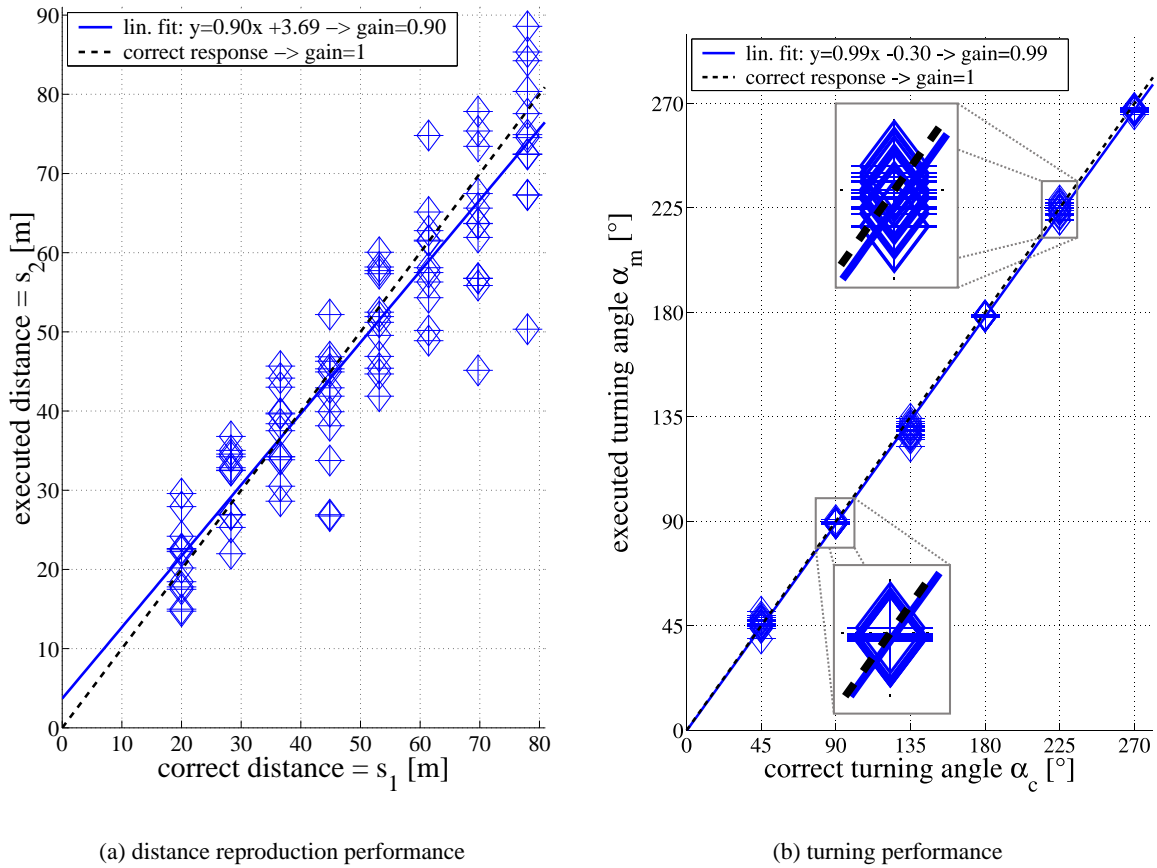
Figure 3: Typical distance reproduction (a) and turn execution performance (b) from one participant. The left and right graphs show the executed distance and turning angle, respectively, plotted versus their corresponding correct values. The distance and angular gain factors are 0.9 and 0.99, respectively, as is indicated in the top inset of each figure. The enlargements in (b) illustrate the extremely small within-subject variability and error for turns, indicating the ease with which the task was performed. The individual data for all participants are summarized in Figures 54 and 55.

to give an estimate of homing accuracy on a trial-to-trial basis and for comparison with the literature. The absolute error for distances was $10.6m \pm 1.7m$ or 23.0% of the distance to be reproduced, whereas the absolute error for turns was merely 5.2%.

6.2.2   Correlation analysis

To investigate the influence of the independent variables individually, we performed pairwise correlation tests between the signed and absolute errors for distances $(s_{3_m} - s_{3_c})$ and turns $(\alpha_m - \alpha_c)$ and the independent variables (see Table 2). The Fisher r-to-Z transformed values of the coefficients of correlation were tested against zero using a two-tailed t-test. The results are summarized in Table 2. Responses were uncorrelated to both translational and rotational velocity. Thus, we can exclude simple timing-based strategies. The signed distance error was negatively correlated to the correct distance (s_1), indicating a compression of the response range. The same was true for turns, but with a much smaller compression (see Figure 13). Absolute error increased for both distances and turns with their corresponding correct values. The absolute distance error can be modeled by a linear regression, revealing a constant absolute error of $b_s = 3.2m$ and a linear contribution with

Independent variable	Correlated with dependent variable			
distance s_1	signed error $s_2 - s_1$	$r = -0.16$	$r^2 = 0.025$	t(8) = 2.4, p = 0.04
	abs. error $|s_2 - s_1|$	$r = 0.31$	$r^2 = 0.097$	t(8) = 5.5, p = 0.0005
	distance s_2	$r = 0.82$	$r^2 = 0.667$	t(8) = 8.9, p < 0.0001
translational velocity v_{s_1}	n.s.			
translational velocity v_{s_2}	n.s.			
translational velocity ratio $v_{s_2}/v_{s_1} = gain_{s_2}/gain_{s_1}$	n.s.			
turning angle α_c	signed error $\alpha_m - \alpha_c$	$r = -0.30$	$r^2 = 0.088$	t(8) = 6.7, p = 0.0002
	abs. error $|\alpha_m - \alpha_c|$	$r = 0.17$	$r^2 = 0.029$	t(8) = 3.0, p = 0.017
	turning angle α_m	$r = 0.999$	$r^2 = 0.998$	t(8) = 9.6, p < 0.0001
turning direction	n.s.			
rotational velocity $\dot{\alpha} = gain_\alpha \cdot \dot{\alpha}_0$	n.s.			

Table 2: Results from the correlation analysis for the TURN&GO experiment.

$a_s = 0.151$ $(|s_2 - s_1| (s_1) = a_s \cdot s_1 + b_s = 0.151 \cdot s_1 + 3.2m)$. The corresponding linear regression for the absolute turning error reveals a much smaller linear contribution of $a_\alpha = 0.024$ $(|\alpha_m - \alpha_c| (\alpha_c) = a_\alpha \cdot \alpha_c + b_\alpha = 0.024 \cdot \alpha_c + 3.4°)$.

To test how well the correct distance or turning angle predicts the observed distance and turning angle, respectively, we performed a similar correlation analysis on them. As expected, the correlation was highly significant for both distances and turns (see Table 2). A r^2 value of 0.67 for distances implies that 67% of the variance in the distance traveled (s_2) can be explained by the distance to reproduce (s_1). For the turning angles, almost the whole variance (99.8%) in angles turned (α_m) can be explained by the angle to turn (α_c), indicating an excellent turning response and a negligible execution error.

6.3   Discussion

The basic translation and rotation tasks in this experiment provide a body of baseline data for more complex navigation behavior in general and for the subsequent triangle completion experiments in particular. The most important conclusion is perhaps what was not to be expected from the literature: That optic flow information alone proved sufficient for untrained participants to execute turns with amazing accuracy. We will argue that the semi-circular display setup and reference frame provided might be the determining factor for this unexpected result. This and other issues will be elaborated upon in detail below.

6.3.1   Turning errors

Contrary to the predictions derived from the literature, participants were able to accurately update rotations (and translations, albeit with reduced accuracy) from optic flow presented on a curved 180° projection screen. Participants had no prior training or explicit feedback whatsoever, but were nevertheless able to accomplish the task relatively well, compared to the literature.

In comparable visual turning experiments using a head mounted display, Bakker et al. (1999, 2001) reported turning errors that were more than ten times larger than in the current experiment (for signed error, absolute error, and between-subject variability). Only within-subject variability was

at a comparable level. Directly after feedback training, errors in the Bakker et al. (2001) study were reduced, but still about three times larger than in the TURN&GO Experiment (and increased on the following day).

The reasons for the observed huge performance differences are not fully understood yet. The main difference between our experiments and the literature is the display setup used, i.e., the half-cylindrical projection screen. Hence we suggest that the display setup and reference frame provided play a mayor role that needs to be investigated in future studies. This hypothesized influence of the FOV is corroborated by comparing the two studies by Bakker et al.: A horizontal FOV of 48° led to systematic overshooting or underestimation of the turns (Bakker et al., 2001), whereas a smaller horizontal FOV of 24° led to systematic undershooting which was about twice as large (Bakker et al., 1999). However, merely using a projection screen instead of an HMD does not necessarily get rid of systematic errors: Using a flat projection screen with a FOV of 45°, Péruch et al. (1997) found a significant undershoot of 16% for rotations. In recent experiments, Schulte-Pelkum, Riecke, von der Heyde, & Bülthoff (2003), Schulte-Pelkum, Riecke, & von der Heyde (2003) demonstrated that projection screen curvature is also a critical parameter for turning angle estimates: Changing screen curvature resulted even in larger performance differences than variations in the physical FOV.

6.3.2   Distance errors

As predicted by the literature, participants were able to integrate velocity and acceleration information derived from optic flow to estimate distances traveled, without any training and irrespective of movement velocity. There was no significant undershoot or overshoot for distances (see Figure 13). However, distances showed a considerable absolute error, which was about four times higher than for the turning task. Furthermore, distances were slightly, but insignificantly, compressed towards stereotyped responses. Compared to the results by Bremmer & Lappe (1999), we found a slight compression, but no general overshoot. The differences might be explained by differences in the experimental paradigm: Bremmer & Lappe did not use an intervening turning task, participants could actively control their velocity in the reproduction task, and had previously accomplished a distance discrimination task.

6.3.3   Conclusions and predictions

We conclude that participants did not use a simple, time-based strategy to estimate angles turned or distances traveled. Turn execution errors and variability were negligible, implying that any potential turning errors in the subsequent triangle completion experiments have to be ascribed to either the encoding process or problems with the mental "computation" of the homing trajectory. If participant had no problems in mental spatial reasoning, distance responses in the subsequent triangle completion tasks should be similar to Experiment TURN&GO (no overall signed error, gain=0.91, and considerable variability). Larger systematic errors, on the other hand, would indicate problems in mental "computation" of the homing trajectory.

7 Experiment 2 : "LANDMARKS"

The second experiment was designed to establish a baseline performance for visual homing, for comparison with the subsequent experiments, which investigated visual navigation performance *without* any stable, salient landmarks. The question here was the following: What is the accuracy of visually based homing when an abundance of salient landmarks in a natural-looking virtual environment are available to be used as navigation aids? If visual cues are sufficient, we would expect perfect performance (i.e., negligible systematic errors and variability).

7.1 Methods

7.1.1 Participants

Five male and two female participants participated in the LANDMARKS Experiment. All of them had earlier completed the TOWN&BLOBS Experiment. Ages ranged from 23 to 30 years (mean: 26.5 years, SD: 2.6 years).

7.1.2 Interaction

Participants could freely move through the virtual environment using mouse buttons as in the previous experiment. The maximum velocity was $v_0 = 5m/s$ for translations and $\dot{\alpha}_0 = 40°/s$ for rotations. These motion parameters were chosen to help reduce the incidence of simulator sickness. Combined rotations and translations were possible, but hardly used by the participants.

7.1.3 Scenery

The experimental landscape was a green open square in a photorealistic 3D model of a small town (see Figure 4). The square was surrounded by an abundance of distinct landmarks (streets, trees, houses etc.).

7.1.4 Procedure

A repeated-measures, within-subject design was used (see Table 3). Each participant was presented with 60 isosceles triangles in random order, corresponding to a factorial combination of 6 repetitions for 5 different angles of the first turn and 2 turning directions. There was no time limit for completing the tasks and no feedback about performance accuracy during the whole experiment. The nomenclature used for the triangle is depicted in Figure 5.

Independent variable	Levels	Values
α=turning angle at 1st corner	5	$\alpha \in \{30°, 60°, 90°, 120°, 150°\}$
turning direction	2	left or right

Table 3: Experimental design for the LANDMARKS experiment. The isosceles triangles had a constant segment length of $s_1 = s_2 = 40m$. The different values for α correspond to correct homing distances $s_{3_c} \in \{20.71m, 40m, 56.57m, 69.28m, 77.27m\}$ and correct turning angles at the 2nd corner $\beta_c \in \{105°, 120°, 135°, 150°, 165°\}$.
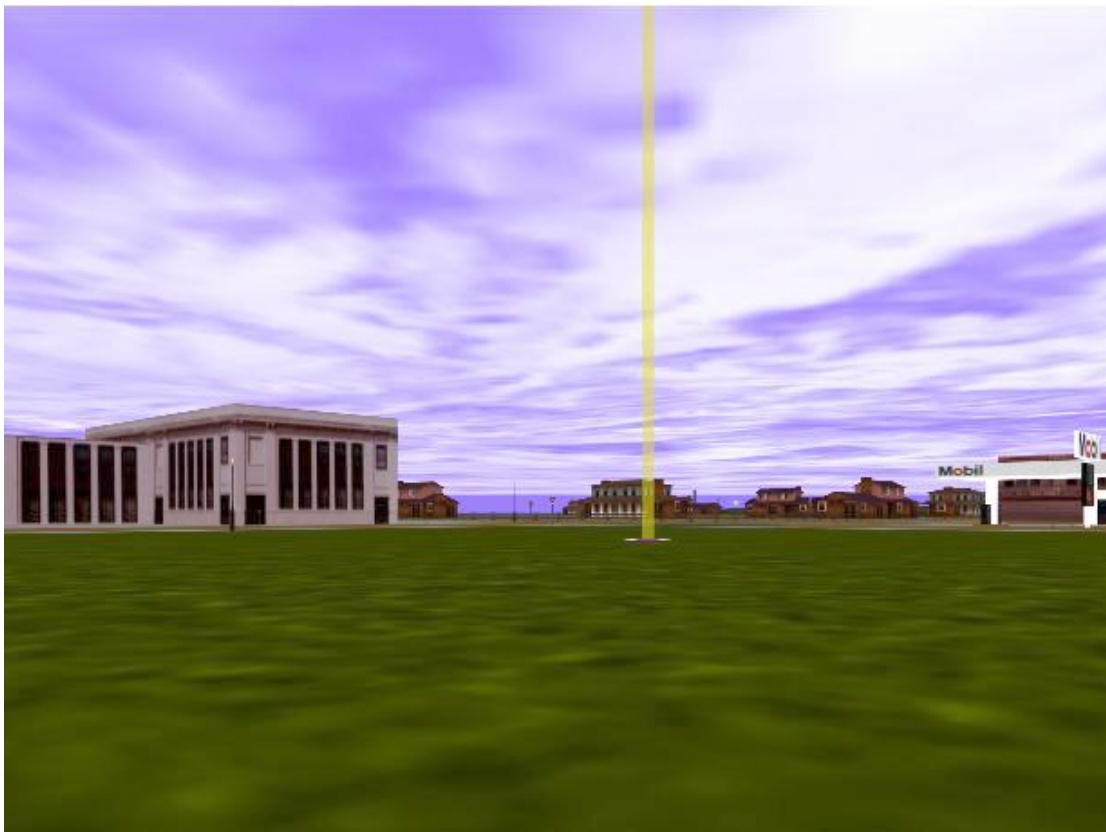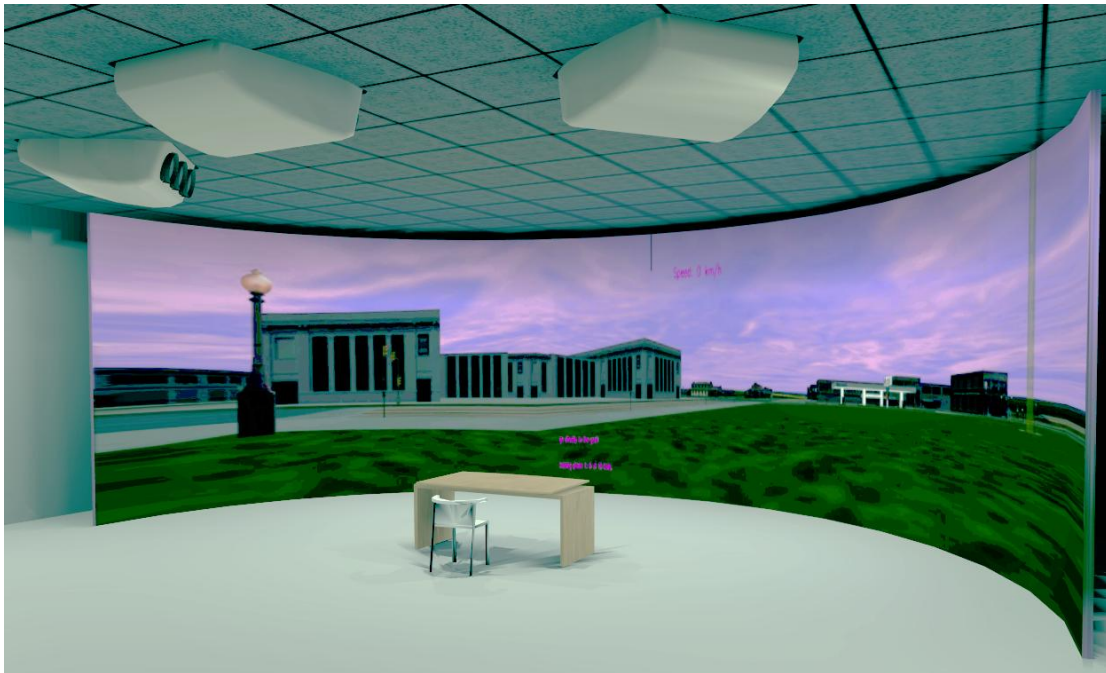
Figure 4: View of the virtual environments lab and the town environment. The yellow cylinder (light beam) represents the first goal, i.e., the first corner of the triangle to be traveled.
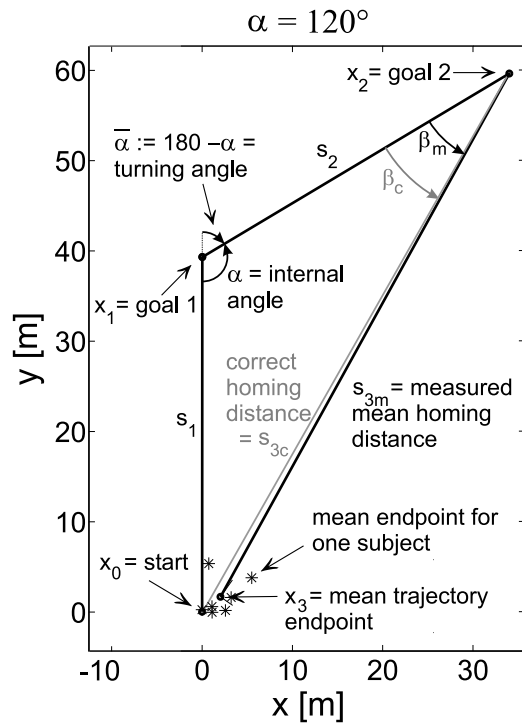
Figure 5: Nomenclature of a triangle to be traveled. The asterisks denote the homing trajectory end points for each participant, pooled over turning direction (left/right).

7.1.4.1   Test phase   Each participant performed one experimental block with 60 trials, lasting about one hour. For each trial, participants had the following task:

1. **Excursion**: At the beginning of each trial, participants were positioned and oriented randomly in the virtual environment, facing the first goal, (i.e., the first corner of the triangle), which was symbolized by a semi-transparent yellow "light beam" (see Figure 4). Participants moved to the yellow light beam, which disappeared upon contact. Then the second goal appeared, (i.e., the second corner of the triangle), which was symbolized by a blue light beam. As the second goal could be outside of the current visual field, the proper turning direction was indicated at the bottom of the projection screen. Participants turned toward the second goal and moved there. Like the first goal, it disappeared upon contact.

2. **Homing task**: After reaching the second goal, the whole scene faded out into darkness for 2s, for compatibility with Experiment TOWN&BLOBS. After that brief dark interval, the actual task was to turn and move directly to the non-marked starting point as accurately as possible. Pressing a designated button recorded the homing end point and initiated the next trial.

7.2   Results and discussion

Homing errors were analyzed using two separate repeated measures 2-way ANOVAs (5 angles $\times$ 2 turning directions) for the signed error of the two dependent variables (turning angle and distances traveled, respectively). None of the factors or any of the interactions were significant ($p > 0.24$ in all cases). For further analysis, the data were consequently pooled over left and right turns. The pooled data are graphically represented in Figure 6, providing a first impression of the homing results[4].

---

[4]The **95% confidence ellipse** is a 2D analogue of the confidence interval (mean $\pm$ two standard errors of the mean). It covers the population center with a probability of 95% and decreases with $1/\sqrt{N}$ with sample size N. The **standard**
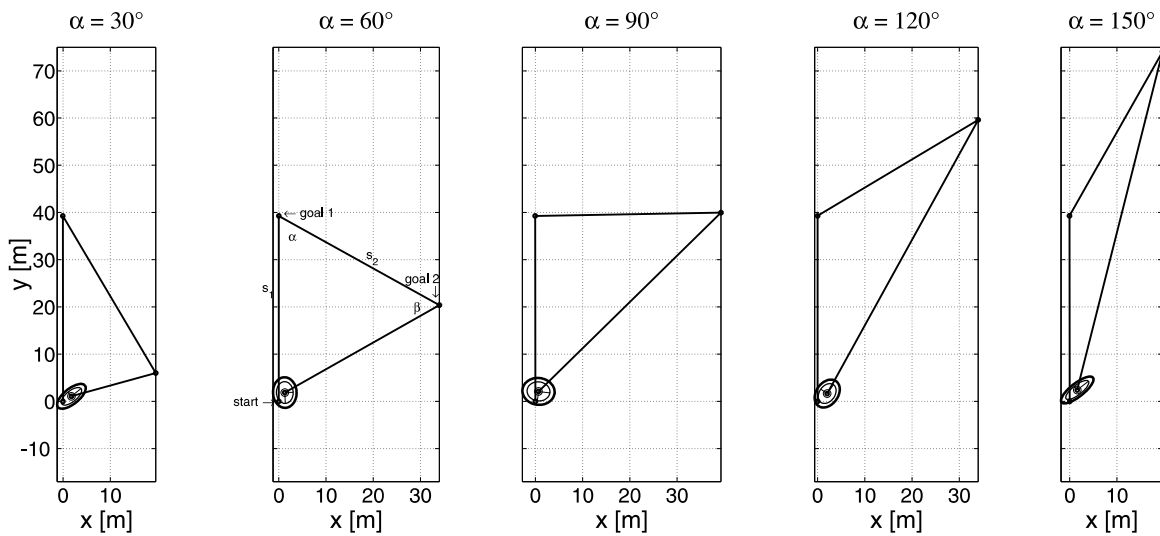
Figure 6: Homing performance in the LANDMARKS experiment. The data is pooled over the turning direction (left/right) as it had no significant influence on homing performance. Plotted are the mean (centroid), the 95% confidence ellipse[2] (outer ellipse with thick line) and the standard ellipse (inner ellipse with thin line) for the homing end points. Note the low variability and negligible systematic errors.
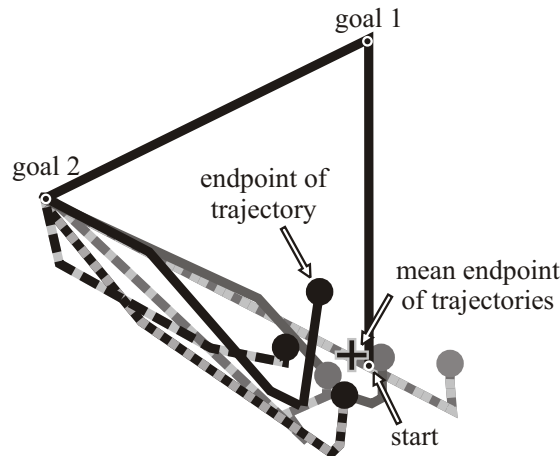
Figure 7: Examples of trajectories for one participant indicating snapshot matching. For the homing task, the participant drove south of the assumed starting point, then turned north and approached it "from behind", until the current view matched the original view from the starting spot. The non-straight trajectories further suggest that piloting is the dominant navigation mechanism, whereas path integration played only a minor role.

Homing performance was excellent, with negligible systematic errors and small between-subject variability.

To quantify that behavior, we again used the gain factor and the signed and absolute error for both measurands (see Figures 13 and 14). Participants slightly undershot the correct homing distance by 1.9m or 4.75%. Turning error, as well as the gain factor for turns and distances, were negligible and did not differ significantly from their correct value (see Figure 13). The absolute error was only 3.3% and 7.2% of the correct turning angle and homing distance, respectively, which is smaller than in

---

ellipse is a 2D analogue of the standard interval (mean $\pm$ one standard deviation). It is used to describe the variability of the data and covers roughly 40% of the data (see Batschelet, 1981, p. 141).

Experiment TURN&GO (5.2% and 23.0%, respectively, see Figure 14). Moreover, between-subject distance variability was largely reduced (see Figures 13 and 14). This performance improvement indicates that participants did indeed take advantage of the landmarks to perform the task. This is corroborated by the questioning after the experiment: When asked about strategies used for homing, all participants reported using configurations of landmarks (scene matching). Some participants even used snapshot matching as a homing strategy: They approached the assumed starting point from "behind" and moved north until the current view matched the initial view from the starting point (see Figure 7 for an example).

We conclude that piloting and especially scene matching led to almost perfect homing performance, and played the dominant role in navigation. However, homing performance was not quite perfect, which might be due to the lack of salient objects close enough to be able to identify the starting position uniquely. We assume that homing accuracy would have improved further, had we provided more salient, nearby landmarks like a location-specific ground texture, and added visibility of the virtual floor directly beneath the participants via a floor display.

8   Experiment 3: "Town&Blobs"

In this experiment, we investigated triangle completion performance *without* reliable landmarks in two different environments: A 3D field of blobs allowing only path integration via optic flow (see Figure 1), and the naturalistic town environment used in the previous experiment, but with landmarks that were only temporarily available (town with "scene swap").

There are three primary questions here: First, can optic flow information alone be sufficient for accurate homing, given a large FOV and the physical reference frame of a curved projection screen? Or will we observe the strong regression towards stereotyped responses found in other studies (Kearns et al., 2002; Klatzky et al., 1990; Loomis et al., 1993; Marlinsky, 1999b; Péruch et al., 1997)?

Second, where do the to-be-expected performance differences between navigation by optic flow and navigation by landmarks (Experiment Landmarks) stem from? To disambiguate between the effect of landmarks (salient reference points) and naturalism of the scene, we included an intermediate condition (town with "scene swap"); it provides naturalism and photo realism of the scene, size cues etc., but removes the landmark-character from the objects, by rearranging them before the return path ("scene swap"). If piloting is the main source for visual navigation, then "scene swap" should reduce performance to the level in the optic flow condition. If, on the other hand, naturalism, familiarity of the environment, or absolute size cues are important for navigation, optic flow performance should be inferior to "scene swap" performance.

Third, at what part of the navigation process do systematic errors occur? Experiment Turn&Go demonstrated negligible turn execution errors and small errors for distance reproduction (slight compression and considerable variability, but no general over- or undershooting). If mental spatial reasoning is easy and error-free, navigation performance should be comparable to the Turn&Go Experiment. Conversely, large systematic errors or variability would suggest difficulties in the mental "computation" of the homing trajectory or in the perception and encoding of angles.

8.1   Methods

8.1.1   Participants

Ten female and ten male naive participants, 17 to 30 years old (mean: 24.2 years, SD: 3.4 years), participated in this experiment. Four participants had to be replaced, as they had extreme difficulties with the experiment. Their behavior showed no correlation with the requirements of the particular trials, e.g., angular and/or distance responses were not correlated with the triangle geometry. Additionally, they had problems understanding the instructions and took much longer to complete the training phase. Only one participant experienced symptoms of simulation sickness and preferred not to finish the experiment.

8.1.2   Scenery

The experiment was performed in two different virtual environments: The simple 3D field of blobs from the first Experiment (Turn&Go) and the more complex town environment from the second Experiment (Landmarks). To exclude object recognition and scene matching as a possible homing strategy in the town environment, all landmarks (houses, streets etc.) in the scene were repositioned or replaced by others during the brief dark interval just before the onset of the return path ("scene swap condition"). The changed landmarks were arranged to form a different-looking, green square of about twice the original size, with the participant located at its center. After a few training trials,

participants reported no longer being confused or disoriented by the scene swap procedure. In the field of blobs environment, all blobs were randomly repositioned before the return path. Using scene swap in the town environment, participants could use piloting during the excursion (to build up a mental spatial representation), but not for the homing task, as there were no objects left indicating where the starting point was.

8.1.3 Procedure

A repeated-measures within-participant design was used (see Table 4). For each block, each participant was presented with 60 isosceles triangles in random order, corresponding to a factorial combination of 6 repetitions for 5 different angles of the first turn and 2 turning directions varied within a block, and 2 scenes varied across blocks. The order of the within-block conditions (angles and turning direction) was randomized, the order of the between-block conditions (scenes) was counterbalanced across participants. There was no time limit for completing the tasks and no feedback about performance accuracy during the test phase. Typically, the test phase lasted about one hour.

Independent variable	Levels	Varied	Values
α=turning angle at 1st corner	5	within block	$\alpha \in \{30°, 60°, 90°, 120°, 150°\}$
turning direction	2	within block	left or right
scene	2	between blocks	3D field of blobs or town environment

Table 4: Experimental design for the TOWN&BLOBS experiment. The isosceles triangles had a constant segment length of $s_1 = s_2 = 40m$. The different values for α correspond to correct homing distances $s_{3_c} \in \{20.71m, 40m, 56.57m, 69.28m, 77.27m\}$ and correct turning angles at the 2nd corner $\beta_c \in \{105°, 120°, 135°, 150°, 165°\}$.

8.1.3.1 Elimination of outliers   Some participants reported not having paid attention for some trials or having accidentally terminated a trial too early. To reliably eliminate those outliers for all participants, we developed the following criterion: There were always six repetitions per experimental condition (triangle geometry). If one of the six end points of those trajectories came to lie outside of a 4.5 σ standard ellipse around the five remaining end points, it was eliminated from the further analysis. A total of 132 trials or 5.5% of the trials were eliminated due to this criterion.

8.1.3.2 Training phase   After reading the experimental instructions, participants participated in a two-phase training session that lasted about 40 minutes. The training phases were similar to the actual experiment, but used additional feedback about the current position and orientation of the observer. Furthermore, triangle geometries were different from the test phase, to ensure that there was no simple direct transfer (e.g., rote learning) or motor learning. Both training phases consisted of ten homing trials each.

1. In the first training phase, compass directions (N, S, E, W) were overlaid on the display to provide a global orientation aid, where "north" was defined by the initial heading for each trial. Additionally, a top down (orthographic) view of the scene was presented on an extra monitor placed next to the participant (see Figure 8). The current position and orientation of the participant was displayed (symbolized by a white arrow) as well as the triangle corner currently visible (goal symbolized by the vertical "light beams").
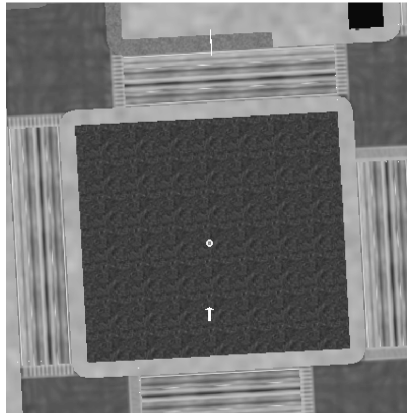
Figure 8: Top-down, orthographic view (here of the town environment) displayed on an auxiliary screen during training phase 1.

2. In the second training phase, the orientation aids were switched off during the navigation phase. After completing each trial, the orthographic view was briefly presented (for 2s) to provide feedback.

The training phase was designed to help inexperienced participants overcome initial disorientation, to ensure a comparable level of proficiency in virtual environment navigation and to avoid the influence of initial learning effects. In pilot experiments, we found that some participants initially had orientation problems in virtual environments without the training phase. This is consistent with Darken & Sibert (1996) and Ruddle, Payne, & Jones (1997), who showed that disorientation in virtual environments can be overcome by additional orientation aids.

8.1.3.3   Test phase   Each participant performed two experimental blocks (one block for each scene, 60 trials per block), in separate sessions on different days. The first block began directly after the training session as described above, the second block was preceded by an identical training session, but only 2×5 instead of 2×10 trials long. Apart from that, the test phase was identical to Experiment LANDMARKS.

8.2   Results and discussion

There are three critical issues addressed in this experiment. First, how well can people visually home without any landmarks? Second, what makes landmarks so useful, their reliability or the naturalism they provide? Third, and maybe most critical, what is the origin of the to-be-expected systematic navigation errors? Each of these issues will be dealt with in detail below.

8.2.1   Systematic errors

Homing errors were analyzed using two separate repeated measures 3-way ANOVAs (5 angles $\times$ 2 turning directions $\times$ 2 scenes) for the signed error of the two dependent variables turning angle and distances traveled, respectively. The ANOVAs revealed a highly significant main effect of the triangle geometry (angle α) on distance error ($F(4,76) = 32.5$, $p < 0.0005$), but not on turning error ($F(4,76)$ $= 0.61$, $p > 0.6$). None of the other factors or any of the interactions came close to significance ($p > 0.25$ in all other cases). In other words, neither the turning direction nor the scenery used had a

significant influence on homing performance. For the further analysis, the data were pooled over both left and right turns and over the two scenes unless indicated differently.
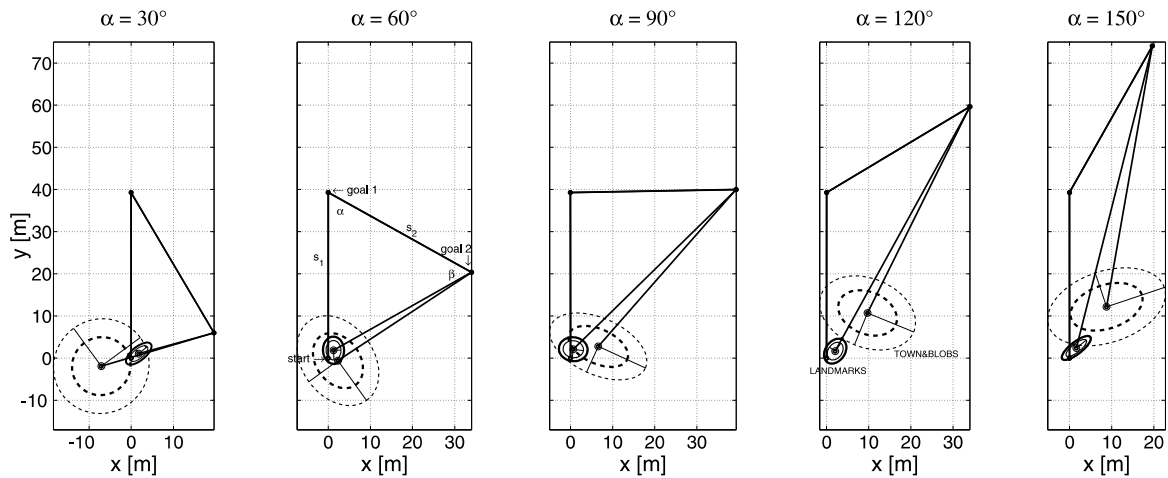
Figure 9: Homing performance in experiment TOWN&BLOBS (larger ellipses with dashed lines) as compared to experiment LANDMARKS (smaller ellipses with solid line). The data is pooled over the independent variables turning direction (left/right) and scenery (town/blobs), as they had no significant influence on homing performance. Plotted are the mean (centroid), the 95% confidence ellipse (inner ellipse with thick dashed line) and the standard ellipse (outer ellipse with thin dashed line) for the homing end points. The ellipses for the LANDMARKS experiment are smaller and include the origin, indicating less variability and more accurate homing performance than in experiment TOWN&BLOBS *without* reliable landmarks. Non-overlapping 95% confidence ellipses indicate significant performance differences (Batschelet, 1981).
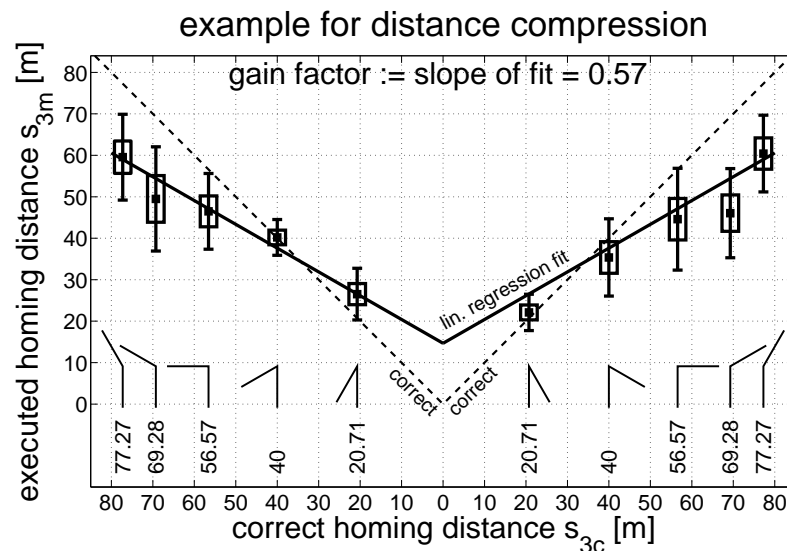
Figure 10: Behavioral response of one representative participant for the town environment. Actual values for distance traveled to complete the triangle (s_{3_m}, see Figure 5) is plotted over its corresponding correct values (s_{3_c}), left for left turns, right for right turns. The symmetry of the plot illustrates the similarity of the response for left and right turns. The mean values over the six repetitions are plotted for each of the ten triangle geometries (symbolized by the little icons below). The boxes refer to the standard error of the mean, the "whiskers" depict one standard deviation.

The pooled data are presented in Figure 9, providing a first impression of the homing results. The mean turning error is small, whereas the main effect of triangle geometry on distance error is obvious:

The shortest homing distance is typically overshot (left plot), whereas larger homing distances are undershot (right plots), indicating a compressed range of distance responses.

To quantify that behavior, the data are plotted differently in Figure 10. It shows one representative experimental block by one participant for the town environment. The homing distance actually traveled is plotted against its corresponding correct value. As for all participants, a linear regression line fits well to the data and summarizes its main aspects: The slope ("gain factor") is less than 1, implying that the range of observed mean homing distances is smaller than the range of correct homing distances. The gain factor in this example is 0.57, indicating a compression of the response range, whereas perfect performance (no compression) would result in a gain factor of 1, indicated by the dashed line going straight through the origin. The y-intercept well above zero indicates a regression (compression) towards mean homing distances larger than zero, and not just an overall scaling between stimulus and response. The general results are summarized in Figures 13 and 14. Averaged over all participants, the distance gain was 0.60 ± 0.07 (standard error of the mean, SE), indicating a general tendency to overshoot short distances and undershoot long distances (see Figure 13). This tendency proved highly significant (two-tailed t-test, $t(19) = 5.6$, $p < 0.0005$). The gain factor for turning angles was 0.91 ± 0.08, which is not significantly below the correct value of 1 ($t(19) = 1.0$, $p > 0.3$). This indicates that, on average, there was no systematic over- or undershooting of turning angles. The signed errors for turns and distances are $-2.8° \pm 3.0°$ and $-0.9m \pm 1.6m$, respectively, indicating a slight but insignificant tendency to undershoot both turns and distances ($t(19) = 0.96$, $p > 0.3$ and $t(19) = 0.56$, $p > 0.5$, respectively). Compared to Experiment LANDMARKS, the only significant difference between sample means was in terms of distance gain ($t(25) = -3.42$, $p < 0.002$). This indicates that the lack of reliable landmarks caused the tendency towards stereotyped homing distances in Experiment TOWN&BLOBS. It further gave rise to a substantial increase in between-subject variability ($F(19,6) = 59.9$, $p < 0.0001$ for turning error, $F(19,6) = 19.9$, $p < 0.002$ for distance error, $F(19,6) = 25.4$, $p < 0.0007$ for angular gain and $F(19,6) = 188.8$, $p < 0.0001$ for distance gain).

8.2.2   Absolute errors

The absolute errors were rather pronounced (see Figure 14), with 14.6% and 30.7% of the correct turning angle and homing distance, respectively. The absolute turning error was more than three times larger than in both Experiment TURN&GO and LANDMARKS ($t(27) = 3.77$, $p < 0.0008$ and $t(25) = 4.03$ $p < 0.0005$, respectively). The absolute distance error was comparable to Experiment TURN&GO, and about four times larger than in Experiment LANDMARKS ($t(27) = 1.10$, $p > 0.2$ and $t(25) = 4.90$ $p < 0.0005$, respectively). Thus, absolute distance error could be explained by the lack of reliable landmarks.

8.2.3   Discussion

The lack of performance differences between the blobs and town environment suggests that participants were not able to take advantage of natural-looking landmarks that are only temporarily available. Hence, naturalism, familiarity of the scene, and absolute size cues did not play a significant role, and piloting was the main source for visual navigation whenever possible.

Path integration based solely on optic flow proved to be sufficient for correct mean turn responses and negligible turn compression for almost all participants. As in the TURN&GO Experiment, we did not find the strong compression towards stereotyped turn responses typically found in the literature (Bakker et al., 1999, 2001; Kearns et al., 2002; Klatzky et al., 1990, 1997; Loomis et al., 1993; Péruch et al., 1997). A detailed comparison to the literature and discussion of potential origins of the observed performance differences will be provided in the general discussion (section 11 of

part II). On the other hand, homing distance showed a considerable compression towards stereotyped responses. Most participants had a tendency to overshoot short distances and undershoot long distances, a phenomenon commonly found in the literature (Klatzky et al., 1997; Loomis et al., 1999). The variability between participants was rather pronounced, though, which might be due to different navigation strategies used. We found no significant learning effect between the first and second block ($p > 0.05$ for two-sided paired t-tests for all six dependent variables), indicating that further learning and task exposure did not improve performance. We know from Experiment TURN&GO that turn *execution* errors are negligible. This suggests that for all four experiments, the observed turning angle directly reflects the turning angle intended by the participant. The same is true for distances traveled[5], but with a reduced precision. Hence, we can use the observed navigation behavior to infer about the intended navigation behavior and the underlying mental representation. Given the negligible turn *execution* error, the considerable absolute turn error and between-subject turn variability in Experiment TOWN&BLOBS indicates that without reliable landmarks, many participants had either problems in correctly *encoding* the turned angle, or in mentally computing the desired homing angle. There is, however, some rather anecdotal evidence suggesting that *encoding* errors for turns are negligible, too. In general, participants were able to estimate turns well even when not actively controlling the motion, e.g., when the experimenter initiated the turn for demonstration purposes *before* the first training phase, and they just observed. Most participants were even able to pinpoint the exact angles $\overline{\alpha}$ turned in Experiment TOWN&BLOBS or during the training phases, indicating a negligible encoding error for turns. Hence, the observed turning errors should be attributed to problems in mental spatial reasoning.

There is no direct evidence on systematic encoding errors for distances traveled, as distances cannot be queried without referring to an absolute or relative scale. However, Experiment TURN&GO presented evidence that participants can reproduce distances fairly well, suggesting that the distance traveled gives a rough estimate of the distance mentally represented and intended to travel. Potential scaling errors in distance encoding and execution were shown to cancel each other out and are thus irrelevant for our reasoning.

We can use this information to understand the origin of the strong distance compression (gain factor of 0.60 ± 0.07) observed in Experiment TOWN&BLOBS. Most participants realized after a few trials that s_1 and s_2 were equal and held constant. This suggests that s_1 and s_2 were encoded to the same, constant value, irrespective of α, and participants knew they were traveling isosceles triangles. This is corroborated by participants' verbal statements. Given that systematic encoding errors for turn are negligible, we can conclude that participants had an essentially correct mental representation of the triangle geometry. The question arising now is where the observed errors in Experiment TOWN&BLOBS, especially the rather pronounced distance compression stems from, given that the mental representation was an isosceles triangle with approximately the correct angle α. An explanation we favor is that participants experienced problems in determining the correct homing response from the mental representation, even though they had all the information needed. Most participants, then, seem to be unable to mentally compute or somehow infer the correct homing distance from a known triangle geometry. This is also the main difference between the distance reproduction task in Experiment TURN&GO and the triangle completion task in Experiment TOWN&BLOBS: For the latter, participants had to use non-trivial mental geometric or spatial reasoning.

---

[5]This is true if one assumes that participants can not only intend and execute the *same* distances as traveled before (as demonstrated in experiment TURN&GO), but also intend and execute *different, scaled* distances. Results from experiment TOWN&BLOBS corroborate this assumption: For isosceles triangles with $\alpha_c = 90°$, most participants knew that the correct homing distance was $s_3 = \sqrt{2} \cdot s_2$, or roughly 1.4 times the distance just traveled. Participants were indeed able to execute this intended distance quite accurately.

9　Experiment 4: "RANDOM TRIANGLES"

Experiment TOWN&BLOBS demonstrated that homing by optic flow or transient landmarks is possible and allows for decent homing performance, apart from a rather pronounced distance compression. A question that remains unanswered is how the simplicity of the triangle geometry (only isosceles triangles with angles α in $30°$ steps) might have influenced homing performance. To address this question, we used the triangle completion paradigm with the 3D field of blobs again, but with novel triangles of completely randomized geometry for each trial. To our knowledge, navigating randomized triangle geometries has never been addressed in the literature. If participants had been able to take advantage of a simple, repetitive, isosceles triangle geometry in Experiment TOWN&BLOBS, we would now expect a clear deterioration in homing performance: Participants should be less certain about the correct homing response and therefore be more conservative in their response, leading to a more pronounced response compression as well as an increased variability and absolute error.

9.1　Methods

9.1.1　Participants

Participants were the same ten participants as in Experiment TURN&GO. There was no reason to expect potential benefits or direct learning transfer, as Experiment TURN&GO did not provide any explicit performance feedback. Furthermore, comparing performance between the first and the second block of Experiment TOWN&BLOBS demonstrated that even exposure to the same task did not improve performance. Hence, different amounts of exposure to VR and VR experiments do not seem to be a critical issue, indicating that comparisons between the experiments presented part of this thesis are legitimate.

9.1.2　Procedure

The experimental procedure was the same as in Experiment TOWN&BLOBS using the 3D field of blobs, but using different triangle geometries for each trial. As before, triangle geometries in the training phase were different from the test phase, to ensure that there was no simple direct transfer (e.g., rote learning) or motor learning possible.

The experimental design is summarized in Table 5. Each participant completed 60 trials. For each trial, values for the length of the first segment, the second segment and the enclosed turning angle were drawn independently, randomly, and without replacement from a set of 60 equally spaced values each. Additionally, the turning direction was chosen randomly. There was no repetition of conditions, which ensured that participants could not memorize individual triangle geometries and utilize them directly in a later trial, as might have been possible in Experiment TOWN&BLOBS.

Independent variable	Levels	Values
s_1= length of segment 1	60 (equally spaced)	$s_1 \in \{20m, 20.90m, \ldots, 73m\}$
s_2= length of segment 2	60 (equally spaced)	$s_2 \in \{20m, 20.90m, \ldots, 73m\}$
α=turning angle at 1st corner	60 (equally spaced)	$\alpha \in \{20°, 24.82° \ldots, 160°\}$
turning direction	2	left or right

Table 5: Experimental design for the RANDOM TRIANGLES experiment.

9.2 Results

As in the previous experiments, the data are analyzed in terms of signed and absolute error as well as gain factor. A correlation analysis reveals further insights about the interrelation between the different independent and dependent variables.

9.2.1 Signed errors

Results are summarized in Figures 13 and 14. Mean turning error and distance error were remarkably small and did not differ significantly from zero or from the results from Experiment TOWN&BLOBS. However, the between-subject variance of the distance error was significantly increased, compared to Experiment TOWN&BLOBS ($F(9,19) = 5.0$, $p < 0.004$), whereas the variance of the angular error remained unchanged ($F(19,9) = 1.7$, $p > 0.4$).

9.2.2 Gain factors

Both angular and distance response showed an obvious compression with gain factors of 0.76 and 0.85, respectively, which was significantly below the correct value of 1 ($t(9) = 5.0$, $p < 0.0008$ and $t(9) = 3.9$, $p < 0.004$, respectively). The angular compression was slightly, but insignificantly more pronounced than in Experiment TOWN&BLOBS ($t(28) = 1.3$, $p > 0.2$). In contrast, distance compression was significantly reduced ($t(28) = 2.6$, $p < 0.02$). Interestingly enough, the variance of both angular and distance gain was significantly reduced, compared to Experiment TOWN&BLOBS ($F(19,9) = 6.0$, $p < 0.009$ and $F(19,9) = 6.5$, $p < 0.007$, respectively).

9.2.3 Correlation analysis

The details and results of pairwise correlation analyses are summarized in Table 6. The analyses revealed a strong correlation between the independent variables s_1, s_2 and s_{3_c} and the observed distance error. For increasing values of s_1, s_2 and s_{3_c}, the distance response shifted from an overshoot to an undershoot, indicating a tendency of the participants to produce medium-sized triangles. The influence of s_1 and s_2 on turning error is best understood by looking at the influence of their ratio (s_2/s_1) or difference ($s_2 - s_1$): For triangles with a shorter second segment ($s_2 < s_1$), turning angles are increasingly overshot. Conversely, turning angles are increasingly undershot for triangles with a longer second segment ($s_2 > s_1$). This highly significant correlation explains about $r^2 = 11.4\%$ of the variance in homing errors.

However, distance and turning errors were not independent from each other: Distance error increased with increasing turning error. Interestingly enough, the turning angle α between the first and second segment did not show any systematic influence on the pattern of homing errors. The strong correlation between distance error and correct homing distance s_{3_c} and between turning error and correct homing angle β_c expresses the distance and turn compression described above.

9.2.4 Absolute errors

Mean absolute errors for turns and distances did not differ significantly from Experiment TOWN&BLOBS ($t(28) = 0.28$, $p > 0.7$ and $t(28) = -1.53$, $p > 0.1$, respectively). Between-subject variability was, however, slightly *decreased* for turns and *increased* for distances ($F(19,9) = 2.9$, $p = 0.053$ and $F(9,19) = 3.5$, $p < 0.01$, respectively).

Correlation between		r	r^2	$t(8)$	p
dist. error	s_1	-0.310	0.096	5.9	0.00027
dist. error	s_2	-0.176	0.031	4.4	0.0017
dist. error	α	-0.007	0.0	0.28	0.78
dist. error	s_2/s_1	0.095	0.009	2.0	0.073
dist. error	$s_2 - s_1$	0.086	0.007	2.0	0.080
dist. error	s_{3_c}	-0.256	0.066	4.0	0.0031
dist. error	β_c	0.015	0.0	0.25	0.80
turn error	s_1	0.263	0.069	5.8	0.00027
turn error	s_2	-0.224	0.050	3.9	0.0037
turn error	α	-0.030	0.001	1.3	0.21
turn error	s_2/s_1	-0.290	0.084	5.5	0.00039
turn error	$s_2 - s_1$	-0.338	0.114	5.4	0.00045
turn error	s_{3_c}	-0.044	0.002	0.74	0.48
turn error	β_c	0.357	0.128	4.3	0.0020
turn error	dist.err.	0.126	0.016	1.9	0.087

Table 6: Results of the correlation analysis for experiment RANDOM TRIANGLES between the error for distances and turns (first column) and the parameters in the second column. The Pearson correlation coefficient, r, and r^2, the coefficient of determination were computed by performing a correlation for each participant's data individually, transforming the resulting r-values (via a Fisher r-to-Z transformation) into Z-values, taking their mean, and transforming the mean back via the inverse transformation (Fisher Z-to-r transformation) into mean r-values. To test whether the correlation coefficients differ significantly from zero ("not correlated"), a two-tailed t-test was calculated for the r-to-Z transformed r-values of the individual participant's data. The resulting significance level is displayed in the last column.

9.3 Discussion

The most striking results from this experiment are the relatively small variability of gain factors and the less pronounced distance compression, compared to Experiment TOWN&BLOBS. This is all the more astonishing, as the variability in signed as well as absolute distance error was significantly increased.

The correlation analyses revealed a regression towards stereotyped responses: For "extreme" triangles (i.e., extreme values of $s_1, s_2, s_{3_c}, s_2 - s_1$, and α_c), participants responded as if those values weren't as extreme. This could be interpreted as a tendency to opt for the "safe bet" for difficult triangle geometries.

However, there was no overall performance deterioration as compared to Experiment TOWN&BLOBS. This suggests that neither motor learning, direct learning transfer between trials, nor the simplicity of isosceles triangles was a determining factor for homing accuracy in Experiment TOWN&BLOBS. Participants were apparently unable to take advantage of the relatively simple and repetitive triangle geometry in Experiment TOWN&BLOBS.

10 Experiment 5: Spatial imagination abilities tests

To investigate whether mental spatial abilities might be a determining factor for homing accuracy, we performed two standard, paper and pencil spatial imagination abilities tests with the participants from Experiment TOWN&BLOBS and RANDOM TRIANGLES and correlated the results with the homing performance. Test 1 was a "Schlauchfiguren-Test" (Stumpf & Fay, 1983), where participants saw in each trial one picture of a tube folded inside a transparent cube, and had to decide from which viewpoint a second picture of the same object was taken (Figure 11, top pictures). Participants were asked to complete 21 trials in twelve minutes. The second test was a "Würfel Erkennen Test", part six of the "Intelligenz Struktur Analyse Test" (ISA, 1998), in which participants had to judge the identity of cubes seen from different directions (see Figure 11, bottom picture). Participants were asked to complete 17 trials in 18 minutes. Responses for both tests were given in a multiple choice-type manner.
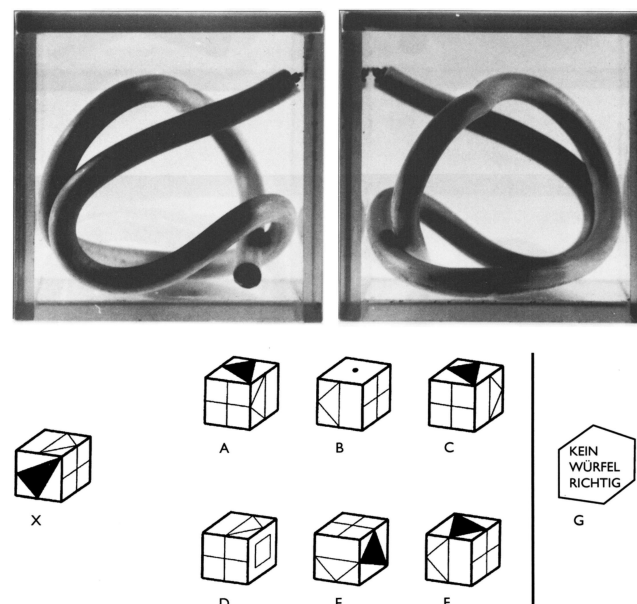
Figure 11: Sample stimulus from spatial imagination abilities test 1 (top) and test 2 (bottom).

A correlation analysis was conducted between the test results (% correct responses) and the absolute error and absolute value of the signed error for turns, distances, angular gain factor, and distance gain factor. We used 14 of the 20 participants from Experiment TOWN&BLOBS and all 10 participants from Experiment RANDOM TRIANGLES. If the mental spatial reasoning phase was indeed the main cause for the observed systematic errors as we proposed in subsection 8.2.3, at least one of the error measures should be negatively correlated with the test performance and none positively. Additionally, we expect a higher correlation for Experiment RANDOM TRIANGLES, which required more complex spatial reasoning. To test these hypotheses, one-sided t-tests were conducted. The results for an α-level of $p < 0.15$ are summarized in Table 7.

Five error measures were significantly correlated ($p < 0.05$) and five more approached significance ($p < 0.1$). All correlations were either negative or negligible ($p > 0.15$), indicating that a good test result coincided with a small error measure and hence a good homing performance. For Experiment RANDOM TRIANGLES, which required more complex mental spatial reasoning, both test results correlated nicely, especially with the distance error measures, and were able to explain up to 62% of the rather large variance (see Table 7).

Measurand	Spatial imagination test	r	r^2	t	p
TOWN&BLOBS					
abs. turn. error	test 2	-0.42	0.17	$t(12) = 1.6$	0.070
abs. dist. error	test 2	-0.36	0.13	$t(12) = 1.3$	0.10
|signed turn. error|	test 2	-0.55	0.30	$t(12) = 2.3$	0.021
RANDOM TRIANGLES					
abs. dist. error	test 1	-0.67	0.45	$t(8) = 2.6$	0.016
abs. turn. error	test 2	-0.48	0.23	$t(8) = 1.5$	0.081
abs. dist. error	test 2	-0.79	0.62	$t(8) = 3.6$	0.0035
|signed turn. error|	test 1	-0.48	0.23	$t(8) = 1.6$	0.080
|signed dist. error|	test 1	-0.66	0.43	$t(8) = 2.5$	0.019
|signed turn. error|	test 2	-0.54	0.29	$t(8) = 1.8$	0.0532
|signed dist. error|	test 2	-0.70	0.49	$t(8) = 2.8$	0.012

Table 7: Correlation analysis for the mental spatial abilities tests. Displayed are the results of the correlation analysis between homing performance in experiments TOWN&BLOBS and RANDOM TRIANGLES and the number of correct trials in two mental spatial abilities test. Only correlations from an α-level below p = 0.15 are displayed.

We conclude that mental spatial ability, as assessed by both tests, correlates positively with homing performance, especially for the more complex task in Experiment RANDOM TRIANGLES. This suggests that mental spatial ability might be a determining factor for homing performance in triangle completion experiments based on path integration. This finding agrees well with our explanation of the homing errors proposed earlier. However, further experiments are needed to corroborate this hypothesis, as the number of participants in this study was rather limited, and we did not test to what degree general intelligence and non spatial abilities might be a contributing factor.

11   General discussion

Before drawing our own conclusions in the final discussion (subsection 11.2), we would like to review the related literature in detail and reevaluate our work in the context of the literature at large. On the one hand, these comparisons allows us to better understand the underlying processes explaining the data and especially the observed differences. On the other hand, they enable us to pinpoint critical factors for good spatial orientation and navigation in real and virtual environments.

11.1   Comparison with previous work

In the context of the present data, the relevant literature might be broken down into four categories: Nonvisual path integration experiments in general (subsection 11.1.1), triangle completion experiments in virtual environments using HMDs (subsection 11.1.2) and projection screens (subsection 11.1.3), and studies scrutinizing the origin of systematic homing errors (subsection 11.1.4).

11.1.1   Non-visual navigation experiments based on path integration

To test simple path integration performance, Klatzky et al. (1990) and Loomis et al. (1993) asked participants to reproduce walked distances and turns while blindfolded (see Klatzky et al., 1997, for a comparison). Turn performance was comparable to Experiment TURN&GO when turns were made within a circular hoop surrounding the participant (gain factor = 0.99) (Klatzky et al., 1990), but decreased for turns performed without the hoop (gain factor = 0.82) (Loomis et al., 1993). Distance reproduction showed a slightly increased compression towards stereotyped responses compared to Experiment TURN&GO (gain factors = 0.75 and 0.81, respectively). This suggests that, at least for elementary rotations and translations, visual path integration performance is by no means inferior to path integration by kinesthetic and vestibular cues from blind walking.

For triangle completion tasks, vestibular and proprioceptive cues from blind walking do not seem to allow for homing without considerable systematic errors (Kearns et al., 2002; Klatzky et al., 1990; Loomis et al., 1993; Marlinsky, 1999b; Sauvé, 1989). Participants typically overturned for small correct turning angles (< 90°) and underturned for large turning angles (> 90°). The same compression towards stereotyped responses was found for distances traveled: Short distances were overshot, large distances undershot. This bias is a commonly found trend in psychophysical experiments (Poulton, 1979; Stevens & Greenbaum, 1966). Loomis et al. (1993), in accordance to Klatzky et al. (1990) concluded that "*not only were there significant signed errors for the average of all subjects but also no single subject came close to exhibiting negligible errors over the 27 triangles. It appears that even for the short paths over which subjects were passively guided here* [2, 4, and 6m segment length, remark by the author]*, the proprioceptive and vestibular cues were inadequate for accurate path integration*"(pp. 83-84). For comparison to our results, data from Loomis et al. (1993) were reanalyzed and plotted in Figures 12, 13, 14, and 56. Mean turning errors were close to zero, but showed a rather large variance which was significantly larger than for Experiment TOWN&BLOBS ($F(36,19) = 3.6$, $p < 0.004$ for isosceles triangles). All other measures were substantially below their correct value, indicating general undershooting and biases towards stereotyped responses (response compression).

Path integration accuracy for blind walking decreases further when proprioceptive cues are reduced to mainly vestibular cues from wheelchair transportation (Marlinsky, 1999b; Sholl, 1989). Several additional factors influence path integration performance, including stimulus context and task specificities (Klatzky et al., 1997; Loomis et al., 1999). For return-to-origin tasks, the number of linear segments and turns increase both response time and error, especially when segments cross each other

(Klatzky et al., 1990; Loomis et al., 1993). Blind triangle completion experiments by Mittelstaedt & Glasauer (1991) revealed an influence of the walking speed: Participants overshot distances for walking speeds faster-than-normal and undershot distances for slow walking speeds. This relation reversed polarity for passive (wheelchair) transportation.

11.1.2   Triangle completion experiments with head mounted display

Kearns et al. (2002) and Duchon et al. (1999) conducted triangle completion experiments in a virtual environment consisting of a large round room with uniformly textured walls and floor. In one condition, ego-motion was controlled using a joystick and visually presented via a non-headtracked head mounted display with a FOV of $60° \times 40°$. Participants' homing performance was sensitive to changes in segment length of the triangle, suggesting that they were able to integrate optic flow from translations to yield the distance traveled. In contrast, participants' mean homing response reflected no sensitivity to variations in turning angle $\overline{\alpha}$: For isosceles triangles with angles $\alpha \in \{60°, 90°, 120°\}$, participants produced the same mean response regardless of actual triangle geometry, acting as if traversing an equilateral triangle. Without the external reference frame of the physical surround, participants seemed to be unable to use the rotational optic flow to extract the turning angle. This effect was not found in the present experiments or the experiments by Péruch et al. (1997), all of which used projection screens. This suggests that the type of display (HMD versus projection screen) and hence the external reference frame available might influence the sensitivity to angles turned.

In another condition (Kearns et al., 2002, exp. 3) participants wore a headtracked HMD and physically walked triangles, with the triangle corners being indicated visually as before. Homing results showed a reduced variability reflecting a higher subjective confidence. However, participants still gave the same stereotyped response irrespective of the turning angle $\overline{\alpha}$. Compared to the tendency
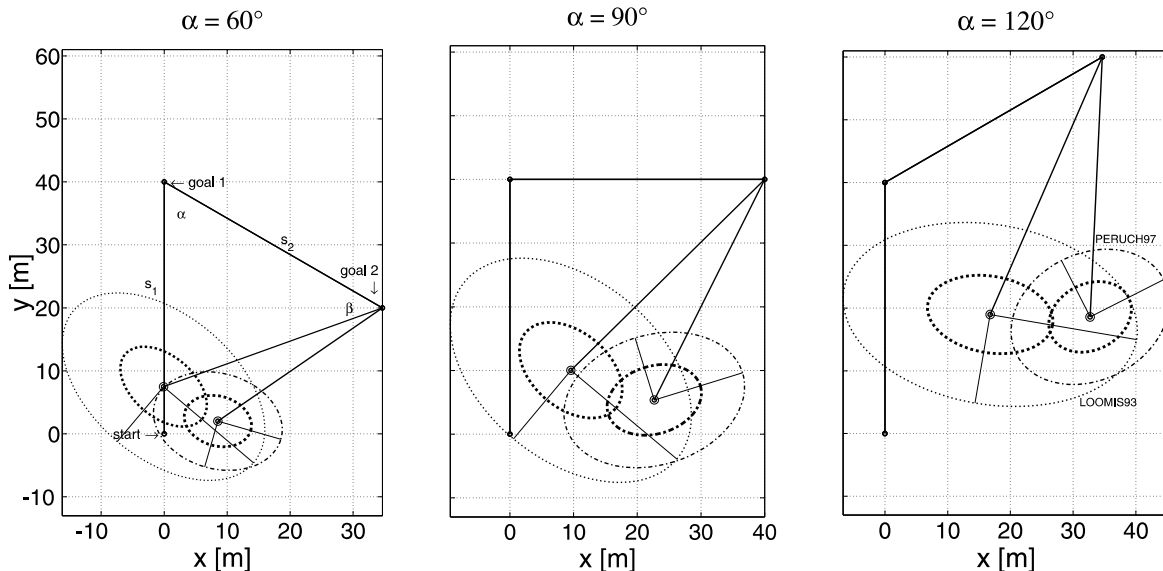
Figure 12: Homing performance under different conditions, plotted as in Figure 6 and 9. Dotted lines represent results for visual triangle completion within a circle of equal cylinders (reanalysis of data from Péruch et al. (1997), data from experiment 1 and 2 pooled) and dash-dotted lines for blind walking (reanalysis of data from Loomis et al. (1993), Experiment 1, triangles with $s_1 = s_2 = 4m$). Data from Péruch et al. and Loomis et al. is scaled to fit the triangles used in our experiments. Their full data set is summarized in Figures 57 and 56, respectively, for convenience. Note the considerable variability and systematic homing errors, especially the general distance undershoot.
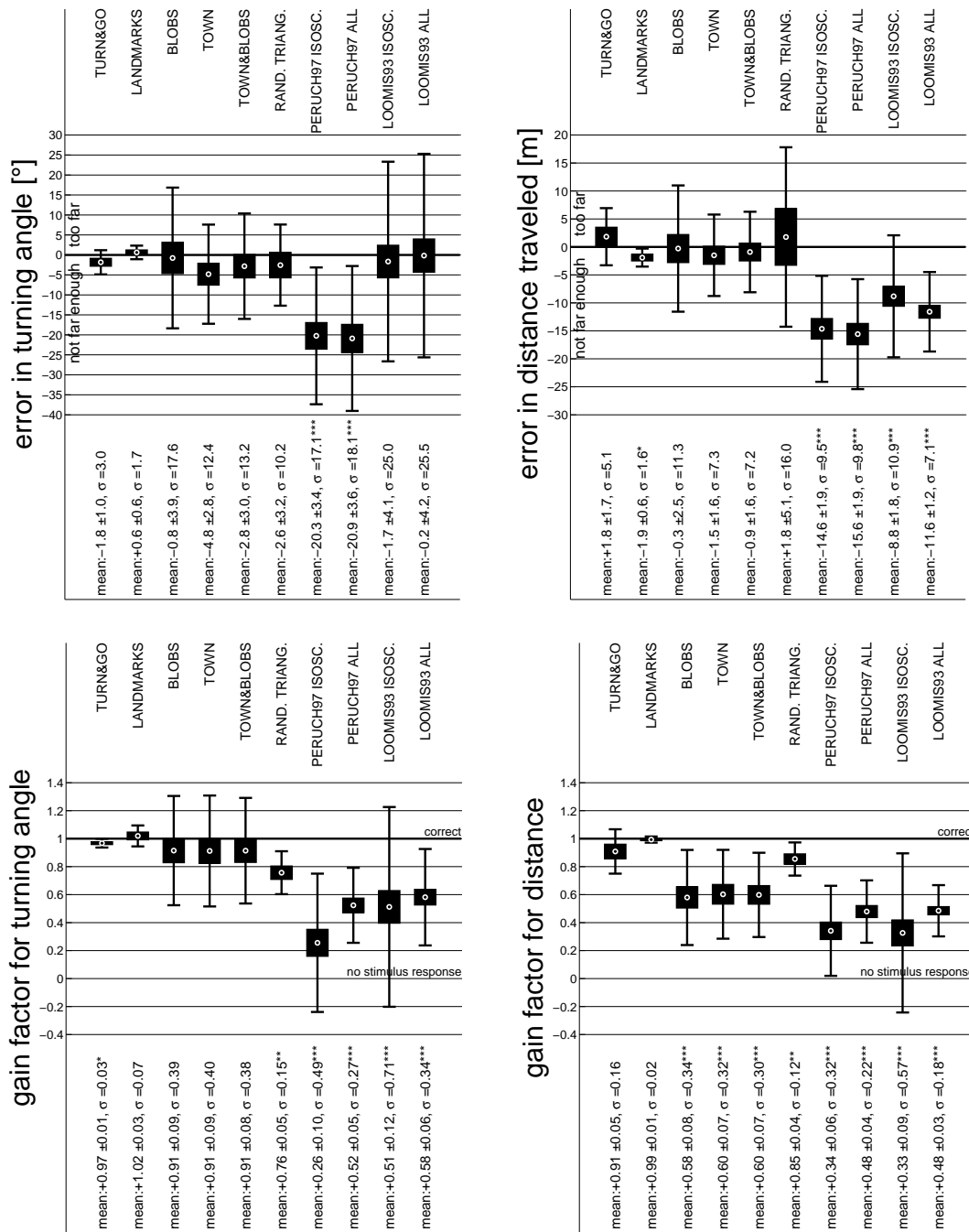
Figure 13: Comparison of navigation performance for the different experimental conditions. At the top of each plot, the experimental conditions are displayed (from let to right): Exp. 1 ("TURN&GO"); Exp. 2 ("LANDMARKS") with reliable landmarks; Exp. 3, using the 3D field of blobs ("BLOBS"), the town environment ("TOWN") and data from both blocks pooled together ("TOWN&BLOBS"); Exp. 4 ("RANDOM TRIANGLES"); reanalysis of data from Péruch et al. (1997) on visual triangle completion within a circle of equal cylinders for isosceles triangles only ("PERUCH97 ISOSC.") and for all triangles ("PERUCH97 ALL"); reanalysis of data from Loomis et al. (1993) on blind walking triangle completion, again for isosceles triangles with $s_2 = 4m$ only ("LOOMIS93 ISOSC.") and for all triangles ("LOOMIS93 ALL"). Data from Péruch and Loomis is scaled to match the triangle size used in our experiments. Below are the plots of the four measures, the center indicating the arithmetic mean. Boxes represent intervals of one standard error of the mean, whiskers represent one standard deviation. The gain factor was defined as the slope of the linear regression fit. At the bottom of each plot, the numeric values of the mean, standard error, and standard deviation are displayed. The stars '*' indicate whether the mean differs significantly (on a 5%, 0.5% or 0.05% significance level, using a two-tailed t-test) from the corresponding correct value, depicted as a thick horizontal line.
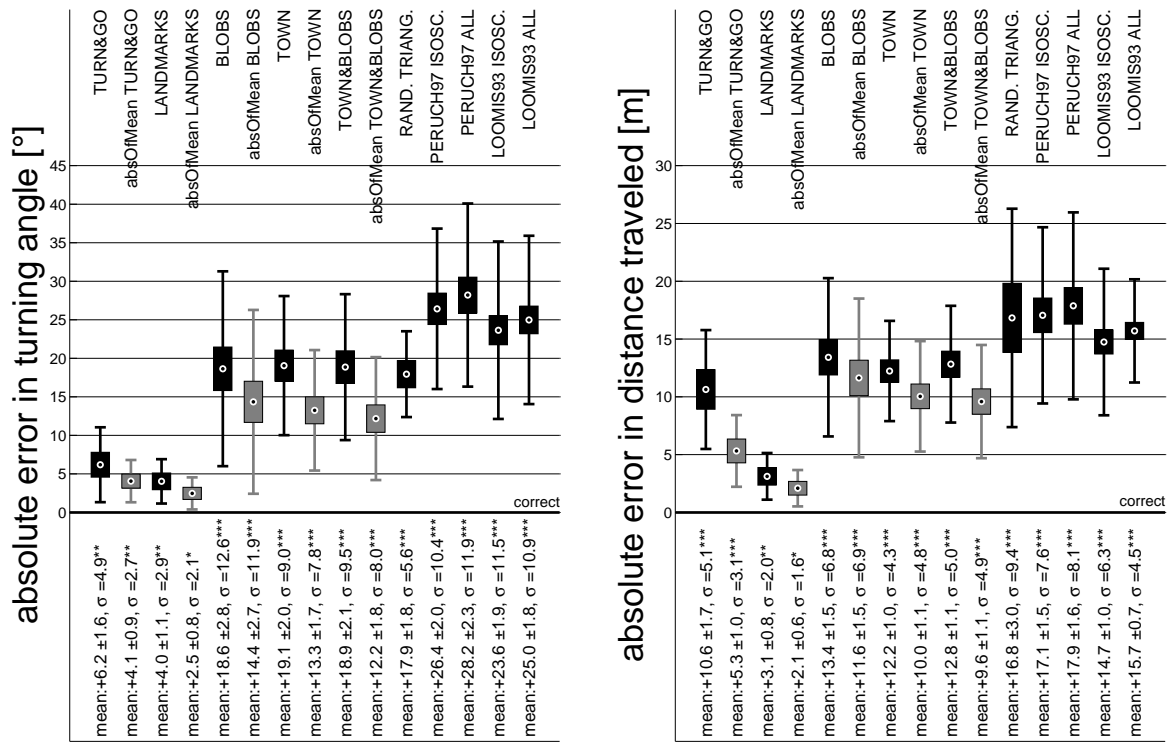
Figure 14: Absolute error for the different experimental conditions, plotted as in Figure 13. For the black boxes, the absolute value is taken as is customary for each individual trial *before* taking the mean over the repetitions of each condition. The gray boxes refer to the same data as the black boxes left of it, but *without* the estimated random error: Here, the absolute value is taken *after* computing the mean over the repetitions of the same condition. The significant differences between the gray and black means indicate that participants' responses had a considerable variability ("random error") within the *same* condition, i.e., participants were unable to produce the same response for the repetitions of the same condition. That is, the difference can be seen as a first rough estimate for the precision of the responses.

to *under*turn by 7.1° (SD: 35.9°) for purely visual navigation in the first condition, physical walking led to a general *over*turning by 19.9° (SD: 27.1°). Removing all visual information except the poles denoting the triangle corners hardly altered participants' responses, indicating that the proprioceptive cues from walking dominated over optic flow information. This overturning and lack of stimulus response for physical rotations was not found for blind walking experiments (Klatzky et al., 1990; Loomis et al., 1993; Marlinsky, 1999b) and can hence not be simply attributed to proprioceptive cues from walking. Consequently, the effect seems to be caused by the visual display presenting the triangle to be traveled (see subsection 11.2).

11.1.3 Triangle completion experiments with projection screen

Loomis et al. (1993) and Klatzky et al. (1990) have shown that kinesthetic and vestibular cues from blind walking are inadequate for accurate path integration as assessed by triangle completion experiments. Péruch et al. (1997) conducted comparable triangle completion experiments in Virtual Reality to investigate human path integration ability based on visual information (optic flow). Participants used a joystick to move within an area surrounded by 16 identical cylinders equally spaced on a circle of 60m diameter. The simulated ego-motion was displayed on a planar projection screen subtending a physical field of view (FOV) of 45° horizontal × 38° vertical. Participants had to complete 27 triangles corresponding to a factorial combination of 3 values for the simulated field of view

(horizontal sFOV = 40°, 60° and 80°) × 9 triangle geometries (3 angles × 3 lengths of the second segment). Interestingly enough, the sFOV had no significant effect on homing performance.

For comparison to our results, data from Péruch et al. (1997) were reanalyzed and plotted in Figures 12, 13, 14, and 57. Participants showed a general undershoot for both turning angles and distances traveled. Results also revealed a strong regression towards stereotyped values for turning angles and distances traveled, especially for isosceles triangles (see Figure 13). All those effects were stronger than in the blind walking studies by Loomis et al. (1993) and Klatzky et al. (1990), suggesting that path integration by optic flow is inferior to path integration by kinesthetic and vestibular cues. The experiments presented in this thesis contradict this notion. They demonstrated equal or superior performance compared to nonvisual path integration.

The most obvious difference in homing results between experiments by Péruch et al. (1997) and all other experiments in Figure 13 is the strong general undershooting of turning angles. This might be related to the turn execution error observed by Péruch et al.: When asked to turn around by 180°, participants responded by turning only 150.4° ± 0.9°, corresponding to a underturn by 16%. A similar general underturning of 15% or 20.3° was observed for isosceles triangles. Could this execution error of underturning by 16% explain the underturn of 15% observed for triangle completion, rather than an encoding error as proposed by the authors? Compared to Experiment TOWN&BLOBS, visual homing performance for isosceles triangles by Péruch et al. (1997) yielded significantly reduced homing performance for all performance measures displayed in Figures 13 and 14 ($|t(44)| > 2.1$, p < 0.05).

The question arises as to where the obvious performance difference between Experiment TOWN&BLOBS and experiments by Péruch et al. (1997) stem from. The execution error of underturning observed by Péruch et al. (1997) can only explain the differences in signed turning errors. The remaining performance differences might be caused by the different experimental procedures (training phase, number of triangles). They might, however, also be due to differences in the VR-setup: Péruch et al. (1997) used a joystick and a planar projection screen with non-matched simulated and physical FOV, whereas mouse-button based navigation and a half-cylindrical projection screen with matched simulated and physical FOV was used for Experiment TOWN&BLOBS. Technical limitations in the study by Péruch et al. might also have reduced overall performance. Further experiments might provide a more definitive answer to this question.

11.1.4   Origin of systematic homing errors

To analyze potential origins of the systematic homing errors, Loomis et al. (1993) and Péruch et al. (1997) applied an "encoding error model". This model was initially proposed by Fujita et al. (1993) to explain their blind walking data, and attributes all systematic homing errors to errors in mentally encoding the distances walked and angles turned. It assumes that the internal representation of the triangle satisfies Euclidean geometry (axiom 1), that distances are coded by just *one* function, (i.e., equal distances or turning angles are encoded as equal, (axioms 2&3)), and that there is no systematic error in either the computation of the homeward trajectory or its execution (axiom 4). Loomis et al. (1993) and Péruch et al. (1997) concluded that a compression in the encoding of turns and distances is the only source of the observed systematic errors. Péruch et al. (1997) argued for a nonlinear compression according to a power function with exponents below 1, whereas Fujita et al. (1993) and Loomis et al. (1993) used a simple linear compression.

For the study by Péruch et al. (1997), there is, however, some evidence that the assumption of no execution error (axiom 4) are not met: Péruch et al. reported a significant systematic undershooting by 16% (29.4° ± 0.9°) for requested simple 180° turns. This indicates a turn execution error, which

in turn violates axiom 4 of the encoding error model. This implies that we cannot simply ascribe all systematic errors to encoding errors.

In our studies, systematic encoding and execution errors were negligible for turns and small or irrelevant for distances, and could by no means explain the observed systematic homing errors. We thus argue that participants in our studies had mainly problems with mentally determining the correct homing response (see subsection 8.2.3). This was confirmed by experiment 5, which showed that participants with good mental spatial abilities had less problems determining the correct homing response from the information available. Furthermore, we found evidence that the mental determination of the homeward trajectory was not void of systematic errors: Axiom 2 predicts that participants knew in experiment TOWN&BLOBS that they were traveling isosceles triangles. This is also corroborated by our questionnaires: Almost all participants consciously *knew* they were traveling isosceles triangles. Geometry tells us that for all isosceles triangles, the final turn has to be between 90° and 180°, and cannot be less than 90° (or the path would not be closed). Five out of 20 participants, however, showed mean final turning angles of less than 90° (for isosceles triangles with $\alpha = \pm 30°$), which contradicts axiom 4 or axiom 1. Hence we have to reject the encoding error model for our data, as at least one axiom is clearly not satisfied. Attempts to nevertheless apply this encoding error model to our data produced nonsensical results that violated trigonometry (negative values for encoded angles or distances).

It remains to be seen whether those systematic errors in the mental spatial reasoning phase also occur in the absence of vision (e.g., blind walking). A lack of generalization to blind walking would have far-reaching implications for the understanding of human spatial reasoning and the design of human-computer interfaces.

11.2   General conclusion

The experiments reported here were designed to investigate human navigation ability based solely on visual path integration. The literature indicates that *"humans are incapable of navigating precisely by path integration alone"* (Loomis et al., 1999, p. 143). See Loomis et al. (1999) and Klatzky et al. (1997) for a review. We found, however, that untrained participants were able to reproduce distances and perform turns with relatively small systematic errors, irrespective of movement velocity (Exp. TURN&GO). Especially for rotations, the systematic errors and variance both within- and between-subject were strikingly small, much smaller than for nonvisual turning (Bakker et al., 1999; Klatzky et al., 1990; Loomis et al., 1993; Marlinsky, 1999a). This finding is in sharp contrast with results from turning experiments by Bakker et al. (1999, 2001): Without feedback training, visual information displayed via a head-mounted display led to turning errors that were more than ten times larger than in experiment TURN&GO (for signed error, absolute error, and between-subject variability). Using a flat projection screen with a small FOV, Péruch et al. (1997) found an undershoot of purely visually displayed rotations by 16%. This suggests that the half-cylindrical projection screen used in the present study is the determining factor for the excellent turning performance observed there.

However, the large FOV of 180° does not seem to be the sole determining factor for turning accuracy, even though increasing the FOV has been shown to facilitate navigation (Alfano & Michel, 1990; Arthur, 2000; Ruddle & Jones, 2001). In a study comparable to experiment TOWN&BLOBS, we found that systematically reducing the FOV while leaving the reference frame of the half-cylindrical projection screen visible only slightly decreased homing performance (Riecke, 1998, Exp. 4). This suggests that the half-cylindrical reference frame provided by the projection screen and the visibility of one's own body plays a critical role for navigation performance. Most participants experienced little difficulties determining egocentric angles between objects presented on the screen. The half-cylindrical reference frame might facilitate the estimation of egocentric angles by suggesting a polar

coordinate system. This hypothesis is corroborated by the fact that we did not find the strong bias towards stereotyped turn responses typically observed for triangle completion experiments (Kearns et al., 2002; Klatzky et al., 1990; Loomis et al., 1993; Péruch et al., 1997). On the other hand, flat projection screens or displays in non-circular rooms typically lead to systematic distortions in the judgment of egocentric angles. HMDs appear to produce even more extreme distortions: Participants showed no sensitivity to turning angles and produced the same response regardless of actual triangle geometry (Kearns et al., 2002). This was found for purely visual navigation as well as head-tracked walking. Further experiments are planned and currently being performed to disentangle the individual contributions of display geometry, FOV, spatial reference frames, and visibility of one's body for spatial orientation in virtual environments.

First results showed that visual turning errors were indeed significantly larger when an HMD instead of a projection screen with the same FOV was used (gain factors of 0.56 and 0.71, respectively) (Schulte-Pelkum, Riecke, von der Heyde, & Bülthoff, 2002). In a second experiment, projection screen curvature was investigated and showed a clear effect on turn production: For a flat projection screen with a FOV of $86° \times 64°$, participants typically overshot the intended turning angle (gain = 1.12), whereas they undershot the intended turn when a curved projection screen with the same FOV was used (gain = 0.84) (Schulte-Pelkum, Riecke, von der Heyde, & Bülthoff, 2003; Schulte-Pelkum, Riecke, & von der Heyde, 2003). Reducing the FOV via blinders to $40° \times 30°$ showed virtually no effect for the curved screen (gain = 0.81 versus 0.84), but some effect for the flat screen (gain = 0.94 versus 1.12).

Contrary to our expectation, most participants were *not* able to take advantage of natural-looking landmarks if they were only temporarily visible, indicating that naturalism of the scene did not play an important role (Exp. TOWN&BLOBS). The reasons for this remain unclear. Longer exposure to virtual environments and the experimental procedures might allow participants to develop more efficient strategies, as was demonstrated in Riecke (1998, Exp. 4). Comparing the first and second block of Experiment TOWN&BLOBS, however, showed no significant learning effect (see subsection 8.2.3). This suggests that mere exposure to an experiment does not necessarily improve performance. Conversely, triangle completion experiments with stable, reliable landmarks demonstrated that piloting by salient landmarks and visual scene-matching plays a dominant role in visual navigation, is used whenever possible and leads to almost perfect homing performance (Exp. LANDMARKS).

It is often claimed that kinesthetic and vestibular cues are necessary for spatial orientation tasks involving rotations of the observer (Bakker et al. (1999), Chance et al. (1998), Klatzky et al. (1998), May et al. (1995), see subsection 5.4). It might well be that purely visually displayed movements do not allow for the rapid, obligatory spatial updating found for physical movements (Farrell & Robertson, 1998; May & Klatzky, 2000; Rieser, 1989; Wang & Simons, 1999), see part III. However, the lack of all nonvisual cues in the present experiments did not prevent participants from executing turns, reproducing distances and performing triangle completion tasks with rather small systematic errors. Extended exposure to virtual environments, unlimited response time, and the spatial reference frame and large FOV provided by the half-cylindrical projection screen might all contribute to the relatively good overall navigation performance. For the triangle completion experiments, initial feedback training might also have improved performance and influenced navigation strategies. Optic flow, presented via a half-cylindrical projection screen, provided nevertheless *sufficient* information to solve the tasks.

For visual turning experiments, Bakker et al. (2001) found that feedback training does improve performance, but they conclude that this improvement can *"especially be attributed to a reduction in bias and not to a reduction of the variability of participants' performance"* (p. 222). In experiment TURN&GO, negligible turning bias was found *without* any training, indicating that there was simply no need to calibrate turns. Distance responses and especially mental spatial reasoning, however,

might indeed have improved due to the training. Further experiments are needed to determine what, if any, influences prior training has on spatial orientation in virtual and real environments.

We can only speculate how our results would transfer to more general navigation tasks. If navigation of more complex, moult-segment or continuous routes is based on the same underlying processes, we would expect that mental spatial abilities are again the determining factor for navigation performance. This would in turn predict that each additional segment or turn increases the cognitive load and thus the navigation error, especially for path configurations that are mentally more difficult to picture. For pure path integration, e.g., when participants continuously update some kind of homing vector, response time for homing should not depend on path complexity. Only if participants build up some form of mental representation of the whole path would we predict that the response time also increases with path complexity. Klatzky et al. (1990) and Loomis et al. (1993) found that additional path segments increase homing error as well as response time. This is incompatible with the homing vector hypothesis and suggests that participants build up some mental representation of the whole path, which is in turn used to determine the homing response. The performance decrease was stronger for segments crossing each other, which might be explained by an increased difficulty in representing the route. It should, however, be remembered that any salient landmark potentially leads to a piloting-based navigation strategy and dominates navigation performance.

Using a Virtual Reality setup proved to be a powerful method to investigate human navigation abilities and investigate the underlying mental spatial processes. The "scene swap" paradigm and the 3D field of blobs allowed us to reduce possible navigation mechanisms to purely visual path integration without any landmarks. Using this paradigm, we were able to demonstrate that purely visual path integration is indeed sufficient for basic navigation tasks like rotations, translations, and homing by triangle completion. Furthermore, display geometry, the reference frame provided by the display boundaries as well as the visibility of one's own body seem to influence navigation strategies and performance and should be carefully considered in designing Virtual Reality interfaces.

Part III

Spatial updating in real and virtual environments

12 Introduction

In the following, we will first discuss the reasons that motivated and guided the study of spatial updating described in part III of this thesis (subsection 12.1). Next, an introduction to spatial updating and a brief literature overview will be provided (subsection 12.2), followed by an outline of the three spatial updating experiments performed (subsection 12.3).

12.1 Motivation

The visual path integration experiments presented in part II demonstrated that optic flow, presented on a half-cylindrical $180°$ projection screen, provides *sufficient* information to perform basic navigation and spatial orientation tasks like rotations, translations, and homing with unexpectedly high accuracy. Contrary to the prevailing opinion, systematic homing errors could be mainly ascribed to difficulties in mentally determining the correct homing response (phase 2 in the 3-stage navigation model described in subsection 5.3), whereas systematic errors in encoding the path traveled and executing the desired trajectory were small for translations and negligible for rotations (phase 1 and 3, respectively). Furthermore, mental spatial abilities were found to determine to a large degree participants' homing performance. Certain specifics of the participants responses puzzled us, however, and gave rise to a number of open questions that were addressed in the subsequent experiments described in part III.

12.1.1 Qualitative errors: left-right confusion

For a few trials, some participants in Experiment TOWN&BLOBS and RANDOM TRIANGLES remembered correctly the angle $|\overline{\alpha}|$ just turned, but reported that they completely forgot which direction ($sign\,(\overline{\alpha})$) they turned. This occurred more often when participants were disturbed or not concentrating, or when they were involved in a conversation in the training phase. This left-right-confusion represents a *qualitative* error, not just a quantitative systematic or random error. These qualitative errors have to our knowledge never been reported in human or animal path integration studies when vestibular and/or proprioceptive turn cues above threshold were available. Apart from those few trials, participants were nevertheless able to accomplish the task rather well, much better than for nonvisual path integration (blind walking, see subsection 11.1).

These qualitative errors suggest that some essential spatial cues were missing in the visual path integration experiments. It seems as if visual turn cues did not provide participants with a robust, intuitive knowledge about the turning direction, but required them instead to consciously concentrate on the performed rotations. This hypothesis was corroborated by the participants' verbal reports after the experiment.

12.1.2 Exceptionally long response time

Another observation puzzled us in the triangle completion experiments without reliable landmarks. Before executing the homing response, participants typically took a rather long time (from several

seconds up to more than a minute) to determine the desired homing response. This suggests that the mental spatial reasoning phase (i.e., phase 2 in the 3-stage navigation model presented in subsection 5.3) was not trivial at all, and that the homing response was by no means based on quick, intuitive decisions as is typically observed for nonvisual path integration studies (e.g., Farrell & Robertson, 1998; May, 2000; Rieser, 1989). This relatively long response time suggests again that homing decisions were based on abstract, computationally expensive cognitive strategies and not on intuitive spatial orientation abilities, which are typically used for more natural path integration situations like walking in darkness.

12.1.3   Cognitive, abstract, and computationally expensive strategies

Furthermore, verbal responses of the participants as well as the observation that mental spatial abilities correlated positively with homing performance indicates that abstract, cognitive strategies and abilities were the basis for the observed navigation behavior (instead of intuitive, more natural spatial abilities). These abstract cognitive strategies used in the visual tasks might explain the relatively high overall navigation accuracy, compared to blind locomotion studies (e.g., Klatzky et al., 1990; Loomis et al., 1993; Sauvé, 1989), see section 11.

It seems like the optic flow and reference frame of the projection screen provided participants *in principle* with all the information necessary (angle turned and distance traveled, derived from integrating movement velocity or acceleration over time) to perform basic navigation tasks, but that this information had to be processed on a rather abstract, highly cognitive level. This high cognitive demand might explain why participants were so easily disrupted by any kind of distraction. This is quite different from the quick, intuitive responses observed from spatial updating studies where participants locomoted without vision, but with vestibular and/or kinesthetic movement signals (Farrell & Robertson, 1998; Klatzky et al., 1998; May & Klatzky, 2000; May, 2000; Presson & Montello, 1994; Rieser, 1989). There, participants seem to have a natural, effortless, and automatic "feeling of where they are and where things are" (i.e., they experience a high degree of "spatial presence", see section 16). This in turn leads to a faster and more intuitive and robust access to spatial information about their immediate surround. So what exactly is missing in the visual cues?

12.1.4   Spatial updating as a prerequisite for intuitive spatial orientation

When moving through space, our sensory inputs somehow automatically transform the "world inside our head" accordingly so as to stay in alignment with the outside world. This "spatial updating" of our egocentric mental spatial representation occurs without conscious effort and is normally "obligatory" in the sense of being largely beyond conscious control and hard to suppress. See subsection 12.2 for an in-depth introduction to spatial updating.

From the literature, we know that spatial updating is typically impaired when proprioceptive and vestibular cues in particular are missing (Klatzky et al., 1998; May & Klatzky, 2000; Presson & Montello, 1994; Rieser, 1989; Simons & Wang, 1998; Wang & Simons, 1999; Wang & Spelke, 2000; Wraga, Creem, & Proffitt, 2003). Qualitative errors seem to occur most often when kinesthetic and/or vestibular cues about ego-turns are missing. Klatzky et al. (1998) and May & Klatzky (2000) for example found that participants completely forgot to update ego-rotations that were not physically performed, i.e., when the corresponding vestibular and proprioceptive cues were missing.

This made us hypothesize that the lack of concomitant vestibular motion cues in the experiments described in part II impaired spatial updating, which in turn might explain the qualitative errors (left-right confusions) observed. Might it be that the visual cues provided sufficient cues to solve simple

navigation tasks *cognitively*, but that those cues were incapable of initiating proper *spatial updating* during simulated ego-motions? If this was true, all VR setups that rely heavily on visual cues for simulating ego-motions while omitting vestibular cues (i.e., most of the existing and affordable VR setups) might face the same problem, namely that they do not allow for "normal" and intuitive navigation, as they do not sufficiently trigger spatial updating. Or are there any conditions where visual information *can* be sufficient for triggering spatial updating, i.e. to automatically and mandatorily transform the world inside our head accordingly during simulated ego-motions? Part III was designed to answer this question.

Benefits from even partial and preliminary answers to this question are twofold. First, they provide insight into the way spatial information is processed and integrated in the human brain. Second, this knowledge can help to understand what part of the spatial information and what stimulus parameters are really essential for good spatial orientation and for initiating proper spatial updating. By knowing what information is critical for enabling excellent spatial orientation, and what cues or simulation parameters are less critical or can even be ignored completely, we are empowered to devise more elegant paradigms for ego-motion simulation. Ultimately, this knowledge can hence initiate first steps towards a lean and elegant ego-motion simulation paradigm, by for example reducing the amount of physical movements to the absolute minimum that is still sufficient for allowing normal spatial orientation.

12.1.5 Main approach

In the experimental methodology used in part II, participants were apparently able to somehow cognitively compensate for the missing vestibular cues and lack of proper spatial updating. Hence, if we want to asses the cues and simulation parameters necessary for "normal" spatial orientation and for initiating proper spatial updating, we need to establish a different experimental paradigm that reduces the usability of abstract or compensatory cognitive strategies to an absolute minimum.

Ideally, one would want to study a mostly autonomous process that is largely automatized and thus does not require any abstract cognitive strategies. Spatial updating is a process that is not exactly autonomous, but nevertheless largely automatized, as it occurs under most natural situations effortlessly and does not seem to require any attention. This automaticity ensures that spatial updating can operate even under conditions of high stress or cognitive load. But how can we distinguish between this "automatic spatial updating" and abstract cognitive strategies?

This could be achieved by choosing a task that renders cognitive strategies difficult to use or very time-consuming. If the task is hard enough such that participants are pushed to their performance limits, they might have to resort to automatized processes to solve the task at all. If participants are at the same time empowered to use quick and automatized processes, they might choose to do this instead of the more cumbersome or time-consuming cognitive strategy. If they are moreover forced to respond quickly by limiting their response time, they might just not have enough time to successfully apply abstract cognitive strategies and prefer to use more intuitive responses. To achieve all of this, we established a rapid pointing paradigm where participants were moved to new positions and asked to point "as accurately and quickly as possible" to different previously-learned targets announced consecutively via headphones, as will be described in detail in subsection 13.2. Additionally, the difficulty of the task was increased by using a relatively large number of pointing targets (12 or 22). Furthermore, quick responses were enforced by limiting the response time allotted.

Using this rapid pointing paradigm, we are empowered to distinguish between the usage of abstract cognitive strategies and natural spatial updating: As spatial updating is largely automatized, it can still operate well under the high task demands and will yield shorter response times and less, if any,

qualitative errors. Only if the available spatial cues are insufficient to enable automatic spatial updating will participants have to resort to using cognitive strategies. As these are not sufficiently automatized (yet), they can be identified by increased response times and qualitative errors. Furthermore, we would expect increased systematic and random errors if the allotted response time is insufficient to allow for the successful application of compensatory cognitive strategies. In this manner, we are able to identify combinations of spatial cues and stimulus parameters that allow for automatic spatial updating and distinguish them from those which do not. This yields on the one hand a deeper understanding of the integration and interaction of spatial information in the human brain. On the other hand, it allows to pinpoint critical spatial cues and simulation parameters required for enabling quick and intuitive spatial orientation in virtual environments.

The following subsection provides an introduction to the spatial updating literature and experimental paradigms used.

12.2   Spatial updating - introduction and literature overview

In this subsection, we will start by introducing spatial updating and discussing its relevance for spatial orientation. Next, the problem of quantifying spatial updating will be discussed, followed by an introduction into experimental paradigms used in the literature and novel paradigms proposed by the author. We will conclude by presenting relevant findings from the spatial updating literature.

12.2.1   Introduction and terminology

Being mobile species, humans as well as most animals constantly change their position and orientation in space. Due to this ego-motion, the spatial relationships between the observer and the surrounding environment changes continuously. After only a few steps, the relations to all nearby objects have changed considerably according to the ego-motion. As long as all relevant objects are directly and constantly visible that might seem uncritical. As soon as some objects are occluded by others, however, it would make sense to have a process that ensures that we still know where everything is even though we cannot currently perceive it directly.

The situation becomes even worse if vision is completely excluded. Imagine for example that you are at home at night when the main fuse blows. You will have to find your way around in complete darkness until you find candles or the fuse box. Even though you know your home quite well and know how it looks like in principle, this knowledge won't help you much if you were not able to know where everything is from your current position, and easily keep track of where everything is during all motions, even in darkness.

Having to consciously keep track of all potentially relevant objects in the surround is virtually impossible and would increase the cognitive load to a level that would not allow for any further complex task. As spatial orientation is a rather frequent and essential task, and involves knowing where one is with respect to the surround, it would be quite inefficient to allocate general cognitive resources for this task. That is, we need a process that automatically keeps track of where relevant surrounding objects are while we locomote, without much cognitive effort or mental load.

This mostly automatic and seemingly effortless process is mostly referred to as "spatial updating" (Amorim & Stucchi, 1997; Farrell & Robertson, 1998; Farrell & Thomson, 1998; Hollins & Kelley, 1988; Klatzky et al., 1998; Loomis, Da Silva, Philbeck, & Fukusima, 1996; Presson & Montello, 1994; Rieser, Guth, & Hill, 1982). It is this process that allows us to locomote in darkness without much cognitive load or constantly bumping into obstacles, by providing quick and intuitively knowledge of where everything is, even during complex motions. Spatial updating allows us to quickly and accurately grasp objects, point or look into their direction, even when they are not directly visible.

All of these ecologically relevant actions are performed in body-coordinates, that is, with respect to an egocentric reference frame. We therefore put forth the hypothesis that spatial updating mainly updates this egocentric (bodily) reference frame, and not the position and orientation of oneself in an allocentric (world-based) reference frame. Only this egocentric updating allows for quick and accurate access in body-coordinates without having to perform additional coordinate system transformations from allocentric to egocentric reference frames.

Automatic spatial updating   Due to genetic inheritance, life-long training and exposure, spatial updating seems to be directly and tightly coupled to our bodily motions. That is, spatial updating occurs seemingly effortlessly and without extra attention (Rieser, 1989; Rieser, Guth, & Hill, 1986). It is thus "automatic" or "automated" in the sense that it occurs automatically during ego-motions, without us having to consciously concentrate on it. The literature often refers to "**automatic spatial updating**" if the mental spatial representation remains aligned with the outside world during ego-motions (see, e.g., Wraga et al., 2003). Vestibular and kinesthetic cues from blindfolded motions, for example, proved to be sufficient to enable automatic spatial updating during rotations as well as translations (Easton & Sholl, 1995; Farrell & Robertson, 1998; May & Klatzky, 2000). This automaticity ensures that the attentional and cognitive demands of spatially updating the egocentric reference frame are minimal, and that spatial updating does not interfere with other cognitive or non-cognitive tasks. As automatic spatial updating seems to be the default process of updating one's egocentric reference frame, it is used interchangeably with spatial updating, and the adjective "automatic" will only be used to emphasize the automaticity. As imagined perspective-taking or imagining ego-motions are a rather intentional and hence a conscious and cognitive process, it lacks automaticity. Strictly speaking, we would consequently not refer to it as spatial updating in the narrower sense. It might be referred to as a more generalized spatial updating of the egocentric reference frame (see Figure 15). Adopting an allocentric reference frame by for example imagining oneself in a spatial context from a bird's eye view lacks the egocentric perspective and would thus not fall under our definition of spatial updating.

Obligatory spatial updating   As it might be rather hazardous if we could somehow forget to update our reference frame under situations of high stress or high cognitive load, it seems sensible to propose a spatial updating process that operates always and irrespective of our attention or conscious decision. Moreover, as it does not seem to make any sense to *not* update our egocentric reference frame during ego-motions, this process should ideally be reflex-like and beyond conscious control.

Spatial updating seems indeed under full-cue conditions so tightly coupled to the motion cues that we cannot help but update the world inside our head when moving through its real counterpart. That is, any perceptually signaled movement seems to be mandatorily incorporated into our representation of our current position and orientation and cannot simply be excluded by volition. Only with great effort can we cognitively compensate for this compulsory updating and try to ignore moving to a different position and orientation in space and, e.g., imagine that we still are at the original location (see, e.g., Farrell & Robertson, 1998; May & Klatzky, 2000). That is, if we perform an ego-turn, proprioceptive, vestibular, visual, and auditory cues somehow initiate a corresponding counter-rotation of the world inside our head, whether we want to or not. Hence, spatial updating can under those conditions be considered mandatory or obligatory in the sense of being hard-to-suppress and thus to a large degree cognitively impenetrable. To reflect this reflex-like phenomenon, we introduced the term "**obligatory spatial updating**". It seems as if the world inside our head has some kind of inertia that makes it stay in alignment with the outside world and prevents it from moving with the head. It thus acts just like a gyro-compass, but not only for rotations but also for translations.
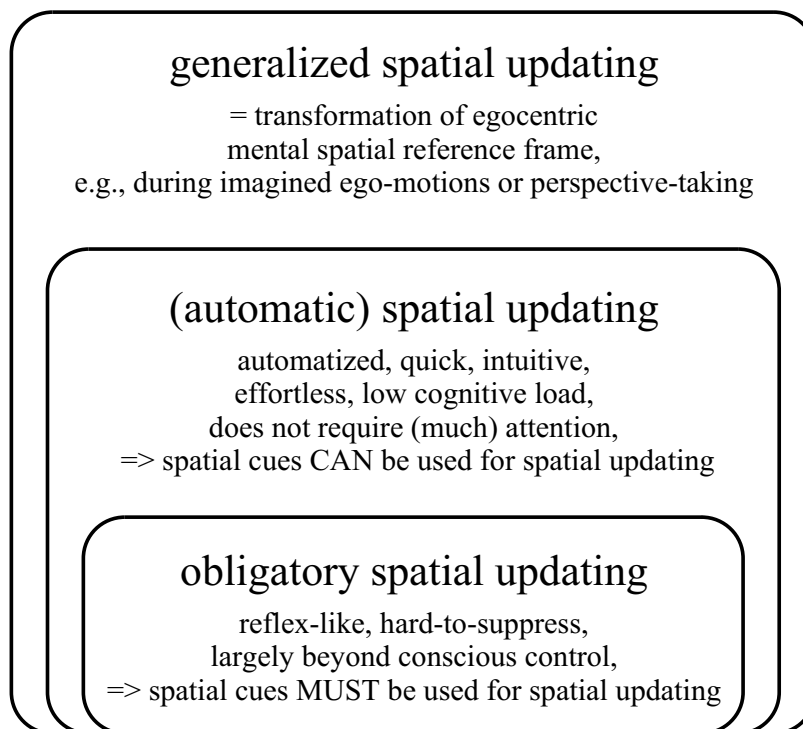
Figure 15: Connection between generalized, automatic, and obligatory spatial updating. At the most general level, **generalized spatial updating** refers to all spatial transformations of our egocentric mental spatial reference frame. This includes mental perspective-taking or consciously updating our egocentric reference frame during imagined ego-motions. **Automatic spatial updating**, which is often referred to as simply spatial updating, is a more specific subset and refers to the largely automatized transformations of our mental egocentric reference frame. Due to this automaticity, both the cognitive load and attentional demands are minimal, if not zero. **Obligatory spatial updating** is a subset of the more general (automatic) spatial updating. It refers to the reflex-like, hard-to-suppress and thus cognitively almost impenetrable phenomenon of perceived spatial cues triggering spatial updating, whether we want to or not. Conversely, spatial cues are called sufficient for triggering obligatory spatial updating if they *must* be used, i.e., if they mandatorily transform our mental spatial reference frame whether we want to or not. Furthermore, spatial cues are called sufficient for enabling (automatic) spatial updating if they *can* trigger this automatic process, but do not necessarily *have* to be used. Two examples might illustrate the difference between automated and obligatory (reflex-like) behavior: When riding a bicycle after extensive practice, keeping balance happens seemingly automatically and without effort. It is, however, not reflex-like (obligatory), as one could still consciously choose to lose balance and fall over. Playing the piano is another example where extensive, year-long practice helps to automate the motions. Even professional piano players can consciously decide to play wrong, however, indicating that the process is not obligatory, but only automated.

Spatial updating performance can, however, be slightly impaired by high cognitive loads, such as counting backwards in steps of seven or three or verbalizing nonsense syllables (Yardley & Higgins, 1998; May & Klatzky, 2000). But even under those conditions, spatial updating still seems to be obligatory in the sense that the mental spatial representation is continuously being updated according to the ego-motions, even though it might not stay perfectly aligned due to accumulating errors induced by the high cognitive load. To specify this, we prefer to use the term "obligatory spatial updating" instead of the unspecific and more general "automatic spatial updating", even though this distinction has to our knowledge never been made in the literature (Farrell & Robertson, 1998, 2000; Farrell & Thomson, 1998; Rieser et al., 1982; Wraga et al., 2003; Yardley & Higgins, 1998). This distinction is illustrated in Figure 15.

On the other hand, certain combinations of spatial cues can be ignored rather easily, and are consequently not sufficient to trigger obligatory spatial updating. Examples of this might include purely

vestibular cues from smooth motions (see Experiment 6 below, section 13). Nevertheless, those cues might under certain conditions be sufficient to enable participants to update their mental reference frame according to the ego-motion. If this is done cognitively, we would expect a considerably increased response time, and would call the underlying process generalized spatial updating, but not automatic spatial updating, as automaticity would imply quick and intuitive responses and consequently short response times. Automatic spatial updating can thus be recognized by quick and intuitive responses and thus short response times and little, if any, qualitative errors. If, for example, the available motion cues *can* be used for quick and intuitive spatial updating, but *don't necessarily have to be used*, we would call the underlying process automatic spatial updating which is not obligatory. If, on the other hand, it is easier to use the available motion cues to update to the new position than to ignore them and act as if one were still be in the original location, we would say that the motion cues *must* be used to update to the new position or orientation, and would hence call the underlying process obligatory spatial updating. It is self-evident that motion cues that *must* be used *can* also be used. That is, obligatory spatial updating implies automatic spatial updating, or, in a logical notation, obligatory spatial updating $\Longrightarrow$ automatic spatial updating. Conversely, automatic spatial updating is a necessary but not sufficient prerequisite for obligatory spatial updating.
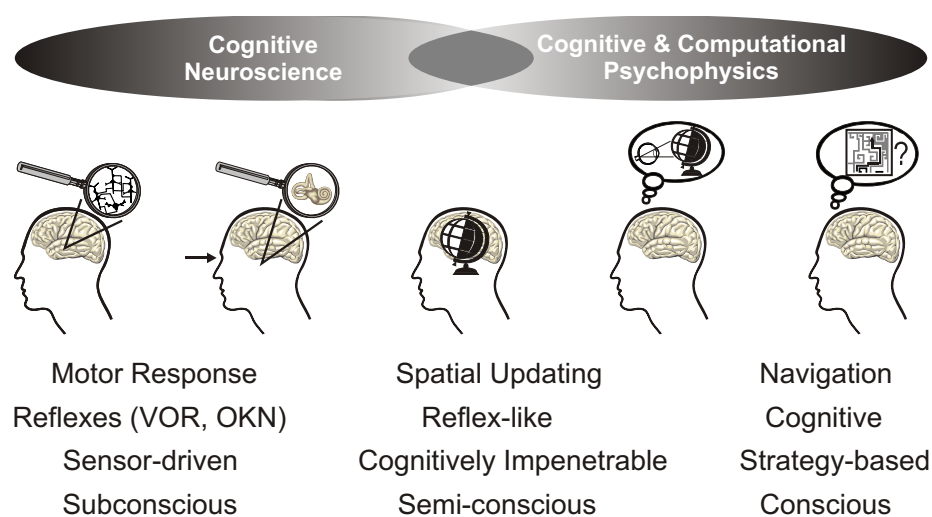
Figure 16: Spatial updating as a link between low-level reflexes and high-level, strategy-based processes.

In summary, spatial updating under full-cue condition can be seen as a reflex-like, largely automatized process that is in most cases beyond conscious control and hard-to-suppress (obligatory). It thus takes an intermediate role between well-known, mostly sensor-driven reflexes like the vestibulo-ocular reflex (VOR) or the opto-kinetic after-nystagmus (OKAN) on the one hand and cognitive, strategy-based processes involved in navigation on the other hand (see part II). This role of spatial updating is illustrated in Figure 16. Spatial updating is furthermore a robust basis for more complex spatial processes like spatial orientation and wayfinding, but has the advantage of being less affected by cognition, strategies, and interpersonal differences. Spatial updating thus allows us to non-invasively study a fundamental, reflex-like process using behavioral measures in a high-level, psychophysical paradigm.

In the following, we will review the methodology and some relevant results of the spatial updating literature.

12.2.2   How can spatial updating be quantified?

Ideally, one might want to use non-invasive brain imaging to investigate the process of spatial updating in humans. Using electrophysiology in rats, "place cells" and "head direction cells" have been identified. Those cells fire when the animal is at a specific location or orientation, respectively, in space (Mittelstaedt, 2000; O'Keefe & Nadel, 1978; Taube, Muller, & Ranck, 1990a, 1990b). As it is not particularly useful to think of individual place cells or head direction cells as encoding any particular location or orientation, Samsonovich & McNaughton (1997) introduced an attractor map concept that is capable of representing coordinates in any arbitrary environment. These attractor maps might represent a neural correlate of spatial updating. If non-invasive methods would reveal similar mechanisms in humans, those methods could in principle be used to quantify and investigate spatial updating. So far, however, non-invasive brain imaging methods are limited in resolution to hardly less than 1 mm. This allows the identification of brain regions that are involved in spatial orientation and navigation, like the intraparietal sulcus (IPS) (Bremmer, Schlack, Shah, Zafiris, Kubischik, Hoffmann, Zilles, & Fink, 2001) or the hippocampus (Maguire, Frith, Burgess, Donnett, & O'Keefe, 1998b; Maguire, Burgess, Donnett, Frackowiak, Frith, & O'Keefe, 1998a; Epstein & Kanwisher, 1998). Recent technological advances in optical imaging managed to achieve a sufficient sub-cellular resolution, but are unfortunately limited to the surface of the brain. As most brain activities related to spatial orientation seem to be located rather far away from the surface (like the hippocampus), existing non-invasive brain imaging technologies are unfortunately insufficient for investigating spatial updating. Indeed, to our knowledge no neural correlates suitable for the online non-invasive investigation of spatial updating or spatial orientation have been identified in animals including humans (Farrell, 1996), even though the right dorsal area seems to be involved in humans (Farrell & Robertson, 2000). Hence, if we want to investigate spatial updating, we need to utilize different methods like behavioral measures, that is, psychophysics. Methods commonly used in the literature as well as novel methods proposed by the author are summarized in the following subsection.

12.2.3   Methodologies and experimental paradigms used in the spatial updating literature

In spatial updating studies, participants are typically moved to a new position or orientation, and have to perform a spatial task from that new position. The general reasoning is the following: Only if the mental spatial representation is already automatically updated can the participant give quick, intuitive, and accurate spatial answers.

Various psychophysical methods are used to quantify spatial updating, all of them include probing the mental spatial reference frame after a real, simulated, or imagined ego-motion. All moving species have to somehow know where *they* are in their surround and where relevant *objects/landmarks* of interest are with respect to them. The most natural, ecologically valid response is thus to ask participants to quickly point towards (currently invisible) objects that were previously learned (see, e.g., Creem & Proffitt, 2000; Rieser & Rider, 1991; Rieser, 1989; Wang & Spelke, 2000; Wraga et al., 2003). Response time, pointing error and pointing variance then quantify the spatial updating process in terms of ease, accuracy, and consistency or configuration error, respectively (see subsection 13.2.5). Other methods include asking participants to name the target that is currently at a specific orientation, e.g., "what is left?" (Carpenter & Proffit, 2001; Wraga et al., 2003), or asking the participants to indicate their new orientation verbally by stating what number on the clock face they are currently facing, given that the initial orientation was 12 o'clock (Yardley & Higgins, 1998). Pointing with body parts like the index fingers or extensions of body parts like canes or short sticks was found to yield the highest accuracy and lowest variability, compared to other methods like rotating dials, drawing, or verbal statements (Haber, Haber, Penningroth, & Novak et al.,

1993; Lehnung, Haaland, Pohl, & Leplow, 2001). Simple navigation tasks like return-to-origin or face-origin paradigms can also be used to test whether the available motion cues were automatically integrated into the perceived ego-position, and hence to investigate spatial updating (May & Klatzky, 2000).

Experimental paradigms in spatial updating studies typically include several of up to four stereotypical spatial updating conditions (see, e.g., Farrell & Robertson, 1998), which are summarized below and in Figure 17.

1. **UPDATE**: After viewing the scene/target objects, participants are moved (often without vision) to a new position. From there, they have to point to the true location of one or several targets (announced, e.g., visually on a screen or auditorily via headphones). If the available spatial updating cues are sufficient, UPDATE performance should not depend on angle turned or distance traveled. This is what is sometimes referred to as automatic spatial updating (Farrell & Robertson, 1998; Wraga et al., 2003). When piloting (landmark-based navigation) is not possible, however, response accuracy declines with length and complexity of the trajectory due to accumulation errors in the path integration process (Klatzky et al., 1990; Loomis et al., 1993).

2. **CONTROL**: Participants are moved to a new position and immediately back to the original position before being probed. This is a simple baseline condition yielding optimal performance. This forth-and-back motion is simple enough that spatially updating the motion should be rather trivial - the motion cues just need to indicate *that* is was a forth-and-back motion, without the need to know the distance moved or anything else. Hence, the observed response might rather reflect the pointing performance for the given *static* spatial cues (e.g., visual display), without too much influence from the motion cues. If the available spatial updating cues are sufficient, UPDATE performance should be about as good as CONTROL performance. That is, participants *can* use the available spatial cues to automatically update to new positions as well as in the baseline (CONTROL) condition. Hence, the available motion cues would enable automatic spatial updating.

3. **IMAGINE**: After viewing the scene/targets, participants are blindfolded. Participants do not move, but instead have to imagine moving to a new position which is announced auditorily. They then have to respond (e.g., point to the named targets) as if they were actually at the new, imagined position. This condition tests whether generalized spatial updating can be consciously performed under reduced or no-cue conditions. On the one hand, this reveals the degree of control we have over transforming our mental spatial representation. On the other hand, IMAGINE conditions can be used to test which spatial cues help for imagined ego-motions and might thus be more critical.

4. **IGNORE**: Participants are moved to a different position, but asked beforehand to ignore that motion and "respond as if you had not moved". That is, participants are asked to imagine that they are still at the original position, facing the initial direction. If the available spatial cues are more powerful in triggering spatial updating and hence transform the world inside our head (even against our conscious will), those motions should be harder to ignore. Spatial updating would then be "obligatory" or "reflex-like" in the sense of consciously hard-to-suppress and hence largely beyond conscious control. Thus, IGNORE tasks can be used to investigate the potential cognitive influence on the reflex-like process of spatial updating under various combinations of instructions, cues and sensory modalities. Compared to UPDATE trials, which quantify how well participants *can utilize* the available spatial cues to spatially update to new orientations/positions (automatic spatial updating), IGNORE trials reveal whether the spatial cues *must be utilized*, i.e., whether they trigger spatial updating even against our own con-
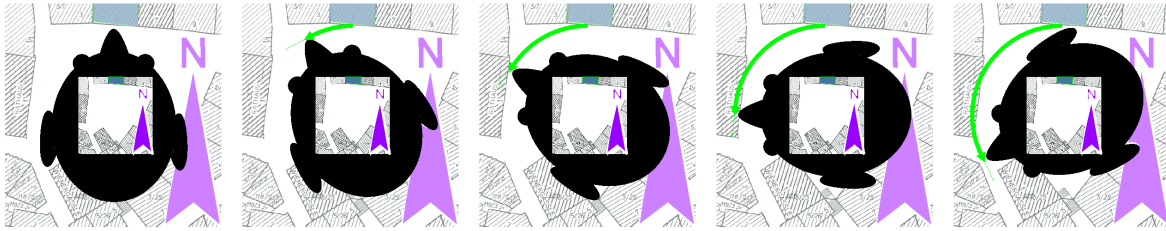
(a) UPDATE condition

(b) CONTROL condition
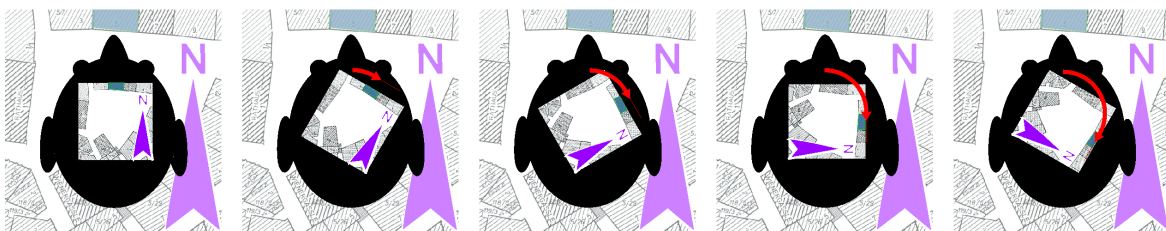
(c) IMAGINE condition
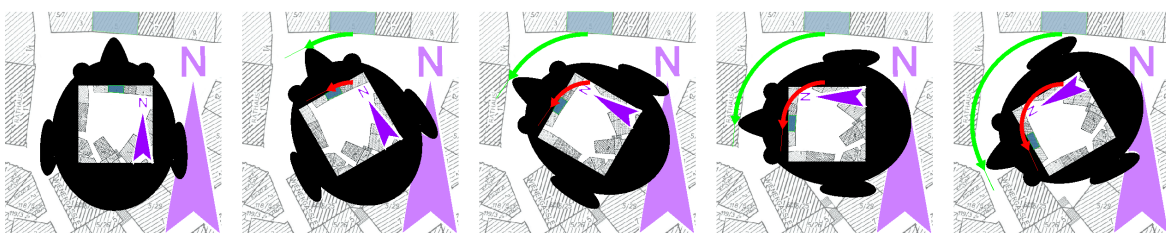
(d) IGNORE condition

Figure 17: Cartoon-like illustrations of the different spatial updating conditions. The black head depicts an observer positioned in a scene (here: the Tübingen market place), represented as the surrounding map. The left plots indicate the respective initial condition, where the observer faces for example north (indicated by the large arrow), and the egocentric mental spatial representation (symbolized by the small map inside the head) is aligned with the surrounding scene. Each row depicts a motion sequence for a left turn in the spatial updating conditions UPDATE (a), CONTROL (b), IMAGINE (c), and IGNORE (d). Note that in the IMAGINE and IGNORE conditions, the mental spatial representation of the surround is no longer aligned with the physical surround.

scious will (obligatory spatial updating). Thus, UPDATE trials quantify automatic spatial updating performance, whereas IGNORE trials quantify the obligatory or reflex-like, cognitively impenetrable component of the spatial updating process triggered by the presented stimuli. If, on the other hand, the task is solved mainly in a non-automatic, highly cognitive or abstract manner, then IGNORE performance should be comparable to UPDATE and especially CONTROL performance.

After IGNORE trials, participants are potentially confused and need to be re-anchored to the real world or correct position and orientation. Therefore, we devised an additional "IGNORE BACKMOTION" condition that comes right after the IGNORE condition, and has to our knowledge never been used in the literature.

5. **IGNORE BACKMOTION**: After each IGNORE trial, participants are moved back to the previous position and orientation. The main purpose of this condition is to avoid potential disorientation that might be induced by the previous IGNORE trial and to re-anchor participants to the previous location by asking them to point and thus probe their mental spatial reference frame. Comparable performance in the IGNORE BACKMOTION and UPDATE condition would suggest that participants were properly re-anchored to that orientation and no longer disoriented by the IGNORE trial beforehand. This is a critical prerequisite for all repeated-measures designs.

Using VR technology, the different sensory modalities can easily be simulated independently, and participants can for example be asked to focus on one specific sensory modality (e.g., vision) while ignoring others (e.g., vestibular, auditory, etc.). See section 13 and Berger, von der Heyde, & Bülthoff (2002) for examples of this approach. Using this paradigm, the individual contributions of the different senses as well as their interaction and integration in the human brain can be investigated. From a human factors point of view, this enables us to understand what factors are critical for enabling automatic as well as obligatory spatial updating. This in turn empowers us to pinpoint critical factors for achieving optimal spatial orientation in virtual environments. These parameters can include VR simulation and setup parameters as well as specifics about relevant sensory modalities and combinations thereof.

In the subsequent experiments, we used rapid pointing tasks to measure the speed and accuracy with which humans can point to objects after being passively moved to different locations in space. This pointing metaphor - much like shooting - has the advantage of allowing the participant only very limited time to perform complex spatial reasoning and utilize abstract mental or geometric strategies, as is often observed in navigation and spatial orientation experiments (see part II and subsection 12.1). Thus, rapid pointing allows us to investigate the expectation of where participants think they are by measuring where they expect objects in their close surround to be with respect to their current position. Using simple geometric deduction, we can then estimate where participants think they are with respect to the surround.

12.2.4 Results and findings from the spatial updating literature

For pointing tasks as well as return-to-origin tasks, kinesthetic and vestibular cues from blindfolded locomotion were found to be automatically incorporated into a configural, map-like representation of one's ego-position with respect to the surround (Easton & Sholl, 1995; Farrell & Robertson, 1998; May & Klatzky, 2000). Instructions to ignore some movements led to considerable errors, indicating the difficulty of consciously influencing spatial updating. Those errors were much greater than those induced by verbal distractions, indicating the inability to ignore physical movements during path integration. This suggests that kinesthetic and vestibular cues from blind locomotion are sufficient to trigger obligatory spatial updating.

Rotations are typically as easy to update as translations, but considerably harder to ignore or imagine (Easton & Sholl, 1995; Farrell & Robertson, 1998; Klatzky et al., 1998; May & Klatzky, 2000; May, 1996; Presson & Montello, 1994; Rieser, 1989). For IGNORE and IMAGINE trials, response latencies typically increase with turning angle, suggesting that participants perform some kind of mental ego-rotation, with a limited rotational velocity (Farrell & Robertson, 1998; Rieser, 1989). For the IGNORE trials, this suggests that participants updated their position automatically and had to "undo" this updating retrospectively to reestablish their original orientation. This cognitive effort involved in imagining or ignoring rotations is similar to difficulties observed when having to use misaligned maps or novel perspectives (May et al., 1995; Shelton & McNamara, 2001; Presson & Hazelrigg, 1984; Roskos-Ewoldsen, McNamara, Shelton, & Carr, 1998). The observed difference between rotations and translations is in accordance with the prevailing opinion that vestibular and kinesthetic cues are indispensable for obligatory updating of ego-rotations (Chance et al., 1998; Klatzky et al., 1998; May et al., 1995). Studies on object recognition and object array recognition demonstrated similar advantages of physical ego-motions (around objects) over object rotations (Carpenter & Proffit, 2001; Simons & Wang, 1998; Simons, Wang, & Roddenberry, 2002; Wang & Simons, 1999; Wraga et al., 2003), see also subsection 17.4.1. This was even found for *imagined* ego-motions versus imagined object rotations (Creem, Wraga, & Proffitt, 2001; Wraga, Creem, & Proffitt, 2000).

In this paper, we will, however, present evidence that visual turn cues *alone can be sufficient* for inducing obligatory spatial updating and hence turn the world inside our head, even without any concurrent vestibular or kinesthetic turn cues (see sections 13, 14, and 15). This was found for photo-realistic visual stimuli from well-known environments including an abundance of salient landmarks, but not for optic flow (see section 15 and May & Klatzky (2000)).

Optic flow information, on the other side, provides *sufficient* information to perform basic navigation tasks like rotations, translations, and homing, at least when presented via a curved 180° projection screen, see part II. More specifically, optic flow provides sufficient information to solve spatial tasks *cognitively*, but *not* to induce *obligatory* spatial updating (i.e., to turn the world inside our head, even against our own conscious will, see Klatzky et al. (1998) and section 15).

Using a return-to-origin paradigm after a simple linear excursion, May & Klatzky (2000) reported that instructions to ignore irrelevant movements lead to systematic errors. Those errors were larger than errors resulting from an increased cognitive load induced by having to count backwards in steps of threes. This data pattern was observed for blindfolded walking as well as joystick-based virtual locomotion in a simple "virtual forest" presented via HMD. That is, vestibular and proprioceptive cues from blind walking as well as purely visual cues without reliable landmarks cannot be ignored completely. Furthermore, the to-be-ignored movements had greater effects than could be attributed to cognitive load per se. This seems to suggest that translatory cues from visual path integration alone are able to induce obligatory spatial updating. The study by May & Klatzky (2000) did not, however, include an UPDATE condition where participants were asked to update the additional movement. Consequently, we can only conclude that an *additional*, to be ignored movement decreases performance compared to a condition *without* that additional motion. As overall performance showed considerable errors even without the additional motion[6], it might well be that merely adding another motion might already explain most of the observed difference. Hence, we cannot judge whether the visual cues were indeed able to induce obligatory spatial updating as defined above (see subsection 12.2.1 and Figure 15). Nevertheless, May & Klatzky (2000) showed clearly that path integration cues from blind walking as well as purely visual translations cannot be ignored completely. Furthermore, the direction of the induced errors is consistent with the interpretation that the to-be-ignored translations were at least partially used for spatial updating. For rotations, however, Klatzky et al.

---

[6]For the blindfolded condition, homing distances were consistently undershot and showed a significant regression towards stereotyped responses (distance gain = 0.72). The HMD condition showed no general undershooting, but the compression was slightly more pronounced (gain = 0.61).

(1998) found that visual cues from optic flow without corresponding physical turns were not automatically incorporated into the perceived ego-orientation. Hence, it seems as if optic flow alone might be able to allow for automatic or even obligatory spatial updating of translations, but not for rotations. The literature on spatial updating including visual cues is rather sparse, however, and allows only for preliminary conclusions. The experiments presented in this paper are an attempt towards providing more definite answers on the potential of visual cues for enabling automatic or even obligatory spatial updating.

Apart from *egocentric reference frames*, which neuroanatomically seem to be closely related to the dorsal stream, humans can use *allocentric reference frames* ("cognitive maps") associated with the ventral stream and responsible for example for landmark-based navigation (piloting) (Farrell, 1996). But even those cognitive maps contain orientation-specific and view-dependent representations, and are thus linked to spatial updating (Hintzman, O'Dell, & Arndt, 1981; Roskos-Ewoldsen et al., 1998; Shelton & McNamara, 1997).

In terms of spatial updating, recent evidence suggests an asymmetry between the nonvisual updating of local and global landmarks. When asked to turn either with respect to the room (local targets) or with respect to surrounding campus buildings (global targets), participants showed an asymmetry when asked to point to targets from the updated environment (i.e., targets from the environment in which they turned) versus targets from the other, non-updated environment (Brockmole & Wang, 2002; Wang & Brockmole, 2003). Participants automatically updated the local targets when moving with respect to the global targets, but global targets were *not* updated automatically for turns with respect to the local environment. This suggests that spatial updating does not necessarily occur for the whole environment, but at least for the local environment, that is potentially more relevant for actions.

Spatial updating is furthermore affected by the nature of the objects constituting the reference frames: The geometry of rooms and consistent scenes generally yields more stable, persistent spatial representations than individual objects or object configurations, and are less affected by disorientation, especially for children (Gallistel, 1990; Gouteux & Spelke, 2001; Wang, 1999; Wang & Spelke, 2000). Views and object arrangements aligned with stable reference frames are moreover easier to imagine and lead to reduced pointing errors than unaligned ones (Shelton & McNamara, 1997, 2001). Having to imagine or ignore ego-rotations seems to involve two conflicting reference frames, a primary one that is being updated automatically during ego-motions, and a consciously adopted, secondary one (Presson & Montello, 1994).

In the remainder of this paper, we will demonstrate that visually presented consistent, landmark-rich scenes are indeed accepted as a stable reference frame and can thus be sufficient to induce obligatory spatial updating of the world inside our head.

12.3   Conclusions and outline of the experiments

Part II demonstrated that purely visual cues enable participants to solve basic navigation problems using rather abstract, highly cognitive strategies. Participants, however, did not seem to have the quick, intuitive, and robust spatial knowledge typically associated with intact spatial updating. Hence, the visual cues provided were apparently insufficient for initiating automatic or obligatory spatial updating during simulated ego-motions. Several features of those experiments might have prevented spatial updating from occurring, including the non-immersiveness of the projection setup and the layout of the virtual environments.

In the subsequent experiments of part III, we will investigate whether spatial updating by visual cues might nevertheless be possible if photorealistic replicas from well-known, landmark-rich scenes are

utilized as visual stimuli. Furthermore, we used more immersive visualization setups like a high-resolution head-mounted display and a purpose-designed and custom-built projection setup, with the goal to increase immersion and spatial presence and thus render the visual cues more powerful.

All these experiments were performed on a motion platform, which allowed the precise control of additional vestibular turn cues. In this manner, we were able to gradually decrease the amount of concurrent vestibular motion cues or even eliminate them completely while still being able to elicit obligatory spatial updating. These results challenge the prevailing opinion that vestibular cues are absolutely indispensable for proper spatial updating and spatial orientation (e.g., Bakker et al. (1999), Chance et al. (1998), Klatzky et al. (1998), May et al. (1995), see also subsection 5.4 and 12.2.4). Our approach thus extends spatial updating work beyond the typically studied nonvisual (blindfolded or imagined) conditions to include high-quality visual cues. Furthermore, using VR technology allows us to disentangle the contribution and interaction of visual and vestibular cues for spatial updating.

As there has hardly been any research on spatial updating using visual cues, current knowledge did not allow for very refined hypotheses, and some of this research is consequently rather exploratory in nature. In fact, we are only aware of one full paper that deals explicitly with spatial updating including complex, non-trivial visual information, and even that is still under review (Wraga et al., 2003), see also subsection 17.4.4. Thus, novel experimental paradigms and methodologies had to be developed in the course of this project and are still being refined.

Experiment 6 ("**REAL WORLD VERSUS VR**", section 13) is designed to establish a new methodology including our rapid pointing paradigm. It compares spatial updating in a real environment and the corresponding virtual replica, thus evaluating our approach of using VR to disentangle the different sensory modalities. The usage of UPDATE, CONTROL, IGNORE, and IGNORE BACKMOTION trials in VR within one experiment allows the investigation of the relevance and interaction of visual and vestibular cues for automatic as well as obligatory spatial updating.

Experiment 7 ("**SIMULATION PARAMETERS**", section 14) uses an even more complex environment with more targets to explore the potential influence of various visuo-vestibular parameters on automatic and obligatory spatial updating. Visual display parameters like the FOV and projection screen versus HMD usage were investigated as well as visuo-vestibular motion parameters like the relation (gain factor) between visual and vestibular motion and the turning amplitude and velocity.

Experiment 8 ("**LANDMARKS VERSUS OPTIC FLOW**", section 15) is aimed at disentangling the contribution of visual landmark information from dynamic visual motion information. This is done by presenting optic flow information only during the motion and pointing phase for half of the trials. The influence of vestibular motion cues is additionally studied by comparing visual motions with and without concurrent physical motions. If all landmarks are removed and the visual motion information is reduced to a mere optic flow pattern, the visual dominance observed in Experiments 6 and 7 is expected to decline and vestibular turn cues are expected to have a stronger effect.

In part IV, section 17, the experimental results from part II and III will be revisited and discussed in the context of the theoretical framework (introduced in section 16) and the literature.

13 Experiment 6: "REAL WORLD VERSUS VR"

13.1 Introduction

Up to now it is rather unclear which sensory cues trigger the process of spatial updating. Under "normal" conditions, visual, auditory, haptic, vestibular and kinesthetic cues are in accordance and give sometimes complementary, but mostly redundant and consistent information about the ego-motion and current position. Within the last centuries and due to advances of technology, however, many "unnatural" situations emerged where the different sensory modalities are no longer in agreement. The earliest examples of this include boats and carriages: When passengers inside have no vision of the external world, visual cues indicate a static surround, whereas vestibular and kinesthetic cues convincingly indicate motion. These situations might have been the earliest incidence of motion-sickness induced by sensory cue conflicts. Later examples for sensory conflict situations include movies, television, and, most recently, Virtual Reality applications. It is commonly found that spatial orientation, and especially spatial updating, largely deteriorates when certain sensory modalities are excluded, reduced, or only insufficiently simulated (Chance et al., 1998; Bakker et al., 1999; May & Klatzky, 2000; Péruch & Gaunet, 1998; Sholl, 1989; Simons & Wang, 1998; Wang & Simons, 1999; Wraga et al., 2003). If information from different sensory modalities are in clear conflict, disorientation, unease, and motion sickness are often observed (Bles, Bos, de Graaf, Groen, & Wertheim, 1998; Chance et al., 1998; Cheung, Howard, & Money, 1991; Draper, Viirre, Furness, & Gawron, 2001; Guedry, Rupert, & Reschke, 1998; Kennedy, Lanham, Drexler, Massey, & Lilienthal, 1997; Stanney, Mourant, & Kennedy, 1998). Especially conflicts between the perceived and expected subjective vertical seem to be critical (Bles et al., 1998).

In our study, we investigated the integration of visual and vestibular cues, two sensory modalities that are essential and most likely sufficient for normal spatial updating. In order to be able to independently control vestibular and visual cues, we used a Virtual Reality setup including a motion simulator (6 degree of freedom (DOF) motion platform) and a head mounted display (HMD). To get a baseline performance of "optimal" spatial updating, we compared VR performance with real world performance in the corresponding real environment. To avoid simulator sickness, motions were selected to be smooth and of relatively low acceleration, and the visual and vestibular vertical was always kept in close alignment. Several goals were pursued in this experiment as listed below:

To establish a rapid pointing paradigm and test its applicability for quantifying spatial updating. If rapid pointing serves as a means of reliably quantifying spatial updating, performance should reveal the typical response pattern observed in the literature: Under full cue conditions, UPDATE performance should be almost as good as CONTROL performance, whereas IGNORE performance should be significantly impaired, reflecting the obligatory aspect of spatial updating.

To compare spatial updating performance in real and virtual environments. We used spatial updating performance in a real environment under full cue conditions as a baseline for optimal performance (block A & B), and successively reduced the amount of useful visual and vestibular information using a virtual replica of the real room (block C-F). If performance in VR is comparable to real world performance, this would validate our approach of using VR to present the stimuli, and would suggest the transferability of results obtained in this VR setup to comparable real world tasks. Potential differences and systematic errors in the VR tasks, conversely, would indicate specific problems in using current VR technology and ideally point to critical factors for the design of VR setups and relevant display and 3D model parameters.

To investigate the importance and interaction of visual and vestibular cues for automatic spatial updating.   VR technology allowed us to vary the amount of visual and vestibular cues independently, revealing their relative importance and interaction. Participants were furthermore asked to focus on one or the other sensory modality, which reveals the relative importance of one modality and the ignorability of the other one. If UPDATE performance is about as good as CONTROL performance, the available cues can be considered *sufficient* or *usable* for quick and accurate spatial updating, that is, they allow for *automatic* spatial updating. Performance differences, on the other hand, would indicate which relevant spatial cues were missing, pinpointing the essential cues for automatic spatial updating.

To investigate the obligatory (reflex-like) character of spatial updating versus the potential cognitive contribution, depending on the available cues.   Comparing IGNORE and UPDATE performance reveals the potential cognitive contribution to spatial updating under the given spatial cues. If the available spatial cues are more powerful in triggering spatial updating and hence turning the world inside our head (even against our conscious will), those turns should be harder to IGNORE. Spatial updating would then be "obligatory" or "reflex-like" in the sense of being consciously hard-to-suppress and consequently largely beyond conscious control. Thus, IGNORE tasks can be used to investigate the cognitive influence on the reflex-like process of spatial updating under various combinations of instructions, cues and sensory modalities. Compared to UPDATE trials, which quantify whether the available spatial information *can* be used for spatial updating (**automatic spatial updating**), IGNORE trials reveal whether that information *must* be used, i.e., whether the spatial cues can trigger spatial updating even against our own conscious decision (**obligatory spatial updating**).

13.2   Methods

13.2.1   Participants

Twelve naive participants (three male and nine female) completed the experiment, with ages ranging from 19 to 33 years (mean: 26.3 ± 0.4 years, SD: 4.8 years). As for all experiments presented in this paper, participants had normal or corrected-to-normal vision and no signs of vestibular dysfunction. Participation was always voluntary and paid at standard rates.

13.2.2   Stimuli and apparatus

13.2.2.1   Scenery and visualization   The pointing stimuli consisted of twelve target objects (the numbers from 1 to 12, arranged in a clock face manner) attached to the walls of the Motion-Lab at eye height (see Fig. 18 and 19). Participants saw either the real room or a photorealistic virtual replica of it (see Fig. 18) presented through a position-tracked head-mounted display (HMD Kaiser ProView XL50, see Fig. 21). The HMD had a resolution of 1024×768 pixel and subtended a physical field of view (FOV) of $40° \times 30°$. The VR-model was presented non-stereoscopically, as stereoscopic cues would be relatively weak for the distances used (2.5-7.5m) (Goldstein, 1996). Furthermore, other depth cues, especially perspective cues (linear perspective, foreshortening, and texture gradient) seem to be more important and more readily usable than cues from stereopsis, which are moreover known to cause visual stress if HMDs and presentation times of more than 10 minutes are used (Surdick, Davis, King, & Hodges, 1997; Mon-Williams & Wann, 1998).

13.2.2.2   Vestibular stimuli and apparatus   For vestibular stimulation, participants were seated on a six degree of freedom Stewart motion platform (Motionbase Maxcue, see Fig. 19). For the
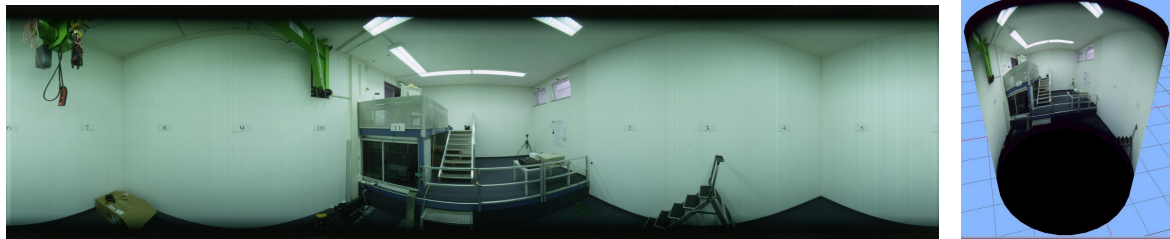
(a) 360° roundshot of the Motion Lab, taken from the standard viewing position of the participant seated on the platform. The scanned roundshot has a resolution of 4096 x 1024 pixel.

(b) Top view of the model

(c) Center view of the model

(d) 40 x 30° center view of the model, as seen by the participant

Figure 18: Building a photorealistic replica of a real room. Displayed is a photorealistic virtual replica of the Motion-Lab, created from a 360° roundshot of the real room (a) that was wrapped onto a cylinder (b). For the experiments, the virtual eye-point is centered in the cylinder (c) and the simulated field of view (sFOV) is set to equal to the physical field of view (FOV) of 40°×30° (d). This provides participants with a highly realistic and undistorted view onto the virtual scene as well as a high degree of immersion.

experiment, however, only rotations around the earth-vertical axis (yaw) were used, as these are the behaviorally most relevant rotations for spatial orientation on the earth's surface. Furthermore, translations seem to be rather easy to spatially update (even for imagined motions), and are hence less interesting for our purpose (see, e.g., Easton & Sholl (1995), May & Klatzky (2000), May (1996), Presson & Montello (1994), Rieser (1989) and subsection 12.2.4).

13.2.2.3 Vibrations Additional vibrations were applied using three special force transducers (shakers), two mounted below the participant's seat, and one below the foot plate (see Fig. 20 (a)). Broad-frequency vibrations were applied during all physical motions in order to yield a more compelling feeling of ego-motion and to mask motion-specific micro-vibration induced by the step motors moving the platform's legs.
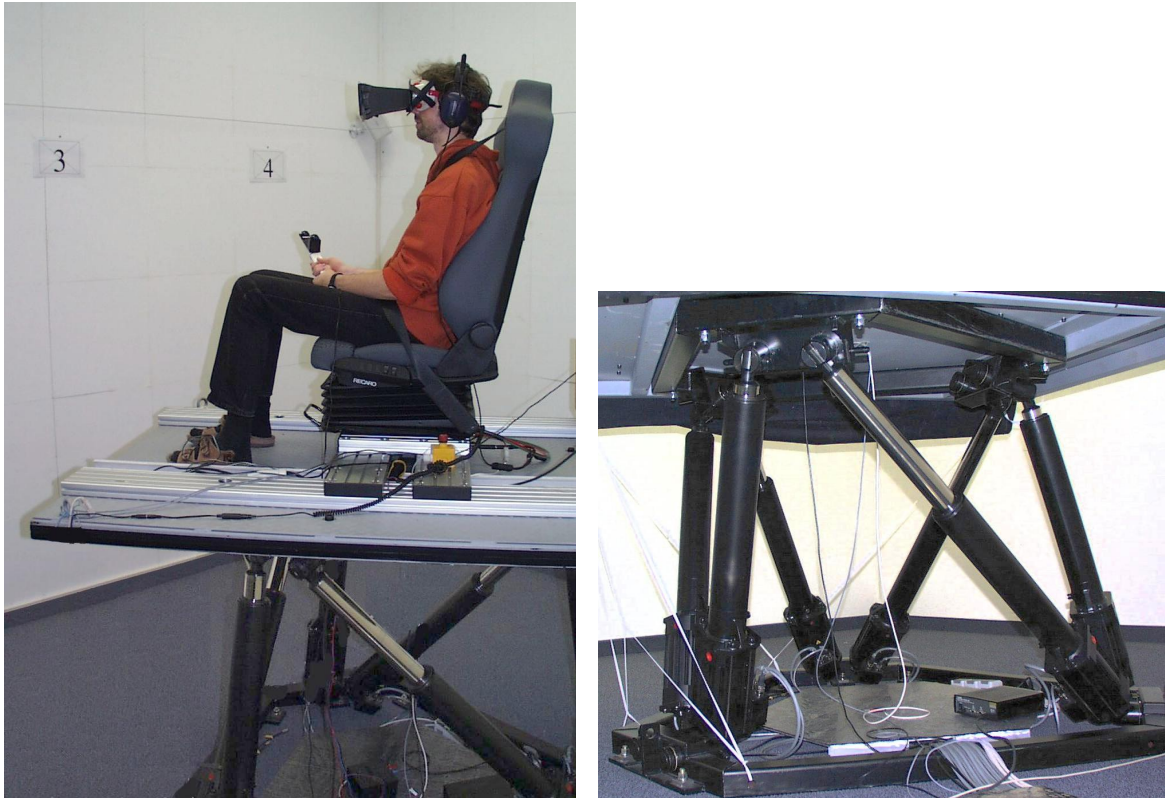
(a) Participant wearing headphones and purpose-designed blinders (vision delimiting cardboard goggles) reducing the FOV to that of the HMD ($40°$x$30°$). The participant is currently pointing towards target '4' using the position-tracked pointer. Note the targets on the wall.

(b) Six degree of freedom motion platform used for vestibular stimulation. The individual legs are electrically driven, which allows for finer control and smoother motions than pneumatic or hydraulic setups.

Figure 19: Experimental setup displaying a participant seated on the motion platform.
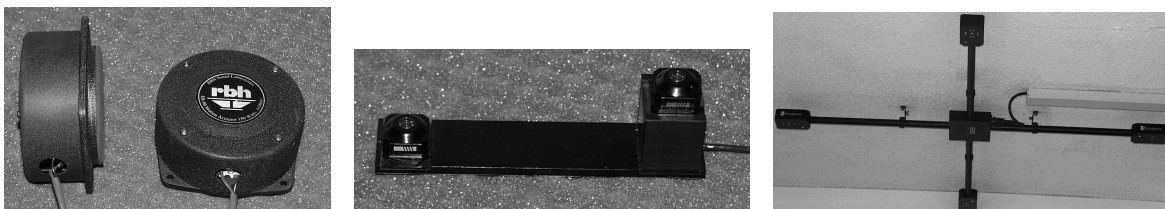
(a) Tactile transducers FX-80

(b) Tracking unit, consisting of two ultrasonic beacons and one inertial cube

(c) Tracker cross-bar with four ultrasonic receivers

Figure 20: Vibration and position tracking setup. (a) Force transducers (RHB Virtual Theater 2 shakers) are used for vibrating the participant's seat and foot plate with frequencies ranging from 10 - 150Hz. (b) & (c) The six degree of freedom position tracker (Intersense IS600-mk2) is used for tracking the position of the HMD and pointing wand.

13.2.2.4 Auditory stimuli Instructions during the experiment were given by a computer-generated voice and were presented via special aviation headphones (Sennheiser HMEC 300, see Fig. 21). These headphones are equipped with active noise cancellation reducing the noise level by more than 25dB. As this was not sufficient to completely eliminate spatial auditory cues from the Motion-Lab, additional broad-band noise was continuously presented via the headphones at a low level. As long-term exposure to white or pink noise can be rather disturbing and reduces the participants' motivation, a special sound file was generated by mixing and equalizing several river sounds. This sound file was never reported as being disturbing and seemed to even increase immersion. This procedure effectively eliminated all spatial auditory cues of the surround without adding unnecessary discomfort.

To mask all auditory motion-specific cues induced by physical platform motions that could have been used by the participants to auditorily estimate the angle turned, additional platform-masking sound was displayed during all simulated motions. The sound consisted of about 20 sound files from different recorded platform motions that were overlaid (mixed) and equalized to yield a broad-band noise masking all auditory cues about the platform motion effectively without having to play it at a loud level.

13.2.2.5 Position tracking The tracking of the pointer and the participant's head was done using a six degree of freedom position tracking system (IS600-mk2 from Intersense, see Fig. 20). The system combines ultrasonic time-of-flight measurements with inertial sensors.

13.2.2.6 Distributed Virtual Reality environment All experiments described in this part of this thesis were performed in the Motion-Lab of the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. A general description of the Motion-Lab and the hard- and software used can be found in von der Heyde (2000, 2001) or online at `www.kyb.tuebingen.mpg.de/bu/projects.html?prj=48`.

13.2.3 Interaction (Pointing)

After each rotation around the earth-vertical axis, the participants' task was to point "as accurately and quickly as possible" to four targets announced consecutively via headphones. Participants were instructed to keep their head still and facing forwards by leaning it against the head rest. The pointing targets were randomly selected to be outside of the FOV of the HMD or the cardboard blinders and within a comfortable pointing range ($|\alpha_{pointer} - \alpha_{straight-ahead}| \in [20°, 99°]$).

Pointing was performed using a purpose-built, six degree of freedom position tracked pointing wand (see Fig. 21). The pointing direction was recorded once the pointer was stabilized in space and had a pitch of less than 70 degrees. This was indicated to the participants by auditory feedback presented via headphones. After each pointing, participants raised the pointer to an upright position (see Fig. 21 (a)), indicating to the computer that the experiment can go on. This upright default position ensured that there was no directional bias and participants had similar pointing response times for all directions, a problem which is often not accounted for in studies using compass-like pointers (e.g., Wraga et al., 2003).

13.2.4 General procedure

A repeated-measures, within-subject design was used, which is summarized in Tables 8 and 10. After a two-stage training phase (see subsection 13.2.8), each participant completed a test phase consisting

(a) Default upright pointer position, indicating that the experiment can go on

(b) Pointer in pointing position

Figure 21: Participant holding a position-tracked pointer and wearing a position-tracked head-mounted display (HMD) and active noise cancellation headphones.

of six blocks of different cue combinations (see subsection 13.2.6 and Table 8), split into two sessions. The first session of three blocks was carried out directly after training, the second session with the remaining three blocks was on a different day, to avoid fatigue effects and obviate the influence of declining alertness. (Even though spatial updating itself might have been automatized under all stimulus conditions and should consequently not require much attention, the repeated rapid pointing tasks and especially the IGNORE conditions seemed to challenge participants quite a bit and require them to be fully alert and concentrating.) To pseudo-balance the blocks among the participants (a full balancing would have required 6! = 720 participants, which was of course unfeasible), they were balanced within the first and second session. More precisely, half of the twelve participants performed blocks B, C, and D in balanced order in the first session, and blocks A, E, and F in the second session. This order was reversed for the other half of the participants.

Each of the six blocks consisted of a total of 30 trials lasting approximately 13 minutes. For each block, the 30 trials were split up into 12 UPDATE trials and six trials each for the CONTROL, IG-NORE, and IGNORE BACKMOTION conditions in pseudo-randomized order (see subsection 13.2.7 and Table 10 for a detailed description). Each trial consisted of the following three parts:

1. **Auditory announcement** indicating whether the upcoming spatial updating condition was an IGNORE trial, an IGNORE BACKMOTION trial, or a "normal" trial (UPDATE and CONTROL trial, see subsection 13.2.7 below for a detailed description);

2. **Motion phase**, which lasted always seven seconds and started as soon as the pointer was in the default (upright) position. The velocity profile was Gaussian, with a peak velocity of twice the mean velocity (see Table 10);

3. **Pointing phase**, consisting of four repetitions of

   (a) auditory target announcement (e.g., "Object 9")

(b) subsequent pointing

(c) raising pointer to upright (default) position

Each consecutive part only started if the pointer was in the upright default position, indicating that the participant was ready and concentrating.

13.2.5 Dependent variables

The pointing data ware analyzed in terms of five dependent variables, revealing different aspects of spatial updating (see below). As pointing data is inherently directional (circular) data, we used circular statistics for computing the dependent variables (see, e.g., Batschelet (1981) for an introduction). Most importantly, in circular statistics angles are not averaged linearly but vectorally, which removes periodicity problems. For example, the arithmetic mean of -179° and +181° degrees is +1°, which does not make sense for directional (circular) data. In circular statistics, on the other hand, angles are inherently represented as directions (unit vectors). Hence, the circular mean of -179° and +181° is the direction of +180°, and is calculated as the direction of the vector mean of two unit vectors pointing in the direction of -179° and +181°. Thus, the circular mean can be thought of as the direction of the center-of-mass. The mean angular deviation, which is the circular statistics analog to the linear standard deviation, is computed from the length of the above-mentioned mean vector, and approaches its linear counterpart asymptotically for small variances (Batschelet, 1981, p. 34 ff.).

1. **Response time**: How easy and intuitive (fast) is the access to our spatial knowledge?
   The response time is calculated from the end of the target pronunciation until the end of the pointing movement. Participants showed consistent differences in their mean response time even for the baseline (CONTROL) condition, ranging, for example, from 0.36s to 1.38s in the CONTROL condition for block A (Real World full FOV). Those subject-specific response time differences were corrected for by computing the *relative* response time $t^{rel}_{n,m,b,s}$, which we define as the response time $t_{n,m,b,s}$ for participant n and trial m in block b and spatial updating condition s, scaled by the mean response time for all participants in the CONTROL condition of that block $\overline{t_{b,Control}} := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}(t_{i,j,b,Control})$, divided by the participant's mean response time in the CONTROL condition of that block $\overline{t_{n,b,Control}} := \frac{1}{M}\sum_{j=1}^{M}(t_{n,j,b,Control})$:

$$t^{rel}_{n,m,b,s} := t_{n,m,b,s} \cdot \frac{\overline{t_{b,Control}}}{\overline{t_{n,b,Control}}} = t_{n,m,b,s} \cdot \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}(t_{i,j,b,Control})}{\frac{1}{M}\sum_{j=1}^{M}(t_{n,j,b,Control})}$$

   , where $t^{rel}_{n,m,b,s}$ is the relative response time for participant n of N=12, trial number m out of M in that condition, block b out of the B=6 blocks and spatial updating condition s out of the S=4 spatial updating conditions. Using this procedure, the mean response time per condition, averaged over all participants, remains the same, but the between-subject response time differences are effectively removed.

2. **Configuration error = Pointing variability**: How consistent is our spatial knowledge of the target configuration? That is, are the angles between landmarks reported consistently?
   The pointing variability is calculated as the mean angular deviation of the signed error, taken over the 4 pointings, and is a measure of the configuration error or the inconsistency when pointing to several targets. The mean angular deviation is the circular statistics analog to the linear standard deviation (Batschelet, 1981, chap. 2.3).

3. **Absolute pointing error**: How accurately do we know where we are with respect to our surround or specific objects of interest?

4. **Absolute ego-orientation error per trial**: Did participants misperceive their ego-orientation? Parts of the absolute pointing error might be confounded with a general misperception of the perceived ego-orientation and might be explained by the latter. For example, if a participant somehow misperceives her orientation by 10°, this might already explain up to 10° of her absolute pointing error. The perceived ego-orientation per trial is estimated by taking the circular mean of the four signed pointing errors per trial (Batschelet, 1981, chap. 1.3).

5. **Ego-orientation error in turning direction**: Did participants misperceive their ego-orientation typically in the direction of motion?

   If they would, that might be explained by some kind of "representational momentum", which describes the systematic tendency for observers to remember an event as extending beyond its actual ending point (Freyd & Finke, 1984; Hubbard & Bharucha, 1988). See also Kozhevnikov & Hegarty (2001), Thornton & Hubbard (2002) for an overview on representational momentum and related findings. Hence, if the moving stimulus (visual and/or vestibular) induces some kind of representational momentum, this might lead to a motion capture in the direction of motion. That is, participants would misperceive their final ego-orientation after a turn in the direction of the motion. One fundamental difference between representational momentum studies and the current study is that the former investigate motion extrapolation for *individual objects* or *object configurations*, whereas the current experiment investigates (real or simulated) motions of *oneself*, not the surround.

   For the UPDATE or CONTROL conditions, we might expect an effect for block F with only vestibular motion cues. We do not, however, expect any representational momentum effect in blocks A-D, which included an abundance of salient visual landmarks. The IGNORE condition, on the other hand, might show an effect for stimuli that are powerful in inducing obligatory spatial updating and hence might also elicit a representational momentum. That is, if participants were not able to fully compensate for the to-be-ignored motion, one would expect a systematic ego-orientation error *in* turning direction. Conversely, if participants were somehow overcompensating, one might expect an ego-orientation error *against* turning direction. Regardless the direction of the effect, there should not be any systematic ego-orientation error in or against turning direction if the presented stimuli are easy to ignore or to compensate for.

13.2.6   Cue combinations (blocks)

In the test phase, each participant was presented with six stimulus conditions (blocks A-F, 15 min. each) in pseudo-balanced order, with different degrees of visual and vestibular information available (see Table 8 for a comparison). Blocks A and B used the real environment under full cues conditions as a baseline for optimal performance. Blocks C-F are the four sensible combinations of useful visual cues (yes/no), useful vestibular cues (yes/no), and resulting visuo-vestibular cue conflict (yes/no), as is indicated in Table 9. Block D was the only one where participants were not turned physically. In blocks A-C, the amplitudes of the visual and vestibular (physical) turns were equal.

Block A: "Real World full FOV"   Participants saw the real Motion-Lab with unrestricted vision. As in all blocks, however, they were not allowed to move their head. This was a baseline condition for optimal spatial updating performance under full cue conditions. Note that auditory spatial cues were excluded at all times.

Block B: "Real World w/ blinders"   Participants saw the real Motion-Lab as in block A, but wore blinders (see Fig. 19 (a)) that restricted their FOV to 40°×30°, in order to match the FOV of the

cue combinations (block)	field of view (FOV)	useful visual cues	useful vestibular cues	cue conflict
Block A: "Real World full FOV"	unrestricted	yes	yes	no
Block B: "Real World w/ blinders"	40° x 30°	yes	yes	no
Block C: "HMD vis. + vest. cues"	40° x 30°	yes	yes	no
Block D: "HMD just vis. cues"	40° x 30°	yes	no	yes
Block E: "HMD constVis. + vest. cues"	40° x 30°	no	yes	yes
Block F: "Blindfolded just vest. cues"	—	no	yes	no

Table 8: Summary of the six different cue combinations (blocks) used in experiment REAL WORLD VERSUS VR.

all eight possible combinations of useful visual cues, vestibular cues, and cue conflict	useful visual cues	useful vestibular cues	cue conflict
nonsensical cue conflict	yes	yes	yes
Block A, B, & C: "vis. + vest. cues"	yes	yes	no
Block D: "HMD just vis. cues"	yes	no	yes
only possible for vestibular loss patients	yes	no	no
Block E: "HMD constVis. + vest. cues"	no	yes	yes
Block F: "Blindfolded just vest. cues"	no	yes	no
no stimulus at all, nonsensical cue conflict	no	no	yes
no stimulus at all	no	no	no

Table 9: Summary of the eight possible logical combinations of useful visual cues (yes/no), useful vestibular cues (yes/no), and resulting visuo-vestibular cue conflict (yes/no). Blocks C-F are the only four sensible combinations. A further block with useful visual cues, but no useful vestibular cues and no resulting cue conflict would only be feasible when using patients with complete vestibular loss. Due to the lack of a sufficient number of available participants with vestibular loss, we did not use this combination, which would nevertheless have been an interesting addendum.

HMD in blocks C-E. Any performance difference between block A and B can thus be ascribed to the reduced peripheral vision (and maybe also the thereby reduced visibility of the pointer).

Block C: "HMD vis. + vest. cues"   Participants wore the head-tracked HMD (see Fig. 21) and saw a virtual copy of the Motion-Lab (see Fig. 18). Block C was designed to be a Virtual Reality replica of block B, to test the influence of using a VR display instead of seeing the real surround. Potential performance deteriorations from block B to block C would indicate limitations of using HMDs as display devices for spatial updating studies. Similar performance, on the other hand, would validate our approach of using VR technology for investigating spatial updating and suggest the transferability of results obtained in similar virtual environments to the real world.

Block D: "HMD just vis. cues"   This was the only block where the participants were *not* turned *physically*, and asked to just use visual information. Apart from the lack of concomitant vestibular information about the turns, this block was identical to block C. This condition is similar to many VR applications, where participants just have visual information about their ego-motion, but no physical motion. Potential differences between block C and D would reveal the relevance and/or importance of vestibular cues for spatial updating.

Block E: "HMD constVis. + vest. cues"   Block E was designed to be the "inverse" of block D: Here, participants were turned physically, but *not visually*, and asked to just refer to vestibular information about the physical turn. More specifically, the HMD was still head-tracked, but the platform motion was subtracted from the current viewing direction, effectively always showing the same view onto the virtual room, as if facing 12 o'clock all the time, irrespective of the platform orientation. Participants were instructed to keep their eyes open at all times even though they were essentially asked to ignore all visually presented cues. As the first trial was always an UPDATE trial which turned participants away from 12 o'clock, the visible (12 o'clock) orientation was wrong for most of the subsequent trials, even for the CONTROL trials.

Block D and E presented participants with cue conflict conditions, where either the vestibular or visual cue, respectively, indicated "no motion", but was to be ignored. Performance differences between block D and E would point towards differences in the way visual and vestibular information is used to accomplish the spatial updating task.

Block F: "Blindfolded just vest. cues"   Participants were blindfolded and had only vestibular information about the turns. Due to the lack of any absolute reference points (landmarks) indicating the current orientation, participants in block E and F were forced to use vestibular path integration to update their current heading over the course of the whole block. Hence, we expected participants to slowly but surely lose track of the proper current ego-orientation due to the accumulation error associated with path integration. Compared to block E with quasi-static visual cues, there was no longer any cue conflict in block F due to the complete lack of any visual cues.

13.2.7   Spatial updating conditions

Four of the five stereotypical spatial updating conditions described in subsection 12.2.3 were used in each block of this experiment. For convenience, these four spatial updating conditions are summarized below. The 30 trials of each block were split up into 12 UPDATE trials and six trials each for of the CONTROL, IGNORE, and IGNORE BACKMOTION conditions in pseudo-randomized order (see Table 10 for a detailed description).

spatial updating condition	turning angles	peak angular velocity	repetitions per block	trials per block
1. UPDATE	±19°, ±38°, ±57°	5.4, 10.9, 16.3°/s	2	12
2. CONTROL	±9.5°, ±19°, ±28.5°	5.4, 10.9, 16.3°/s	1	6
3. IGNORE	±19°, ±38°, ±57°	5.4, 10.9, 16.3°/s	1	6
4. IGNORE BACKMOTION	±19°, ±38°, ±57°	5.4, 10.9, 16.3°/s	1	6

Table 10: Summary of the four different spatial updating conditions. Due to limitations of the platform turning range, the maximum heading deviation from straight ahead (12 o'clock) was ± 57°. The movement time was always set to seven seconds.

1. **UPDATE**: From the current orientation, participants are simply rotated to a different orientation. From there, they have to point consecutively to four targets announced via headphones. If the available cues are sufficient for enabling automatic spatial updating, UPDATE performance should not depend on the angle turned.

2. **CONTROL**: Participants are rotated to a new orientation and immediately back to the original one before being asked to point. This is a baseline condition yielding optimal performance:

If the available spatial updating cues are sufficient, UPDATE performance should be about as good as CONTROL performance (**"automatic spatial updating"**).

3. **IGNORE**: Participants are again rotated to a different orientation, but asked beforehand to *ignore* that motion and "respond as if you had not moved", i.e., participants are asked to imagine that they are still facing the previous direction. If the available spatial cues are more powerful in triggering spatial updating and hence turn the world inside our head (even against our conscious will), those turns should be harder to IGNORE. Spatial updating would then be "obligatory" or "reflex-like" in the sense of largely beyond conscious control and consciously hard-to-suppress (**"obligatory spatial updating"**).

4. **IGNORE BACKMOTION:** After each IGNORE trial, participants are rotated back to the previous orientation. The main purpose of this condition is to avoid potential disorientation that might have been induced by the previous IGNORE trial and to re-anchor participants to the previous orientation by asking them to point and thus probe their mental spatial reference frame. Comparable performance in the IGNORE BACKMOTION and UPDATE condition would suggest that participants were properly re-anchored to that orientation and no longer disoriented by the IGNORE trial beforehand.

13.2.8 Training phase

After reading the instructions and being given a demonstration of the experimental procedure by the experimenter, participants performed a two-phase training session. The goal of this training session was to answer possible questions and familiarize participants with the pointing task, the experimental procedure, and the four different spatial updating conditions (see subsection 13.2.7 above).

In the first training phase (24 trials lasting approximately 20 minutes), participants were seated on the motion platform, saw the real Motion-Lab, and were turned consecutively to different headings (similar to block A). Just as in the test phase, participants were asked after each motion to point "as accurately and quickly as possible" to four targets announced consecutively via headphones. To train participants on pointing accuracy, additional feedback about the pointing direction was given by a laser pointer attached to the pointing wand. This feedback was only available during the first training phase.

In the second training phase (12 trials lasting approximately 10 minutes), participants wore the HMD and had no additional feedback about the pointing direction (similar to block C). The goal of this session was mainly to further familiarize participants with the Virtual Reality setup and the experimental procedure.

13.2.9 Data analysis

13.2.9.1 Initial performance onset correction   Pre-experiments had shown that participants often show a decreased performance for the first trial of each block. To reliably remove this initial performance onset phase for all participants, an additional UPDATE trial was included before the 30 other trials for each block and later removed before any further data analysis.

13.2.9.2 Response time limitation   To motivate participants to point as quickly as possible, pointing response time was limited to four seconds. For response times larger than that, that pointing record was removed. Using this procedure, a total of 16 pointings or 0.19% of all pointings were excluded from further analysis.

13.2.9.3   Tracker offset correction   Due to occasional technical problems with the position tracker, the pointing records for a few blocks showed a considerable constant yaw offset, that is, a drift towards the left or right direction. To reliably eliminate those technical artifacts, all pointing records for each block were corrected for that offset by subtracting the mean yaw offset from all CONTROL conditions of that block. For blocks E and F, this procedure might also have corrected for drifts in participants perceived ego-orientation, which, unfortunately, could not be avoided.

13.3   Results and discussion

To get a first impression of the results, the data for block A ("Real World full FOV") are plotted in Figure 22 for the five dependent variables. Figures 22 (a) - (d) clearly show the typical response pattern for spatial updating: UPDATE performance is comparable to CONTROL performance, whereas IGNORE performance is considerably worse. To be more exact, UPDATE trials were approximately 50ms slower than CONTROL trials, whereas the other four measurands showed no difference. IGNORE BACKMOTION performance was as good as UPDATE performance in all five dependent variables, indicating that participants were properly re-anchored to the surround and no longer disoriented by the IGNORE trial beforehand. In sum, we found for the full cue condition the typical response pattern known from the spatial updating literature. Hence, our method of using computer-tracked rapid pointing as a means of quantifying spatial updating proved quite successful so far.

For a detailed analysis, we will first present the data for all six cue combinations (blocks) for the baseline (CONTROL) condition in subsection 13.3.1. This provides a baseline performance for each block, and already pinpoints critical differences between the different cue combination in the easiest (CONTROL) task. In subsection 13.3.2, the sufficiency of the different cue combination for automatic spatial updating will be investigated by comparing UPDATE to CONTROL performance: If UPDATE is almost as easy as CONTROL, this would indicate that the available cues are sufficient for enabling *automatic* spatial updating, that is, the cues *can* be used. Next, the cognitive penetrability of spatial updating given the six different cue combinations will be examined in subsection 13.3.3 by comparing IGNORE and UPDATE performance. *Obligatory* spatial updating would occur if and only if the available cues cannot be ignored and *must* be used, i.e., if ignoring is considerably harder than updating. We will conclude by analyzing potential learning effects, turning angle effects, and pointing order effects in subsection 13.3.4. The results of the statistical analyses are compiled in Table 11. For reference, the full data set of this experiment is displayed in Figures 58 and 59, compiled per cue combination (block) and spatial updating condition, respectively.

13.3.1   Baseline (CONTROL) performance

The forth-and-back motion of the CONTROL condition is simple enough that spatial updating of the motion is more or less trivial. Hence, the observed response should rather reflect the pointing performance for the given *static* spatial cues (e.g., visual display), without too much influence from the motion cues, as they just need to indicate *that* it was a forth-and-back motion, without the need to know the angle turned or anything else. No matter how good spatial updating works, CONTROL trial should still reveal decent performance: If the motion cues were sufficient for enabling spatial updating, the forth-and-back motion should be easy to update. If, on the other hand, the motion cues were completely insufficient, spatial updating would not occur and participants should have no problem either pointing as if they were still at the same orientation. That is, potential performance differences between the different cue combinations (blocks) should indicate differences in the usability of the currently available *static* spatial information without too much influence from the dynamic motion

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
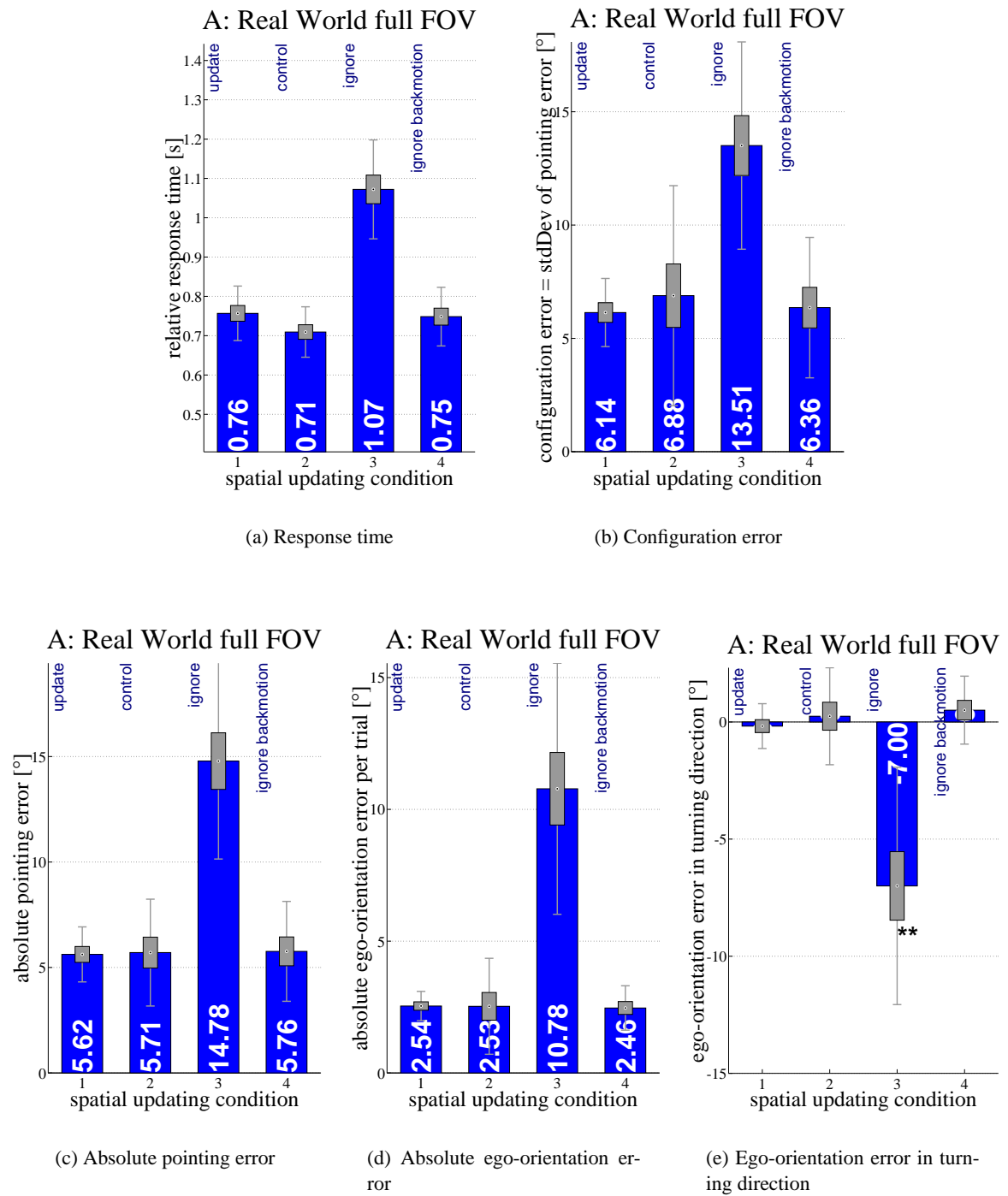
Figure 22: Pointing performance in experiment REAL WORLD VERSUS VR showing the typical response pattern for spatial updating: UPDATE performance is comparable to CONTROL performance, whereas IGNORE performance is considerably worse. Performance in block A (Real World full FOV) is plotted for the five dependent variables, each for the four different spatial updating conditions. The bars represent the arithmetic mean, which is also numerically indicated by the white numbers at the bottom of each bar. Boxes and whiskers denote one standard error of the mean and one standard deviation, respectively. The asterisks '*' in plot (e) indicate whether the mean differs significantly from zero (on a 5%, 0.5% or 0.05% significance level, using a two-tailed t-test).

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
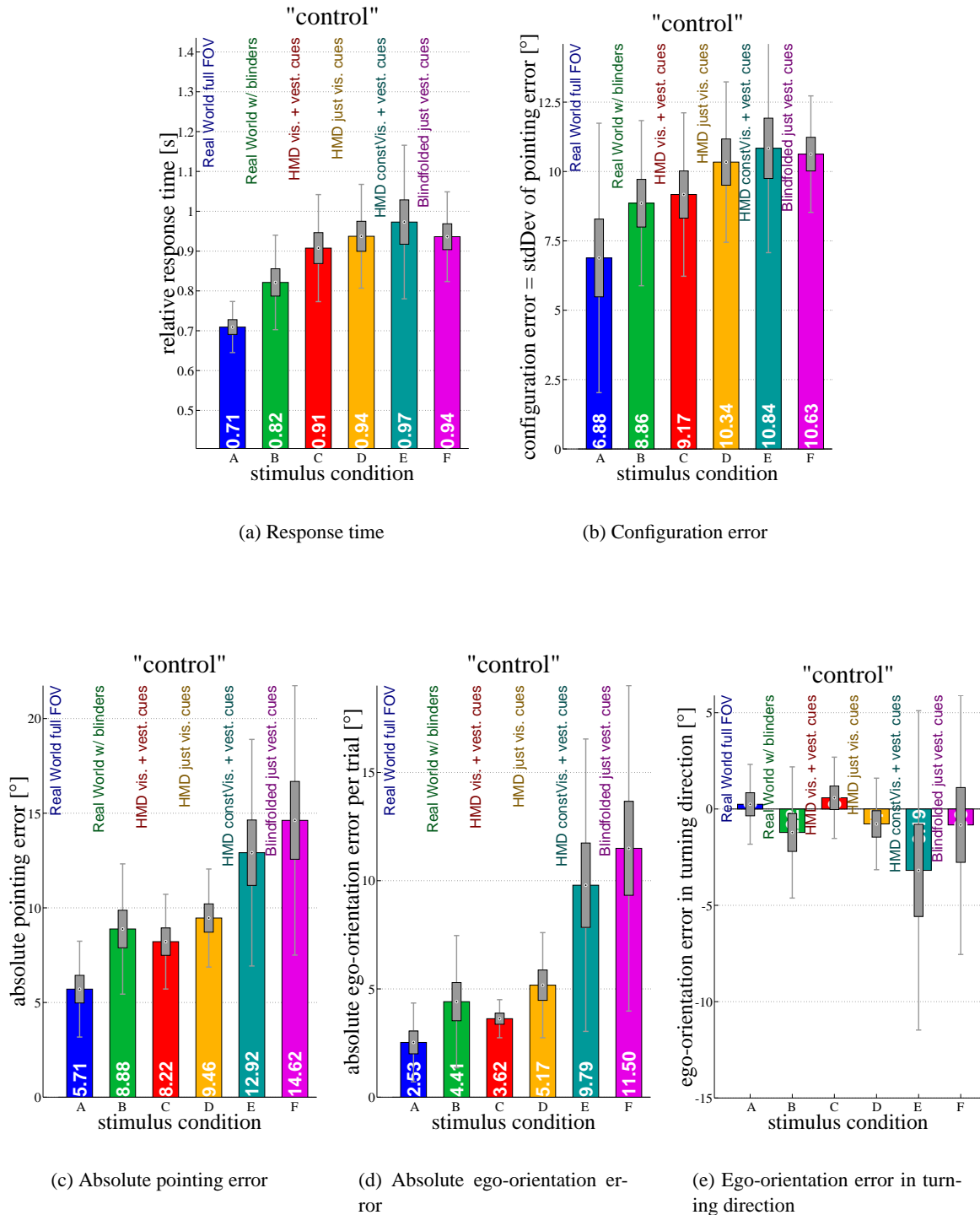
Figure 23: Baseline spatial updating performance for experiment REAL WORLD VERSUS VR. Baseline (CONTROL) performance is plotted for the five dependent variables, each for the six different cue combinations. Note the FOV effect even for the simple baseline task (block A vs. B).

cues. The influence of the dynamic motion cues will be revealed in subsection 13.3.2, where the non-trivial UPDATE condition is compared to the pseudo-static CONTROL task.

The CONTROL data are summarized in Figure 23, the corresponding t-tests are compiled in Table 11. Different questions guided the choice of cue combinations and will be discussed in the following subsections by comparing CONTROL performance between adjacent blocks.

13.3.1.1   Influence of FOV (block A vs. B)   Comparing real world performance with unrestricted vision (block A) versus constrained FOV (block B, see Figure 23) reveals a clear performance decrease for limiting the FOV to $40° \times 30°$. Conversely, participants somehow benefitted from an unrestricted FOV by showing a shorter response time and smaller configuration error, absolute pointing error as well as absolute ego-orientation error even in the rather simple baseline task. The difference in absolute pointing error (of approximately $3°$) might be largely explained be the difference in absolute ego-orientation error (approximately $2°$). That is, participants were considerably worse in judging their ego-orientation when vision was delimited by blinders. As there were more than enough salient landmarks to judge the current ego-location even with a limited FOV, we suggest that participants experienced problems in judging their current head-orientation perfectly when peripheral vision of their body and the surround was eliminated. This hypothesis is corroborated by the observation of participants' head orientation in the reduced FOV conditions: Even though they were instructed to always keep their head facing forwards, they often tended to have a slightly off-center head orientation, which did not seem to happen in the full FOV condition. Potential concurrent effects of the visibility of the pointer, however, cannot be excluded from the current data and await further experiments.

13.3.1.2   Real world versus Virtual Reality performance (block B vs. C)   Participants in block B saw the real world through a restricted FOV, whereas they saw in block C the same view presented through a HMD with the same FOV as the blinders. Figure 23 (a) reveals a small but insignificant response time increase of approximately 90ms for using the HMD in block C (cf. Table 11). This suggests that the HMD condition might be perceived as slightly harder than the real world condition. Some of the response time difference, however, might also be caused by small visualization delays in the HMD condition. All other measures showed essentially the same performance and did not differ significantly, indicating that information displayed via HMD allows for the same spatial accuracy and ego-orientation perception.

13.3.1.3   Influence of vestibular turn cues (block C vs. D)   Omitting all vestibular turn information and just displaying visual turn cues in block D did not significantly reduce performance, compared to block C *with* vestibular turn cues (see Table 11). Hence, vestibular cues seem to play only a minor role for the simple CONTROL trials.

13.3.1.4   Influence of (missing) useful visual cues (blocks C-F)   Providing *only* vestibular turn cues while having to ignore the quasi-static visual cues in block E increased response time, configuration error, and ego-orientation error in turning direction only slightly and insignificantly (see Table 11). The absolute pointing error and absolute ego-orientation error, however, were considerably increased, indicating that participants tended to lose track of their correct ego-orientation without useful visual cues. This effect was slightly but insignificantly more pronounced for block F where participants were blindfolded. The lack of useful reference points in conditions E and F can explain the increased absolute ego-orientation error, as participants were constrained to using path integration, and hence lost track of their correct ego-orientation after several consecutive turns. For

larger overall turning angles, and without the tracker offset correction used (see subsection 13.2.9.3), these ego-orientation errors would most likely be considerably larger.

13.3.1.5 Summary and conclusions Taken together, the results of the CONTROL condition demonstrated the importance of a large FOV and of useful reference points for quick and accurate knowledge of where the surrounding target objects were. Removing visual landmark information in block E and F reduced the available cues to path integration by vestibular cues, which led as expected to considerable misjudgments of the proper self-orientation. Maybe most critical for the further analysis and experiments, VR performance was - apart from a slightly increased response time - as good as real world performance, provided that the FOV was matched. This validates our approach of using VR technology for studying spatial tasks, and suggests the transferability to comparable real world situations.

13.3.2 Automatic spatial updating

In this subsection, automatic spatial updating will be investigated by analyzing the difference between UPDATE and CONTROL performance for the different cue combinations. Subtracting CONTROL performance from UPDATE performance is an attempt to separate *dynamic* effects (i.e., UPDATE effects due to spatial updating) from baseline (CONTROL) differences most likely due to differences in the *statically* available information. In this manner, we compare spatial updating to different orientations to the supposed-to-be trivial updating forth-and-back to the same orientation. The previous subsection revealed that accurate forth-and-back updating is already rather non-trivial when only vestibular turn information is available, or when to-be-ignored visual stimuli are present.

The literature on blindfolded spatial updating suggests a slight response time increase of approximately 100ms for UPDATE trials (e.g., Farrell & Robertson, 1998; May, 2000), and a considerable increase in pointing error (e.g., from 15° to 24° in the study by Farrell & Robertson (1998)). Such a pointing error increase might be explained by path integration errors, which should be compensated for by the useful landmarks in the visual conditions (A-D) of this experiment. Hence, we do not expect any major pointing error increase in those conditions.

13.3.2.1 Conditions with useful visual information (blocks A-D) For all blocks with useful visual landmarks (A-D), response times in the UPDATE trials were consistently increased by approximately 50ms, compared to CONTROL performance (see Figure 24 (a)). This difference was significant for the two real world conditions (blocks A & B, $p < 0.05$), but only marginally significant for the two HMD conditions due to the increased between-subject variability (blocks C & D, $p < 0.1$). This is less than the 100ms expected from the literature, indicating that updating to new orientations is almost as easy as updating forth-and-back to the same orientation. This in turn suggests that spatial updating using landmark-rich visual cues is not trivial, but still quite easy and intuitive. The response time increase of 50ms is lower than the value typically found in the literature for nonvisual spatial updating, suggesting that the uncertainty of nonvisual path integration might have contributed to the increased response time there.

The differences in terms of configuration error, absolute pointing error as well as absolute ego-orientation error were all less than 1°, indicating that visually-assisted updating to new orientations was virtually as accurate as baseline performance. The differences between UPDATE and CONTROL performance for configuration error in block B and absolute ego-orientation error in block C were small but significant, and cannot be convincingly explained by the current data. The ego-orientation error in turning direction was negligible for blocks A-C with vestibular turn cues, but approximately

		response time		configuration error		absolute pointing error		absolute ego-orientation error		ego-orientat. error in turn direction	
		t(11)	p	t(11)	p	t(11)	p	t(11)	p	t(11)	p
CONTROL											
influence of FOV	A vs. B	-2.65	0.022*	-1.35	0.2	-3.46	0.0054*	-2.13	0.057m	1.27	0.23
real world vs. VR	B vs. C	-1.56	0.15	-0.62	0.55	0.802	0.44	0.971	0.35	-1.55	0.15
influence of vest. turn cues	C vs. D	-0.481	0.64	-2.1	0.06m	-2.18	0.052m	-2.13	0.056m	1.3	0.22
const. vest. vs. const. vis. cues	D vs. E	-0.427	0.68	-0.323	0.75	-1.79	0.1	-2.12	0.057m	0.953	0.36
const. vis. vs. vs. no vis. cues	E vs. F	0.49	0.63	0.159	0.88	-0.679	0.51	-0.704	0.5	-0.638	0.54
vis. turn vs. const. vis. cues	C vs. E	-0.9	0.39	-1.11	0.29	-2.24	0.047*	-3.05	0.011*	1.48	0.17
vis. turn vs. no vis. cues	C vs. F	-0.551	0.59	-1.41	0.19	-3.15	0.0092*	-3.52	0.0048**	0.645	0.53
just vis. vs. just vest. turn cues	D vs. F	0.0344	0.97	-0.241	0.81	-2.53	0.028*	-2.84	0.016*	0.0272	0.98
UPDATE-CONTROL											
influence of FOV	A vs. B	-0.549	0.59	0.0658	0.95	0.927	0.37	-0.569	0.58	-1.16	0.27
real world vs. VR	B vs. C	0.263	0.8	-1.63	0.13	-2.54	0.027*	-0.925	0.37	0.999	0.34
influence of vest. turn cues	C vs. D	-0.0606	0.95	0.926	0.37	0.745	0.47	0.915	0.38	-1.82	0.096m
const. vest. vs. const. vis. cues	D vs. E	-1.6	0.14	-3.03	0.011*	-3.41	0.0058*	-1.98	0.073m	0.285	0.78
const. vis. vs. vs. no vis. cues	E vs. F	0.957	0.36	0.787	0.45	0.433	0.67	0.0564	0.96	0.928	0.37
vis. turn vs. const. vis. cues	C vs. E	-1.98	0.073m	-1.86	0.089m	-2.31	0.041*	-1.55	0.15	-0.905	0.38
vis. turn vs. no vis. cues	C vs. F	-1.85	0.091m	-1.04	0.32	-1.27	0.23	-1.24	0.24	0.467	0.65
just vis. vs. just vest. turn cues	D vs. F	-1.65	0.13	-1.8	0.099m	-1.35	0.2	-1.31	0.22	1.59	0.14
IGNORE-UPDATE											
influence of FOV	A vs. B	-0.256	0.8	0.946	0.36	0.532	0.61	0.479	0.64	-0.387	0.71
real world vs. VR	B vs. C	1.93	0.079m	0.00579	1	0.839	0.42	1.2	0.26	-0.756	0.47
influence of vest. turn cues	C vs. D	-0.599	0.56	0.35	0.73	0.00649	0.99	-1	0.34	-0.243	0.81
const. vest. vs. const. vis. cues	D vs. E	4.95	0.00043***	2.9	0.014*	3.1	0.01*	2.33	0.04*	-0.521	0.61
const. vis. vs. vs. no vis. cues	E vs. F	-2.6	0.025*	-0.694	0.5	-0.808	0.44	-0.618	0.55	-2.41	0.035*
vis. turn vs. const. vis. cues	C vs. E	6.43	4.9e-05***	2.71	0.02*	2.97	0.013*	2.39	0.036*	-0.769	0.46
vis. turn vs. no vis. cues	C vs. F	3.42	0.0057*	3.72	0.0034**	2.15	0.054m	1.38	0.2	-3.38	0.0062*
just vis. vs. just vest. turn cues	D vs. F	4.46	0.00097**	3.21	0.0083*	2.83	0.016*	1.93	0.08m	-2.21	0.049*

Table 11: Tabular overview of the paired two-tailed t-tests for the different comparisons in experiment REAL WORLD VERSUS VR. t-values are displayed with 3 digit precision, p-values for $\alpha = 0.05\%$ with 2 digit precision. Trailing zeros are omitted. The asterisks '*' indicate whether the two conditions differ significantly from each other (on a 5%, 0.5% or 0.05% level). An 'm' indicates that the difference is only marginally significant ($p < 0.1$).

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction

Figure 24: Automatic spatial updating performance for experiment REAL WORLD VERSUS VR, quantified as the difference between UPDATE and CONTROL performance. If updating to new orientations is harder than for baseline forth-and-back motions, UPDATE performance should be worse than CONTROL performance, resulting in a positive offset from zero in the above difference plots. This was the case for both conditions that relied on vestibular cues (blocks E & F). A zero or small offset, conversely, indicates that the available dynamic motion cues and static visual cues are sufficient to enable automatic spatial updating. This was observed for all conditions where participants could rely on visual landmark cues (blocks A-D).

2.6° for just visual turn cues in block D, suggesting that the lack of concurrent vestibular turn stimuli caused the slight direction-specificity of the ego-orientation error. Overall UPDATE errors, however, were still rather small in all dependent measures (see also Figure 59), indicating the ease of visually assisted automatic spatial updating.

13.3.2.2 Conditions without useful visual information (blocks E & F)   For the blindfolded condition (block F), the response pattern changed somewhat: Response times for UPDATE trials were significantly increased by more than 100ms. This is about the amount expected from the literature (Farrell & Robertson, 1998; May, 2000), corroborating the hypothesis that blindfolded spatial updating to new locations is not as quick and easy as for forth-and-back motions. However, the absolute response time in the UPDATE condition was only approximately 1.05s (see Figure 59), which is still considerably faster than response times typically observed in the literature: There, response times for pointings after blindfolded rotations differ considerably, with values ranging from 1.6s (May, 2000) and 1.7s (Farrell & Robertson, 1998) over 1.8-3.2s Rieser (1989) up to more than 3s (Creem & Proffitt, 2000; Presson & Montello, 1994). A recent study on visually assisted spatial updating in VR reported even response times between 8 and 12s (Wraga et al., 2003) (see subsection 17.4.4 for a detailed discussion of this study). This considerable difference between out results and the literature indicates the ease and intuitive usability of our pointing device, validating our rapid pointing metaphor.

Configuration error, absolute pointing error, and absolute ego-orientation error were only marginally increased, indicating that the consistency of the mental spatial representation did not suffer from the non-visual ego-motion. The absolute error measures for the UPDATE condition were only about a fourth higher than for the CONTROL task (see Figure 58), indicating that the main cause of the absolute pointing and ego-orientation errors is the accumulating path integration error from the consecutive turns and not so much the updating to new orientations. That is, one simple turn can probably be updated rather well, but the sequence of many turns lead to the accumulation of path integration errors which is visible in the absolute error data. The ego-orientation error in turning direction showed consequently a large variability, but no overall effect.

Block E with additional but to-be-ignored visual information showed slightly more pronounced differences between UPDATE and CONTROL performance, especially for response times, which were increased by more than 200ms. This indicates a severe difficulty in ignoring the visual stimulus, even though it was known to be totally irrelevant. However, the configuration error for UPDATE trials was only moderately increased by approximately 2°, indicating that the mental spatial representation was still rather consistent. The other variables show virtually no difference to the blindfolded condition in block F.

In a way, the UPDATE trials in block E (constVis. + vest. cues) can be seen as UPDATE trials for the *vestibular* stimulus and IGNORE trials for the *visual* stimulus. Conversely, the IGNORE condition of block D (HMD just vis. cues) can be seen as well as a UPDATE condition for the (constant) *vestibular* stimulus and an IGNORE condition for the *visual* stimulus. Figure 58 shows indeed virtually the same impaired performance for the two conditions where the visual cues were to be ignored and the vestibular ones to be trusted (block D IGNORE versus block E UPDATE). This was consistently observed for all five dependent variables. Especially the increased response time and configuration error for those conditions indicate a strong visual dominance over the vestibular cues: Even when explicitly intending to trust the vestibular cues more than the visual cues, participants were apparently unable to suppress the visual cues.

13.3.2.3 Summary and conclusions   The data revealed the relative ease and accuracy of automatic spatial updating when one is provided with meaningful visual landmarks arranged into a con-

sistent scene. Blindfolded spatial updating showed response time differences (Update - Control) similar to those observed in the literature (Farrell & Robertson, 1998), validating our methodology. Additional conflicting visual information was rather hard to ignore and increased response times further. Maybe the most relevant outcome was the comparability of spatial updating for real and virtual environments. This demonstrates the power and usability of VR for investigating spatial updating. Furthermore, our rapid pointing paradigm yielded overall response times that were considerably and consistently smaller than all values we found in the literature. On the one hand, this proves the ease and intuitiveness of our rapid pointing methodology. On the other hand, it allows the investigation of early processes in spatial updating that might not have been accessible before. This might be a critical issue in many spatial updating studies: If, for example, participants in the study by Wraga et al. (2003) (see also subsection 17.4.4) need more than seven times longer for pointing (8-12s) than for verbal responses (1.1-1.5s), this might be a critical issue, since response times of more than 8s might allow more than enough time for any kind of mental spatial task, like mental rotations, cognitive strategies etc. It is consequently at least debatable whether Wraga et al. (2003) measured in fact automatic spatial updating performance and not some kind of rather cognitive mental spatial abilities. Furthermore, such long response times might increase their variability to a level where differences in the order of 100ms (which is the typical difference found between Update and Control trials (e.g., Farrell & Robertson, 1998; May, 2000)) might not be visible any more.

13.3.3   Obligatory spatial updating

In this subsection, we will analyze the obligatory nature of spatial updating initiated by different combinations of visual and vestibular cues. The reasoning is as follows: If and only if spatial updating is obligatory (i.e., largely beyond conscious control) will ignoring the turn stimuli be considerably harder than updating them as usual. That is, the difference between Ignore and Update trials would then be considerably above zero. The corresponding data are summarized in Figure 25 and will be discussed in detail below.

13.3.3.1   Conditions with useful visual information (blocks A-D)   Both real world conditions demonstrate essentially the same obligatory nature of the turn stimuli, without any influence of the FOV: Ignore response times were increased by more than 300ms, and all error measures were greatly increased, too (see Figure 25). The considerable increase in configuration error indicates that the mental representation of the previous (to-be-remembered) orientation was less consistent and could not be remembered properly. The increase in absolute pointing error and absolute ego-orientation error can to a considerable part be explained by a direction-specific misperception of the correct ego-orientation in the opposite direction of the ignore motion (see also Figure 59). That is, participants were apparently unable to correctly remember their previous orientation in the Ignore trials and pointed as if the former orientation was being rotated in the *opposite* direction, i.e., further away than it actually was. This phenomenon is somewhat counterintuitive and conflicts with a motion capture or representational momentum explanation (see page 62), which would predict a misperception in the direction of the motion, not against. It seems like participants were trying to overcompensate the actual rotation by pointing as if having turned further than they actually did.

The VR conditions (blocks C&D) demonstrated comparable obligatory spatial updating, even though the difference between Ignore and Update trials was slightly but insignificantly less pronounced. A look at Figure 59 reveals that the ego-orientation error against turning direction in the Ignore condition is greatest for the real world conditions (block A (7°) and block B (6.3°)), slightly smaller for block C with HMD (4.4°), and negligible for block D without vestibular cues. Even though only the differences between the real world conditions (blocks A and B) and the purely visual condition
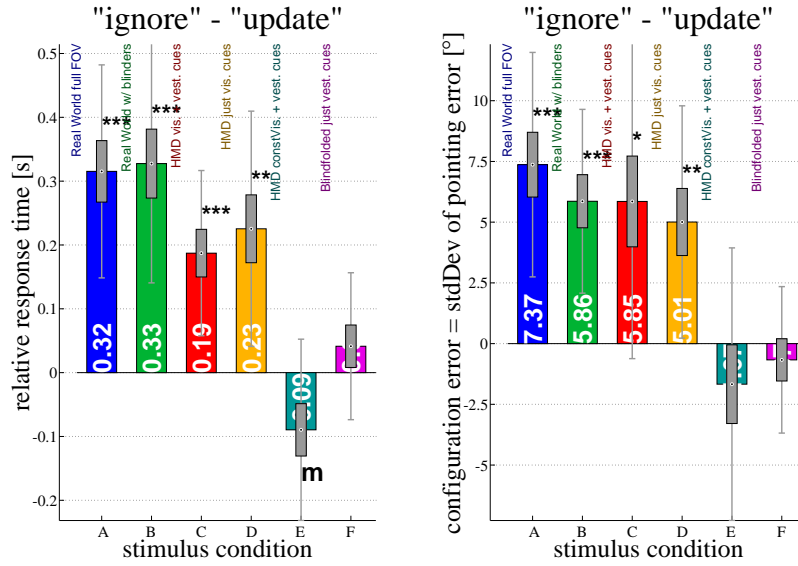
(a) Response time
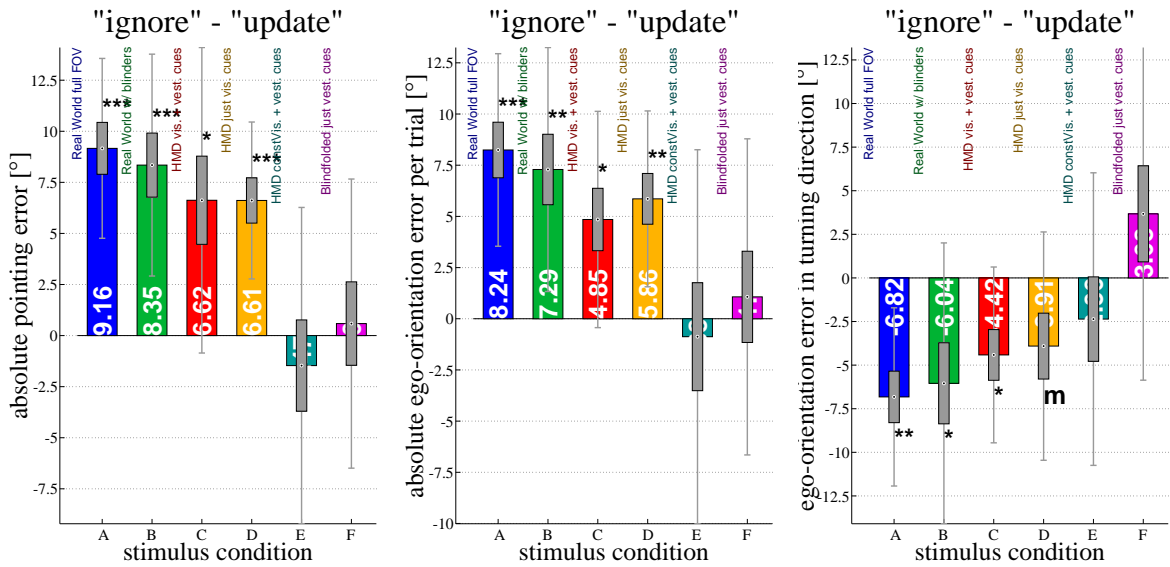
(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error
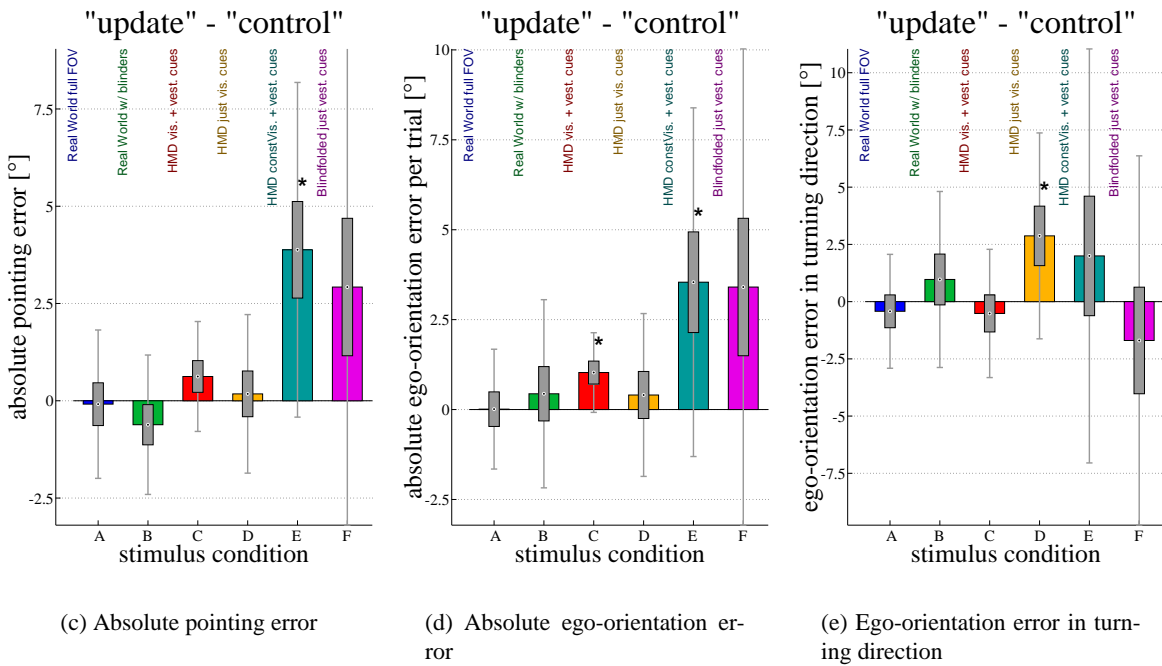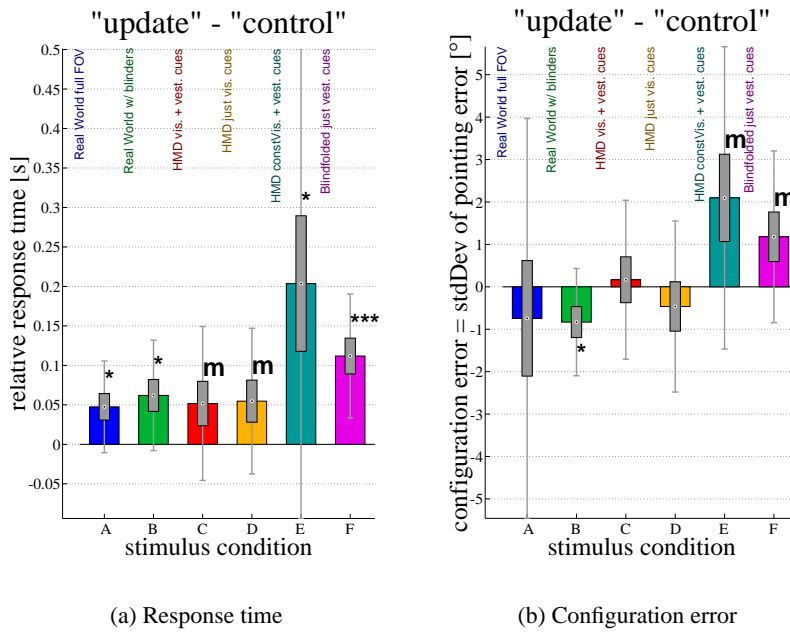
(e) Ego-orientation error in turning direction

Figure 25: Obligatory spatial updating performance for experiment REAL WORLD VERSUS VR, quantified as the difference between IGNORE and UPDATE performance. Values above zero indicate that ignoring is considerably harder than updating, implying obligatory spatial updating. This was the case for all conditions with useful visual cues (block A - D).

(block D) reached significance ($t(11) = 3, p = 0.012^*$ and $t(11) = 2.27, p = 0.044^*$ respectively), the ego-orientation error against turning direction seems to be an interesting variable that has to our knowledge previously been neglected in spatial updating studies. The direction of the effect, however, is in the direction opposite of the one that might be predicted from the representational momentum literature. Consequently, the explanations proposed by the literature are not directly applicable to our results.

13.3.3.2   Conditions without useful visual information (blocks E & F)   Figure 25 reveals that vestibular turn cues without assisting visual turn cues were essentially as easy or hard to Ignore as to Update. This was found consistently for all five dependent measures. That is, smooth vestibular turn cues alone are clearly incapable of inducing obligatory spatial updating and turn the world inside our head against our own conscious will. In block E with the visual stimulus indicating generally the wrong orientation, but no turn, ignoring ego-turns is almost 100ms faster than updating them, suggesting that ignoring was actually perceived as easier than updating.

The observed ease of ignoring vestibular cues from blindfolded turns in block F was rather surprising, as the literature indicates that blindfolded motions should be much harder to Ignore than to Update. Farrell & Robertson (1998) found for example a response time increase from 1.7s to 3.3s, accompanied by a moderate increase in absolute pointing error from approximately 24° to 31°. Several differences in the experimental paradigms and setups used might be able to explain some of the observed differences:

Our hand-help pointing wand enabled generally considerably smaller response times than those observed in the literature (see subsection 13.3.2.2 and (Creem & Proffitt, 2000; May, 2000; Farrell & Robertson, 1998; Presson & Montello, 1994; Rieser, 1989; Wraga et al., 2003)). This suggests that our pointing paradigm might be easier and more intuitive to use, which enables us to better investigate the quick process of spatial updating. The overall small response times in our experiments, however, do by no means explain the ease in ignoring purely vestibular turn stimuli. The smooth motions used were clearly above detection threshold, but the accelerations and velocities reached might still be considerably below the values typically used in the literature. Furthermore, pointing targets in our study were attached to the walls of the room and hence embedded in a consistent, natural scene. In this manner, participants probably did not update or image the position of *individual* targets or target arrays, but most likely updated the scene and room geometry as a whole, which is known to be more reliable and less prone to disorientations (Wang & Spelke, 2000). This is different from many spatial updating studies which used arrays of objects that were not well embedded or part of a consistent scene, which might yield to different effects (Carpenter & Proffit, 2001; Easton & Sholl, 1995; Farrell & Robertson, 1998, 2000; Farrell & Thomson, 1998; May & Klatzky, 2000; May, 1996; Presson & Montello, 1994; Rieser et al., 1982; Yardley & Higgins, 1998; Rieser & Rider, 1991; Rieser, 1989; Wang & Spelke, 2000). Finally, the repeated turns without in between visibility of the scene in our study might also have contributed to the ease of ignoring the motion.

13.3.3.3   Summary and conclusions   Comparing Ignore and Update performance revealed obligatory spatial updating for all conditions with useful visual information (blocks A-D). That is, visual cues alone, even without any concurrent vestibular stimuli, can be sufficient for turning the world inside our head, even against our own conscious will. This clearly indicates reflex-like, cognitively hardly penetrable spatial updating by visual cues alone. Moreover, a strong visual dominance over the vestibular cues was observed, even in the conditions where participants were explicitly asked to ignore the visual stimulus completely and just trust the vestibular cues (block E).

Smooth vestibular turn cues without any assisting visual cues, on the other hand, were clearly incapable of triggering reflex-like obligatory spatial updating (block F). This outcome is to our knowl-

edge unprecedented and not predicted by the literature. Low accelerations and velocities and a highly consistent scene that is easier to mentally picture might all contribute to this apparent contradiction. Further experiments, however, are needed to understand this fundamental difference and pinpoint the exact condition under which vestibular cues might indeed be sufficient for initiating obligatory spatial updating as is typically claimed by the literature.

13.3.4 Further analyses

13.3.4.1 Learning effect To test whether participants' performance depended on the amount of exposure to the task, a correlation analysis was performed between performance and session number. No performance feedback was given during any of the test phases. That is, potential improvements would be due to practice effects, implicit learning, or improved strategies. The results of the correlation analysis, however, revealed no significant correlation for any of the dependent parameters and spatial updating conditions (all $r^2 \, 's \leq 0.032$, all $t's(70) \leq 1.51$, all $p's \geq 0.068$).

13.3.4.2 Turning angle effect If spatial updating was non-automatic and thus effortful and requiring much cognitive effort like mental spatial rotation, we would expect that smaller turns should be easier and lead to better UPDATE performance than larger turns. This should be reflected in increased errors and especially response times for larger turns. A correlation analysis, however, revealed no significant performance decrease with turning angle for any of the dependent variables and cue combinations (blocks) ($p > 0.05$). This suggests that spatial updating was performed *during* the motion, and not afterwards. Furthermore, the lack of a turning angle effect argues against higher cognitive processes like mental spatial rotations performed after or during the actual turn. Together with the rather low overall response times, this suggests that spatial updating was indeed automatic.

If, on the other hand, participants in the IGNORE condition performed some kind of mental back-rotation as is often claimed, response times in the IGNORE condition should be positively correlated with turning angle. Correlation analyses showed such an effect only for the absolute pointing error and absolute ego-orientation error in block B (Real World w/ blinders, $r = 0.38$, $r^2 = 0.14$, $t(11) = 2.48$, $p = 0.031^*$ and $r = 0.29$, $r^2 = 0.085$, $t(11) = 2.44$, $p = 0.033^*$, respectively). None of the other correlations reached significance for any of the dependent variables and cue combinations (blocks) ($p > 0.05$). This lack of consistent response time increase with turning angle argues against the mental rotation hypothesis.

13.3.4.3 Pointing order effect If participants were disoriented or otherwise confused by the motion cues or task requirements, one would expect the first pointing of each trial to be worse than the later pointings. For the IGNORE trials, one might expect the opposite effect in the visual conditions if the stimuli are dominant enough to make it hard to imagine the previous location. To quantify these effects, correlation analyses were performed between the two dependent variables which were computed on an individual pointing basis, namely response time and absolute pointing error, and the pointing number.

UPDATE performance did not correlate significantly with pointing number for any of the cue combinations (blocks), indicating again that updating was being performed *during* the rotation, not afterwards. That is, the mental spatial reference frame was already aligned with the new orientation when being asked to point, which is the essence of what automatic spatial updating is good for.

IGNORE performance showed no significant effect in absolute error, but a significant increase in response times for both real world conditions ($r = 0.089$, $r^2 = 0.008$, $t(11) = 2.23$, $p = 0.048^*$ for unrestricted vision in block A and $r = 0.11$, $r^2 = 0.011$, $t(11) = 2.40$, $p = 0.035^*$ for limited

FOV in block B). Blocks C-F showed no such effect. This indicates that only the real world stimulus was strong enough to make it hard to keep in mind the *previous* orientation over the time of the four pointings while having to see the *current*, to-be-ignored view. This effect corroborates earlier observations that ignoring real world stimuli seems somewhat harder than ignoring VR stimuli (see the response time analysis and representational momentum discussion in subsection 13.3.3.1).

13.4   Summary and conclusions

The rapid pointing paradigm was well received by all participants and enabled response times and accuracies below the values typically observed in the literature, indicating the ease and intuitive usability of our pointing device. For all conditions with useful visual cues, the typical response pattern for obligatory spatial updating was observed: UPDATE performance was almost as good as CONTROL performance, whereas IGNORE performance was considerably worse. This suggests that our rapid pointing method allowed indeed to reliably quantify spatial updating.

The response pattern was found irrespective of concurrent vestibular motion cues, indicating that visual cues alone were sufficient to elicit reflex-like obligatory spatial updating. Furthermore, performance in VR was about as good as performance in its real world counterpart (as long as the FOV was the same). That is, a simulated, photorealistic view onto a consistent, landmark-rich environment was as powerful in turning our mental spatial representation against our own conscious will as a corresponding view onto the real world. This highlights the power and flexibility of using highly photorealistic VR for investigating human spatial orientation and spatial cognition. Last but not least, it validates our VR-based experimental paradigm, and suggests the transferability of results obtained in this VR setup to comparable real world tasks.

14 Experiment 7: "SIMULATION PARAMETERS"

14.1 Introduction

The previous experiment showed the typical response pattern for spatial updating, and examined extreme cases of either full or no information in the visual and vestibular domain. The purpose of Experiment SIMULATION PARAMETERS is to explore what other visuo-vestibular parameters might be critical for spatial updating. Parameters of interest included visual display parameters like FOV and projection screen versus HMD usage, but also visuo-vestibular motion parameters like the relation between visual and vestibular motion and the turning amplitude and velocity. Due to the number of potentially relevant parameters, the nature of this experiment is rather exploratory. The main purpose was consequently to scan the realm of parameters to identify critical parameters that are worth investigating in more depth in later experiments.

14.2 Methods

As the rapid pointing paradigm developed for the previous experiment proved quite successful, the same paradigm was used again in this experiment. To obviate potential criticism, the visual scene was exchanged: Instead of using 12 regularly arranged target objects attached to the wall of the rectangular Motion-Lab, we used a considerably larger, more complex, and less regular environment with almost twice as many target landmarks arranged irregularly (the Tübingen market place, see Figure 28).This should render abstract cognitive strategies (like using symmetries and counting targets etc.) virtually impossible, thus forcing participants to resort to automatized spatial updating whenever possible. Furthermore, participants are pushed close to their performance limit, which might yield even clearer results.

The physical turning angles in the previous experiment were somewhat limited due to physical constraints of the motion platform, which restricts the heading to $\pm 57°$ from straight ahead. As larger turning angles might, however, be more difficult to update and hence lead to clearer performance differences between conditions, we cheated in this experiment by using *visual* turns which were considerably larger than the simultaneous *vestibular* (physical) turn. That is, we introduced different gain factors between the physical and visual motion, and asked participants to use the visual cues for reference - which they did anyway before even being instructed. Pre-experiments had shown that vestibular/visual gain factors down to 1/4 are accepted and typically pass unnoticed by participants when they are involved in an engaging task like rapid pointing. This is in accordance with the literature indicating that the vestibular system is flexible and perceptually easily re-calibrated by sufficient visual cues (e.g., Ivanenko, Viaud-Delmon, Siegler, Israël, & Berthoz, 1998). In this manner, we were able to extend the visual motion range to $\pm 228°$, which includes turns of more than 360°.

14.2.1 Participants

A group of 6 female and 2 male naive participants took part in Experiment SIMULATION PARAMETERS. Ages ranged from 17 to 38 years (mean: 27.9 ± 0.8 years, SD: 4.8 years). All participants had been living in Tübingen for several years and were familiar with the Tübingen market place.

14.2.2 Stimuli and apparatus

Stimuli and apparatus were identical to the previous experiment, section 13, apart from the differences described below.
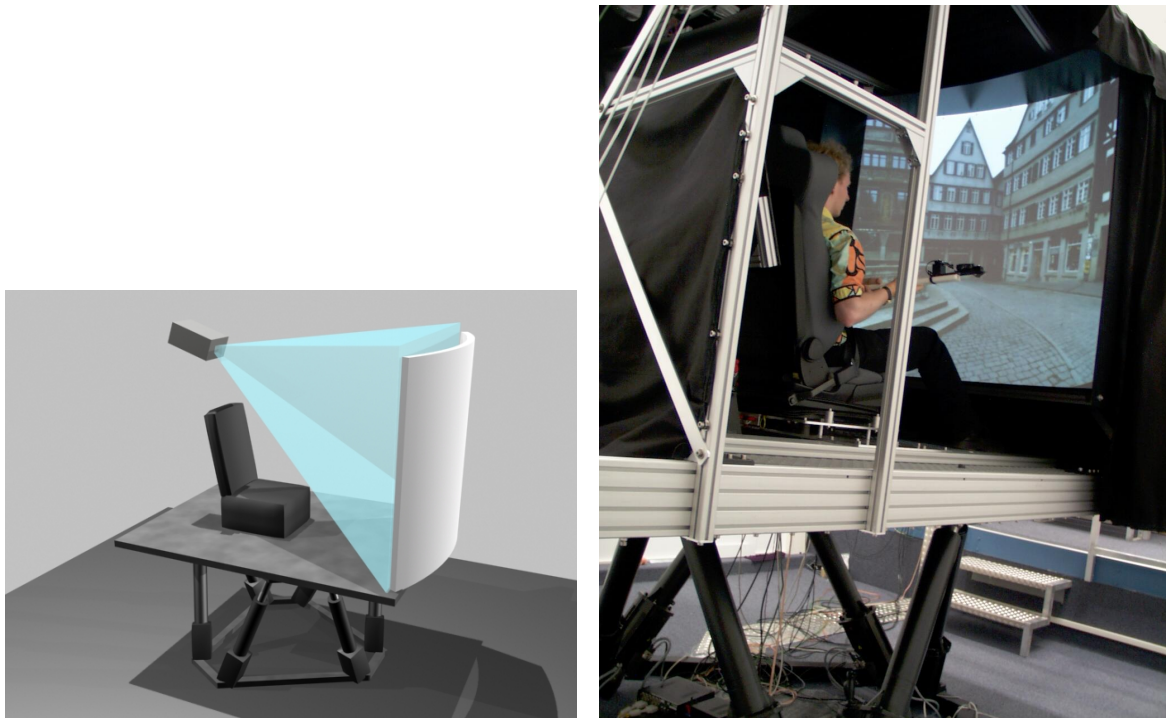
(a) Schematic experimental setup showing the six degree of freedom motion platform (Motionbase Maxcue) and the projection setup.

(b) Participant sitting on the motion platform and facing the curved projection screen, which displays a view of the Tübingen market place. The physical field of view is 84°x63° and matches the simulated FOV.

Figure 26: Projection setup mounted on top of the motion platform.

(a) Participant wearing position-tracked head-mounted display (40°x30° FOV, 1024x768 pixel) and active noise cancellation headphones.

(b) Participant wearing headphones and blinders (vision delimiting cardboard goggles) reducing the FOV to that of the HMD (40°x30°).

(c) Position-tracked pointer in the default position (upright) and pointing position.

Figure 27: Display devices and pointing apparatus used in experiment SIMULATION PARAMETERS.

14.2.2.1   Visualization   Three different visualization conditions were used in this experiment:

1. An **HMD** condition (see Fig. 27), which was comparable to block C of the previous experiment, section 13.

2. A **projection screen** condition, where participants were seated in front of a projection screen mounted on top of the platform as described below (see Fig. 26)

3. A **blinders** condition, where participants were again seated in front of the projection screen, but wore cardboard goggles limiting the FOV to that of the HMD ($40° \times 30°$) (see Fig. 27). The blinders were the same as in the previous experiment, but used to limit the view onto the computer-rendered image on the projection screen, not the real surround.

The projection setup was purpose-designed by the author and Markus von der Heyde mainly to allow for a larger FOV and to avoid drawbacks often associated with HMDs: HMDs are known to lead to a number of problems, first of all discomfort (so-called Virtual Reality-induced symptoms and effects (VRISE)). Commonly found symptoms include fatigue, headaches, eye strain, and blurred vision (Cobb, Nichols, Ramsey, & Wilson, 1999; Hettinger et al., 1996; Howarth & Costello, 1997; Mon-Williams & Wann, 1998; Stanney et al., 1998). Further drawbacks associated with HMDs include distortions of the perceived space as well as impaired spatial orientation performance (Arthur, 2000; Bakker et al., 1999, 2001; Hettinger et al., 1996; Kearns et al., 2002; Nelson et al., 1998), see also subsection 6.3.1, 11.1.2, and 11.2. The whole projection setup is mounted on top of the motion platform (see Fig. 26) in order to allow for optimal viewing conditions and immersion. As the curved projection screen used for the experiments in part II seemed advantageous over flat projection screens (see discussion in subsection 11.2), we used again a curved projection screen. As the whole projection setup had to be mounted on top of the platform, the size, FOV, and curvature of the screen had to be considerably reduced, however. This resulted in a curved projection screen of 1.68m width $\times$1.42m height, with a curvature radius of 2m. This yields a physical FOV of $84° \times 63°$ for the participant seated at a distance of about 1.14m. A modified wide-angle LCD video projector (Sony VPL-PX 21 with wide angle lens VPL-FM 21) is mounted above and behind the head of the participant and projects a computer-rendered 1024×768 pixel image non-stereoscopically onto the screen (see Fig. 26). The video projector is mounted using a purpose-designed two-stage vibration absorbing system to avoid potential damage due to extreme platform motions.

To enhance immersion and to reduce the influence of the external reference frame of the physical surround (Motion-Lab) as much as possible, the projection screen is surrounded by a black frame and the whole projection setup is surrounded by light-proof black curtains on all sides. Furthermore, the participant is wearing active noise-canceling headphones as described in subsection 13.2.2.

14.2.2.2   Scenery and pointing targets   A photorealistic virtual replica of the Tübingen market place was used as the visual stimulus (see Fig. 28). We refrained from using the real scene as comparison, as the previous experiment demonstrated already comparable performance in a real scene and it's photorealistic virtual replica. This suggests that our approach of using photorealistic virtual environments is feasible, and that we do not need further real world - VR comparisons to validate our approach.

From this market place, 22 salient landmarks were selected as pointing targets and marked by little red dots (see Fig. 28). The landmarks were irregularly spaced, at an angular distance between about 8° and 23°, with a mean angular distance of approximately 16°. The layout of the Tübingen market place and the target configuration was irregular and without any symmetry that could have been used to "cheat" or use strategies other than normal spatial updating to solve the task (see Figure 28 (e)).

(a) High resolution 360° roundshot of the Tübingen market place.

(b) Photorealistic model of the Tübingen market place, created by wrapping a 360° roundshot photograph onto a cylinder. This creates an undistorted view for the observer positioned in the center of the cylinder.

(c) Full 84°x63° view of the market place, displaying the landmarks 'Lammhofpassage', 'Briefkasten', 'Kreissparkasse', 'Marktschenke', 'Bäckerei', and 'foto-markt', indicated by little red dots.
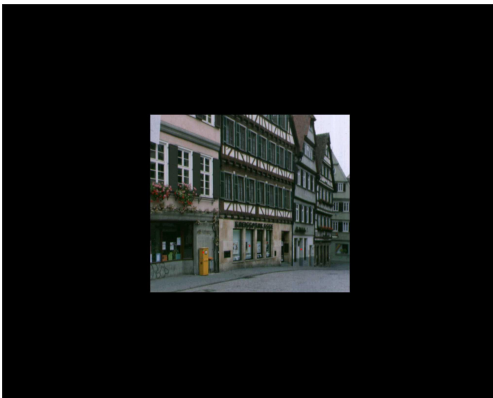
(d) Same view as in (c), but with the reduced FOV of 40°x30° (blinders or HMD conditions).
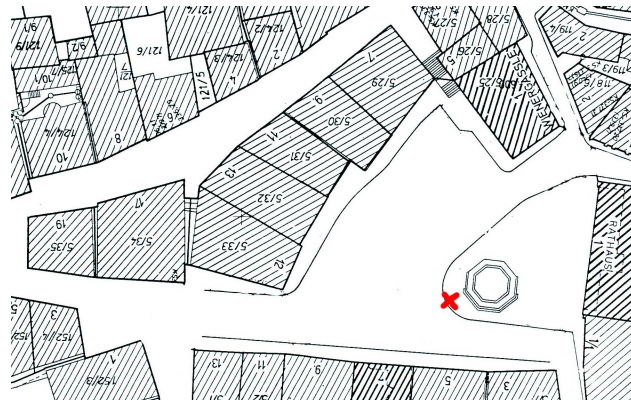
(e) Map of the Tübingen market place. The viewing position is marked by a red cross.

Figure 28: Building a 360° roundshot model. The virtual replica was created by wrapping a 360° round shot photograph of the Tübingen market place (top Figure (a)) onto a cylinder (bottom Figure (b)). This creates an undistorted view for the observer positioned in the center of the cylinder (Figures (c) and (d)). The virtual observer is positioned near the fountain, as indicated by the red cross in Figure (e).

In most spatial updating studies in the literature, however, target configuration and/or room geometry were quite simple: Typically, a rather limited number of four to eight targets is used, which are often arranged regularly or along the cardinal directions (front-back-left-right), and are typically embedded in just a simple and featureless symmetrical room (Carpenter & Proffit, 2001; Creem & Proffitt, 2000; Easton & Sholl, 1995; Farrell & Robertson, 1998, 2000; May & Klatzky, 2000; May, 1996; Presson & Montello, 1994; Rieser, 1989; Wang & Spelke, 2000; Wraga et al., 2003; Yardley & Higgins, 1998). Hence, it seems quite possible that participants did not always update the room or target configuration properly. Instead, they might for example have used simpler strategies like reversing left and right directions instead of updating 180° turns, updating individual targets analytically instead of using normal, holistic spatial updating, or counting objects instead of updating turns (e.g., take the third target to the left instead of updating a 135° turn).

The asymmetric scene and target configuration in this experiment made all of these strategies impracticable. Furthermore, the abundance of landmarks ensured that participants could not update all targets individually (4-8 targets seem to be the maximum number of items that can be kept in mind individually) or rote-learn all relative angular distances between all targets (which would add up to $\sum_{n=1}^{N-1=21} n = 231$). Instead, participants had to update the whole scenery to deduct individual target locations, which is exactly what we intended and what "normal" spatial updating is about.

14.2.3 Procedure

As before, a repeated-measures, within-subject design was used. The experimental design and the cue combinations used for each block are summarized in Tables 12 and 13 and described in more detail in subsection 14.2.4. Each participant completed 8 blocks of different cue combinations in pseudo-balanced order. Each block consisted of 32 trials and lasted about 18 minutes. The 32 trials were split into 12 UPDATE trails, and 6 trials each for the other spatial updating conditions (CONTROL, IGNORE, and IGNORE BACKMOTION). To avoid clustering, the order of the spatial updating conditions was fixed to six repetitions of the sequence CONTROL → UPDATE → IGNORE → IGNORE BACKMOTION → UPDATE. This sequence is depicted in Figure 29 for one exemplary block (D). The blocks were performed in two or three sessions on different days to avoid fatigue effects and obviate the influence of declining alertness.

spatial updating condition	visual turning angle α	trials per block
1. UPDATE	$80° \leq |\alpha| \leq 456°$	12
2. CONTROL	$80° \leq |\alpha| \leq 114°$	6
3. IGNORE	$80° \leq |\alpha| \leq 228°$	6
4. IGNORE BACKMOTION	$80° \leq |\alpha| \leq 228°$	6

Table 12: Summary of the four different spatial updating conditions per block in experiment SIMULATION PARAMETERS. Due to limitations of the platform turning range, the maximum physical (vestibular) heading deviation from straight ahead was ± 57°. For blocks A, B, C, and J, the visual turning angle range was four times smaller than in the above table, for block D half as large. Turning angles were pseudo-randomized to cover the range described above.

In a follow-up study, the same participants performed three more control conditions, first block I (jump-condition), followed by 2 more blocks (J and K) in balanced order (see subsection 14.2.4 and Table 13). For the sake of comparability, the results of the follow-up study are presented together with the original eight blocks. This comparison might seem critical due to potential learning or practice effects. The learning effect analysis in subsection 14.3.4, however, revealed only a significant learning effect in terms of response time. This response time decrease was found for both the
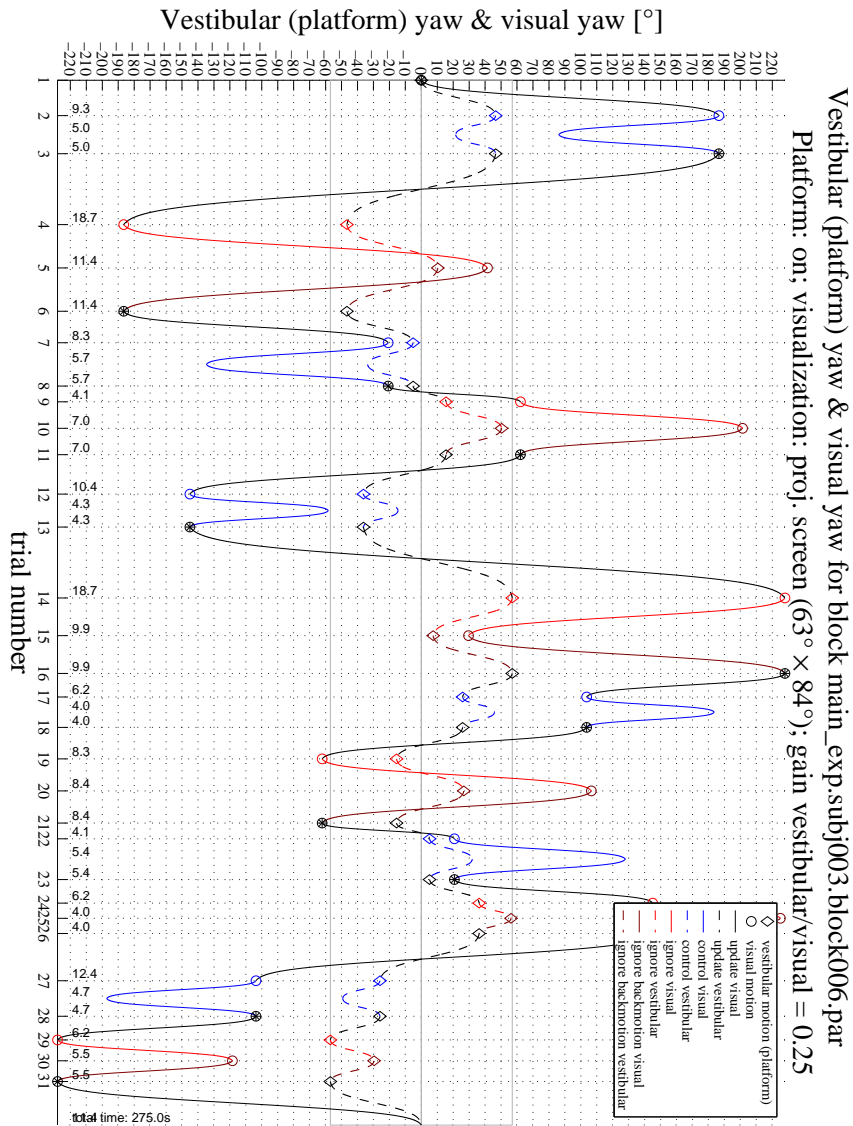
Figure 29: Vestibular (platform) motion and visual motion for one representative block (D) of experiment SIMU-LATION PARAMETERS. Depicted are the vestibular (platform) and visual yaw angle, demonstrating the sequence of the spatial updating conditions (CONTROL → UPDATE → IGNORE → IGNORE BACKMOTION → UPDATE etc.) and the effect of the gain factor. In this block, the vestibulo-visual gain factor was $g = 1/4$, indicating that the visual motions were four times as large as the physical (vestibular) ones. Pointings occurred at all circles and diamonds of the trajectory.

UPDATE and IGNORE conditions, but not the CONTROL condition. This suggests that careful comparisons between the first eight blocks and the three subsequent blocks might be legitimate for the other four dependent variables and for the CONTROL condition in general.

14.2.4 Stimulus conditions

The different stimulus conditions for the eleven blocks were chosen to allow for comparisons between different visual display parameters like FOV and projection screen versus HMD usage on the one hand and visuo-vestibular motion parameters like gain factors and turning amplitude and velocity on the other hand. The parameter combinations are compiled in Table 13 and motivated below in more detail.

block	visualization setup	field of view (FOV)	gain vest./vis.	angular range (visual)	mean visual turn velocity	cue conflict
A	HMD	$40°\times30°$	1	$[-57°, +57°]$	$20°/s$	no
B	blinders w/ proj. screen	$40°\times30°$	1	$[-57°, +57°]$	$20°/s$	no
C	proj. screen	$84°\times63°$	1	$[-57°, +57°]$	$20°/s$	no
D	proj. screen	$84°\times63°$	0.5	$[-114°, +114°]$	$20°/s$	gain
E	blinders w/ proj. screen	$40°\times30°$	0.25	$[-228°, +228°]$	$20°/s$	gain
F	proj. screen	$84°\times63°$	0.25	$[-228°, +228°]$	$20°/s$	gain
G	blinders w/ proj. screen	$40°\times30°$	0	$[-228°, +228°]$	$20°/s$	yes
H	proj. screen	$84°\times63°$	0	$[-228°, +228°]$	$20°/s$	yes
I	proj. screen	$84°\times63°$	0	$[-228°, +228°]$	jump	yes
J	proj. screen	$84°\times63°$	0.25	$[-57°, +57°]$	$20°/s$	gain
K	proj. screen	$84°\times63°$	0.25	$[-228°, +228°]$	$80°/s$	gain

Table 13: Summary of the 8+3 different cue combinations (blocks) used in experiment SIMULATION PARAMETERS.

The HMD condition (block A) was chosen to allow for comparisons to the previous experiment, which used the same HMD and the same yaw range, but a different scene. Block B and C used the projection screen with and without blinders, and were apart from this identical to block A. This allows for comparisons between the different visualization systems (HMD vs. blinders) and FOV (projection screen with and without blinders).

Blocks D-H were designed to investigate larger turning angles, which are often assumed to be more difficult to update. To do this, gain factors $g_{vestibular/visual} < 1$ had to be introduced, as the motion platform used has a limited turning range of $\pm 57°$. Pre-experiments had shown that gain factors have only limited, if any, influence on spatial updating performance, and that participants were easily re-calibrated to gain factors as low as $g = 1/4$. If engaged in a challenging task like rapid pointing, they did not even notice the gain factors $g \neq 1$. Block D used an intermediate gain factor of $g = 1/2$, block E and F reduced the gain factor to $g = 1/4$. In block G and H, the gain factors were set to zero, to investigate spatial updating performance if no concurrent vestibular turn stimuli were presented at all. To investigate the influence of the FOV, participants wore blinders in blocks E and G, and the results were compared to blocks F and H, respectively, with unrestricted vision.

The three control conditions (blocks I-K) were run in one separate session after the first eight blocks were completed. They were designed to answer some of the questions not sufficiently addressed by blocks A-H. It seemed as if stimulus parameters and movement specifics had little if any influence on spatial updating performance in blocks A-H and the previous experiment as long as sufficient visual

cues were provided. This raised the question whether a smooth, continuous spatial updating, similar to processes known as mental spatial rotation, can fully explain the observed results. To address this issue, a "jump" condition was introduced in block I. In this condition, participants were presented with new views *without* any continuous motion in between, as if they were teleported directly to the new orientation. This was similar to a slide-show type presentation. If motion parameters have any influence on visually assisted spatial updating, this jump to new orientation should disrupt spatial updating performance considerably and improve ignore performance. Similar results for continuous and discontinuous (jump-like) spatial updating, on the other hand, would suggest that the mere view of the new orientation is sufficient to somehow teleport the mental spatial representation to the new orientation and re-anchor the mental reference frame almost instantaneously.

Block J used a gain factor of $g = 1/4$ and a visual yaw range of $\pm 57°$, resulting in a vestibular (physical) yaw range of only $\pm 14.25°$. This allowed the disambiguation of effects of gain factor and turning angle by comparing the results to block B and F. Block K was designed to elucidate the effect of movement velocity by using visual rotational velocities that were four times higher than in the other conditions. Apart from that, block K was identical to block F.

14.2.5   Interaction (Pointing)

The pointing paradigm was similar to the previous experiment (section 13), apart from a few changes described below. As before, pointing targets were selected to be outside of the current FOV. Due to the increased FOV in this experiment, pointing targets were now selected to be further away from straight ahead ($|\alpha_{pointer} - \alpha_{straight-ahead}| \in [42°, 110°]$). The allotted pointing response time was increased to nine seconds, as the scene and target configuration were considerably more complex. As the target names in this experiment had different lengths, but were unique after the second syllable, the response time computation was adapted accordingly by defining $t = 0$ to be at the mean pronunciation time of the first two syllables (1.43s). Furthermore, the summed response time between the first target announcement and the forth and last pointing was announced acoustically just before the next trial. Pilot experiments had shown that this performance feedback effectively motivates participants to perform as good as they could. It was further found to decrease boredom and keep participants alert by enhancing the game-like character of the experiment.

14.2.6   Training phase and course of the experiment

Due to the complexity of the experiment, participants completed an extended training phase as described below. Purpose of the training phase was to familiarize participants with the experimental requirements and the rapid pointing procedure in particular, the different spatial updating conditions, and the VR setup. The various phases of the training are listed below in chronological order.

1. **Landmark pre-test**: To test how well participants were acquainted with the Tübingen market place, they were asked to name as many landmarks (salient objects) on the Tübingen market place as they could. Participants were able to name between 8 and 17 landmarks (mean: 11.4), indicating that they were already quite familiar with the environment used.

2. **Demonstration of the experiment**: To give a first impression of the requirements and procedures of the experiment, the experimenter performed several trials and explained the different parts of the experiment as they occurred.

3. **Real world training**: Participants were seated on a swivel chair and held the pointing wand with attached laser pointer in their hands. After being rotated by the experimenter to different orientations, participants were asked to point "as accurately and quickly as possible" to

Figure 30: Picture of landmark Fotomarkt used in the landmark picture training phase. Note the little red dot in the center of the image indicating the exact target location.

different objects in the real Motion-Lab. In this manner, participants were familiarized with the pointing task and the different spatial updating conditions. They were trained until they were able to point with an accuracy of roughly $4°$ (corresponding, e.g., to a 20cm offset for a distance of 2.9m).

4. **Landmark picture training**: Participants were seated in front of a computer screen and listened to the computer-generated voice pronouncing target landmarks. About one second after each landmark announcement, a picture of that landmark appeared on the computer screen, with a little red dot indicating the target position (see Fig. 30). Participants were asked to familiarize themselves with the different targets and were allowed to take as much time as they needed. They initiated the next target pronunciation by pressing a designated key. In this manner, all 22 targets were shown in random order two times. In a third block, only the landmark pictures were shown on the monitor and participants were asked to name all visible landmarks. The landmark picture training ensured that participants were familiar with all landmarks and could recognize them quickly and easily from both the view and the computer-generated target announcement. The random picture sequence ensured that participants did not learn or code landmarks in a sequential manner, which could have introduced order artifacts or cardinal directions.

5. **Landmarks training in scene context**: The purpose of this training phase was to train participants to quickly locate each target in the scene context of the simulated market place. They saw a view of the market place presented on the monitor, similar to Figure 28 (c), and could smoothly change the simulated orientation (yaw) using designated keys. After the computer-generated voice announced a target, participants were asked to quickly change the simulated (yaw) orientation to face that target. Each target was announced two times in random order.

6. **Written instructions**: To ensure that participants understood the experimental requirements properly, they were given written instructions to read while the experimenter started the VR simulation.

7. **Main experiment in 11 blocks**: Before the first block, participants performed a few practice trials of that block until they felt comfortable with the task and did not have any further questions.

8. **Landmark post-test**: This test was identical to the landmark pre-test. All participants were now able to name all 22 landmarks easily and in the correct order.

9. **Interview**: After the experiment, participants were questioned about strategies used and potential problems and possible improvements of the experiment.

14.3   Results and discussion

To provide a first impression of the results, the data for the HMD condition (block A) are summarized in Figure 31. As in the previous experiment, Figures 31 (a) - (d) demonstrate the typical response pattern for obligatory spatial updating: UPDATE and IGNORE BACKMOTION performance are almost as good as baseline CONTROL performance, whereas IGNORE performance is considerably decreased.

Compared to the results in the similar HMD condition (block C: HMD vis. + vest. cues) of the previous experiment, however, overall performance is slightly decreased: Configuration error, absolute pointing error, as well as absolute ego-orientation error are all decreased by 4-5°. Furthermore, mean response times in the CONTROL condition were increased by 250ms, indicating that the task was considerably harder, even for the supposedly simple baseline task. In the previous experiment, mean response times between participants varied between 0.40s and 1.88s in the CONTROL condition for block C, which is slightly below the 0.66s to 2.23s observed in the CONTROL condition of this experiment for the otherwise comparable block A (HMD, g=1, +-57deg).

Apart from possible overall differences in the participant populations used, differences in the experimental procedures might account for the observed performance differences. Most prominently, the scenery used in Experiment SIMULATION PARAMETERS was considerably more complex and without any potentially helpful symmetry or regularity. Furthermore, the number of targets was almost doubled, and targets were arranged irregularly without any symmetry. Such a complex environment has to our knowledge never been investigated in the spatial updating literature, where the usage of simple and regular setups or surrounding rooms with no more than eight targets is the prevailing standard (Creem & Proffitt, 2000; Creem et al., 2001; Easton & Sholl, 1995; Farrell & Robertson, 1998, 2000; Klatzky et al., 1998; May & Klatzky, 2000; May, 2000, 2000; Presson & Montello, 1994; Rieser, 1989; Shelton & McNamara, 2001; Simons & Wang, 1998; Wang & Simons, 1999; Wang & Spelke, 2000; Wraga, Creem, & Proffitt, 1999b, 1999a; Wraga et al., 2000, 2003; Yardley & Higgins, 1998). Nevertheless, participants in our study were able to successfully update to new orientations and had a hard time ignoring the turn, indicating that spatial updating was still working and the rapid pointing approach was still a successful paradigm.

For a detailed analysis, and similar to the structure in the previous experiment, we will first present the baseline (CONTROL) performance for all blocks in subsection 14.3.1. The sufficiency of the cues for automatic and obligatory spatial updating will subsequently be analyzed in subsection 14.3.2 and 14.3.3, respectively. As this experiment is exploratory in nature, the line of argument will be rather qualitative, focusing on the interesting and significant effects. For reference, however, the corresponding statistical analyses of interest are summarized in Tables 14, 15, and 16. Due to the small number of participants, the power of the tests is of course limited, and only strong effects might be clearly visible. Nevertheless, this procedure seems sufficient to scan the realm of potentially relevant parameters and identify the ones that are worth being investigated in more detail in future studies. The full data set of this experiment is displayed in Figures 60, 61, and 62, compiled per cue combination (block) and spatial updating condition.

14.3.1   Baseline (CONTROL) performance

The results for the baseline (CONTROL) condition are summarized in Figure 32, with the corresponding statistical analyses in Table 14. In general, CONTROL performance showed a considerable

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
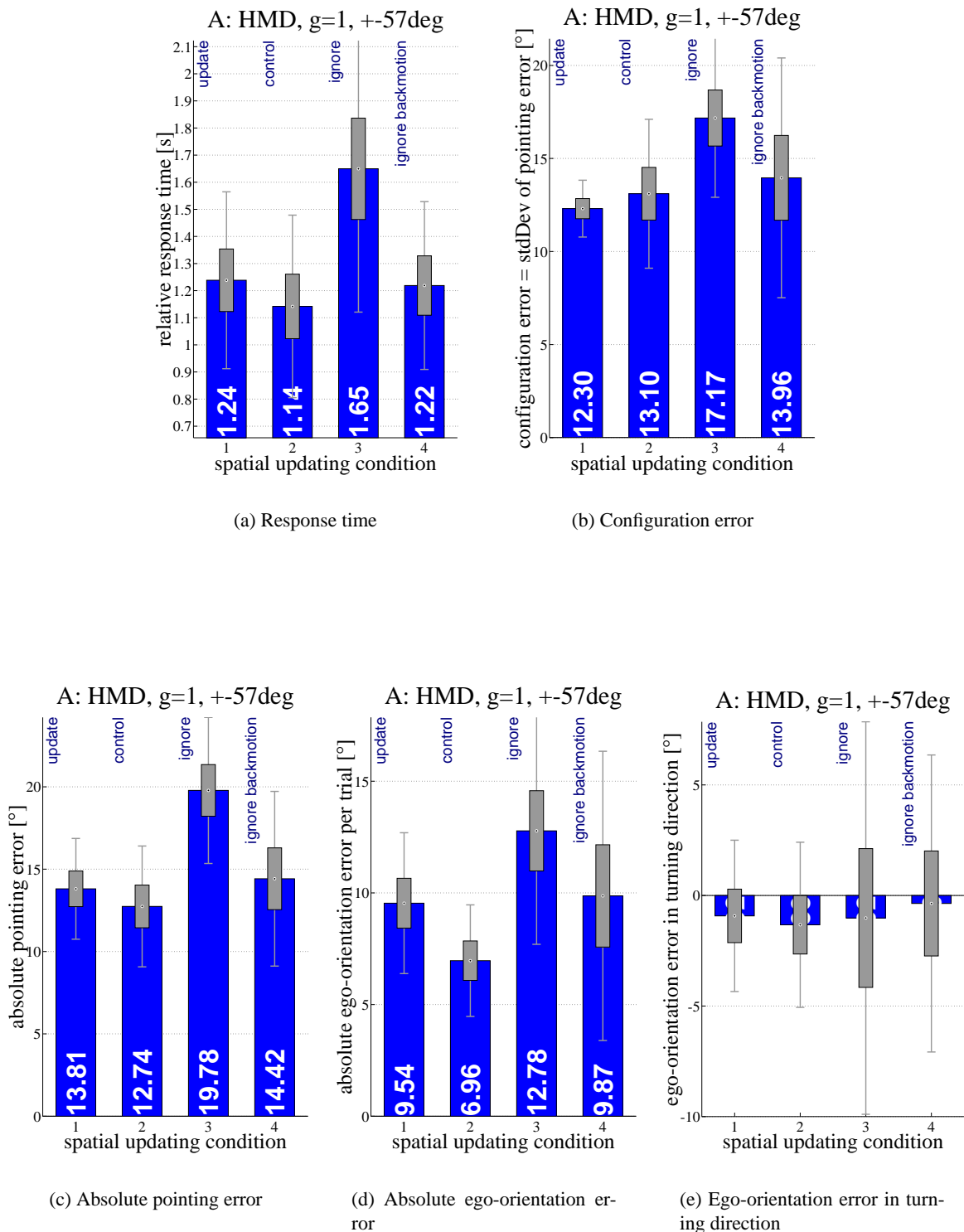
Figure 31: Pointing performance in Experiment SIMULATION PARAMETERS showing the typical spatial updating pattern. Performance in the HMD condition (block A, with a gain of $g = 1$, and a visual as well as vestibular yaw range of $\pm57°$) is plotted for the five dependent variables, each for the four different spatial updating conditions as indicated at the top of each plot.
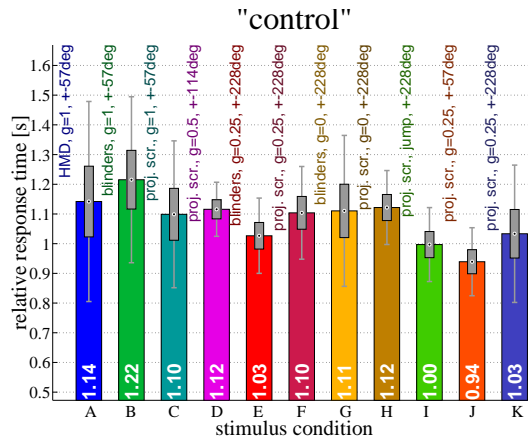
(a) Response time
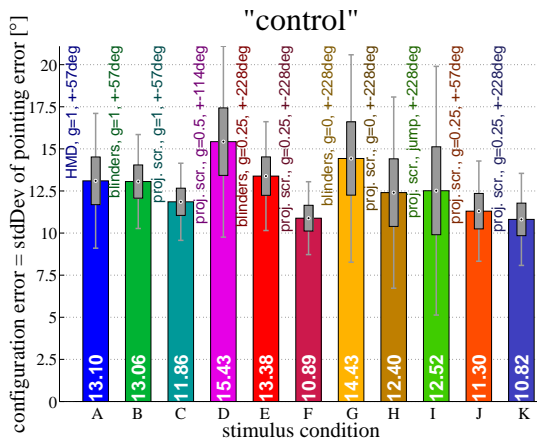
(b) Configuration error
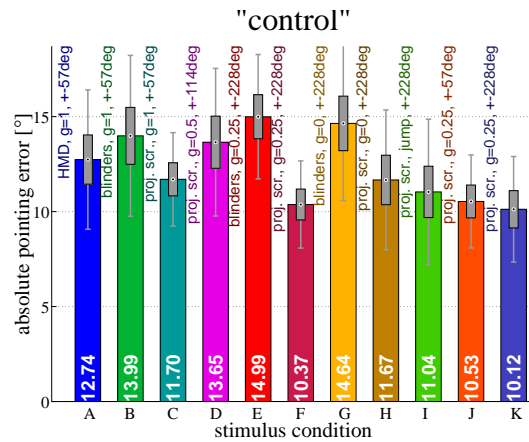
(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction

Figure 32: Baseline spatial updating performance for Experiment Simulation Parameters. Baseline (Control) performance is plotted for the five dependent variables for the eleven different cue combinations (blocks).

between-subject variability, indicated by the often rather large standard deviations, but only few differences between blocks. That is, neither the difference between HMD and blinder usage, nor the gain factor, turning angle, or movement velocity produced any clear performance differences for the simple forth-and-back motions (cf. Table 14). Even the jump-condition (block I), where participants saw an intervening view of a new orientation before seeing the same view again, did not decrease performance. Only the FOV had a consistent effect on pointing accuracy (blocks B vs. C, E vs. F, and G vs. H). Reducing the FOV via blinders increased configuration error, absolute pointing error, and absolute ego-orientation error by up to 5°. This reduction in pointing accuracy in the baseline trials is even more pronounced than the reduction by 2-3° in the previous experiment, where the FOV was reduced from an *unlimited* FOV to the same blinders-limited FOV of $40° \times 30°$. This suggests that the more complex target configuration used in the current experiment did indeed lead to clearer differences between the different cue combinations. Compared to the previous experiment, however, we did not observe any clear response time differences. This might indicate that the response time advantage for an unlimited FOV in the previous experiment was caused by the peripheral vision of the targets to point to. It could, however, also be caused by performance trade-offs between response time and pointing accuracy, as only the response time was fed back after each trial, and participants might have focused more on achieving small response times than on pointing accurately and consistently.

14.3.2 Automatic spatial updating

As in the previous experiment, automatic spatial updating was investigated by analyzing the difference between UPDATE and CONTROL performance for the different cue combinations. The data are compiled in Figure 33, with the corresponding statistical analyses being summarized in Table 15. Figure 33 (a) reveals a small increase in response time especially for the HMD condition (block A). All other dependent measures showed only marginal differences of typically less than 2.5° between UPDATE and CONTROL trials (Figures 33 (b) - (e)). Furthermore, there were apparently no differences between the different cue combinations whatsoever, which is corroborated by the pairwise t-test presented in Table 15. Taken together, this indicates that for all cue combinations tested, automatic spatial updating was almost as easy and accurate as baseline performance. Hence, we can conclude that photorealistic visual cues from a consistent, landmark-rich environment are under a wide range of simulation parameters sufficient to enable quick and accurate spatial updating. That is, neither the difference between HMD and blinder usage, nor the absence of any vestibular turn cues or any of the parameters FOV, gain factor, turning angle, movement velocity, or discontinuous (jump-like) updating (block I) reduced automatic spatial updating performance consistently. This is in accordance with the previous experiment, which also showed decent automatic spatial updating performance for all conditions with useful visual information (blocks A-D) and no clear effect of the stimulus conditions for the difference between UPDATE and CONTROL trials. This indicates again the power and usability of good visual stimuli for spatial updating, and points towards a visual dominance.

In the UPDATE trials of the jump condition (block I), the reference frame of the new orientation had to be instantiated before being able to perform the pointing task. Hence, the response time difference of approximately 30ms between UPDATE and CONTROL trials in the jump condition (block I) might be interpreted as the time needed to establish or re-anchor a new egocentric reference frame. If this interpretation was true, 30ms would indeed be rather quick for such a complex task like changing the orientation of the egocentric reference frame of the surrounding scene, even though the scene is well-known and highly trained. This would call for a highly automatized updating of the egocentric reference frame. To address this effect directly, however, one might want to compare UPDATE trials to

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
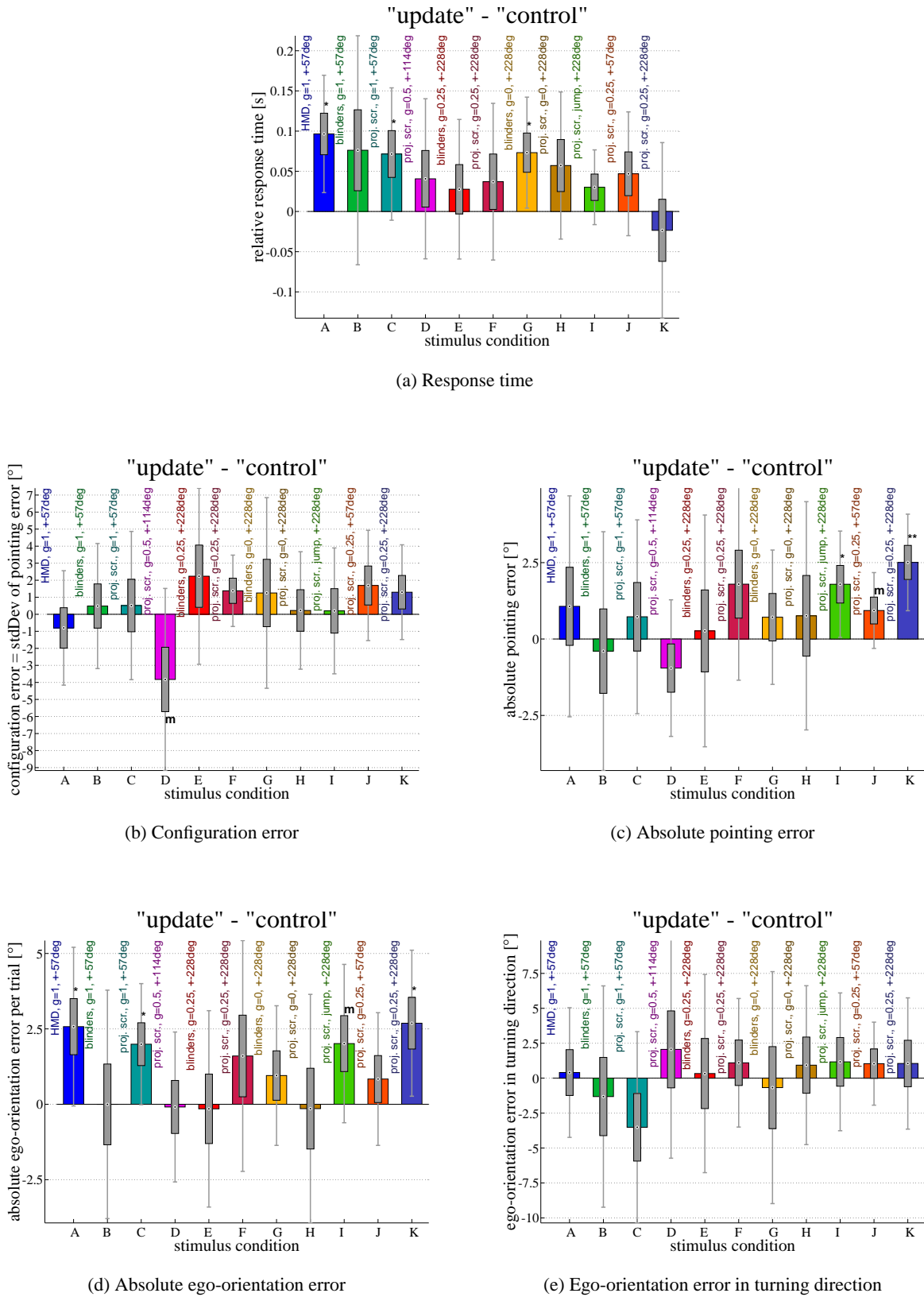
Figure 33: Automatic spatial updating performance for experiment SIMULATION PARAMETERS. Difference between update and control performance for the five dependent variables. If updating to new orientations is harder than for baseline forth-and-back motions, UPDATE performance should be worse than CONTROL performance, resulting in a positive offset from zero in the above plots. A zero or small offset, conversely, indicates that the available dynamic motion cues and static visual cues are sufficient to enable automatic spatial updating.

conditions with constant static visual information, i.e., without the in-between flashing of a different orientation as was done in the CONTROL condition.

14.3.3 Obligatory spatial updating

As in the previous experiment, the power of the available turn cues for turning our mental spatial representation was investigated by asking participants to ignore all turn stimuli and respond as if still being at the previous orientation. Only if spatial updating was indeed obligatory in the sense of being reflex-like and hard to suppress will IGNORE performance be considerably worse than UPDATE performance. The differences between IGNORE and UPDATE trials are graphically presented in Figure 34, the corresponding statistical analyses are compiled in Table 16. A first look at the figures shows that under all conditions, ignoring the turn stimuli was considerably harder than using them to UPDATE as usual. That is, spatial updating was always obligatory and reflex-like. There were, however, systematic differences between the different cue combinations (blocks) which will be elaborated upon as follows. Each subsection is concerned with answering one main question, with the order of the questions being the same as the order of the statistical tests in Table 16.

14.3.3.1 HMD versus blinders Using the HMD instead of blinders (block B vs. A) made the ignore task considerably easier and more accurate. All five dependent variables point towards this direction, suggesting that visual cues presented via HMD are easier to ignore and consequently less convincing in triggering obligatory spatial updating and turning the mental spatial representation against our own conscious decision. This provides first evidence that curved projection screens, even when viewed through a vision delimiting device, might be more suitable for presenting ego-motions and enabling good spatial orientation in virtual environments than HMDs .

14.3.3.2 Influence of FOV Comparing block E and F reveals that turns presented *without* blinders were considerably harder to ignore (increased response times) and resulted in an increased configuration error, compared to turns with blinders-restricted FOV. That is, the increased FOV seems to render spatial updating more obligatory and reflex-like. This effect was less pronounced for purely visual turns (blocks G vs. H). Smaller turns and a gain factor of 1, on the other hand, did not result in any consistent FOV effect (blocks B vs. C). Due to the limited number of participants in this study, only preliminary conclusions are justified, and further experiments are needed to corroborate the notion that turns presented via a larger FOV lead to more obligatory spatial updating.

14.3.3.3 Influence of visuo-vestibular gain factors and vestibular cues Comparing spatial updating for different gain factors led to no clear effect, suggesting that the presence of concurrent vestibular turn stimuli is not required as long as a consistent, landmark-rich visual scene is presented (blocks B vs. E, B vs. G, E vs. G, F vs. H, and B vs. J, see Table 16). This might be interpreted in the direction that "good" visual cues alone *can* be fully sufficient for initiating obligatory spatial updating and hence enabling good spatial orientation. Again, further experiments are needed to corroborate this effect. The subsequent Experiment (LANDMARKS VERSUS OPTIC FLOW) is designed to address this issue more directly.

14.3.3.4 Influence of turning angle and movement velocity Comparing blocks with different yaw ranges shows increased response times, configuration errors, absolute pointing errors and absolute ego-orientation errors for larger turns (blocks C vs. F, C vs. H, and J vs. F). Not all of these

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

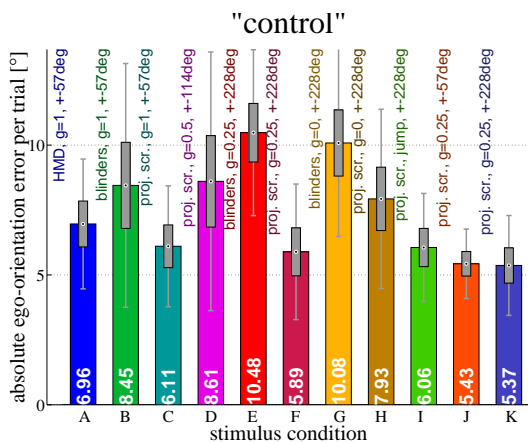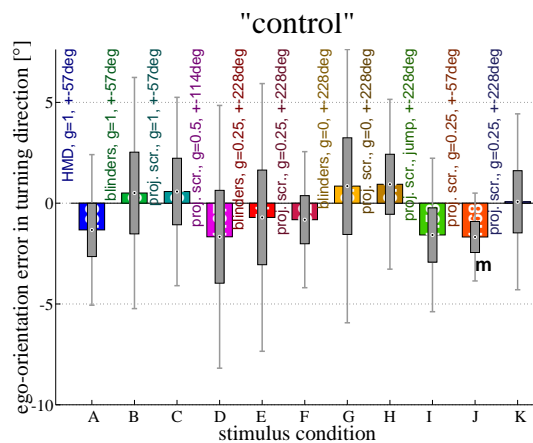(e) Ego-orientation error in turning direction
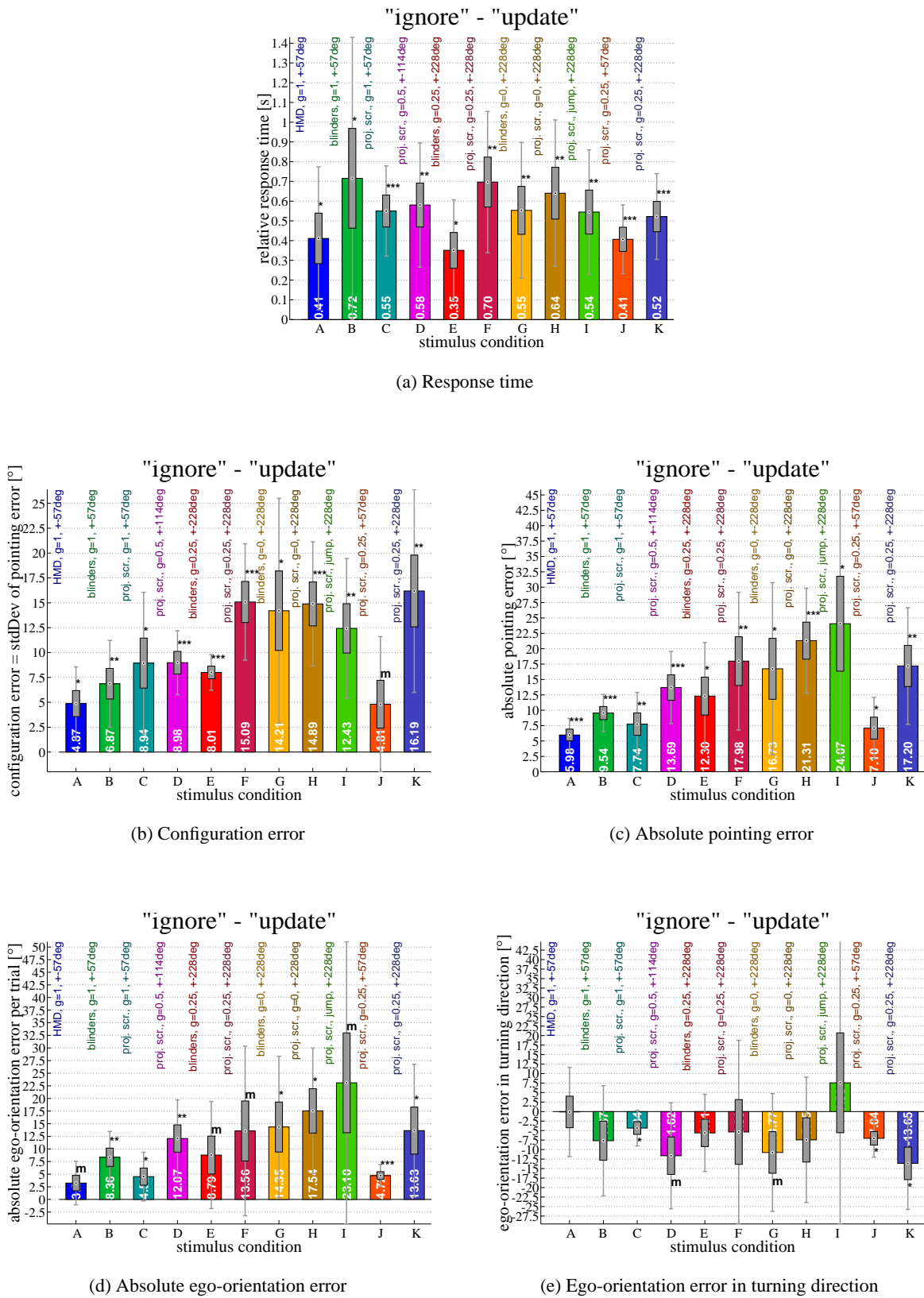
Figure 34: Obligatory spatial updating performance for experiment Simulation Parameters. The difference between Ignore and Update performance is plotted for the five dependent variables. Values above zero indicate that ignoring is considerably harder than updating, implying obligatory spatial updating.

differences reached statistical significance (see Table 16), which might of course be due to the limited number of participants. It nevertheless gives a first hint that the turning angle might be a critical variable, as larger turns seem to be considerably harder to ignore than smaller turns.

Comparing block J and K reveals considerable performance deficits in all five measurands for block K, where both turning angles and response times were four times larger than in block J, but the average movement duration was the same. Movement velocity is most likely not the cause of the difference, as block F and K differed only in movement velocity and showed virtually the same performance[7]. This corroborates the above argument that larger turning angles are indeed harder to ignore and hence lead to more obligatory spatial updating.

14.3.3.5   Continuous versus discontinuous (jump-like) motions and misjudged ego-orientations

Block I investigated spatial updating performance without any explicit or apparent motion cues whatsoever. That is, participants were immediately presented with a new view, similar to a slide-show type presentation. We have already seen that this jump-like visual presentation of new orientations yielded the same baseline performance and was as efficient in enabling automatic spatial updating as comparable conditions (e.g., block K). But will it be sufficient to trigger obligatory spatial updating as well?

Comparing IGNORE and UPDATE performance for block I and K reveals the same response time increase of more than 500ms for the IGNORE task, indicating that the IGNORE task is indeed quite hard and rather time-consuming. The configuration error was slightly less pronounced in the discontinuous (jump) condition, whereas the absolute pointing error as well as the absolute ego-orientation error were more pronounced. However, none of these differences reached statistical significance (see Table 16). Taken together, this implies that discontinuous, slide-show type presentations of new orientations are virtually as successful in triggering obligatory spatial updating as are continuous motions to the new orientation. This finding was rather unexpected and awaits further investigations to allow for a comprehensive explanation.

Analyzing the direction of the ego-orientation error (Figure 34 (e)) provides first insights into the underlying processes: The most pronounced difference between block I and K was indeed in terms of the signed ego-orientation error, which was more than 13° against the motion direction for the fast but continuous motions in block K, whereas the jump-like condition showed no such asymmetry (see also Figure 64). That is, only the continuous motions seem to be able to induce a consistent misjudgment of the previous, to-be-remembered ego-orientation in the direction opposite to the turning direction.

For the continuous motions in block J and K, where participants were highly trained on the task, this ego-orientation error against turning direction explains a considerable part of the absolute pointing errors and absolute ego-orientation errors. This implies that even though participants might have improved their overall pointing accuracy somewhat, they were nevertheless unable to correctly remember their previous orientation, especially when exposed to rapid and large turns. This occurred even though they had already been exposed to the IGNORE task for several hours and were highly trained with explicit feedback, as they were always moved back to the previous orientation in the following IGNORE BACKMOTION trial.

The most obvious hypothesis about the underlying processes would be that participants tried to remember the previous orientation by remembering the landmark they were facing. Many participants reported in fact using this strategy. The observed systematic ego-orientation error of about 13° was, however, almost as large as the mean angular separation of the targets (about 16°), indicating that

---

[7]The observed difference in response time is most likely due to the previously-mentioned overall improvement in response time over the course of the experiment.

participants' judged previous ego-orientation was closer to the wrong landmark (i.e., the landmark *adjacent* to the one faced) than to the correct one (i.e., the landmark previously faced). Hence, the underlying processes remain unclear. No matter what the underlying reasons are, however, the Ignore trials did show significant ego-orientation errors against turning direction in both this experiment and Experiment Real World versus VR. As the representational momentum literature has to our knowledge so far only been concerned with object or object array motions, the observed self-motion effect might be an interesting extension that could provide further insights into the underlying mechanisms and awaits further investigation.

14.3.4　Learning effect

Due to the exploratory nature of this experiment and limited number of participants, further analyses were confined to the learning effect. This was critical, as a potential overall learning or practice effect over the course of the experiment would make comparisons between the first eight blocks (which were pseudo-balanced among the participants) and the subsequent three control blocks (J-K) illegitimate. To quantify potential performance improvements over time within the first eight blocks of the experiment, correlation analyses were performed between all five dependent variables and the block order for the Update, Control, and Ignore conditions. Only the response time was significantly negatively correlated with block number. This was found for both the Update and Ignore conditions, but not the Control condition[8]. As the only consistent overall learning effect was a decrease in response time in the Update and Ignore conditions, comparisons between the first eight blocks and the three subsequent blocks seem legitimate for the other four dependent variables and for the Control condition in general.

To double-check the comparability, the last two blocks of the control study (i.e., the tenth and eleventh block (either J or K)) were also compared using 2-sided t-tests. The statistical analysis showed no significant effect of block order for any of the dependent variables ($p > 0.05$), suggesting that all order effects and learning curves had settled by that time.

14.4　General discussion and conclusions

The purpose of this experiment was to explore the influence of various motion and simulation parameters on baseline spatial orientation performance (Control condition), as well as automatic and obligatory spatial updating performance (Update and Ignore condition, respectively).

The baseline (Control) condition revealed only benefits for an increased FOV, which allowed for smaller configuration errors and pointing errors in general. Automatic spatial updating, however, was completely independent of all parameters varied. That is, as long as participants were presented with a photorealistic view onto the well-known Tübingen market place, they could readily adopt the new orientation and knew immediately where the currently invisible landmarks were. It did not matter how the scene was presented, how large the FOV was, or how far they were moved. Even the complete absence of any concurrent vestibular turn stimuli or the discontinuous, jump-like presentation of the new orientation did not prevent participants from being able to indicate quickly and accurately where the surrounding objects of interest were. We conclude that photo-realistic visual stimuli from a well-known environment including an abundance of salient landmarks are under a wide range of simulation and motion parameters sufficient to enable automatic spatial updating and hence allow us to turn the world inside our head, irrespective of vestibular cues.

---

[8] $r = -0.22, r^2 = 0.047, t(62) = 1.75, p = 0.042^*$ for the Update condition and $r = -0.23, r^2 = 0.051, t(62) = 1.82, p = 0.037^*$ for the Ignore condition.

This response pattern changed when participants were asked to ignore all turn cues and point as if not having turned. In all stimulus conditions, simulated movements were obligatory to the extent that they were considerably harder to IGNORE than to UPDATE. However, this reflex-like component of "turning the world inside our head" even against our conscious decision was much more pronounced for larger turns and when presented via a large FOV. Actually, all blocks with the same movement range and the same, large FOV (i.e., blocks F, G, H, and K) showed virtually the same performance. That is, if the FOV is large enough, other factors like movement velocity, gain factor, and even discontinuous (jump-like) versus continuous motions do not seem to influence spatial updating performance.

On the other hand, reducing the FOV via blinders led to decreased baseline performance and made the IGNORE task easier. This suggests on the one hand that for small FOVs, some critical information might be missing. On the other hand, turns presented through a rather limited FOV of $40° \times 30°$ seem to be less capable in inducing obligatory, reflex-like spatial updating. This is in accordance with findings from the previous experiment that showed a similar advantage for large FOVs in all five performance measures. For smaller turning angles, which were shown to be easier to ignore, the FOV effect was less pronounced, suggesting that a small FOV can be compensated for if the task is not too hard.

Apart form the well-known smooth spatial updating induced by *continuous* movement information, we found also a *discontinuous*, jump-like spatial updating that allowed participants to quickly adopt a new orientation without any explicit motion cues. These slide-show type presentations of new orientations were even sufficient in triggering obligatory, reflex-like spatial updating. This finding is to our knowledge totally unexpected and unprecedented in the literature.

In sum, the comparison between the large number of cue combinations (blocks) provides preliminary evidence that a large FOV offers considerable advantages for baseline spatial orientation performance and enables excellent automatic as well as obligatory spatial updating. A reduced FOV, especially when using an HMD, seems to render the visual motion stimulus less convincing, as the simulated motions were easier to ignore. This argues against the usage of HMDs and any display with rather limited FOVs for applications involving simulated movements of the observer.

Combining the results from this experiment with the findings from the previous experiment, we can conclude that photorealistic visual stimuli of consistent, landmark-rich scenes are clearly sufficient for enabling excellent automatic spatial updating as well as triggering obligatory spatial updating, especially when presented through a large FOV. This was found irrespective of concurrent vestibular turn cues. Even discontinuous, slide-show like presentation of new orientations was sufficient and yielded virtually the same, excellent performance. On the other hand, vestibular cues from smooth turns alone were clearly incapable of initiating obligatory spatial updating, as they could as easily be ignored as they could be used to deliberately update to the new orientations.

This result conflicts with the prevailing opinion that vestibular cues are required or even sufficient for proper updating of ego-turns. Several factors might explain this difference, primarily the immersiveness of our visualization setup and the abundance of natural landmarks in a well-known, consistent environment. Furthermore, the smooth consecutive motions used might have reduced the salience of the vestibular stimuli and rendered them virtually irrelevant. Using higher accelerations or jerky motions might increase the salience of the vestibular cues to a level where they are required or even sufficient for obligatory spatial updating. A different method would be to reduce the salience of the visual cues to a level comparable to the vestibular cues, which have only velocity and/or acceleration information to rely upon. The subsequent Experiment (LANDMARKS VERSUS OPTIC FLOW) pursues this approach by removing all landmarks and presenting optic flow information only, thus reducing the visual cues to mere velocity information.

CONTROL		response time		configuration error		absolute pointing error		absolute ego-orientation error		ego-orientat. error in turn direction	
		t(7)	p	t(7)	p	t(7)	p	t(7)	p	t(7)	p
HMD vs. blinders	A vs. B	-1.18	0.28	0.0362	0.97	-0.949	0.37	-0.891	0.4	-0.897	0.4
FOV effect	B vs. C	0.691	0.51	0.747	0.48	1.15	0.29	1.38	0.21	-0.0337	0.97
	E vs. F	-1.94	0.093m	1.88	0.1	3.29	0.013*	3.61	0.0086*	0.0376	0.97
	G vs. H	-0.11	0.92	0.615	0.56	1.39	0.21	1.12	0.3	-0.0339	0.97
gain factor & turn angle (blinders)	B vs. E	1.49	0.18	-0.251	0.81	-0.873	0.41	-1.46	0.19	0.371	0.72
	B vs. G	0.901	0.4	-0.678	0.52	-0.406	0.7	-1.02	0.34	-0.107	0.92
gain factor & turn angle (proj. scr.)	C vs. F	-0.0483	0.96	1.17	0.28	1.2	0.27	0.141	0.89	0.588	0.58
	C vs. H	-0.396	0.7	-0.266	0.8	0.0216	0.98	-1.13	0.29	-0.162	0.88
just gain factor	E vs. G	-0.829	0.43	-0.583	0.58	0.229	0.83	0.245	0.81	-0.413	0.69
	F vs. H	-0.289	0.78	-0.856	0.42	-1.19	0.27	-2.98	0.02*	-1.15	0.29
	B vs. J$	2.28	0.057m	1.13	0.29	1.73	0.13	1.71	0.13	0.896	0.4
just turn angle	F vs. J$	2.79	0.027*	-0.306	0.77	-0.119	0.91	0.383	0.71	0.649	0.54
turn velocity	F vs. K$	1.19	0.27	0.0571	0.96	0.187	0.86	0.468	0.65	-0.383	0.71
turn vel. & turn angle	J vs. K	-1.16	0.28	1.21	0.27	0.731	0.49	0.0684	0.95	-1.04	0.33
turn angle & jump	I vs. J	1.11	0.31	0.42	0.69	0.302	0.77	0.687	0.51	0.0666	0.95
jump vs. cont.	I vs. K	-0.383	0.71	0.564	0.59	0.543	0.6	0.937	0.38	-0.634	0.55

Table 14: Tabular overview of the paired two-tailed t-test for the different comparisons for the baseline (CONTROL) condition in experiment SIMULATION PARAMETERS. Block comparisons marked with a '$' are comparisons between one of the first eight conditions and one of the three control conditions performed afterwards. Hence, potential order or learning effects cannot be fully excluded, even though those effects did not reach statistical significance (see subsection 14.3.4). Note that the only consistent effect was a small benefit for an increased FOV (E vs. F).

UPDATE - CONTROL		response time		configuration error		absolute pointing error		absolute ego-orientation error		ego-orientat. error in turn direction	
		t(7)	p	t(7)	p	t(7)	p	t(7)	p	t(7)	p
HMD vs. blinders	A vs. B	0.491	0.64	-0.775	0.46	0.952	0.37	2.06	0.079m	1.1	0.31
FOV effect	B vs. C	0.109	0.92	-0.0125	0.99	-0.523	0.62	-1	0.35	0.623	0.55
	E vs. F	-0.247	0.81	0.409	0.69	-0.754	0.48	-0.865	0.42	-0.244	0.81
	G vs. H	0.432	0.68	0.367	0.72	-0.0324	0.98	0.76	0.47	-0.445	0.67
gain factor & turn angle (blinders)	B vs. E	1.5	0.18	-0.634	0.55	-0.396	0.7	0.135	0.9	-0.406	0.7
	B vs. G	0.0674	0.95	-0.432	0.68	-0.914	0.39	-0.811	0.44	-0.192	0.85
gain factor & turn angle (proj. scr.)	C vs. F	1.69	0.14	-0.667	0.53	-0.777	0.46	0.243	0.81	-1.86	0.11
	C vs. H	0.376	0.72	0.182	0.86	-0.0176	0.99	1.31	0.23	-1.58	0.16
just gain factor	E vs. G	-1.66	0.14	0.318	0.76	-0.313	0.76	-0.85	0.42	0.251	0.81
	F vs. H	-0.484	0.64	0.983	0.36	0.802	0.45	1.85	0.11	0.0632	0.95
	B vs. J$	0.655	0.53	-0.743	0.48	-0.873	0.41	-0.541	0.61	-0.744	0.48
just turn angle	F vs. J$	-0.351	0.74	-0.215	0.84	0.761	0.47	0.651	0.54	0.0342	0.97
turn velocity	F vs. K$	1.34	0.22	0.0548	0.96	-0.439	0.67	-0.569	0.59	0.019	0.99
turn vel. & turn angle	J vs. K	1.26	0.25	0.326	0.75	-2.24	0.06m	-1.3	0.23	-0.0062	1
turn angle & jump	I vs. J	-0.517	0.62	-0.626	0.55	1.07	0.32	0.786	0.46	0.0718	0.94
jump vs. cont.	I vs. K	1.34	0.22	-0.602	0.57	-0.945	0.38	-0.621	0.55	0.0451	0.97

Table 15: Tabular overview of the paired two-tailed t-test for automatic spatial updating performance in experiment SIMULATION PARAMETERS. Note that there were no significant differences in terms of automatic spatial updating whatsoever.

IGNORE - UPDATE		response time		configuration error		absolute pointing error		absolute ego-orientation error		ego-orientat. error in turn direction	
		t(7)	p	t(7)	p	t(7)	p	t(7)	p	t(7)	p
HMD vs. blinders	A vs. B	-2.28	0.057m	-0.852	0.42	-2.01	0.084m	-3.03	0.019*	1.3	0.24
FOV effect	B vs. C	0.672	0.52	-0.87	0.41	0.978	0.36	1.5	0.18	-0.567	0.59
	E vs. F	-3.28	0.014*	-3.44	0.011*	-1.32	0.23	-0.912	0.39	-0.0269	0.98
	G vs. H	-1.8	0.12	-0.247	0.81	-0.652	0.54	-0.384	0.71	-0.396	0.7
gain factor & turn angle (blinders)	B vs. E	1.8	0.12	-0.658	0.53	-0.85	0.42	-0.0831	0.94	-0.452	0.66
	B vs. G	1.09	0.31	-1.45	0.19	-1.27	0.25	-1.11	0.31	0.461	0.66
gain factor & turn angle (proj. scr.)	C vs. F	-2.05	0.08m	-2	0.086m	-2.8	0.026*	-1.72	0.13	0.131	0.9
	C vs. H	-0.856	0.42	-1.56	0.16	-4.12	0.0045**	-3.19	0.015*	0.473	0.65
just gain factor	E vs. G	-2.77	0.028*	-1.52	0.17	-0.653	0.53	-0.752	0.48	0.947	0.38
	F vs. H	0.655	0.53	0.0553	0.96	-0.658	0.53	-0.492	0.64	0.189	0.86
	B vs. J$	1.4	0.2	0.663	0.53	1.07	0.32	1.88	0.1	-0.123	0.91
just turn angle	F vs. J$	1.89	0.1	2.7	0.031*	2.2	0.064m	1.48	0.18	0.179	0.86
turn velocity	F vs. K$	2.08	0.076m	-0.285	0.78	0.171	0.87	-0.0101	0.99	0.851	0.42
turn vel. & turn angle	J vs. K	-1.21	0.26	-2.88	0.024*	-3.05	0.019*	-1.95	0.093m	1.34	0.22
turn angle & jump	I vs. J	1.39	0.21	2.19	0.065m	2.4	0.047*	1.81	0.11	1.07	0.32
jump vs. cont.	I vs. K	0.352	0.74	-1.17	0.28	0.978	0.36	1.09	0.31	1.49	0.18

Table 16: Tabular overview of the paired two-tailed t-test for obligatory spatial updating performance in experiment SIMULATION PARAMETERS. Note that the only consistent effects were in terms of presentation device (HMD versus blinders), FOV, and turning angle.

15 Experiment 8: "LANDMARKS VERSUS OPTIC FLOW"

15.1 Introduction

The previous experiments demonstrated that continuously visible landmarks (either from the surrounding Motion-Lab or a replica of the Tübingen market place) alone are under a wide range of display and motion parameters clearly sufficient for enabling automatic spatial updating as well as eliciting obligatory spatial updating. That is, participants were unable to successfully ignore or suppress the visual rotation of the scene. Concurrent vestibular motion cues showed little, if any, effect. Only a large FOV and turning angles that extend well beyond the FOV had a clear effect, by rendering the visual motion stimulus harder to ignore and thus more obligatory. That is, spatial updating of the surround induced by large visual turns presented via a large FOV was more reflex-like and cognitively less penetrable.

The previous experiments, however, left two critical issues unresolved that will be addressed in the current experiment:

1. First, what empowers the visual cues to trigger obligatory spatial updating and enable automatic spatial updating? Is it the landmarks forming a consistent, well-known scene, or merely the visual motion stimulus, similar to the vection induced by a rotating optic drum? To tackle this issue, we compared spatial updating induced by a rotation of a consistent visual scene (LANDMARKS condition) to spatial updating induced by a mere rotating optic flow pattern (OPTIC FLOW condition, where only optic flow was visible during the motion and pointing phase).

   If the mere visual rotation stimulus is sufficient, OPTIC FLOW performance should be comparable to LANDMARKS performance, at least in terms of response time and configuration error. The other pointing error measures are expected to increase somewhat due to path integration errors and the lack of reliable landmarks usable for position fixing.

   If, on the other hand, the visual motion stimulus is rather irrelevant, and the visibility of the visual scene alone is the critical issue, we expect OPTIC FLOW performance to decrease in all dependent measures. If it is mainly the static visibility of the visual scene that aligns the mental reference frame, we expect a performance decrease for the OPTIC FLOW condition even in the supposedly simple baseline (CONTROL) condition.

   The excellent performance in the jump condition of the previous experiment suggests that the latter might be the case, whereas the vection literature might predict at least some contribution of the visual motion stimulus (see below).

2. Under which condition might the vestibular motion cues have a relevant contribution to automatic or obligatory spatial updating?

   The previous experiments have shown that vestibular cues did not play a significant role when landmark information was continuously available. This might be explained by the visual cues (based mainly on landmarks, i.e., piloting) being much more reliable than the vestibular cues (being based on path integration by integrating the sensed acceleration or velocity only). Reducing visual strategies to path integration by presenting only optic flow information during the motion and pointing phase might render the fidelity and reliability of the visual and vestibular system more comparable. This might reduce the visual dominance observed before and lead to a significant contribution of vestibular cues for spatial updating. This will be investigated by presenting only visual turn cues in half of the trials, and comparing performance to the other half of the trials that included additional vestibular cues from physical turns.

So what happens if all visual reference objects (landmarks) are removed during the motion and pointing phase and replaced by optic flow stimuli? That is, if spatial updating via piloting is no longer feasible and participants are bound to using path integration?

A large body of data suggests that many lower organisms are capable of performing path integration from nonvisual cues, often with amazing accuracy (Etienne, Maurer, & Séguinot, 1996; Gallistel, 1990; Maurer & Séguinot, 1995; Mittelstaedt & Mittelstaedt, 1982; Müller & Wehner, 1988; Seguinot, Maurer, & Etienne, 1993; Wehner, Michel, & Antonsen, 1996). Studies on human path integration using blindfolded walking demonstrated considerable systematic as well as random errors for all tasks that are more complex than simple translations or rotations. Performance, however, was well above chance (see Klatzky et al. (1997) and Loomis et al. (1999) for a review). Studies on path integration using just visual cues suggest that humans are able to extract angles turned and distances traveled from optic flow information. Performance, however, seems to be highly dependent on the display device used, especially for more complex navigation tasks (see, Bakker et al. (1999, 2001), Beall & Loomis (1997), Bremmer & Lappe (1999), Kearns et al. (2002), Loomis & Beall (1998), Péruch et al. (1997), Riecke (1998), Schulte-Pelkum et al. (2002), Warren et al. (2001), Warren & Wertheim (1990), Wartenberg et al. (1998) and part II). Spatial updating studies using optic flow information are rather sparse, and suggest that optic flow might be sufficient to enable automatic or even obligatory spatial updating for translations (see May & Klatzky (2000) and subsection 12.2.4 for a discussion of their results). For rotations, however, optic flow information, at least when presented via HMD, seems to be insufficient for spatial updating, and physical rotations seem to be required (Klatzky et al., 1998). In a control condition, Klatzky et al. disoriented participants before the experiment, such that they no longer knew their physical orientation in the experimental room. This procedure improved the updating of visually presented turns considerably, suggesting that the knowledge about the egocentric reference frame of the room interferes with spatial updating by optic flow.

Apart from the few studies on visually induced spatial updating, there is a long tradition of investigating vection (visually induced illusory ego-motion), which might possibly be closely linked to spatial updating. It has long been known that a rotating optic flow stimulus is capable of inducing the sensation of self-rotation (circular vection) (Fischer & Kornmüller, 1930; Mach, 1922), see, e.g., Dichgans & Brandt (1978), Warren & Wertheim (1990), Wertheim (1994b, 1994a) for an introduction). Both the FOV and the region of retinal stimulation (foveal or peripheral) were found to affect this illusory self-motion, and their respective importance seems to be determined by a number of factors including the frequency content of the visual stimulus. See Palmisano & Gillam (1998) and Wolpert (1990) for a review on this issue. Quantification methods include subjective reports (e.g., about the onset of vection) as well as behavioral measures (like body sway). The percept of circular vection, however, seems to never occur immediately with the stimulus onset, but rather gradually after a variable latency of 4-6 seconds (Wertheim, 1994b). Eventually, the observer perceives herself as moving and the visual stimulus as stationary. This self-motion perception might, however, be intermittently interrupted by sudden "drop outs", where the percept switches to a stationary self and a moving visual stimulus. See, e.g., Lathan, Wall, & Harris (1995), Mergner & Becker (1990), Mergner & Rosemeier (1998), Mergner & Glasauer (1999), Schweigart, Mergner, Evdokimidis, Morand, & Becker (1997) for explanations of this behavior. Even though vection seems to be closely related or even a prerequisite for spatial updating, the vection onset latency and the possible drop outs suggest that optic flow alone might not be sufficient to immediately and reliably induce obligatory spatial updating.

Here, we ask on the one hand whether optic flow with or without concurrent vestibular turn cues *can* be used for spatial updating, that is whether optic flow is sufficient for enabling **automatic spatial updating**. This would imply that UPDATE and CONTROL performance are comparable, at least in terms of consistency and response time.

One the other hand, the IGNORE conditions were used to reveal whether optic flow *must* be used, that is, whether optic flow (with or without concurrent vestibular turn cues) is capable of triggering **obligatory spatial updating**. The literature and the experiments reported in part II suggest that optic flow might be feasible for enabling automatic spatial updating, but not for initiating obligatory spatial updating, as participants tend to make qualitative errors, especially when being distracted or asked to respond quickly (see (Klatzky et al., 1998) and subsection 12.1).

15.2   Methods

In general, the methods and simulation parameters used in this experiment were similar to the previous experiment. To get the clearest possible differentiation between the different conditions, we used rather fast visual motions ($80°/s$), the full FOV of the projection screen ($84°\times63°$), and a large yaw range of $[-228°, +228°]$, just as in block K of the previous experiment (proj. scr., g=1/4, $\pm$228deg). The differences to block K of the previous experiment are described below. All other parameters remained unchanged.

15.2.1   Participants

A group of 13 female and 4 male naive participants took part in Experiment LANDMARKS VERSUS OPTIC FLOW. Ages ranged from 15 to 45 years (mean: 25.7 ± 0.4 years, SD: 7.4 years). All participants had normal or corrected-to-normal vision and no signs of vestibular dysfunction.

15.2.2   Stimuli and apparatus

(a) Photorealistic model of the Tübingen market place, created by wrapping a 360° roundshot photograph onto a cylinder. This creates an undistorted view for the observer positioned in the center of the cylinder.
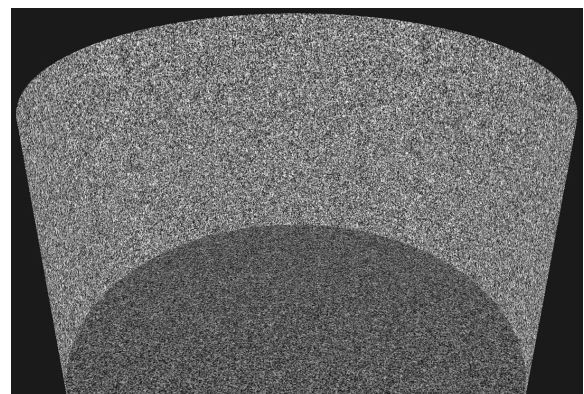
(b) Optic flow stimulus used to remove all landmark information during the motion and pointing phase in half of the trials. The stimulus consisted of a purpose-generated greyscale fractal texture.

Figure 35: Visual stimuli used for Experiment LANDMARKS VERSUS OPTIC FLOW.

The scenery used in the LANDMARKS condition was the same photorealistic replica from the Tübingen market place as in the previous experiment (see Figure 35 (a)). In the OPTIC FLOW condition, all landmarks were removed during the motion and pointing phase and replaced by a simple optic flow display (see Figure 35 (b)). After each pointing phase, the corresponding view of the Tübingen market place became again visible for several seconds until the next trial started. This allowed participants to re-anchor to the correct orientation before the next trial. Participants never knew in advance whether the next motion would be with landmarks or just with optic flow. This ensured that participants could not prepare differently for the next trial or utilize condition-dependent strategies. Visualization and pointing were the same as in the previous experiment.

15.2.3 Procedure

As for all experiments described in this thesis, a repeated-measures, within-subject design was used, which is summarized in Table 17 and 18. Each participant completed three sessions that were identical apart from the quasi-randomization of turning angles and cue combinations. All three sessions were performed on the same day, with intermittent breaks to avoid fatigue effects and obviate the influence of declining alertness. Each session consisted of 42 trials and lasted about 18 minutes. As several participants showed a reduced performance for the first one or two trials of each session, the first two trials (which were always UPDATE trials) were removed from each session and for all participants. The remaining 40 trials were split up into 16 UPDATE trails, and 8 trials for each of the other spatial updating conditions (CONTROL, IGNORE, and IGNORE BACKMOTION), see Table 17.

Spatial updating condition	Visual turning angle α	Trials per cue combination (A, B, C, & D) and session	Trials per session	Trials altogether
1. UPDATE	$80° \leq |\alpha| \leq 456°$	4	16	48
2. CONTROL	$80° \leq |\alpha| \leq 114°$	2	8	24
3. IGNORE	$80° \leq |\alpha| \leq 228°$	2	8	24
4. IGNORE BACKMOTION	$80° \leq |\alpha| \leq 228°$	2	8	24

Table 17: Summary of the four different spatial updating conditions used in Experiment LANDMARKS VERSUS OPTIC FLOW.

Stimulus condition	Visual cues	Vestibular turn cues
A	LANDMARKS	yes (PLATFORM ON, $g = 1/4$)
B	LANDMARKS	no (PLATFORM OFF, $g = 0$)
C	OPTIC FLOW	yes (PLATFORM ON, $g = 1/4$)
D	OPTIC FLOW	no (PLATFORM OFF, $g = 0$)

Table 18: Summary of the four different stimulus conditions used in Experiment LANDMARKS VERSUS OPTIC FLOW.

All four cue combinations (A-D) were used within each session. To avoid clustering, the order of the spatial updating conditions and vestibular conditions was fixed to the sequence CONTROL → UPDATE → IGNORE → IGNORE BACKMOTION → UPDATE for the conditions with concurrent vestibular turn cues (g=1/4) and afterwards the same sequence for the conditions without vestibular turn cues (g=0). This sequence was repeated four times, as is depicted for one exemplary session in Figure 36. The OPTIC FLOW and LANDMARKS conditions were quasi-randomized within each session, to yield an equal number of trials in all four conditions (A-D, see Table 17). In the landmark
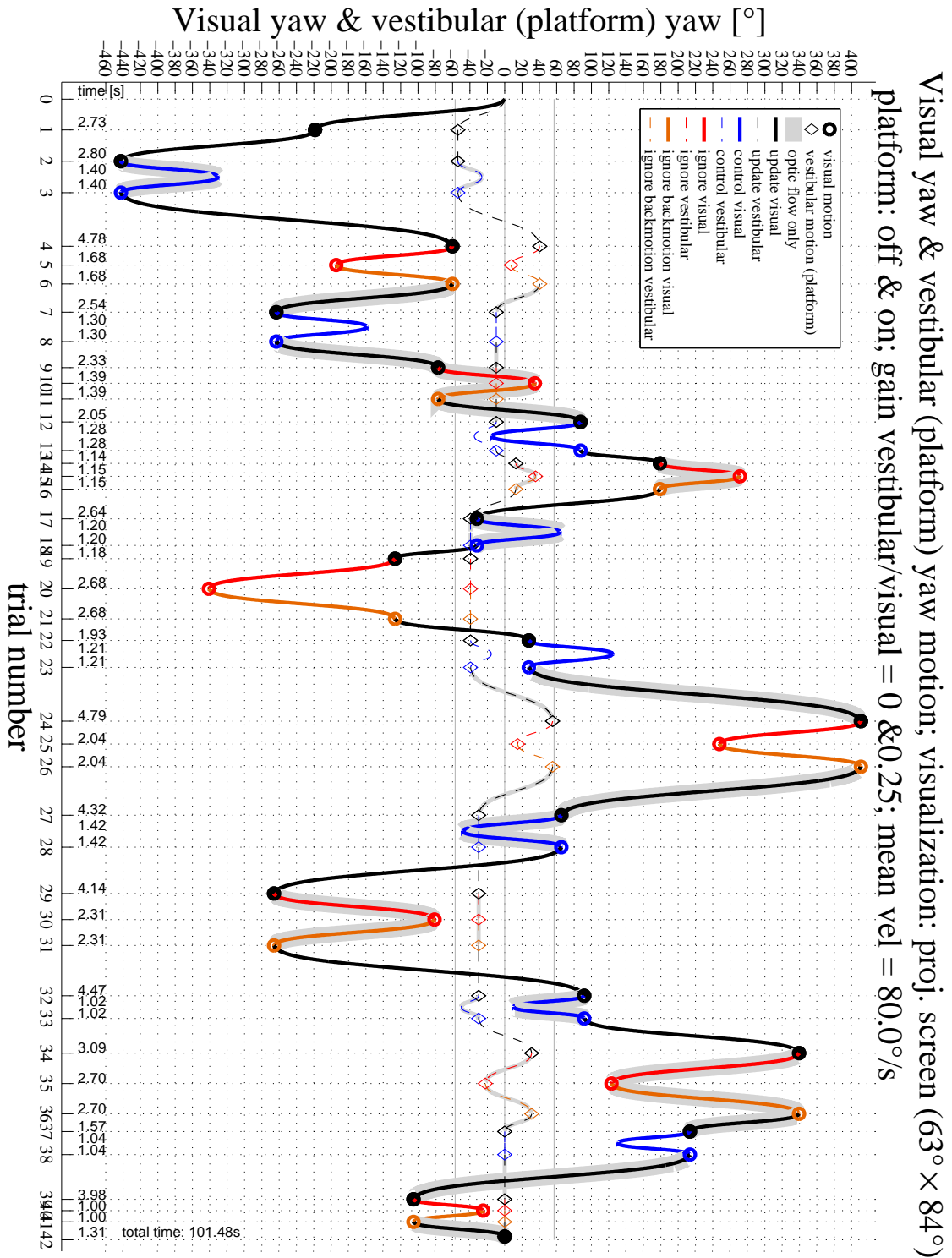
Figure 36: Example of the vestibular (platform) motion and visual motion for one of the three sessions of Experiment LANDMARKS VERSUS OPTIC FLOW. Depicted are the vestibular (platform) and visual yaw angle, demonstrating the sequence of the spatial updating conditions (CONTROL → UPDATE → IGNORE → IGNORE BACKMOTION → UPDATE etc.), cue combinations, and the effect of the gain factor 1/4 and 0. Trajectories highlighted in gray were OPTIC FLOW trials, the other ones LANDMARKS trials. Pointings occurred at all circles and diamonds of the trajectory.

pre-test, participants were able to name between 6 and 22 landmarks on the Tübingen market place (mean: 13.3), indicating that they were already quite familiar with the environment used.

15.2.4   Data analysis - periodicity correction

The data analysis was essentially the same as in the previous experiment, apart from an angular ambiguity correction due to the Optic Flow condition. In the Update trials of the Optic Flow condition, some participants consistently over- or underestimated the angle turned. One participant, for example, always responded as if the angle turned was 1.5 times larger than it actually was (see Figure 37). That is, he showed a considerable ego-orientation error in turning direction. For a 100° turn, for example, he responded as if the turn was 150°, that is, he showed a consistent ego-orientation error of +50° in turning direction. For a 200° turn, the ego-orientation error was approximately +100°. For a 400° turn, the ego-orientation error was +200°, which might also be seen as 200°-360° = -160°. This ambiguity due to the 360° periodicity of angles can be removed if one assumes that each participant has a rather consistent over- or underestimation of the angle turned. For smaller turning angles, where the periodicity problem is not yet present, participants showed indeed a rather consistent gain factor $g = \alpha_{perc}/\alpha_{corr}$ between perceived and correct turning angle, which was 1.5 for the above-mentioned participant.
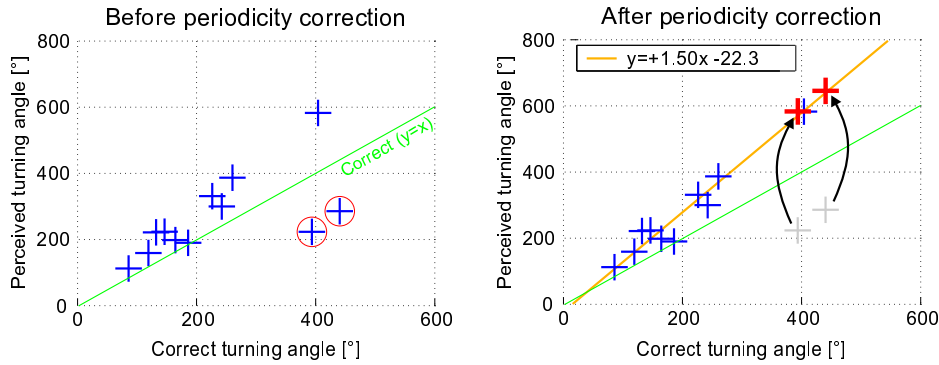
Figure 37: Example illustrating the periodicity correction for the Update trials of participant *ebhc* in condition D (Optic Flow, Platform Off).

To reliably eliminate those periodicity ambiguities, the following algorithm was applied to all Update trials. For each cue combination and participant, there were 12 repetitions in the Update condition, with pseudo-randomized turning angles. The main idea of the periodicity correction algorithm was to resolve the 360° ambiguity by selecting those perceived turning angles $\alpha'_{n_{perc}} = \alpha_{n_{perc}} \pm m \cdot 360°, (m = 0, 1, 2, \ldots)$ that minimized the variance of the gain factors $g = \alpha_{perc}/\alpha_{corr}$ for those 12 Update trials. This algorithm is illustrated for one representative participant in Figure 37. That is, the algorithm assumes a consistent gain factor and selects the angles correspondingly. As the 360° ambiguity was only present for larger turns (>180°), the gain factors $g_n = \alpha_{n_{perc}}/\alpha_{n_{corr}}$ for each of the $n = 1 \ldots 5$ lowest turning angles were accepted as they were. The remaining trials were then sorted in order of ascending absolute turning angles $|\alpha_{corr}|$. For the next-larger turning angle $\alpha_{n+1_{corr}}$, values of 360° were added or subtracted to $\alpha_{n+1_{perc}}$ until the variance over the first $n + 1$ gain factors $Var(g_{i=1\ldots n+1})$ was minimal. This criterion was applied for each of the further turning angles. In this manner, the ego-orientation errors of a total of 15 trials or 3.7% of the 408 Update trials in the two Optic Flow conditions were corrected. This algorithm implies that the ego-orientation error for single trials could be larger than 180° if the gain factor was sufficiently different from one.

The absolute pointing errors, however, were not corrected, as this would be inconsistent with the prevailing usage of absolute pointing errors. Consequently, the absolute ego-orientation errors and ego-orientation errors in turning direction could end up being larger than the absolute pointing errors, as only the ego-orientation errors were allowed to be larger than 180°.

15.3 Results and discussion

To give a first impression of the results, the data for condition A (LANDMARKS, PLATFORM ON) are displayed in Figure 38. Figures 38 (a) - (d) demonstrate as expected the typical response pattern for spatial updating: UPDATE and IGNORE BACKMOTION performance are almost as good as baseline CONTROL performance, whereas IGNORE performance is considerably worse. Compared to block K of Experiment SIMULATION PARAMETERS, which had essentially the same stimulus parameters, performance is generally decreased, however: Response times are increased by approximately 150ms, configuration errors and absolute pointing errors are increased by 4-6°. The extended exposure to the task in block K of the previous experiment or differences in the participant population might both have contributed to this unexpected difference, as might have the intermittent OPTIC FLOW trials in this experiment.

As in the previous experiments, we will continue by presenting first the baseline (CONTROL) performance (subsection 15.3.1), followed by the analysis of automatic and obligatory spatial updating in subsections 15.3.2 and 15.3.3, respectively. The details of the statistical analysis are compiled in Figure 19, and the full data set is presented in Figures 63 and 64 for reference.

15.3.1 Baseline (CONTROL) performance

Pointing results for the baseline forth-and-back motion are summarized in Figure 39. CONTROL performance for both LANDMARKS conditions (A&B) were equally good and virtually indistinguishable, indicating that the absence or presence of concurrent vestibular motion stimuli was completely irrelevant for the task. Removing all landmark information during the motion and pointing phase in the OPTIC FLOW conditions (C&D) impaired baseline performance consistently in all dependent variables: Not only was the task harder (indicated by the response time increase of approximately 100ms), but also information for accurate and consistent pointing seemed to be lacking (increased absolute pointing error and configuration error, respectively). Interestingly enough, virtually the whole increase in absolute pointing error might be explained by a corresponding large increase in absolute ego-orientation error. That is, even for the simple forth-and-back motions of the CONTROL condition, participants experienced considerable difficulties in remembering the correct previous orientation. The absolute ego-orientation error was even larger than the mean angular distance between the targets ($\approx 16°$), indicating that participants' sense of ego-orientation was severely impaired due to the optic flow stimulus. This ego-orientation error was slightly direction-specific only in the OPTIC FLOW, PLATFORM OFF condition (see Figure 39 (e)), and was in the same order and magnitude as in the visual conditions of Experiment REAL WORLD VERSUS VR, whereas Experiment SIMULATION PARAMETERS revealed no such direction-specific effect.

Comparing the two OPTIC FLOW conditions reveals a small but significant benefit for the PLATFORM OFF condition (D) in terms of configuration error. This might make sense, as participants' task was to point after the forth-and-back motion from the same orientation as before, and in condition D only visual cues indicated any motion at all. The additional vestibular motion cues in condition C apparently disrupted the consistency of the mental spatial representation slightly, indicated by the small but consistent increase in configuration error.

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
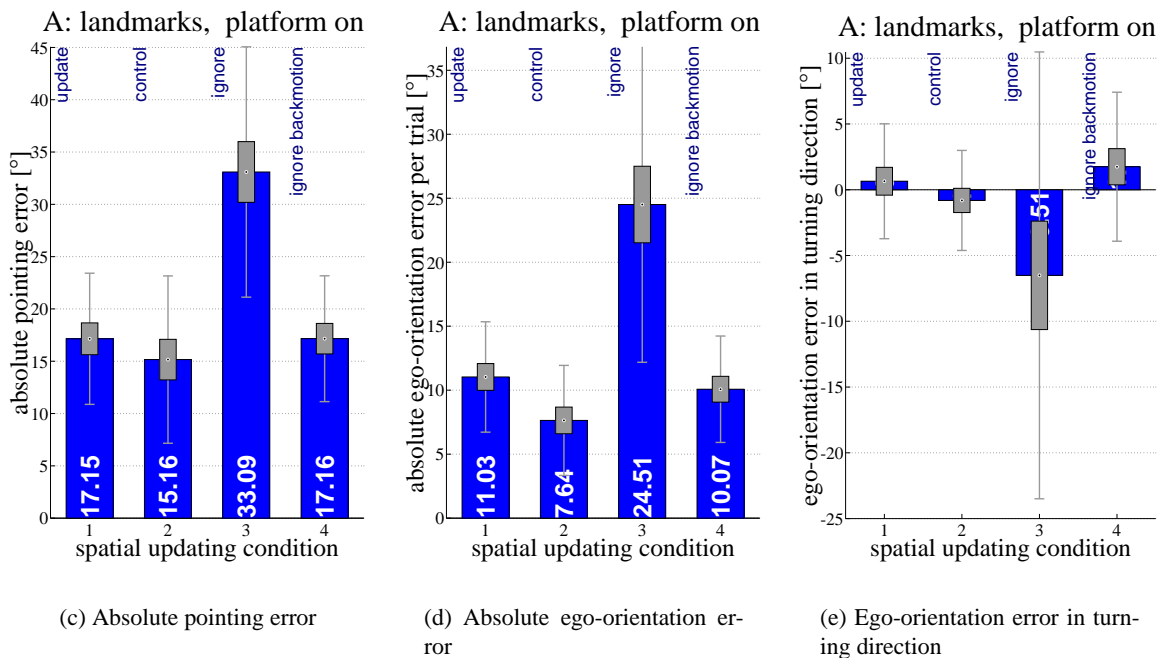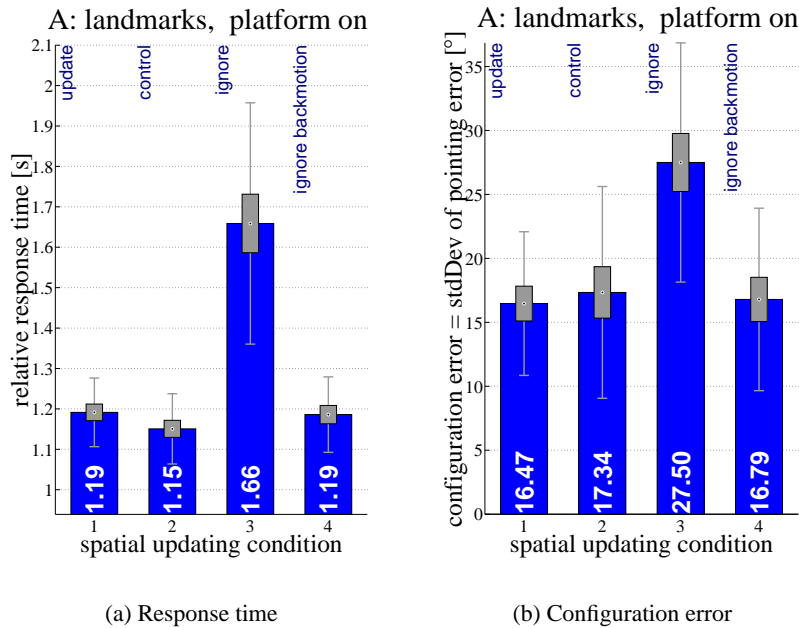
Figure 38: Pointing performance in Experiment LANDMARKS VERSUS OPTIC FLOW showing the typical spatial updating pattern. Performance in condition A (landmarks, platform on) is plotted for the five dependent variables, each for the four different spatial updating conditions.

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction
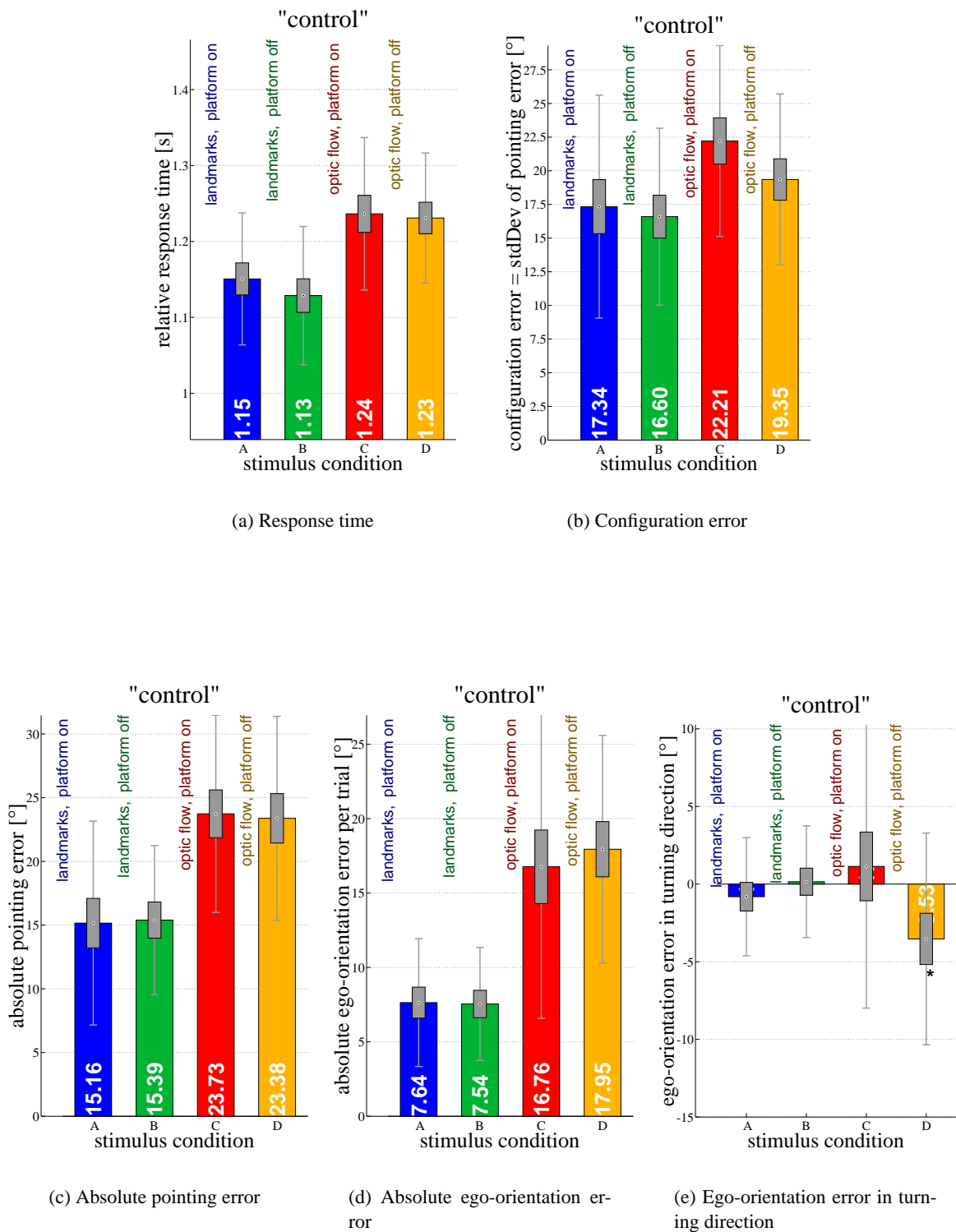
Figure 39: Baseline spatial updating performance in Experiment LANDMARKS VERSUS OPTIC FLOW. Baseline (CONTROL) performance is plotted for the five dependent variables for the four different cue combinations. Note the overall performance decrease in the OPTIC FLOW trials even for the relatively simple baseline (CONTROL) condition.

Conditions compared		Response time		Configuration error		Absolute pointing error		Absolute ego-orientation error		Ego-orientat. error in turn direction	
		t(16)	p	t(16)	p	t(16)	p	t(16)	p	t(16)	p
CONTROL											
influence of vestibular cues	A vs. B	0.819	0.43	0.763	0.46	-0.253	0.8	0.0955	0.93	-1.02	0.32
	C vs. D	0.19	0.85	2.71	0.015*	0.227	0.82	-0.401	0.69	2.14	0.048*
landmarks vs. optic flow	A vs. C	-2.66	0.017*	-2.27	0.037*	-4.26	0.0006**	-4.12	0.00079**	-0.791	0.44
	B vs. D	-3	0.0086*	-1.91	0.074m	-5.12	0.0001***	-5.76	2.9e-05***	1.77	0.096m
UPDATE - CONTROL											
influence of vestibular cues	A vs. B	0.0935	0.93	0.000887	1	-0.0581	0.95	-0.438	0.67	-0.742	0.47
	C vs. D	-1.27	0.22	-3.49	0.003**	0.529	0.6	0.981	0.34	2.2	0.043*
landmarks vs. optic flow	A vs. C	-3.68	0.002**	-0.747	0.47	-5.81	2.6e-05***	-5.59	4.1e-05***	-6.61	6e-06***
	B vs. D	-4.01	0.001**	-3.5	0.0029**	-7.15	2.3e-06***	-6.77	4.5e-06***	-3.7	0.0019**
IGNORE - UPDATE											
influence of vestibular cues	A vs. B	-1.3	0.21	-1.59	0.13	-0.0807	0.94	0.302	0.77	0.918	0.37
	C vs. D	0.923	0.37	1.13	0.28	-1.47	0.16	-1.65	0.12	-2.09	0.053m
landmarks vs. optic flow	A vs. C	5.64	3.7e-05***	4.18	0.0007**	6.85	3.9e-06***	6.69	5.2e-06***	6.35	9.6e-06***
	B vs. D	5.75	3e-05***	6.87	3.8e-06***	6.35	9.6e-06***	6.14	1.4e-05***	2.7	0.016*

Table 19: Tabular overview of the paired two-tailed t-tests for the different comparisons.

To sum up, landmark cues allowed for good baseline performance, whereas OPTIC FLOW performance was consistently decreased. This decrease in the supposedly simple baseline task suggests that it is not so much the knowledge about the correct orientation but the static visibility of the visual scene that aligns the mental reference frame properly and allows for optimal pointing performance. Optic flow information, however, still allowed for decent baseline performance far from chance. Additional vestibular motion cues proved irrelevant for the LANDMARKS condition, and even decreased pointing consistency slightly in the OPTIC FLOW condition.

15.3.2 Automatic spatial updating

The comparison of UPDATE and CONTROL performance in this subsection reveals whether the available spatial cues *can* be used for automatic spatial updating to new locations if participants are asked to do so. Figure 40 summarizes this automatic spatial updating performance. As before, both LANDMARKS conditions show virtually the same excellent updating performance, irrespective of vestibular cues: Response times were increased by less than 40ms, indicating that automatic spatial updating to new orientations was almost as easy as baseline performance. The configuration error remained unchanged, indicating that the consistency of the mental spatial representation did not suffer from the fast turns. The small increase in absolute pointing error of roughly 2° was probably caused by the increase in absolute ego-orientation error of roughly 4°.

For the OPTIC FLOW conditions, the response pattern changes drastically: Response times are increased by more than 200ms, indicating that automatic spatial updating was impaired without landmarks. Furthermore, both absolute pointing and ego-orientation error are increased by more than 30°. This indicates that participants were rather uncertain about the angle just turned and their new orientation. Due to the lack of landmarks, participants were forced to use path integration to estimate the angle turned, which apparently resulted in considerable misestimations of more than 50° (see absolute ego-orientation error plot in Figure 64). Furthermore, there was a considerable general overestimation of the angle turned, indicated by the considerable ego-orientation error in turning direction. This overestimation was substantially more pronounced for condition C with additional vestibular motion cues (50.6°, see Figure 64) than for condition D without vestibular motion cues[9] (32.6°). The direction of this effect was unexpected, as one might rather predict that additional vestibular motion cues should improve the ego-motion perception. The additional vestibular cues from physical motions apparently increased the perceived turning angle, even though the physical turning angles were only 1/4 of the corresponding visual turning angles. The consistent overestimation of turning angles in both OPTIC FLOW conditions was rather surprising, as participants received feedback about their current orientation from the market place scene that became visible after each OPTIC FLOW trial. Nevertheless, most participants were apparently unable to use this feedback to recalibrate their turn perception. The demanding task and the quick turns might both have contributed to this unexpected effect.

The observed angular overestimation errors are considerably larger than the almost negligible errors for actively producing turns in Experiment TOWN&BLOBS in section 6. A recent experiment on turn execution using optic flow presented on the same screen as in the current experiment, however, resulted also in a consistent overestimation of the angles turned (Schulte-Pelkum et al., 2002). This suggests again that the display setup and the FOV might be a critical variable. Further experiments are needed, however, to pinpoint the critical factors for unbiased turn perception via optic flow.

Two further differences between the two OPTIC FLOW conditions were apparent, indicating a small benefit from the additional vestibular cues when the reliability of the visual cues is reduced. On the

---

[9] Comparing UPDATE performance between condition C and D using a paired t-test shows a significant difference ($t(16) = -3.09$, $p = 0.0071^*$).

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

(e) Ego-orientation error in turning direction

Figure 40: Automatic spatial updating performance in Experiment Landmarks versus Optic Flow. In the Landmarks condition, the difference between Update and Control performance was only minimally above zero, indicating that automatic spatial updating was rather easy and accurate. In the Optic Flow conditions, however, automatic spatial updating was considerably impaired, indicated by the clear offset from zero.

one hand, response times in the PLATFORM ON condition were roughly 50ms smaller than in the PLATFORM OFF condition. This effect was apparent in both the difference between UPDATE and CONTROL performance and the UPDATE performance itself (see Figure 64), but did not reach statistical significance[10]. On the other hand, the PLATFORM OFF condition showed a substantially larger configuration error in the UPDATE condition than in the CONTROL condition. In the PLATFORM ON condition, however, UPDATE and CONTROL trials showed virtually the same configuration error. Comparing condition C and D reveals a significant difference both in the difference between UPDATE and CONTROL performance (see Table 19) and in the UPDATE performance itself (see Figure 64, $t(16) = 2.28$, $p = 0.036^*$). That is, the lack of vestibular motion cues in condition D increased the configuration error for turns to new orientations. This suggests that the mental spatial representation of the surround was slightly less consistent when vestibular motion cues were missing. Even under those conditions, however, the configuration error of $25°$ was quite far from chance. In the light of results by Wang & Spelke (2000), this effect could be interpreted as an increased disorientation after optic flow-induced turns without concurrent vestibular motion cues. It is to be noticed that reduced vestibular motion in condition C with a gain factor of $g = 1/4$ were sufficient to prevent this apparent disorientation, and that the full vestibular motion cues ($g = 1$) were not needed. That is, our approach of using gain factors $g < 1$ was quite successful and seems legitimate, even though we can only hypothesize from Experiment SIMULATION PARAMETERS that veridical vestibular cues ($g = 1$) should not change the results for condition C and A.

In summary, photo-realistic landmarks embedded in a consistent scene proved again sufficient for enabling automatic spatial updating, irrespective of concurrent vestibular cues. Vestibular turn cues became only relevant when visual cues were reduced to a mere optic flow pattern. That is, only when visual cues were apparently insufficient for solving the task easily did vestibular cues become more important. And even then, vestibular cues had only a moderate effect by preventing the configuration error from increasing.

As an increase in configuration error has been observed for participants that were previously disoriented (Wang & Spelke, 2000), one might argue vice versa that the presence of concurrent vestibular turn cues in Experiment LANDMARKS VERSUS OPTIC FLOW (condition C) prevented the slight disorientation observed in the PLATFORM OFF condition (D). This hypothesis is, however, rather speculative and awaits further exploration.

15.3.3   Obligatory spatial updating

As in the previous experiments, the difference between IGNORE and UPDATE performance was used to investigate the reflex-like component of spatial updating under the different conditions (see Figure 41 and Table 19). As expected from the previous experiments, IGNORE performance was considerably impaired in all measures for both LANDMARKS conditions. Especially the large increase of approximately 500ms in terms of response time indicates that spatial updating was for both LANDMARKS conditions obligatory in the sense of being hard-to-suppress and consequently reflex-like.

Comparing IGNORE performance for condition A and B with and without vestibular turn cues shows only a small but insignificant increase in configuration error for the PLATFORM OFF condition. Apart from that, IGNORE performance was virtually identical, indicating again that vestibular cues were irrelevant for the task, and that visual landmark cues alone are sufficient to render spatial updating obligatory.

15.3.3.1   Influence of optic flow   The OPTIC FLOW conditions C and D, however, showed no obligatory spatial updating whatsoever: IGNORE performance yielded response times that were at

---

[10]$t(16) = 1.39$, $p = 0.18$ for comparing the UPDATE trials of condition C and D, see also table 19

(a) Response time

(b) Configuration error

(c) Absolute pointing error

(d) Absolute ego-orientation error

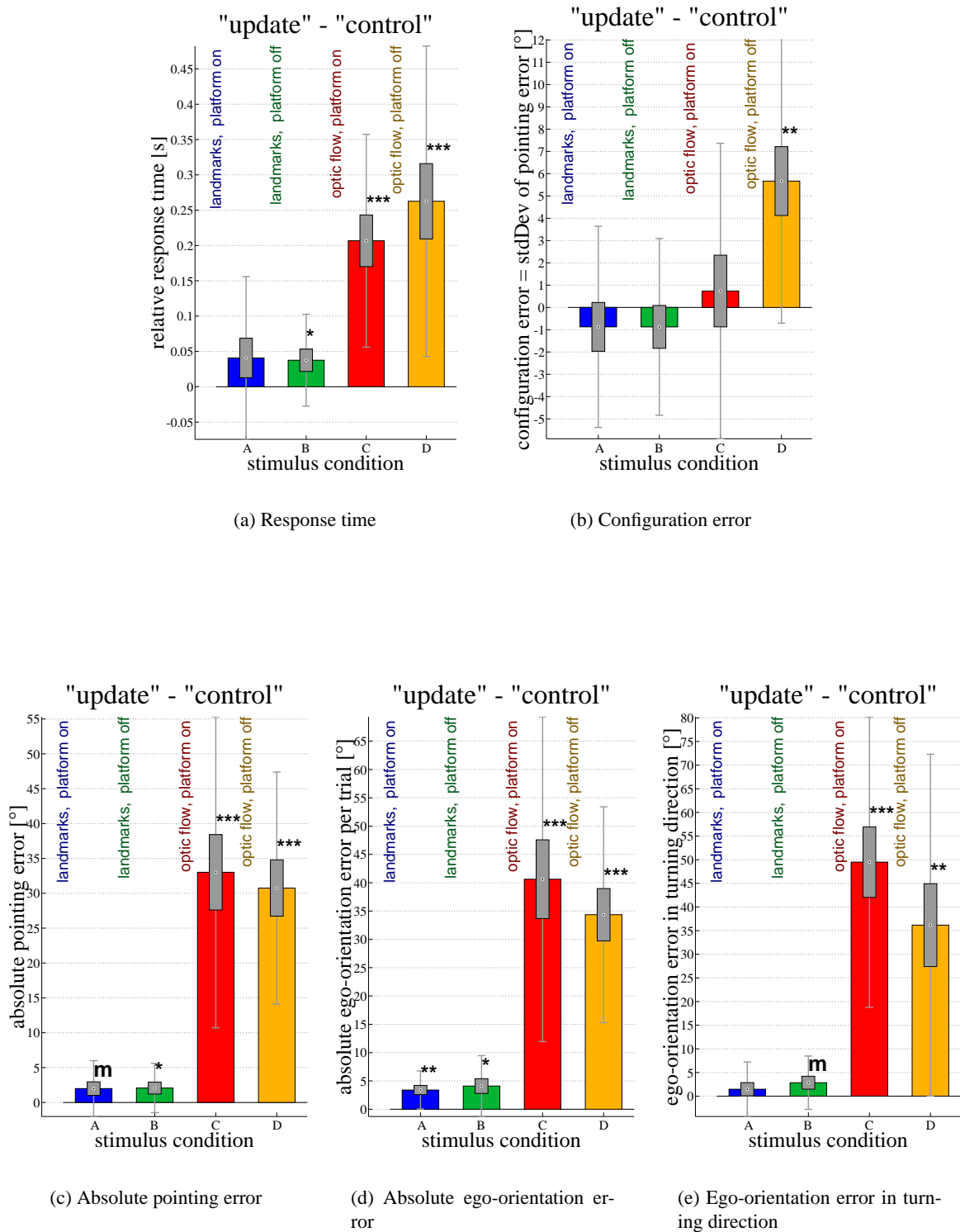(e) Ego-orientation error in turning direction
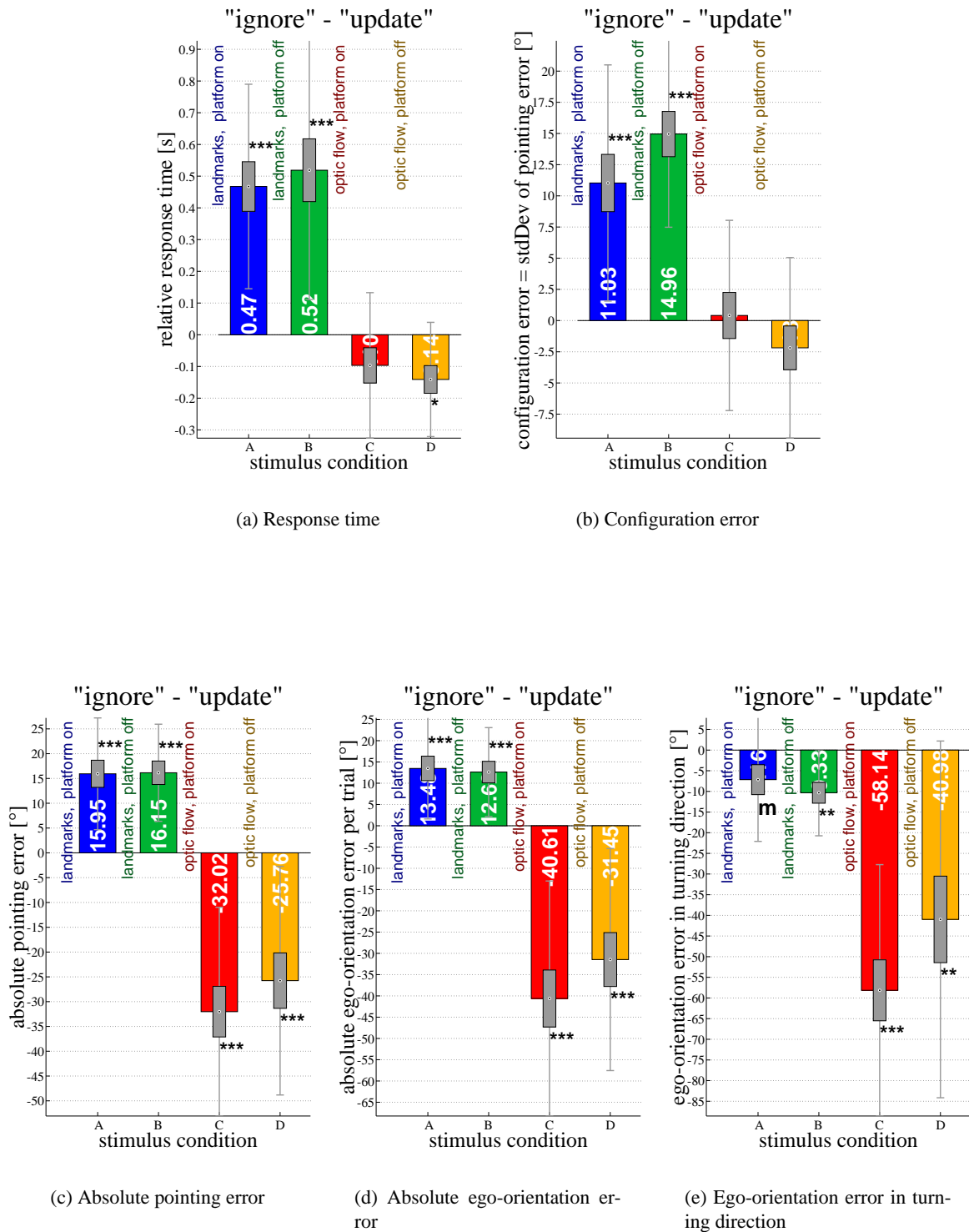
Figure 41: Obligatory spatial updating performance in Experiment LANDMARKS VERSUS OPTIC FLOW. For the LANDMARKS conditions, the differences between IGNORE and UPDATE performance measures were positive for Figures (a) - (d). That is, ignoring a turn was considerably harder than updating it as usual, implying obligatory spatial updating. For the OPTIC FLOW conditions, however, the offsets from zero were negative, indicating that ignoring a turn was actually easier and more accurate than updating it. Hence, optic flow information proved completely insufficient for inducing obligatory spatial updating.

least 100ms *smaller* than in the UPDATE conditions. Furthermore, absolute pointing error and the ego-orientation errors were significantly *reduced* in the IGNORE trials. Taken together, this suggests that ignoring OPTIC FLOW turns was actually much easier and more accurate than updating them. Only the configuration error shows virtually no effect, suggesting that it did not matter for the consistency of participants' spatial representation whether they were instructed to IGNORE a turn stimulus or use it to UPDATE to the new orientation. That is, the consistency of the mental representation seemed completely unaffected by the optic flow stimulus. This makes sense, as optic flow can only indicate motions, but has no further relation whatsoever to the scene.

The different presentation of the data in Figure 63 shows more clearly that IGNORE performance was typically in between CONTROL and UPDATE performance, and often as good as CONTROL performance. That is, having to UPDATE an optic flow-induced turn is considerably harder than having to IGNORE it. On the other hand, response times for the IGNORE trials were considerably longer than for the CONTROL trials[11]. This is rather interesting, as participants had essentially the same task in both the CONTROL and IGNORE trials, namely having to point as if still being at the previous location, without any useful static visual cues. This effect is most puzzling in the PLATFORM OFF condition (D), where virtually the only difference between CONTROL and IGNORE trial was the optic flow displaying either a forth-and-back motion or just a forth motion, respectively. Nevertheless, the optic flow simulating a forth motion considerably impaired participants' performance in terms of response time and configuration error[12]. That is, participants performed significantly better when the optic flow stimulus was *consistent* with their task of pointing as if being at the previous location (CONTROL trials). A conflicting optic flow motion, on the other hand, disrupted performance considerably (IGNORE trials). Hence, optic flow information *cannot* easily be ignored without noticeable performance decreases. It was, however, still much harder to use optic flow information to UPDATE to new orientations than to IGNORE it and act as if still being at the same position.

To sum up, the rotating optic flow stimulus did indeed prove to have an effect on the mental spatial representation as assessed by rapid pointing tasks. Even though the presentation times were probably not sufficient to enable the full-fledged precept of vection (perception of ego-motion induced by optic flow), it was nevertheless not trivial to simply ignore the visual stimulus altogether and act as if still being at the same location. On the other hand, the optic flow information was clearly *not* sufficient to enable easy automatic spatial updating to new orientations. This was most clearly indicated by the considerable increase in response time for the UPDATE trials.

15.3.3.2 Influence of vestibular cues in the OPTIC FLOW conditions

We hypothesized above that the irrelevance of vestibular cues in the LANDMARKS conditions might be due to the fact that the visual cues are considerably more reliable, as they provide an abundance of salient landmarks usable for position-fixing. Hence, we put forth the idea that vestibular cues are largely dominated by visual landmarks, and that vestibular cues should become important when visual cues are reduced to a comparable reliability by removing all landmark information in the OPTIC FLOW conditions. That is, vestibular cues might render the turn harder to ignore if visual information is reduced to mere velocity information extracted from optic flow.

Comparing condition C and D, however, reveals no contribution of vestibular cues for both response time and configuration error. The difference plots for the other three dependent variables in Figure 41 shows actually a tendency towards *easier*-to-ignore turns in the PLATFORM ON condition, which is an effect in the direction opposite of the one expected. This tendency, however, did not reach statistical significance (see Table 19). Besides, the method of analyzing the difference between IGNORE

---

[11]Comparing response times for IGNORE and CONTROL trials yields significant differences both for condition C ($t(16) = 2.29, p = 0.036^*$) and for condition D ($t(16) = 3.16, p = 0.0061^*$).

[12]Comparing IGNORE and CONTROL trials for condition D without vestibular motion cues yields significant differences in terms of response time ($t(16) = 3.16, p = 0.0061^*$) as well as configuration error ($t(16) = 2.29, p = 0.036^*$).

and Update performance might have exaggerated this effect, and a look at the absolute Ignore values in Figure 63 and 64 reveals that there is indeed only a minimal tendency in this direction: The Optic Flow, Platform On condition yielded somewhat smaller errors for the absolute pointing errors, absolute ego-orientation errors, and direction-specific ego-orientation error than the Optic Flow, Platform Off condition. These differences, however, were far from statistical significance ($t(16) = 1.48, p = 0.16, t(16) = 1.41, p = 0.18$, and $t(16) = -0.18, p = 0.86$, respectively). In summary, our initial hypothesis proved wrong, and vestibular motion cues were by no means able to render motions harder to ignore, not even when all visual information was reduced to mere velocity information from optic flow.

15.3.4 Further analyses

15.3.4.1 Learning effect From the previous experiments, we expected to see either no performance improvement over time at all (as in Experiment Real World versus VR), or an improvement just for the response time (as in Experiment Simulation Parameters). As all cues combinations were used *within* each session for Landmarks versus Optic Flow, the correlation analysis could be performed for the trial number across sessions, which is expected to yield a much clearer result than the correlation with session number as was done in the previous experiments. The results of the correlation analyses are compiled in Table 20. From all five dependent variables only the response time showed any significant improvements over the course of the experiment. The feedback of the pointing time after each trial might have contributed to this effect.

Spatial updating condition	Stimulus condition	r	r^2	$t(16)$	p
Update	B: Landmarks, Platform Off	-0.245	0.06	-2.80	0.013*
Update	C: Optic Flow, Platform On	-0.225	0.051	-2.77	0.014*
Update	D: Optic Flow, Platform Off	-0.331	0.109	-3.82	0.0015**
Control	D: Optic Flow, Platform Off	-0.436	0.190	-3.12	0.0066*
Ignore	A: Landmarks, Platform On	-0.389	0.151	-3.06	0.0075*
Ignore	B: Landmarks, Platform Off	-0.377	0.142	-3.38	0.0038**
Ignore Backmotion	A: Landmarks, Platform On	-0.304	0.092	-2.36	0.032*

Table 20: Significant results of the correlation analysis for the learning effect in Experiment Landmarks versus Optic Flow. The correlation analyses were performed between all five dependent variables and the accumulative trial number across the three sessions. From all five dependent variables tested, only the response time showed any significant correlations ($p \leq 0.05$) which are displayed in the above table. This indicates an overall improvement in terms of response time, but none of the other dependent variables.

It is interesting to note that the correlations tended to be stronger and reach higher significance in the supposedly harder conditions: For the Update condition, the learning effect was stronger for the conditions without Landmarks and without concurrent vestibular motion cues. For the Ignore condition, on the other hand, performance improvements were only found for the Landmarks conditions, which are known to be hard to ignore. For the Optic Flow conditions, no significant performance improvement was found, corroborating the previous findings that ignoring optic flow was easy from the very beginning, and did not get any easier over time.

We have seen earlier that participants in both Optic Flow conditions had considerable problems in estimating the turning angle and consistently overestimated the turning angle. The lack of significant improvements for both ego-orientation error measures in the Optic Flow conditions suggests consequently that participants were apparently unable to recalibrate their turn magnitude perception. This is all the more astonishing as explicit feedback about the new orientation was provided

by the scene becoming visible after each OPTIC FLOW trial. We presume that participants were so involved and challenged by the rapid pointing task that they had no cognitive or other resources left to successfully recalibrate their turn perception.

Furthermore, the lack of performance improvements in terms of configuration error indicates that participants had already a good and consistent mental representation of the scene from the beginning. Else, the continuous visibility of the scene in the LANDMARKS conditions could have been used to improve the mental spatial representation of the scene. Hence, the pre-knowledge of the real scene and the extended training phase were apparently sufficient to provide participants with a consistent representation.

15.3.4.2   Pointing order effect   If participants were disoriented or otherwise confused by the motion cues or task requirements, one would expect the first pointing of each trial to be worse than the later pointings. To quantify these effects, correlation analyses were performed between the response time and the pointing number. The results are summarized in Table 21 and discussed in detail below.

Spatial updating condition	Stimulus condition	r	r^2	$t(16)$	p
UPDATE	C: OPTIC FLOW, PLATFORM ON	-0.141	0.020	-3.46	0.0032**
UPDATE	D: OPTIC FLOW, PLATFORM OFF	-0.129	0.017	-2.55	0.021*
IGNORE	C: OPTIC FLOW, PLATFORM ON	-0.291	0.085	-4.25	0.00061**
IGNORE BACKMOTION	A: LANDMARKS, PLATFORM ON	-0.122	0.015	-2.44	0.027*
IGNORE BACKMOTION	B: LANDMARKS, PLATFORM OFF	-0.243	0.059	-3.87	0.0013**
IGNORE BACKMOTION	C: OPTIC FLOW, PLATFORM ON	-0.283	0.080	-5.15	0.00010***
IGNORE BACKMOTION	D: OPTIC FLOW, PLATFORM OFF	-0.262	0.068	-4.59	0.00030***

Table 21: Results of the correlation analysis for the pointing order in Experiment LANDMARKS VERSUS OPTIC FLOW. Displayed are the results of the correlation analysis between response time and the pointing number. Only correlations that were significant on at least a $p = 0.05$ level are displayed.

UPDATE   If the available cues are sufficient for enabling automatic spatial updating, UPDATE performance should be quick and easy from the first pointing on, and should not improve considerably for the later pointings. If, on the other hand, the available cues are insufficient for automatic spatial updating, UPDATE performance should be worse for the first pointings and might improve for the later ones if participants can somehow cognitively or otherwise compensate for the missing cues by, e.g., mental imagery. The correlation analyses reveal a significant negative correlation for both OPTIC FLOW conditions, indicating that the available cues were indeed insufficient for automatic spatial updating. The LANDMARKS conditions, on the other hand, showed no such effect, indicating that the visual scene was indeed sufficient to enable automatic spatial updating, irrespective of vestibular cues.

CONTROL   The CONTROL trials should be so easy that all pointings are equally fast. The correlation analysis shows indeed no significant correlations, corroborating our assumption that the baseline task is rather easy and does not depend too much on the forth-and-back motion.

IGNORE   For the IGNORE trials, the only significant correlation was for the OPTIC FLOW, PLATFORM ON condition. Later pointings were apparently easier than earlier pointings, suggesting that

participants could somehow compensate for the to-be-ignored turn stimulus. Why this effect did not reach significance for the OPTIC FLOW, PLATFORM OFF condition ($p = 10.7$) remains unclear and awaits further investigations.

IGNORE BACKMOTION   The IGNORE BACKMOTION were designed to re-anchor participants to the previous orientation after the potentially confusing IGNORE trials. The larger the confusion and the harder the re-orientation, the harder should be the first pointings. Hence, we expect a strong negative correlation for the conditions where participants were confused or needed some time to re-anchor to the previous orientation. The correlation analysis shows indeed significant negative correlations for all four stimulus combinations. This effect is stronger in the OPTIC FLOW conditions, probably because participants needed some additional time to correctly remember the previous orientation (which was still invisible until after the four pointings).

15.3.4.3 Map drawings   After the experiment, eleven of the 17 participants were asked to sketch the geometric layout of the Tübingen market place. The resulting sketches are displayed in Figure 42. Even though all participants were able to easily name all 22 landmarks in the proper order, most of them were unable to depict the main geometric features of the scene correctly. This is all the more astonishing as all of them had been living in Tübingen for several years. Note also the tendency towards right angles or even depicting the whole market place as one big square. Only two or three participants were able to capture the overall geometry correctly.

15.4 Summary and conclusions

Experiment LANDMARKS VERSUS OPTIC FLOW was mainly concerned with answering two questions:

1. First, what empowers visual cues to trigger obligatory spatial updating and enable automatic spatial updating? Is it the landmarks forming a consistent, well-known scene, or merely the visual motion stimulus, similar to the vection induced by a rotating optic drum?

2. Will the vestibular motion cues have a relevant contribution to automatic or obligatory spatial updating if the available visual strategies are reduced to path integration via optic flow?

The previous experiments have shown that vestibular cues did not play any significant role as long as useful visual landmark information was available. This was fully corroborated by Experiment LANDMARKS VERSUS OPTIC FLOW, where additional vestibular motion cues proved again completely irrelevant for all conditions with available landmarks. To answer the first question, visual motion information was separated from visual landmark information by presenting only optic flow in half of the trials. This led to a consistent performance decrease in all dependent measures for the CONTROL and even more so for the UPDATE trials. That is, OPTIC FLOW information proved clearly insufficient for enabling quick and accurate automatic spatial updating. Furthermore, updating optic flow turns was considerably *more* difficult than ignoring them, indicating that optic flow information is by no means able to trigger obligatory spatial updating.

This finding confirms results from Klatzky et al. (1998) who used a different experimental paradigm: After being exposed to a purely visually presented two-segment path (just like in the excursion of the triangle completion experiments presented in part II), participants were asked to quickly turn physically to face the origin, just as they would if they had physically walked the path and were at the end of the second segment. Participants responded as if they updated the translations for the
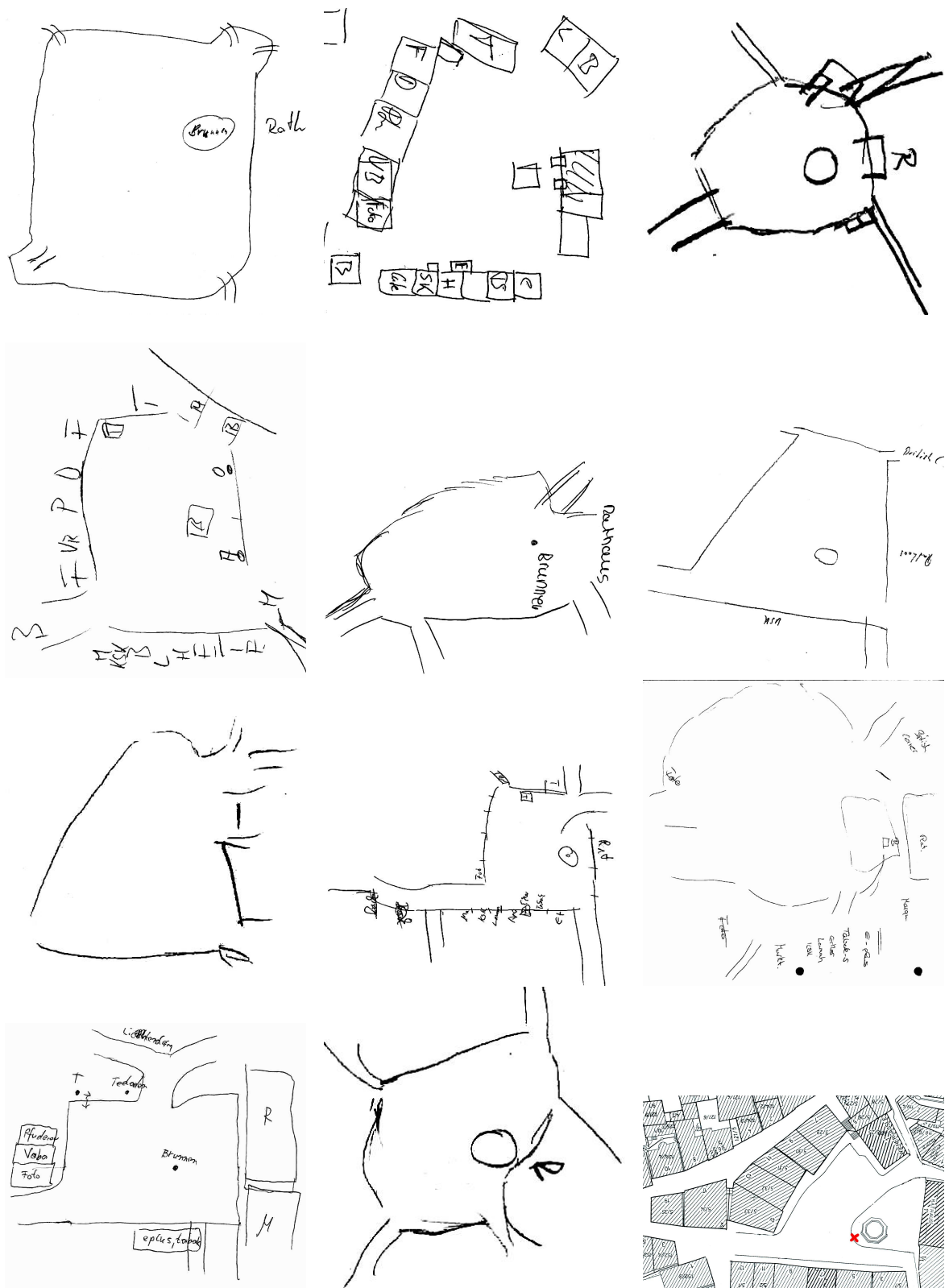
Figure 42: Map sketches from eleven of the 17 participants. The bottom right figure displays the correct scene layout of the Tübingen market place. Note that only very few participants were able to capture the irregular overall geometry of the scene.

two linear segments (s_1 and s_2) properly, but completely "forgot" to update the in between turn ($\overline{\alpha}$). Only if the turn was performed physically did they update their heading properly. The same behavior occurred when participants were asked to imagine walking the excursion or when watching another person walk. Klatzky et al. conclude that "*simulated optic flow was not by itself sufficient to induce spatial updating that supported correct turn responses*" (p. 293). This is in agreement with our conclusions that optic flow was insufficient for obligatory as well as automatic spatial updating.

Does this mean that an optic flow stimulus has no influence whatsoever on the mental spatial representation? Actually not. Even though for the OPTIC FLOW conditions, IGNORE performance was much better than UPDATE performance, it was still not as good as CONTROL performance, where participants had essentially the same task of pointing as if still being in the same orientation. Hence, the rotating optic flow stimulus did indeed have some effect on the mental spatial representation as assessed by rapid pointing. Even though the presentation times were probably not sufficient to induce a convincing perception of ego-motion (vection), it was nevertheless not possible to simply ignore the optic flow rotation altogether and respond as if still being at the same location (IGNORE task).

This is in agreement with informal reports from some participants that the optic flow stimulus evoked some kind of vection, at least for the larger turns. As vection was not explicitly addressed in our experiments, however, we can only speculate that vection might be a necessary prerequisite for continuous spatial updating. Future experiments comparing introspective and behavioral measures of vection with spatial updating performance as assessed by rapid pointing might allow us to determine the causal relation between vection and spatial updating.

A point of critique often mentioned is that the IGNORE task could be interpreted as a pure memory task, without any relation to spatial updating whatsoever. This would imply that we did not measure spatial updating performance, but rather performed some kind of spatial memory task. That is, the presented visual stimulus in the IGNORE tasks acted merely as an unspecific distractor that did not influence the mental spatial representation in any specific manner. If this argument was true, any view different from the correct one should impair IGNORE performance equally and unspecifically. There is an abundance of evidence to refute this critique. First of all, if it was mainly a memory task and the visual stimulus just an unspecific distractor, IGNORE performance should not depend on the angle turned, and should not have any relation to the turning direction. Experiment SIMULATION PARAMETERS, however, showed a clear turning angle effect. Moreover, there was a direction-specific ego-orientation error against turning direction in all conditions with continuous visual motion. Only the discontinuous jump condition did not show this effect. Experiment LANDMARKS VERSUS OPTIC FLOW provides further evidence against the memory argument. If it was just the absence of the correct visual stimulus that impaired IGNORE performance, and the unmatched visual stimulus acted mainly as an unspecific distractor, the simple OPTIC FLOW display (which is incapable of inducing instantaneous spatial updating as observed in the jump condition) should disrupt IGNORE performance as much as the LANDMARKS display. The data, however, disproved this hypothesis convincingly. Last but not least, only if the presented optic flow stimulus had any specific updating effect on participants' egocentric reference frame would one predict IGNORE performance to be harder than CONTROL performance. Conversely, if it was merely a memory task, and/or the effect of the optic flow was unspecific, IGNORE trials should rather lead to improved performance, as they were specifically announced before the simulated turn, whereas the CONTROL trials were not. As already pointed out in the previous paragraph, however, IGNORE performance was significantly impaired compared to CONTROL performance. In sum, the experiments performed demonstrated convincingly that even the optic flow display had a clear and distinct influence on participants' egocentric reference frame used for performing the pointing tasks, thus refuting the "unspecific memory task" critique once and for all.

To address the second question: The visual dominance observed in all conditions with useful visual landmark information can most likely be explained by the visual landmark cues being much more

reliable than the vestibular cues, which are based on path integration and hence do not allow for position fixing via landmarks. We hypothesized that this qualitative difference between the visual and vestibular cues might have caused the observed visual dominance. Hence, reducing visual strategies to path integration by presenting only optic flow information during the motions might render the fidelity and reliability of the visual and vestibular system more comparable and lead to a significant contribution of vestibular cues for spatial updating. This was investigated by presenting only visual turn cues in half of the trials, and comparing performance to the other half of the trials that included additional vestibular cues from physical turns.

We found, however, only a small but significant benefit from additional vestibular motion cues for the UPDATE trials of the OPTIC FLOW condition. And even then, vestibular cues had only a moderate effect by preventing the configuration error from increasing. In the context of the literature, where increased configuration errors were found for participants that were previously disoriented (Wang & Spelke, 2000), one might speculate vice versa that the additional vestibular turn cues prevented the slight disorientation observed for the condition with just optic flow. Further experiments, however, are needed to test this rather speculative hypothesis.

For the CONTROL trials of the OPTIC FLOW condition, additional vestibular motion cues even decreased pointing consistency slightly but significantly. The IGNORE trials showed no benefit from additional vestibular motion cues whatsoever, and if anything, a slight but insignificant tendency towards easier-to-ignore turns.

In summary, our initial hypothesis proved wrong, and vestibular motion cues were by no means able to render motions harder to ignore, not even when all visual information was reduced to mere velocity information from optic flow.

Part IV

Theoretical framework and general discussion

The experiments presented in this paper provided a number of novel and often unexpected insights into navigation, spatial orientation, and spatial updating. A number of processes and information sources were found to influence performance, including optic flow, reliable and unreliable landmarks, mental spatial reasoning, reference frames, and display parameters. But how do all those results fit together? Are they just isolated findings or is there a way to merge them to provide a unifying "big picture"?

Ideally, one would like to have some kind of comprehensive framework in which to incorporate the most important findings. Such a framework could provide a number of advantages: First, it might allow for a coherent representation of the experimental paradigms and results. Second, it could help to structure and clarify our reasoning and discussions. Perhaps one of the most important advantages is that it might allow for a deeper understanding of the underlying processes and mutual dependencies. Last but not least, it could suggest novel experiments and experimental paradigms, allow for testable predictions, and stimulate scientific discussion.

In order to be comprehensive enough to include at least all the experiments presented in this thesis, this framework should include a number of processes related to spatial orientation (e.g., spatial updating, ego-motion perception, landmark identification and spatial learning) as well as memory components (e.g., landmark memory, allocentric spatial memory, and egocentric reference frames). As we were unable to find any framework in the literature that seemed adequate and met our requirements, we decided to develop our own framework.

Our main guiding principle was to develop a consistent and systematic approach by trying to understand the logical and functional dependencies between related items. For example, it seems evident that ego-motion perception cannot occur without some kind of motion perception. That is, intact ego-motion perception seems to be logically dependent on intact motion perception. Conversely, if we observe intact ego-motion perception, we can conclude that motion perception must also be intact, which can be represented as *"ego-motion perception ⇒ motion perception"* using standard logical notation (see Table 22).

This framework is of course a "work in progress" and will never be finished in the sense of being able to explain everything related to, e.g., spatial orientation. Hence, the framework presented in the following section can be thought of as a snapshot and a first step towards a more comprehensive model. Nevertheless, we are confident that it meets already most of the requirements stated above. In section 17, we will test this model by applying it to the different experiments presented in this thesis, thus revisiting and discussing them in the light of one unifying framework. We will conclude by applying the framework to selected experimental results from the literature in subsection 17.4.

16 Qualitative modeling of spatial orientation processes using logical propositions

Our main goal for this section is to understand issues and terms related to spatial orientation, spatial updating, and spatial presence by analyzing their logical and functional relations. Here, we present first steps towards a logically consistent framework that describes and relates the associated items. This is done by trying to determine a set of necessary prerequisites and sufficient conditions. The underlying logic of our model suggests novel experimental paradigms that can pinpoint critical factors for good spatial orientation. More specifically, we were able to derive novel paradigms that allows one to quantify spatial presence and spatial updating, and disambiguate between continuous and instantaneous spatial updating (see subsection 16.2). "Spatial presence" can be understood as the consistent "gut" feeling of being in a specific spatial context, and intuitively and spontaneously knowing where one is with respect to the immediate surround. (See, e.g., (Regenbrecht, 1999) for an extensive review). As we will argue later, spatial presence might be a critical factor for achieving and understanding spatial updating and consequently also for quick and intuitive spatial orientation. Hence, any reliable quantification method that extends beyond the typically used subjective questionnaires might be quite helpful. Furthermore, analyzing experimental results in the context of a logical framework might allow for a deeper understanding of the underlying processes and could help in adapting and refining the framework.

The framework and some of the experiments were inspired by the following observation: In most Virtual Reality (VR) situations involving simulated movements of the observer, people feel lost or disoriented after only a few simulated motions. In comparable real world situations, however, spatial orientation is typically rather robust and effortless. This suggests that some critical prerequisites of good spatial orientation are missing in most VR simulations, even though they might look great and were rather costly. Comparing experiments in real world and VR nevertheless offers the opportunity to test what was missing in a given simulation. Thus, VR can be used as a flexible research tool for investigating spatial orientation processes. By comparing the necessary and/or sufficient conditions for good spatial orientation, our logical framework can assist in analyzing and understanding why spatial orientation fails in certain situations. By focusing on the hereby determined essential spatial cues and display parameters, one should ultimately be able to design convincing ego-motion simulators without having to simulate all sensory cues veridically.

16.1 Introduction

Before going into more detail, we would first like to present the overall structure and main components of the framework. The framework in its reduced form is graphically represented in Figure 43. *Spatial behavior*[13] and *spatial perception* are the main components of the action-perception cycle and constitute the top and bottom part of the framework, respectively. Meaningful *spatial behavior* is essentially based and logically dependent on *spatial perception*[14], and is mediated by several possible *spatial orientation processes*. At the bottom part of the framework, we distinguish mainly between two branches, a *relative motion branch* on the left side and an *absolute location branch* on the right side.

The *relative motion branch* (which is on the left in Figure 43) is based on path integration of perceived motions. It is responsible for generating the perception of ego-motion (vection) and the continuous updating of the self-location in space. Being based on path integration, sensory cues stem mainly

---

[13]Items of the framework are set in italics for convenience.

[14]Conversely, *spatial perception* is also influenced by *spatial behavior*, but does not logically require any *spatial behavior*, as is most obvious in the extreme case of locked-in patients.
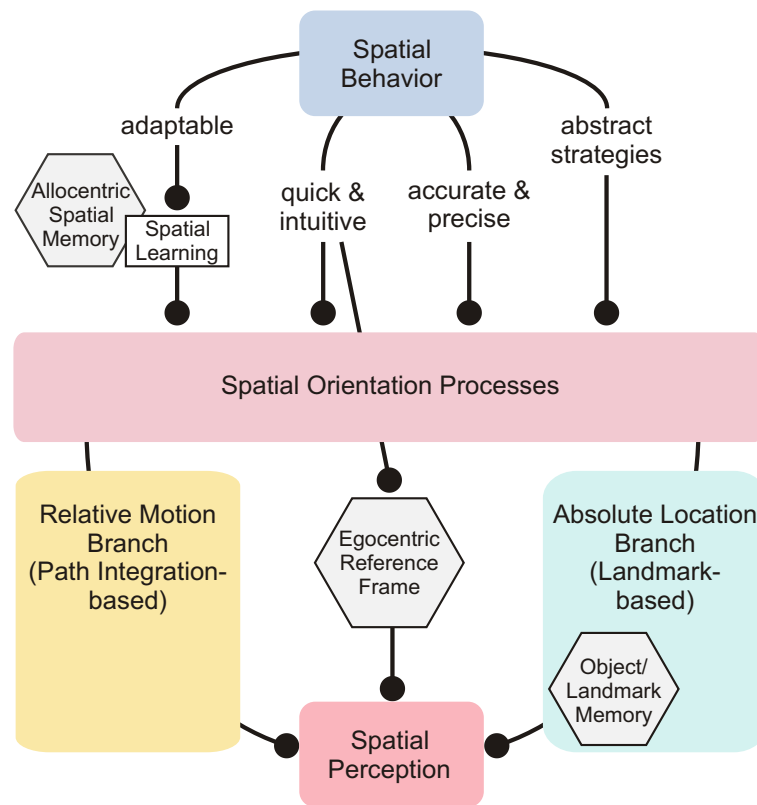
Figure 43: Overview of the model.

from vestibular and proprioceptive information and from optic flow. The *absolute location branch* (on the right) constitutes an alternative approach to finding one's way around, by using landmarks as reference points. *Object/landmark memory* is involved in the recognition of salient features in the environment.

At the top of the model, we distinguish between four different aspects or properties of *spatial behavior* (**adaptable, quick & intuitive, accurate & precise**, and **abstract strategies**). These different aspects of *spatial behavior* seem to depend logically on different underlying spatial orientation processes and data structures, as will be discussed in detail in section 16.3. **Adaptable** *spatial behavior* is, in addition, based on *spatial learning*, which is closely related to *allocentric spatial memory*.

In addition to the left and right branch, we propose a central pathway (in the center) that is responsible for robust and automated spatial orientation. That is, if we want to know where we are without having to think much about it, we need a process that allows for **quick & intuitive** spatial orientation and prevent us from getting lost, even when we do not constantly pay attention or have other obligations. To achieve this, some automated process (called "automatic spatial updating" or just "spatial updating") needs to always update our *egocentric mental reference frame* of the surround during ego-motions, such that it stays in close alignment with the physical surround (see part III).

In the following, the complete framework will first be introduced by describing each item briefly, categorizing it, and stating its hypothesized functional connections. We will continue by discussing some implications for the quantification of spatial updating and spatial presence, and by hypothesizing about further logical connections. That is, this framework will be used to generate hypotheses which can both guide future research and be experimentally tested.

Name	Statement	Operator	Meaning of statement
simple statements			
assertion	A		A is true
negation	$\neg A$	not	A is false
compound statements and sentential connectives			
disjunction	$A \vee B$	or	either A is true, or B is true, or both
conjunction	$A \wedge B$	and	both A and B are true
implication (conditional)	$A \Rightarrow B$	if ..., then	if A is true, then B is true
equivalence (biconditional)	$A \Longleftrightarrow B$	if and only if ..., then	A and B are either both true or both false

Table 22: Operators and statements as used in propositional logic.

Ideally, the final version of this framework should describe the **logical and functional relationships** between all related terms. As a first step, all terms introduced in this framework are grouped by their coarse classification into GOAL/DESIRED PROPERTY, DATA, or PROCESS. Note that "GOAL/DESIRED PROPERTY" is an attribute of the system described by the framework and not of the framework itself. The logical connections (arrows) between terms are meant to be understood in the mathematical sense, and we use the syntax from propositional logic as summarized in Table 22. Note that if A implies B, this is equivalent to saying that $non - B$ implies $non - A$ ($A \Rightarrow B \iff \neg B \Rightarrow \neg A$). A is therefore a *sufficient* but not a *necessary* prerequisite for B. In other words, B is a *necessary* but *not sufficient* prerequisite for A (contraposition). Please note also that the **information flow** is in most cases in the opposite direction, i.e, from B to A. That is, B is typically "more general" and does include (in the mathematical sense) the more specific A. The difference between logical implications and information flow is illustrated in Figure 44, using the simple example of the well-known action-perception loop.

(a) Information flow representation   (b) Logical connectors representation
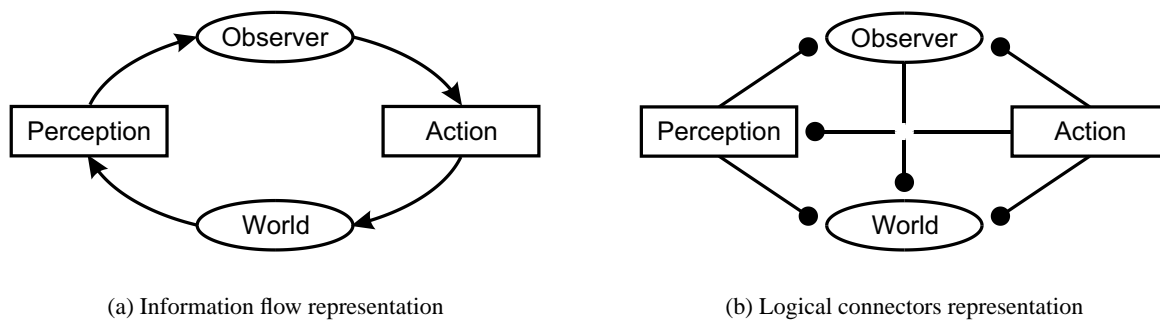
Figure 44: Action-perception loop, adapted to illustrate the difference between the typically used information flow arrows and our logical connections. (a) In the information flow paradigm, the *observer* obtains information about the surrounding *world* through *perception*. At the same time, the *world* is influenced by and receives information about the *observer* through her/his *actions*. (b) Using logical notations, the graphic looks quite different: The *world* at the bottom is the necessary prerequisite for the *observer* as well as her/his *action* and *perception*, indicated by the logical connectors ending at the *world* box. The opposite is true for the *action* box: All connections to it start there, indicating that any meaningful *action* requires an *observer* that is acting, a *world* (s)he is acting upon, and *perception* of the world, or else the behavior would be at random. Last but not least, *perception* implies and logically requires some perceiving entity, represented here as the *observer*.

Note that the individual items of the framework are not meant to be understood as simple yes-or-no decisions, such as "either spatial updating works, or else it does not". As human spatial orientation is like most mental processes highly complex and error-tolerant, this would oversimplify things. Rather, we would like to propose a more qualitative interpretation of the logical connections for this

framework, much like a fuzzy logic approach. In this manner, $A \Rightarrow B \iff \neg B \Rightarrow \neg A$ would imply that, e.g., "if B is impaired, so is A", or "if A works well, so does B". Furthermore, "if B does not work or exist at all, A is also substantially impaired or defunct".

16.2 Continuous versus instantaneous spatial updating

In order to convincingly explain the results from the spatial updating experiments in sections 13-15 in the context of this framework, we need to refine our concept of spatial updating. That is, we would like to distinguish between the classical *continuous spatial updating* known from the blindfolded spatial updating literature and *instantaneous spatial updating* that is able to explain, e.g., the jump condition in Experiment SIMULATION PARAMETERS. (See section 14 and subsections 17.2.3 and 17.2.4).

But let us first revisit spatial updating in general. When moving, all spatial relationships between ourselves and the environment change. Despite the tremendous amount of changes, we feel immersed in the current surround, naturally experience spatial presence, and know where we are. Hence, some robust process needs to continuously update these self-to-world relationships as we move: This general process of spatial updating operates typically reflex-like and automatic, and is responsible for both "continuously transforming the world inside our head" (*continuous spatial updating*, see below and also part III) and "aligning the world inside our head with the outside world" using landmarks (*instantaneous spatial updating*). That is, spatial updating can be thought of as the spatial transformation process operating on the egocentric mental spatial representation. In this manner, *continuous spatial updating* is the process of continuously and incrementally (smoothly) transforming our egocentric reference frame, whereas *instantaneous spatial updating* is the immediate, and if need be discontinuous ("jump"- or "teleport"-like) process. Whereas the continuous process might have some limitations in terms of transformation speed (e.g., a limited mental rotation speed), the instantaneous one probably does not.

In the literature, spatial updating is typically investigated in situations that have no reliable landmarks usable for position-fixing, e.g., by blindfolding participants or displaying optic flow only (e.g. Farrell & Robertson, 1998; Klatzky et al., 1998; May & Klatzky, 2000; May, 2000; Presson & Montello, 1994; Rieser et al., 1982; Rieser, 1989; Simons & Wang, 1998; Wang & Simons, 1999). Only recently were visual landmark cues integrated in human spatial updating research (Christou, Tjan, & Bülthoff, 1999; Wang & Spelke, 2000; Wraga et al., 2003). Without any available landmarks, only relative or motion information can be used for spatial updating. In blindfolded navigation studies, for example, velocity and acceleration information from the vestibular systems can be used to continuously update the mental spatial representation using path integration. Even proprioception provides only relative movement information (e.g., the number of steps traveled). That is, all body senses provide only relative information, and are thus prone to accumulation errors during path integration. Nevertheless, the literature shows clearly that vestibular and proprioceptive cues are, under many conditions, sufficient to enable automatic spatial updating. As this process is essentially based on path integration, any interruption or impairment due to, e.g., high cognitive load or distractions could potentially yield a completely misaligned mental reference frame, which is essentially useless. Hence, it seems natural to propose a spatial updating process that operates continuously and autonomously, and thus needs to be highly automated. This is what we refer to as *continuous spatial updating*. Any discontinuity in a spatial updating process based on relative spatial information or motion cues would critically disrupt its usability and reliability. Hence, one might argue that this continuous spatial updating should also be obligatory in the sense of reflex-like and hard-to-suppress. However, experimental evidence presented in this paper indicates that smooth rotational vestibular cues alone without any additional proprioceptive cues are clearly insufficient for obligatory spatial updating, even though this seems to contradict the prevailing opinion.

As continuous spatial updating alone is based on path integration and leads to exponentially increasing alignment errors over time, it seems sensible to propose a second process that can re-anchor the potentially misaligned mental reference frame to the physical surround. We would like to introduce the term *instantaneous spatial updating* to refer to this process. To give an example, imagine the following: You are at home at night when the main fuse blows. You will have to walk around in darkness until you manage to find the fuse box or some light source. When walking around in complete darkness, we become increasingly uncertain about our current ego-position. That is, we still have some intuitive feeling of where we are, but we would not bet high on the exact location. The situation changes as soon as we can perceive the location of a known landmark. This instantaneous position fixing could occur via different sensory modalities: Auditorily, for example the phone could be ringing. Haptically, we might touch or run into the kitchen table. Visually, somebody else might already have replaced the fuse, or lightning might have lit the room for a fraction of a second. That is, any clearly identifiable spatial cue (landmark) could re-anchor our mental reference frame instantaneously, without much cognitive effort or time needed. This process of re-aligning or re-anchoring the mental reference frame to the surround is what we refer to as *instantaneous spatial updating*.

When locomoting under full-cue conditions, this instantaneous spatial updating probably occurs at any instance in time and is thus indistinguishable from continuous spatial updating, as both processes are in close agreement and complement each other. Moreover, they can be considered as a mutual back-up system for the case that one of them fails or does not receive sufficient information. As pointed out earlier, blind navigation is a prototypical example where continuous spatial updating needs to bridge the potentially large gap between possible re-alignments using instantaneous spatial updating. Conversely, being a passenger in a bus driving over cobblestones, for example, the constant jerks and vibrations render *continuous spatial updating* by vestibular and proprioceptive cues utterly useless for navigation. Hence, the visual cues will most likely be used for constant *instantaneous spatial updating*. This redundancy of having two spatial updating processes running in parallel whenever possible is thus quite useful and improves the overall robustness of spatial orientation.

In the earlier sections of this thesis, we used the prevalent terminology for spatial updating, which does not distinguish between continuous and instantaneous spatial updating. For a deeper understanding of the underlying processes explaining our experimental results, however, this distinction became critical and will be incorporated in the conceptual framework presented in the following.

Our distinction between continuous and instantaneous spatial updating bears some resemblance to Kosslyn's distinction between "shift transformations" and "blink transformations", respectively (Kosslyn, 1994). Shift transformations are responsible for smooth and seemingly continuous transformations of mental images like object translations and rotations. If, however, "an image object must be transformed a large amount, the image may be allowed to fade and a new one is generated" (Kosslyn, 1994, p. 402), which Kosslyn refers to as "blink transformation". Note that shift and blink transformations refer to mental object image transformations that are continuous and discontinuous, respectively. Continuous and instantaneous spatial updating, on the other hand, refer to the transformation of the complete mental egocentric spatial reference frame, which involves a change in the observer's position or orientation. Furthermore, spatial updating is normally automated and reflex-like (obligatory), whereas Kosslyn's image transformations are typically deliberate, cognitive processes (i.e., neither automatic nor obligatory). These fundamental differences might explain the often found advantage of self motions over object motions for the updating of physical as well as imagined rotations (see Simons & Wang (1998), Wang & Simons (1999), Simons et al. (2002), Wraga et al. (1999a, 2000, 2003) and subsection 17.4.1.

16.3   Framework

The framework is graphically represented in Figure 45 and will be introduced in detail below. It covers on the vertical axis items ranging from low-level processes like spatial perception at the bottom to high-level processes like spatial behavior at the top. On the horizontal axis, the range spans from reflexive to cognitive control of behavior. While this model is a working hypothesis, the experiments in this thesis provide some experimental evidence for it. In other words, we will hypothesize about further connections that are plausible and helpful in interpreting experimental results, but not yet well-grounded in experimental data. These hypothesized connections, however, suggest novel ways of quantifying spatial updating and spatial presence by measuring the adjacent, logically related items of the framework. An exhaustive analysis would unfortunately go beyond the scope of this paper.

16.3.1   Goals and desired system properties

In the paragraphs below, we will introduce three goals or desired system properties that can be seen as a motivation and prerequisite for successful spatial behavior.

Overall goal guiding this framework: *Spatial Orientation*   All moving organisms have the goal of finding, for example, food, shelter or a path through the world without constantly getting lost. All of these tasks critically rely on spatial orientation. Hence, our framework has to follow this global aim of spatial orientation as a critical boundary condition for successful spatial behavior. Homing is one prominent example from the literature. The ability to find the way back to the origin of an excursion can be found in most moving species – from ants to humans (Klatzky et al., 1997; Maurer & Séguinot, 1995; Mittelstaedt & Mittelstaedt, 1982).

Additional goals guiding this framework: *Consistency* and *Continuity*   Perception is in many respects *continuous* in space and time. Furthermore, the different sensory modalities are typically found to contribute to one *consistent* percept of the world. That is, the relation between oneself and the surrounding real world is spatiotemporally continuous and consistent. Unless we navigate through computer-generated worlds, we are neither teleported in space or time (discontinuity) nor do we perceive ourselves to be at several places at the same time (inconsistency). Both *consistency* and *continuity* of the self-to-world relation should therefore be additional desired properties in our framework. Conversely, any kind of inconsistency or discontinuity potentially reduces spatial orientation abilities and should thus be avoided in the design of VR applications. In general, organisms might also use this continuity of perception to deduce high spatiotemporal correlations in order to statistically learn properties of the world (Bayesian approach). Hence, it seems plausible to include both *consistency* and *continuity* in the framework. spatiotemporal continuity is also an important prerequisite when we learn new objects (Wallis & Bülthoff, 2001). This aspect has been successfully implemented in a machine vision recognition system (Bülthoff, Wallraven, & Graf, 2002; Wallraven & Bülthoff, 2001).

16.3.2   Processes and data structures

In the following, we will try to guide the reader sequentially through this model in a bottom-up manner: We will start with the most fundamental processes and data structures and gradually work our way up until we have all the main ingredients enabling good spatial orientation, which is our overall guiding goal. After briefly describing and categorizing each term, we will state the most

relevant logical and/or functional connections to the aforementioned terms. Finally, some extensions and debatable hypotheses are put forward to be discussed in a larger context. Figure 45 shows the complete overview. As the complete model is rather complex, we advise the reader to focus on the terms and relations that have been introduced up to that point. We will start by describing the path integration-based left branch of the framework.
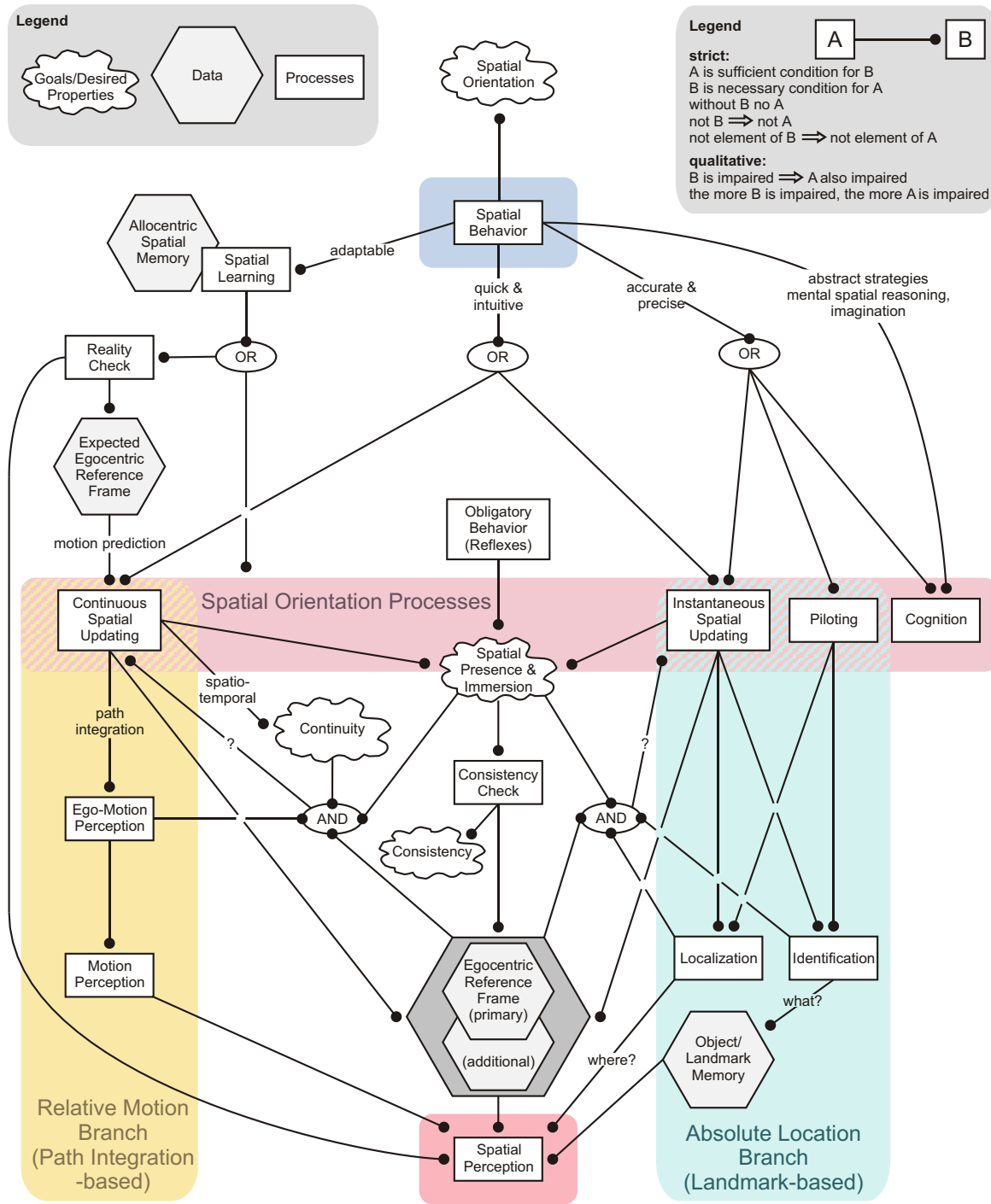

Figure 45: Conceptual framework, as described in the text.

Spatial Perception [PROCESS] Physical stimuli can be perceived in multiple dimensions and modalities. We group here all kinds of perception, regardless of their sensory modality (e.g., visual, auditory, haptic, kinesthetic etc.) into *spatial perception* if the percept covers some spatial aspect of the stimulus. For the purpose of the overall framework, we do not need or intend to refine this rather coarse and low-level definition of *spatial perception*. Its main purpose is to constitute the basis and necessary prerequisite for the whole framework.

Motion Perception [PROCESS] When we perceive temporal changes of spatial stimuli, we can have the percept of motion. For example, closely listening to a mosquito can tell us whether it moves or not (auditory *motion perception*). Another example is the perception of visual motion from optic flow using simple Reichardt-detectors (Reichardt, 1961). *Motion perception* depends logically on *spatial perception* in the sense that we cannot perceive any motion if we cannot perceive spatial cues: (*motion perception* $\Rightarrow$ *spatial perception*) $\iff$ ($\neg$ *spatial perception* $\Rightarrow \neg$ *motion perception*). Furthermore, only if continuous changes in space occur over time can we perceive motion. (Under certain conditions, however, small spatial jumps can be perceived (interpreted) as "apparent motion").

Ego-Motion Perception [PROCESS] If perceived motion is interpreted as self-motion of the observer and not just as a motion of some entity relative to the (stationary) world or observer, we call this phenomenon *ego-motion perception*. Whenever we move through the world, we typically have the percept of ego-motion. The classical example for illusory *ego-motion perception* is the visually induced vection (feeling of ego-motion) that can be achieved by presenting a rotating optic flow pattern in an optic drum for several seconds (see, e.g., Dichgans & Brandt, 1978; Fischer & Kornmüller, 1930; Mach, 1922). Obviously, without perceiving any motion in any modality, one would not feel any ego-motion. Therefore, we can state: *ego-motion perception* $\Rightarrow$ *motion perception*.

Egocentric Reference Frame [DATA] An *egocentric reference frame* can be understood as a mental representation of the "world in our head", as seen from the first-person perspective. This mental model is thought to contain at least the immediate surround or scene. We do not assume any preferred storage format like body schemes or specific coordinate systems. Even if this mental model does not explicitly exist, it nevertheless makes sense to store somewhere the existing knowledge of the immediate surround from the egocentric perspective, as this is the perspective from which we interact with the environment by grasping objects, moving towards them etc.

Incoming information from several modalities can code multiple *egocentric reference frames*. The most prominent or salient one, on which the majority of sensory inputs agree, is called the primary *egocentric reference frame,* which can be in conflict with additional (secondary) reference frames indicated by other sensory input. In most VR applications, for example, at least two competing *egocentric reference frames* are present: On the one hand, the intended or simulated one, that is, the reference frame of the virtual environment. On the other hand, participants are embedded in the physical reference frame of the simulation room. Hence, the *egocentric reference frames* depend critically on *spatial perception*: *egocentric reference frame* $\Rightarrow$ *spatial perception,* because without (typically multi-modal) perception we would not have the basis for the perceived egocentric perspective. This connection is not further specified here, but is supposed to cover the dependency on multiple modalities.

Consistency [GOAL/DESIRED PROPERTY] As stated in the introduction, we propose the overall goal of a spatiotemporally consistent relation between oneself and the surround.

Consistency Check [PROCESS]   In connection to an existing *egocentric reference frame* and the overall goal of *consistency*, we propose the notion of a *consistency check*: At any moment, we should have one and only one consistent mental reference frame that defines our perceived ego-position in the world. That is, both an *egocentric reference frame* and *consistency* are necessary prerequisites for a *consistency check*. Conversely, without the overall goal of *consistency* and the existence of the data structure (*egocentric reference frame*) there would be no process checking for consistency: *Consistency check ⇒ egocentric reference frames* and *consistency check ⇒ consistency*.

This *consistency check* is related to *spatial presence & immersion*: When directly perceiving the real world, we typically feel spatially present. Total spatial presence can thus be considered the "default". If the perceived stimuli can be consistently embedded in the primary reference frame, everything is fine, and spatial presence (intenseness of being there) will be high. If, on the other hand, the perceived stimuli cannot be consistency embedded into the same primary reference frame, the intensity of the primary reference frame might be reduced and "breaks in presence" (BIP, see Slater (2002)) can occur. For example, if you are in the midst of a dream and the telephone rings, you will either incorporate the ringing into your dream, or else you will probably wake up. That is, either the primary reference frame (the dream) continues to dominate the secondary reference frame (the physical surround), or a break in presence (and sleep) will occur and and you will wake up. In that moment, the primary and secondary reference frame will be swapped, and the real world will take over. The equivalent can occur in VR simulations: Any event from the physical surround that cannot be integrated into the virtual world competes with the simulation and will be detected by the *consistency check*, thus disturbing presence.

Spatial Presence & Immersion [GOAL/DESIRED PROPERTY]   *Spatial presence* can be regarded as the consistent feeling of being in a specific spatial context, and intuitively knowing where one is with respect to the immediate surround. *Immersion*, on the other hand, could be seen as the subjective feeling of being fully drawn into that spatial context. For the sake of simplicity, however, we do not distinguish between *spatial presence* and *immersion* in this framework and therefore put them into the same box in Figure 45.

Spatial presence & immersion requires the functioning and positive outcome of the *consistency check* of the *egocentric reference frame*: If there is no agreement on one single (consistent) reference frame at a time, we cannot be fully immersed in the spatial situation (Regenbrecht, 1999) (*spatial presence & immersion ⇒ consistency check*). Furthermore, without the knowledge of some egocentric spatial reference frame, we would obviously not be able to immerse into anything (*spatial presence & immersion ⇒ consistency check ⇒ egocentric reference frame*).

In Virtual Reality applications, we can perceive high *spatial presence & immersion* only if the simulated world is consistently accepted as the only reference frame. That is, in order to be fully immersed and spatially present in the simulated world, one has to "forget" about the physical reference frame of the simulator (which would constitute a second, conflicting reference frame) or else the consistency check would detect a conflict.

If one wishes to logically distinguish between *spatial presence* and *immersion*, we would propose to see *immersion* as a logical prerequisite for *spatial presence*, in the sense of *spatial presence ⇒ immersion*. That is, no *spatial presence* without *immersion*. This proposition is in agreement with the so-called "book problem" in presence research (e.g., Schubert, 2002): When reading a book, the reader can be drawn into the book and feel immersed without feeling spatially present at the described location (but not the other way around). It appears to us as if immersion might be closely related to the well-studied phenomenon of "flow" states (Csikszentmihalyi, 1991). These are enjoyable states of consciousness where one is so completely focused and concentrated on one activity that it amounts to absolute absorption.

Obligatory Behavior (Reflexes) [PROCESS]   Here, we would like to introduce something which can actually be measured directly: the process of *obligatory behavior (reflexes)*, which cannot easily be voluntarily suppressed. For example, people with fear of heights cannot help but be afraid if they stand close to an abyss. The same is true for fear of flight or fear of narrow spaces[15]. For example, people with arachnophobia (fear of spiders) might not like to look at pictures of spiders, but that would most certainly not elicit any spatial response like running away. Only if the spider is in a spatial context and crawling towards them would they react spatially by trying to escape. In sum, *obligatory behavior* in this context is meant to refer to compulsory behavior that is elicited by a spatial context or situation. That is, it would seem most natural for us to dodge away if an unknown object flies at high speed towards our head.

One critical point in those situations is to believe the actual danger - that is to feel immersed and spatially present: *Obligatory behavior* ⇒ *spatial presence & immersion*. Without the immersion and spatial presence, the obligatory response is not elicited. This means for example that people with fear of heights do not feel that fear if the are not fully immersed into the situation of, e.g., standing at the edge of a cliff (Regenbrecht, 1999). Conversely, if we observe intact reflexive behavior, the participant was spatially present and immersed. That is, *spatial presence & immersion* can be quantified indirectly be measuring obligatory spatial behavior.

It is to be noted, however, that for phobic people, merely *imagining* a fear-inducing situation can elicit all characteristics of a panic attack. Here, we would argue that they feel fully immersed in their *imagined* environment. This suggests that in extreme cases, our framework can operate on purely imagined space, too.

Continuity [GOAL/DESIRED PROPERTY]   As mentioned in the introduction, one of the overall desired properties of perception is the apparent *continuity* of the perceived stimulus in particular and the world in general (at least for self-initiated ego-motions). We propose that this property can be seen as the guiding goal of the overall system.

Continuous Spatial Updating [PROCESS]   When we move, all spatial relationships between ourselves and the environment change. Nonetheless, we feel immersed in the current surround and naturally experience spatial presence. Apparently, some robust process continuously updates these self-to-world relationships as we move: This *continuous spatial updating* process refers to the incremental transformation of our *egocentric reference frame* based on relative positional and rotational information. That is, it can operate without any landmarks, by incrementally updating the *egocentric reference frame* using perceived velocity, acceleration, and relative displacements. Blindfolded walking with ears muffled is the stereotypical example for this process.

More specifically, convincing *ego-motion perception*, *spatial presence & immersion*, *egocentric reference frame* as well as *continuity* are necessary prerequisites for *continuous spatial updating*. Simply put, we cannot update any ego-position if we cannot perceive its changes (*continuous spatial updating* ⇒ *ego-motion perception*). This part is often understood as path integration. Furthermore, we cannot update to a new location in space if we are not already spatially present at any location beforehand and possess a corresponding *egocentric reference frame* - otherwise there would be nothing to update (*continuous spatial updating* ⇒ *spatial presence & immersion* and *continuous spatial updating* ⇒ *egocentric reference frame*). Finally, a continuous update is only possible if the sequential changes are continuous in time and space (*continuous spatial updating* ⇒ *continuity*). Without

---

[15]We refer to such phobias as *obligatory behavior (reflexes)* as they are largely beyond conscious control – else they would not constitute a problem for the phobic person. This reflexive character of phobias is also the reason why they typically do not disappear without proper therapeutic treatment.

continuous spatial updating, the *egocentric spatial reference* frame would become increasingly misaligned, which would eventually lead to a discontinuity the next time *instantaneous spatial updating* re-aligned the *egocentric reference frame* (see below).

Expected Egocentric Reference Frame [DATA]   Executing all possible behaviors in order to test their potential outcome is very inefficient. A more efficient approach would be to automatically predict and imagine what we would perceive if we would perform a certain movement. In this manner, we generate an expectation of what we should perceive if we had actually performed that motion. Moving in space is in this sense very predictable by the organism and therefore we hypothesize: *expected egocentric reference frame ⇒ continuous spatial updating* in the sense that without *continuous spatial updating* one would not be able to predict the changed percept of the world.

Reality Check [PROCESS]   Once we have an expectation of what we ought to perceive for a given motion, we can compare the actual percept to the predicted one. That is, we need both an *expected egocentric reference frame* and *spatial perception* to allow for the *reality check* (*reality check ⇒ expected egocentric reference frame* and *reality check ⇒ spatial perception*). If they match, everything is fine, and the reality check process will probably not come to consciousness or require any attention. If not, this might require some attention or action, that is, we might for example want to look again to make sure that everything is okay or allocate some cognitive resources to resolve the mismatch or act appropriately. An example might help to illustrate this point.

If we walk on ice and slip, the outcome of our behavior and motion (slipping) no longer matches the expectation (walking). The *reality check* detects this discrepancy and brings it to consciousness and alerts us. This is necessary to respond appropriately and prevent one from falling.

This double-checking is the obvious connection to *spatial perception*. One rather far-fetched hypothesis would be to propose: *spatial perception ⇒ reality check*, implying that we can only perceive *if* we expect and maybe even *what* we expect. Naturally, this cannot be sufficient to explain perception, but it sheds a new light on change blindness results - even considerable changes in our surround go unnoticed if we do not expect them to occur (Simons & Levin, 1997).

Spatial Learning [PROCESS]   If the *reality check* encounters an unexpected event, there might be something we could learn from this discrepancy. Since the organism cannot predict everything right from the start, its internal prediction model needs to be developed through learning. Many learning algorithms as understood in the neurosciences require an error signal, which can be defined as the difference between stimulus and prediction.

As we are concerned here with *spatial* behavior only, we would like to constrain ourselves to *spatial learning*. *Spatial learning* can be seen as the process of building up and modifying spatial knowledge, i.e., the process which operates over time on the *allocentric spatial memory* (see below). We hypothesize that spatial learning requires either a *reality check* or at least one of the four spatial orientation processes (*spatial learning ⇒ (reality check ∨ any spatial orientation process)*).

Four examples might help to illustrate this point. Homing experiments without landmarks (Loomis et al., 1999; Klatzky et al., 1997) are the stereotypical example for learning how to find home based on relative motion information and *continuous spatial updating* only (left branch). There are no compelling real world examples where only *instantaneous spatial updating* is used for spatial learning. Rapid serial presentation of images of an unknown scene might be a way to test if *instantaneous spatial updating* can nevertheless be used for spatial learning. When driving through unknown environments, landmark based large-scale navigation (*piloting*) is probably the predominant spatial orientation process that helps us to learn the new environment. An example involving

higher cognitive spatial orientation processes is learning an environment from abstract knowledge like maps.

Allocentric Spatial Memory [DATA]   Through *spatial learning*, we can acquire *allocentric spatial memory*, e.g., spatial memory in the form of a "cognitive map" allowing for novel shortcuts (see, e.g., Poucet, 1993; Tolman, 1948; Trullier, Wiener, Berthoz, & Meyer, 1997) . Therefore, *spatial learning* can be seen as an ongoing process operating on the knowledge stored in *allocentric spatial memory*. We would like to state that learning and memory are tightly coupled, require one another and thus cannot be strictly separated. We express this as a direct coupling (equivalence on the logical, but of course not on a functional level) between *spatial learning* and *allocentric spatial memory*.

Object/Landmark Memory [DATA]   Having described the path integration-based left branch of the framework, we will now discuss the more static, absolute location-based branch. *Object/landmark memory*, which is the most basic data structure in our framework, contains knowledge about objects and landmarks *without* their spatial context or relationships. This data structure is needed for, e.g., object recognition (see below). We do not assume any preferred storage format, but presume that we cannot built up any knowledge of spatially extended objects or landmarks without some kind of *spatial perception* (*object and landmark memory* $\Rightarrow$ *spatial perception*).

Identification [PROCESS]   Having the ability to store knowledge about objects and landmarks, it makes sense to demand some recognition process which can identify objects, in order to label them as individuals and potentially recognized them later. This *identification* process can be seen as the "what path" in the perception model by Mishkin, Ungerleider, & Macko (1983). The logical relation here is as follows: *identification* $\Rightarrow$ *object and landmark memory*. In other words, if one cannot remember any objects, it should not be possible to recognize and to identify them later.

Localization [PROCESS]   As soon as we perceive anything spatially, we can localize it even without necessarily being able to identify it. That is, the *localization* process does not assume any attribution of identity. One could compare this to the "where path" in the Mishkin et al. (1983) model of perception. The logical relation between these two terms is: *localization* $\Rightarrow$ *spatial perception*. In other words, without any *spatial perception* we could have no *localization* process (i.e., $\neg$ *spatial perception* $\Rightarrow$ $\neg$ *localization*).

Instantaneous Spatial Updating [PROCESS]   As introduced in subsection 16.2, *instantaneous spatial updating* refers to the process of re-aligning or re-anchoring the mental spatial reference frame to the surround using position-fixing via landmarks (*instantaneous spatial updating* $\Rightarrow$ *egocentric reference frame*).

This process can be triggered by, for example, haptic, auditory, and, probably most frequently, visual landmarks. *Instantaneous spatial updating* is thus critically depending both on the *localization* and *identification* process: *Instantaneous spatial updating* $\Rightarrow$ *localization* process means that it would not make sense to re-anchor the mental reference frame if we were not sure about the exact coordinates to use. Moreover, *instantaneous spatial updating* $\Rightarrow$ *identification* means that it would not make sense to re-anchor the mental reference frame if we could not recognize anything familiar that told us where we were. Furthermore, we propose that *spatial presence & immersion* is a necessary prerequisite for automatically triggering *instantaneous spatial updating*, just as it was for *continuous spatial updating* (*instantaneous spatial updating* $\Rightarrow$ *spatial presence & immersion*).

Piloting [PROCESS]   Position- or recognition-based navigation (also called *piloting*) uses extero-ceptive information to determine one's current position and orientation. Such information sources include visible, audible or otherwise localizable and identifiable reference points, so-called land-marks (i.e., distinct, stationary, and salient objects or cues). This implies *piloting ⇒ localization* and *piloting ⇒ identification*. Many studies have demonstrated the usage and usability of different types of landmarks for navigation purposes, (see Golledge (1999), Hunt & Waller (1999) for an extensive review). Piloting allows for correction of errors in perceived position and orientation through refer-ence points (position fixing) and is thus well-suited for large-scale navigation. Piloting mechanisms often used include scene matching or recognition-triggered responses. Compared to *instantaneous spatial updating*, *piloting* is neither reflex-like nor automated, and does not require any aligned ego-centric reference frame. Note that no higher cognitive processes are needed for piloting, as even simple robots can use for example snapshot-based piloting for navigation (Franz, Schölkopf, Mallot, & Bülthoff, 1998).

Spatial Orientation [GOAL/DESIRED PROPERTY]   The main overall goal of the system de-scribed by the framework is in this context, as stated above, proper *spatial orientation*, which is essentially the ability (not the behavior itself) to easily find one's way around.

Spatial Behavior [PROCESS]   Last but not least, we seem to have all basic ingredients to define *spatial behavior* as behavior performed in space and time and at the same time relying on spatial knowledge about the world.

First of all, it seems plausible to assume *spatial behavior ⇒ spatial learning*: Without learning spatial knowledge, we would not be able to adapt to new situations and find our way around in a novel or changing environment. That is, we propose that *spatial learning* is required for the **adaptability** of *spatial behavior*.

As spatial behavior (especially in animals) is typically quick and intuitive, many of the required computational processes need to be largely automated. Hence, we propose that automatic spatial updating is a necessary prerequisite for **quick & intuitive** *spatial behavior*. Therefore, we propose that quick and intuitive *spatial behavior ⇒ continuous spatial updating* and/or *spatial behavior ⇒ instantaneous spatial updating*. Consequently, quick and intuitive *spatial behavior* should not be possible without either *continuous spatial updating* or *instantaneous spatial updating* or both being operational. As both continuous and instantaneous spatial updating logically imply *spatial presence & immersion*, we hereby indirectly claim that *spatial behavior ⇒ spatial presence & immersion*. In other words, when we do not feel ourselves at a specific location and orientation, we cannot interact with the world in a natural and effortless manner. Hence, we proposed indirectly that *spatial presence & immersion* are required for quick and intuitive *spatial behavior*.

For the consistency of this model, we would like to exclude for the time being behavior that can be modeled by simple direct coupling of perception and action, without any spatial knowledge (e.g., Braitenberg vehicles (Braitenberg, 1984)). Instead, we limit our view of *spatial behavior* as such being motivated by and thus depending on good *spatial orientation*. Without *spatial orientation* we are not able to perform the required *spatial behavior* (*spatial behavior ⇒ spatial orientation*). Consequently, *spatial behavior* can be used to measure and evaluate the successful *spatial orientation* in psychological experiments.

Obviously enough, spatial behavior should be most accurate and precise if we can recognize and localize unique reference points. As *instantaneous spatial updating* as well as *piloting* are the two processes relying on the *localization* and *identification* of such landmarks, we propose that at least one of them has to work for us to have accurate and precise spatial behavior. Hence, we propose

that **accurate & precise** *spatial behavior* ⇒ *instantaneous spatial updating* or *spatial behavior* ⇒ *piloting*.

Having identified specific items that are required for different aspects of *spatial behavior* (accurate & precise, adaptable, and quick & intuitive spatial behavior), we are enabled to analyze spatial or experimental situations accordingly: If the observed spatial behavior is, for example, accurate and precise, but response times are long and participants report not having much of an intuitive spatial orientation, we could conclude that *piloting* (the landmark-based static right branch of the framework) is intact, whereas *continuous spatial updating* as well as *instantaneous spatial updating* are probably largely impaired. This might in turn, for example, be due to the lack of convincing *spatial presence & immersion*.

This case is elaborated upon in more detail in subsection 17.1.1 in the context of Experiment LAND-MARKS. Note that cues for *instantaneous spatial updating* can potentially stem from all sensory modalities that provide useful landmarks. Landmark cues can obviously be visual, but also auditory (e.g., the phone ringing when being disoriented in darkness), haptic (e.g., touching a door knob), or maybe even olfactory (e.g., the smell of milk boiled over on the kitchen stove).

Conversely, if the observed spatial behavior is quick & intuitive but lacks accuracy and precision, we would argue that automatic *continuous spatial updating* was working, but neither *instantaneous spatial updating* nor *piloting* were intact. Thus, the central and left relative motion-based part seem to be intact, whereas the absolute location-based right branch is not. Examples for this case include blindfolded walking, getting lost in deep forest, and of course visually induced vection in an optic drum. Note that sensory cues that might allow for *continuous spatial updating* include vestibular cues (accelerations), proprioceptive cues (e.g., from walking), but also visual or auditory cues from optic or acoustic flow, respectively.

16.4 Where does cognition fit into the model?

So far we have attempted to lay out a consistent framework based on logical connections between related items. The contribution of higher cognitive processes or strategies has so far not been taken into consideration. Moreover, especially the lower part of the framework seems to be largely beyond conscious control: For example, even if we might consciously decide to do so, it is virtually impossible to influence *identification* (not recognize your friend's face) or *ego-motion perception* (consciously elicit the convincing sensation of ego-motion).

So where does cognition fit into this model? By its very nature, cognition is flexible and versatile and consequently cannot simply be represented as one box logically dependent on other boxes. Rather, cognition might be considered as an optional process that can be resorted to if the partly automated framework fails or does not allow for the desired spatial behavior. That is, we have conscious access to, for example, the lower items of the framework (*motion perception*, *localization*, and *identification*), even though we cannot consciously control them. Hence, we can for example consciously question *motion perception* to cognitively derive the simulated displacement, even though we might not perceive any *ego-motion*. See subsections 17.1.2 and 17.1.3 for a more detailed discussion of this phenomenon. We are, however, unable to use this abstract knowledge about the simulated turning angle to intentionally evoke the percept of convincing ego-motion. That is, the lower items in the framework can be queried, but are nevertheless to a large degree beyond conscious control.

Cognition [PROCESS] Ultimately, this leads to a fourth connection to *spatial behavior*: Higher cognitive processes (*cognition*) can be used to develop novel strategies to solve a complex navigation

problem, or to use mental spatial reasoning or spatial imagination to derive the desired spatial behavior (see subsections 8.2.3, 11.1.4, 10, 17.1.2 and 17.1.3). For example, finding the shortest route in a subway system might require rather advanced cognitive processing.

Cognition can consequently be considered a necessary condition for *spatial behavior* based on non-automated **abstract strategies**, **mental spatial reasoning**, and **imagination**. This can be represented in the framework as **mentally mediated** *spatial behavior* ⇒ *cognition*. Due to the inherent flexibility of *cognition*, however, there are no other fixed links to *cognition*. Rather, cognition can be used to flexibly query the desired information from most or maybe even all of the other items of the framework. Hence, if we observe spatial behavior that is neither quick & intuitive nor very accurate & precise, we could argue that the behavior might have been based on abstract cognitive strategies. As mental geometric reasoning can lead to quite accurate and precise spatial behavior, we propose *cognition* as a third possibility for achieving accurate and precise *spatial behavior* (apart from *instantaneous spatial updating* and *piloting*): **Accurate & precise** *spatial behavior* ⇒ *(cognition* ∨ *piloting* ∨ *instantaneous spatial updating*).

16.5   Ways to measure spatial presence and immersion

Until very recently, quantifying presence and immersion has been typically attempted using highly subjective and introspective methods like questionnaires (Hendrix & Barfield, 1996a, 1996b; IJsselsteijn, de Ridder, Freeman, Avons, & Bouwhuis, 2001; Lessiter, Freeman, Keogh, & Davidoff, 2001; Regenbrecht, 1999; Regenbrecht & Schubert, 2002; Schloerb, 1995; Witmer & Singer, 1998). These methods were an important first step towards understanding the nature and relevance of presence and immersion for many applications, but share certain undesirable side-effects. All introspective measures have to somehow explicitly question the participant, which in itself can reduce presence and immersion. Questionnaires in particular do not allow for online measures in the spatial context, as they are used after the exposure. In the following, we would like to sketch novel quantification methods that do not rely on introspection but rather on psychophysical measures. They complement the existing methodologies and might allow for more sensitive and reliable online measures even without the participant noticing the measurement. How those results relate to subjective measures still remains an open question.

Having embedded *spatial presence & immersion* into a logical framework allows us to devise new quantification methods by either measuring all necessary prerequisites or, even more elegantly, measuring at least one sufficient condition. As we have seen in the previous section, spatial presence is embedded into a collection of processes with useful and testable properties. We found three sufficient but not necessary prerequisites of spatial presence: *continuous spatial updating, instantaneous spatial updating*, and *obligatory behavior*. In addition, we have one necessary, but not sufficient, prerequisite (*consistency check*). Having laid out the logical framework, we can now use this prerequisite to measure presence: The degree of mismatch between the primary egocentric reference frame and other potentially conflicting reference frames becomes a proposed measure for spatial presence. The actual measurands are the reference frames from different modalities and the potential mismatch between them by appropriate psychophysical methods.

Furthermore, certain spatial behaviors seem impossible without sufficient *spatial presence & immersion*. Measuring the functioning of *obligatory behavior* is a feasible and currently discussed method to quantify *spatial presence & immersion*. In the same line of reasoning, effortless *continuous* or *instantaneous spatial updating* cannot occur without sufficient *spatial presence & immersion*. Following the logical chain further up in our model, we see that spatial updating (continuous or instantaneous) is a necessary prerequisite for quick and intuitive *spatial behavior*. Conversely, the observation of such quick and intuitive *spatial behavior* implies automatic spatial updating and consequently also *spatial presence & immersion*. Those examples represent indirect measures of spatial

presence that can readily lead to novel experiments complementing current presence research. In part III, for example, we developed a rapid pointing paradigm that does not allow participants enough time to use *piloting* or *cognition*. In that manner, possible spatial orientation processes were reduced to *spatial updating*. Since the usage of at least one of the two *spatial updating* processes implies *spatial presence & immersion*, the results indirectly reflect the degree of *spatial presence & immersion*.

16.6   Further hypotheses about logical relations

So far we tried to sketch a clear chain of logical connections which can be summarized as *spatial behavior* $\Rightarrow$ *spatial perception*, which is plausible per se (see Figure 44). In addition to some assumptions we had to make in laying out our string of arguments, we would now like to introduce two hypothetical additional loops. We propose that *spatial presence & immersion, continuity*, *ego-motion perception* and an *egocentric reference frame* **together** are sufficient to enable proper *continuous spatial updating* (*spatial presence & immersion $\wedge$ continuity $\wedge$ ego-motion perception $\wedge$ egocentric reference frame $\Rightarrow$ continuous spatial updating*). In other words, continuous spatial updating should work if all four prerequisites are true. Conversely, if we observe impaired *continuous spatial updating*, then we can conclude that at least one of the prerequisites is violated. ($A \wedge B \wedge C \wedge D \Rightarrow E$ is equivalent to $\neg E \Rightarrow \neg A \vee \neg B \vee \neg C \vee \neg D$).

Taken together with the previously established logical connections (*continuous spatial updating $\Rightarrow$ spatial presence & immersion $\Rightarrow$ consistency check $\Rightarrow$ egocentric reference frame*) $\wedge$ (*continuous spatial updating $\Rightarrow$ continuity $\Rightarrow$ egocentric reference frame*) $\wedge$ (*continuous spatial updating $\Rightarrow$ ego-motion perception*), we can furthermore conclude the following: if any of the four prerequisites is violated, *continuous spatial updating* would be rendered impossible or at least largely impaired ($\neg A \vee \neg B \vee \neg C \vee \neg D \Rightarrow \neg E$). Together with the above argument, this leads to the following equivalence: $\neg E \Longleftrightarrow \neg A \vee \neg B \vee \neg C \vee \neg D$, which is the same as saying that $E \Longleftrightarrow A \wedge B \wedge C \wedge D$. In other words, this means that instead of measuring *continuous spatial updating*, we can measure *consistency check $\wedge$ spatial presence & immersion $\wedge$ egocentric reference frame $\wedge$ ego-motion perception*.

Furthermore, as *spatial presence & immersion* implies both *consistency check* and *egocentric reference frame*, measuring *continuous spatial updating* is equal to measuring *spatial presence & immersion $\wedge$ ego-motion perception*. This opens up many interesting experimental investigations. For example, *spatial presence & immersion* can be quantified by measuring *continuous spatial updating* and *ego-motion perception* and vice versa.

A very similar second loop is located in the absolute location-based right part of the framework. Experiment SIMULATION PARAMETERS showed that merely presenting an image of a new orientation in the "teleport" condition without any motion information whatsoever can be sufficient to trigger obligatory spatial updating. Therefore, we propose that *spatial presence & immersion $\wedge$ egocentric reference frame $\wedge$ localization $\wedge$ identification $\Rightarrow$ instantaneous spatial updating*. Following the same reasoning as before, this opens up the possibility to measure *instantaneous spatial updating* instead of *spatial presence & immersion $\wedge$ localization $\wedge$ identification*. Even more pragmatically, one could use standard psychophysics to measure the latter two of the conditions (*localization $\wedge$ identification*) as well as the new method of quantifying *instantaneous spatial updating* introduced in SIMULATION PARAMETERS in order to quantify *spatial presence & immersion* in quasi-static situations.

16.7   Discussion

So far, we have not attempted to relate each item in the framework to the corresponding functional relations and information flow. Many of the proposed connections may indeed be closely linked to corresponding processing steps and neural connections in the human brain. Most of the boxes might also be considered as being localized in specific brain regions. There is for example a large body of literature arguing that the hippocampus is critically involved in path integration as well as landmark-based navigation and cognitive maps in animals including humans (Berthoz, 1997; Maguire et al., 1998b, 1998a; McNaughton et al., 1996; Mittelstaedt, 2000; O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978; Poucet, 1993; Samsonovich & McNaughton, 1997). Furthermore, ego-motion perception seems to be closely linked to the intraparietal sulcus (IPS) in humans and the equivalent area (ventral intraparietal area (VIP)) in macaque monkeys (Bremmer, Klam, Duhamel, Ben Hamed, & Graf, 2002; Bremmer et al., 2001). Trying to associate all the individual boxes and logical connections of the current framework with corresponding neural substrate would be a challenging, as well as a promising, endeavor. It goes, however, well beyond the scope of this paper.

So, can we measure spatial presence now? As Wijnand IJsselsteijn, one of the leading researchers in the presence community phrased it: "Presence needs to be unambiguously operationalised, and subdivided into its basic components in order for it to be measurable in a way that will make sense." (IJsselsteijn, 2002). In our paper, we attempted this by embedding spatial presence into a logical framework. This allowed us to operationalize spatial presence through a set of necessary and/or sufficient conditions for spatial presence. Instead of subdividing spatial presence itself into its basic components, however, we analyzed related processes. This allowed us to generate a number of testable predictions and measurement paradigms for spatial presence itself as well as for related issues like continuous and instantaneous spatial updating.

We are aware that many factors can potentially affect spatial presence. Examples include the (actual or assumed) ability to explore the virtual surround, to interact with it, and to predict the outcome of one's actions (Regenbrecht, 1999; Regenbrecht & Schubert, 2002; Schubert, Friedmann, & Regenbrecht, 2001). Narrative components and dramatic effects are further factors that have been shown to enhance spatial presence (Regenbrecht, 1999). None of these factors alone, however, seems to be absolutely required in the sense of spatial presence logically implicating that factor (e.g., *spatial presence $\nRightarrow$ dramatic effects*). Thus, those and many other potentially influential factors are missing in our framework. The same is of course true for the other items of the framework.

We hope that the proposed framework will stimulate the scientific discussion and help to clarify our reasoning and discussions, especially when such loosely defined terms as spatial presence, immersion, or spatial updating are involved. Only future research, however, will enable us to rigorously test the proposed logical framework and refine or extend it where appropriate.

In summary, we embedded current terminology from the field of spatial orientation in a functional and logical framework. This framework covers aspects ranging from spatial perception over allocentric and egocentric spatial memory up to spatial behavior. Finally, we used this framework to generate hypotheses which can guide future research and can be experimentally tested.

17 Summary of the experiments, applications of the framework, and conclusions

In this section, we will discuss all experiments described in this paper in the context of the theoretical framework described in the previous section. On the one hand, this serves as a first test for the applicability and consistency of the framework. On the other hand, analyzing the logical connections for each experiment might allow for a deeper understanding of the underlying processes and hence of the experimental outcome. In the following, the individual experiments will be briefly summarized before revisiting and discussing them in the light of the framework. For each subsection, the graphical representation of the framework will be adapted to capture the items and connections that are intact or not.

17.1 Navigation and spatial orientation experiments

In this subsection, the framework will be applied to the navigation and spatial orientation experiments in part II. First, the triangle completion experiment with reliable landmarks will be investigated (subsection 17.1.1, Experiment LANDMARKS), followed by the triangle completion experiment with only temporarily available landmarks (subsection 17.1.2, block TOWN of Experiment TOWN&BLOBS). We will conclude with the experiments that were based on optic flow usage (subsection 17.1.3, Experiment TURN&GO, RANDOM TRIANGLES, and block BLOBS of Experiment TOWN&BLOBS).

17.1.1 Navigation experiments with reliable landmarks: LANDMARKS

In Experiment LANDMARKS (section 7), participants performed triangle completion tasks in front of the half-cylindrical projection screen which presented a highly consistent and reliable virtual scene of a town. As expected, participants used piloting and scene matching, which allowed for almost perfect homing performance. That is, the *spatial behavior* of triangle completion was accurate and precise, indicating that at least the landmark-based right part of the framework must have been used.

Regarding the basic components of the lower part of the framework, it is reasonable to assume that *spatial perception, motion perception, egocentric reference frame, localization, object/landmark memory* and *identification* were all operational (see Figure 46).

Note however, that two competing *egocentric reference frames* were present: On the one hand, the intended or simulated one, that is, the reference frame of the virtual environment. On the other hand, participants were embedded in the physical reference frame of the simulation room. That is, they were seated in front of an immobile table, hearing the low whirr of the air conditioning from behind, and seeing the surrounding static room with a static screen (see Figure 1, page 11). This static reference frame is both *perceived* to be static as well as consciously *known* to be immobile.

Hence, our model predicts that the *consistency check* should detect this obvious conflict between the primary, intended reference frame of the Virtual Reality simulation and the secondary reference frame of the physical setup. Consequently, *spatial presence & immersion* should be somewhat impaired, and in turn also *continuous* and *instantaneous spatial updating* (cf. Figure 46). Finally, *spatial behavior* should thus be less quick and intuitive.

In sum, from the bottleneck of the four potential mechanisms that can in principle be used for spatial orientation (*continuous spatial updating, instantaneous spatial updating, piloting*, and *cognition*), the two spatial updating mechanisms are to some degree excluded, leaving us with *piloting* and *cognition*. *Piloting* is definitely possible, as it is according to our framework only depending on *localization* and *identification*, which are both assumed to be intact. *Cognition* is computationally more demanding
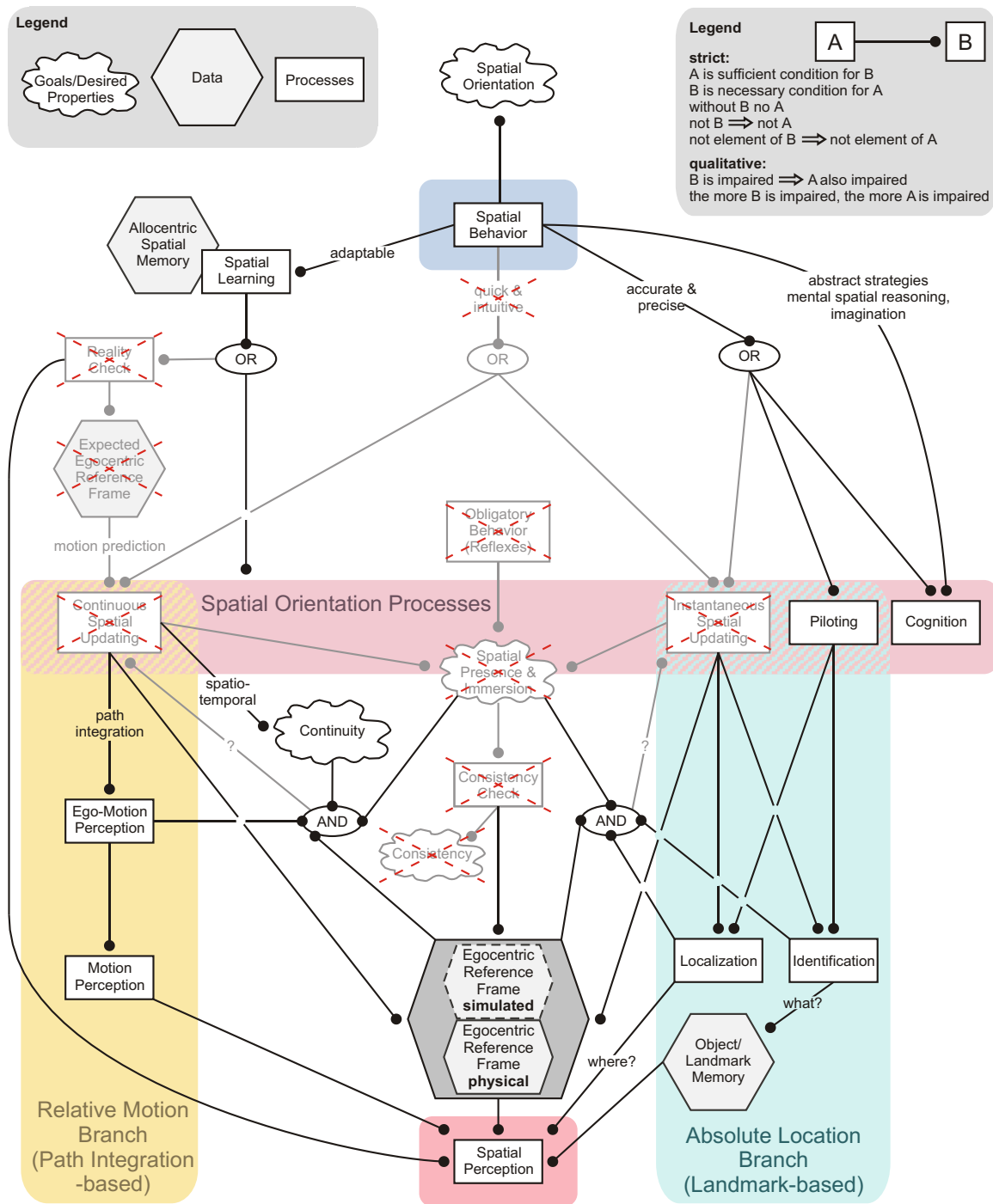
Figure 46: Conceptual framework applied to navigation experiments with reliable landmarks. The items and logical connections (arrows) that are intact for those experiments are drawn a solid black. Items that are impaired are gray-shaded and crossed out. Partial impairment is emblematized by dashed crosses, complete impairment by solid crosses. Note that only the landmark-based right part of the framework is operational. The path integration-based left part, which is based on integrating relative motion, is largely impaired due to the conflicting reference frames and probably also the lack of convincing ego-motion sensation.

than *piloting*, which suggests that *piloting* might be the prevailing mechanism used. The observed trajectories and verbal reports indicate indeed that piloting and especially scene matching were used by many participants. As there were typically no landmarks close enough to the starting position that could be used to identify it uniquely, however, some participants might have resorted to more abstract strategies and mental spatial reasoning to yield a higher homing accuracy. According to verbal reports, some participants did indeed use complex cognitive strategies like imagining straight lines connecting opposite landmarks, and using the intersection point of two imaginary lines to identify the starting position. Hence, we can conclude that both *piloting* and *cognition* were possible and indeed used in Experiment LANDMARKS, which explains the observed highly accurate and precise *spatial behavior*.

17.1.2 Navigation experiments with unreliable landmarks: TOWN&BLOBS

After successfully applying the framework to piloting-based navigation in the previous subsection, we will now proceed to the navigation experiment with only unreliable landmarks, that is, the TOWN part of Experiment TOWN&BLOBS. As in Experiment LANDMARKS, participants had to perform triangle completion tasks in the town environment. Before participants took the return path, however, all landmarks were exchanged and rearranged in a brief dark interval to form a different-looking scene. This scene swap paradigm effectively rendered all landmark-based homing mechanisms impossible, which resulted in turn in a considerable overall performance decrease. Averaged over participants, the mean turn response was still quite good and showed only small systematic errors. The distance responses, however, showed a considerable regression towards stereotyped responses, indicated by a gain factor of 0.6. Furthermore, both within- and between-subject variability were rather high. Using the three-stage navigation model presented in subsection 5.3, we argued that participants had essentially all the information needed (negligible encoding errors), but experienced considerable problems in mentally computing the desired homing response. As systematic execution errors were small, we concluded that the observed systematic navigation errors were mainly caused by the considerable errors in the mental spatial reasoning phase.

Next, the conceptual framework will be applied to see if it agrees with these rather complex results or might even explain the underlying processes. Up until the scene swap, the analysis is the same as for Experiment LANDMARKS, subsection 17.1.1 (see Figure 46): *Continuous* as well as *instantaneous spatial updating* are impaired due to the reduced *spatial presence & immersion* induced by the reference frame conflict. This reduces possible navigation mechanisms to *piloting* and *cognition*.

The scene swap, however, disrupts the spatiotemporal continuity of the scene, as all landmarks were exchanged and could consequently no longer serve as unique reference points. That is, the whole simulated *egocentric reference frame* had suddenly vanished and was replaced by a different one. This abrupt switch to a different reference frame implies an additional temporary lack of consistency, which in turn temporarily decreases *spatial presence & immersion*. As the scene remains spatiotemporally continuous and consistent until the next trial, *spatial presence & immersion* are expected to recover sooner or later to almost the same level as in the LANDMARKS Experiment.

More critically, the scene swap permanently replaced all landmarks that could have been used for identifying the origin of locomotion. Hence, any previously-established *allocentric spatial memory* of the original scene was now utterly useless (see Figure 47). Furthermore, *instantaneous spatial updating, piloting,* and all *cognition*-based strategies relying on landmarks were rendered impossible (as intended), suggesting that *spatial behavior* was no longer *accurate & precise*.

Participants still performed well above chance, however, so how did they perform the task at all? There seem to be at least two feasible strategies, both being based on *cognition*:
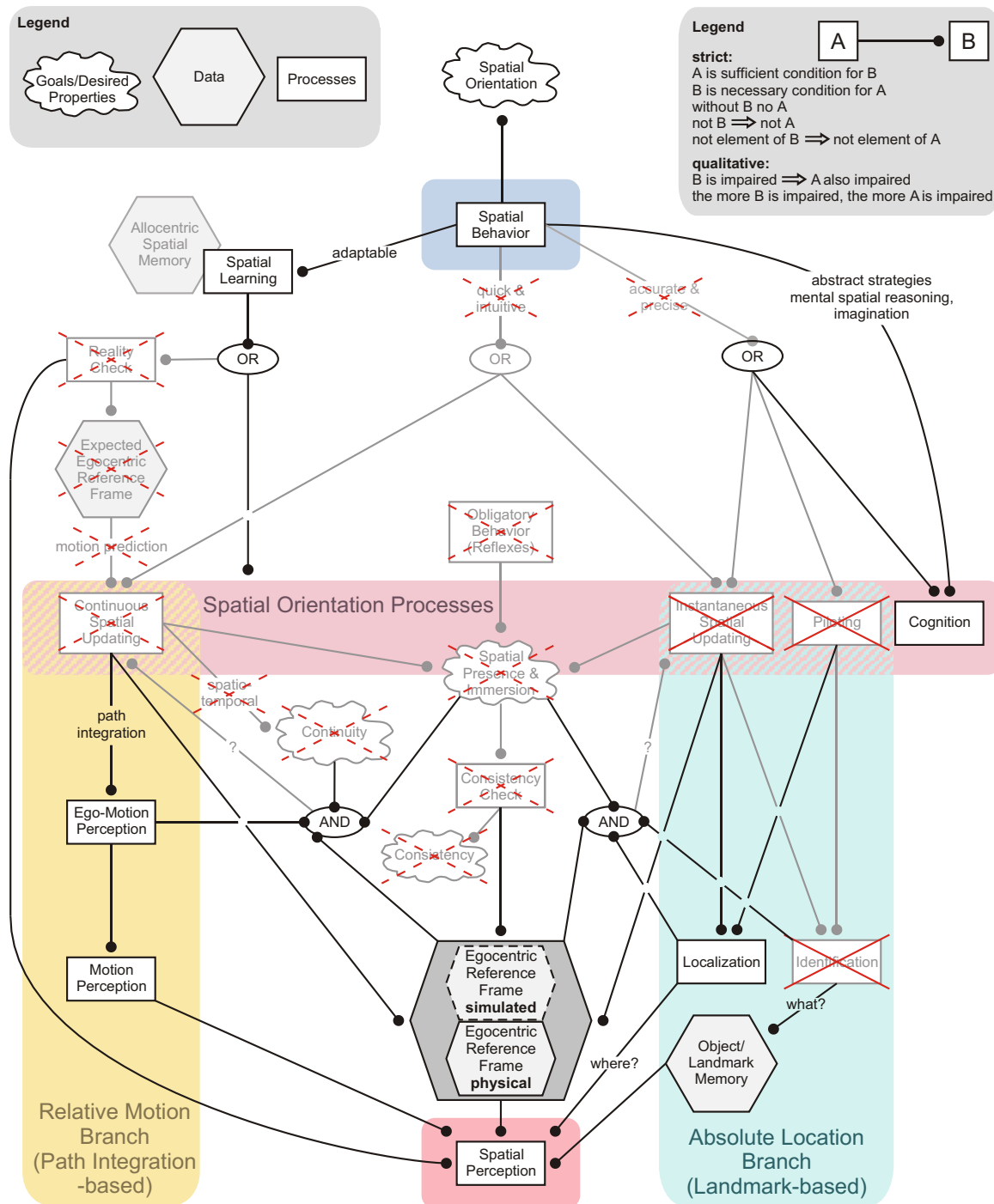
Figure 47: Conceptual framework applied to navigation tasks without reliable landmarks, using the example of the homing task of the TOWN part of the navigation Experiment TOWN&BLOBS with only temporarily available landmarks. Note that the landmark-based right branch of the framework (namely *instantaneous spatial updating* and *piloting*) is completely dysfunctional due to the scene swap. Furthermore, all previously built-up *allocentric spatial memory* became utterly useless after the scene swap. In addition, the upper parts of the path integration-based left branch (*Continuous Spatial Updating*) are impaired due to the conflicting reference frames. Hence, participants had to resort to *cognitive* strategies to solve the task.

1. As the origin of navigation was defined in the original reference frame before the scene swap, participants could on the one hand have tried to somehow transfer the representation of the origin from the first to the second scene. Probably the easiest way to do this would have been to use the consistent scene before the scene swap to derive some kind of homing vector, and transfer only this homing vector to the new scene. This usage of the original scene for solving the task was only possible in the TOWN condition of Experiment TOWN&BLOBS, but not in the BLOBS condition. Hence, if this strategy was used in the TOWN condition, performance should be better than in the BLOBS condition or at least different. There were, however, no performance differences between the two conditions, suggesting that this strategy was most likely not used.

2. On the other hand, participants might have merely used the geometry of the excursion path to derive the homing response. That is, they could have encoded the lengths s_1 and s_2 of the excursion and the enclosed angle α and thus mentally compute the correct turning angle $\overline{\beta}$ and homing distance s_3. This strategy is feasible with just *motion perception* and path integration by optic flow, and could thus explain the virtually identical performance in the TOWN and BLOBS conditions. This abstract geometric strategy would explain the rather long response times and qualitative errors discussed in subsection 12.1. It furthermore agrees with our earlier argument that the observed systematic errors can be mainly ascribed to the mental spatial reasoning phase (see subsections 5.3 and 11.1.4).

17.1.3   Navigation experiments based on path integration: TURN&GO, TOWN&BLOBS, and RANDOM TRIANGLES

In this subsection, the framework will be applied to the three navigation experiments that were based on path integration by optic flow: Experiment TURN&GO, the BLOBS condition of Experiment TOWN&BLOBS, and Experiment RANDOM TRIANGLES. The baseline Experiment TURN&GO investigated the usability of mere optic flow information for reproducing distances and executing turns. Turn execution was virtually void of systematic errors, and distance reproduction showed only a slight tendency towards stereotyped responses (gain = 0.91). We concluded that larger systematic errors in the other navigation experiment should thus be ascribed to errors in the encoding phase or mental spatial reasoning phase, but not the execution phase. The BLOBS condition of Experiment TOWN&BLOBS showed virtually the same performance as the TOWN condition as discussed above. Experiment RANDOM TRIANGLES investigated homing for completely randomized triangle geometries, to investigate the influence of the simplicity of the triangle geometry. Homing performance was by no means inferior to the TOWN&BLOBS performance with simple, isosceles triangles, indicating that participants did not take advantage of the simplicity of the triangle geometry.

In all those optic flow experiments, only the basic bottom part of the framework seems fully intact, as is sketched in Figure 48. That is, *spatial perception*, *motion perception*, and *localization* should all be fully functional. On the landmark-based right part of the framework, *object/landmark memory* and *identification* are already rather useless, as the scene is devoid of any landmarks that could be identified. Consequently, both *instantaneous spatial updating* and *piloting* can be excluded as possible spatial orientation mechanisms. On the other hand, the *egocentric reference frame* is rather sparse due to the lack of any landmarks. Together with the conflicting reference frame of the physical surround, this suggests that both *spatial presence & immersion* and *continuous spatial updating* are considerably impaired.

As both the relative motion-based left and absolute location-based right branch of the framework are largely impaired, participants could neither use intuitive spatial orientation skills (i.e., *automatic spatial updating*) nor *piloting*, but were forced to use information from low-level *motion perception*
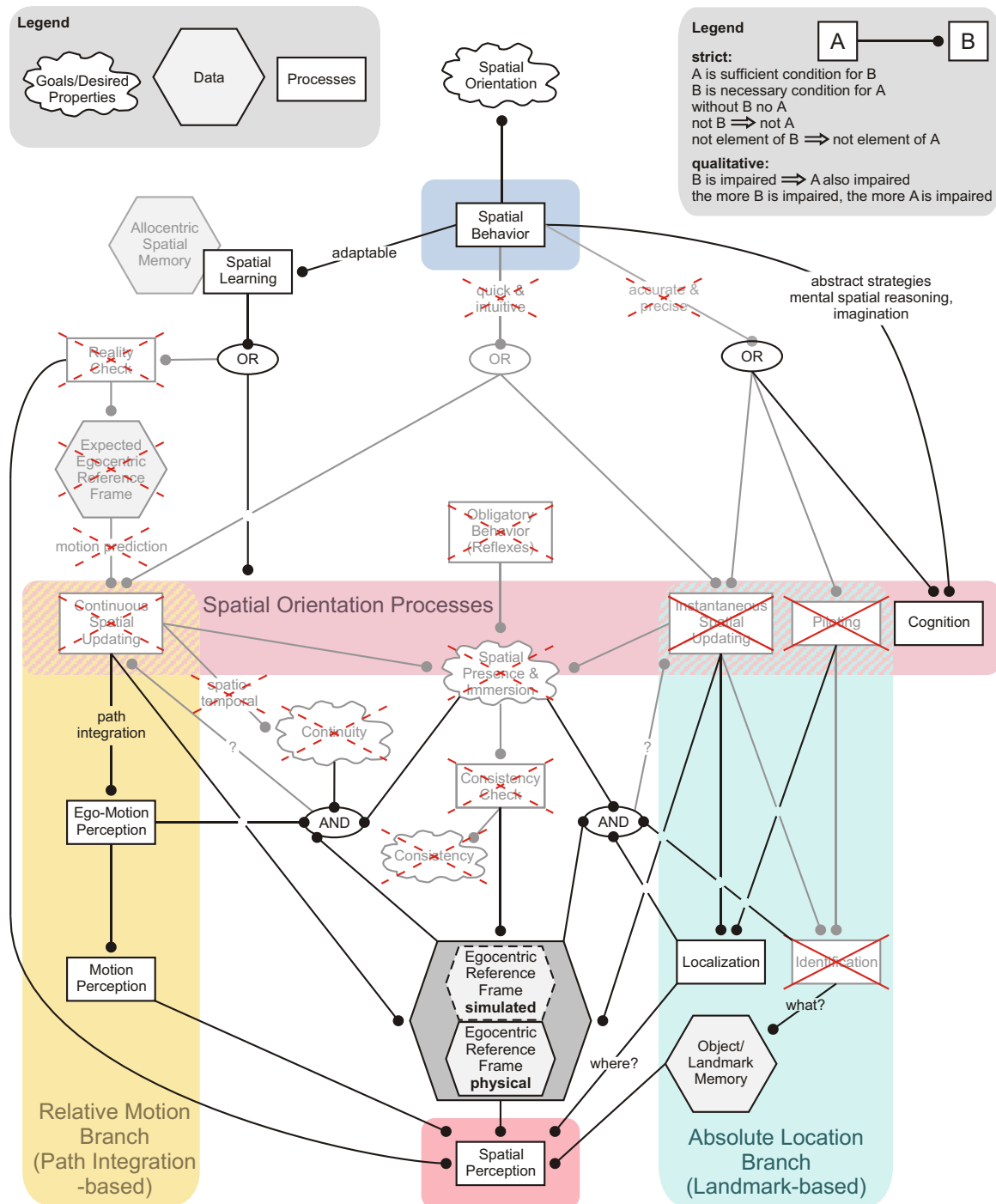
Figure 48: Conceptual framework applied to navigation experiments based on path integration via optic flow. The complete lack of any landmarks renders the landmark-based right branch dysfunctional. Consequently, the framework looks the same as in Figure 47 for unreliable landmarks. As before, the conflicting reference frames impair *continuous spatial updating* and hence the upper parts of the path-integration based left branch, leaving only cognitive strategies as a possible navigation mechanism.

to solve the task analytically, as described in the previous subsection. *Motion perception* provides information about the distances traveled and angles turned, and *cognition* (mental spatial reasoning) can be used to derive the correct homing response. This conclusion is in agreement with participants' verbal statements about the strategies used. All but one participant reported having tried to imagine a top-down, orthographic view of the triangle to derive the homing response or even to mentally compute the correct homing angle and distance. Only one participant tried to continuously update a homing vector. She turned out to be a biologist familiar with the animal homing literature and homing vector hypothesis. None of the participants reported that intuitive spatial orientation was by any means sufficient to solve the task well.

17.1.4 Conclusions

Even though the framework is still in a preliminary state and not thoroughly grounded in experimental evidence, it was nevertheless possible to apply it successfully to the navigation experiments described in part II. Revisiting the experiments in the context of a unifying framework allowed for a clearer understanding of the critical issues and underlying processes. The path integration-based right part of the model was generally impaired due to the obvious lack of *consistency* between simulated and physical reference frame. This resulted in a reduced *spatial presence & immersion* and consequently also an impaired *continuous spatial updating* as well as *instantaneous spatial updating*. This in turn explains why participants were unable to successfully use their intuitive spatial orientation skills and respond quickly.

Scene swap or optic flow usage, on the other hand, marred the absolute location-based right branch of the framework, thus rendering all landmark-based strategies unfeasible. Being deprived of their natural and automatized spatial orientation skills, participants had to resort to abstract cognitive strategies based on mental spatial reasoning on the basis of visual motion perception. This in turn might explain the long response times and qualitative errors (left-right confusions) discussed in subsection 12.1.

17.2 Spatial updating experiments - CONTROL and UPDATE conditions

In the following, the framework will be applied to the spatial updating experiments described in part III. In this subsection, the CONTROL and UPDATE conditions will be considered, followed by the IGNORE conditions in subsection 17.3. First, the conditions with useful landmarks and continuous motion information will be investigated (subsections 17.2.1), followed by the conditions without useful landmarks (subsection 17.2.2) and without any motion information ("teleport" condition, subsection 17.2.3).

17.2.1 Full cue conditions - Conditions with useful landmarks and continuous motion information

In the real world conditions of Experiment REAL WORLD VERSUS VR (block A & B), participants saw continuously the physical surround of the Motion-Lab, and there were no cue conflicts or multiple reference frames involved for the UPDATE and CONTROL trials. Together with the observed excellent pointing performance, this suggests that the full framework was operational, without any impairments. This is depicted in Figure 49. The blinders-restricted FOV in condition B reduced the pseudo-static CONTROL performance in all dependent measures, suggesting that the statically available information did not allow for the same excellent performance as in block A with unrestricted vision. As spatial updating performance was reduced, and *continuous spatial updating* is not expected to have a strong effect on the CONTROL condition, we conclude that *instantaneous spatial*
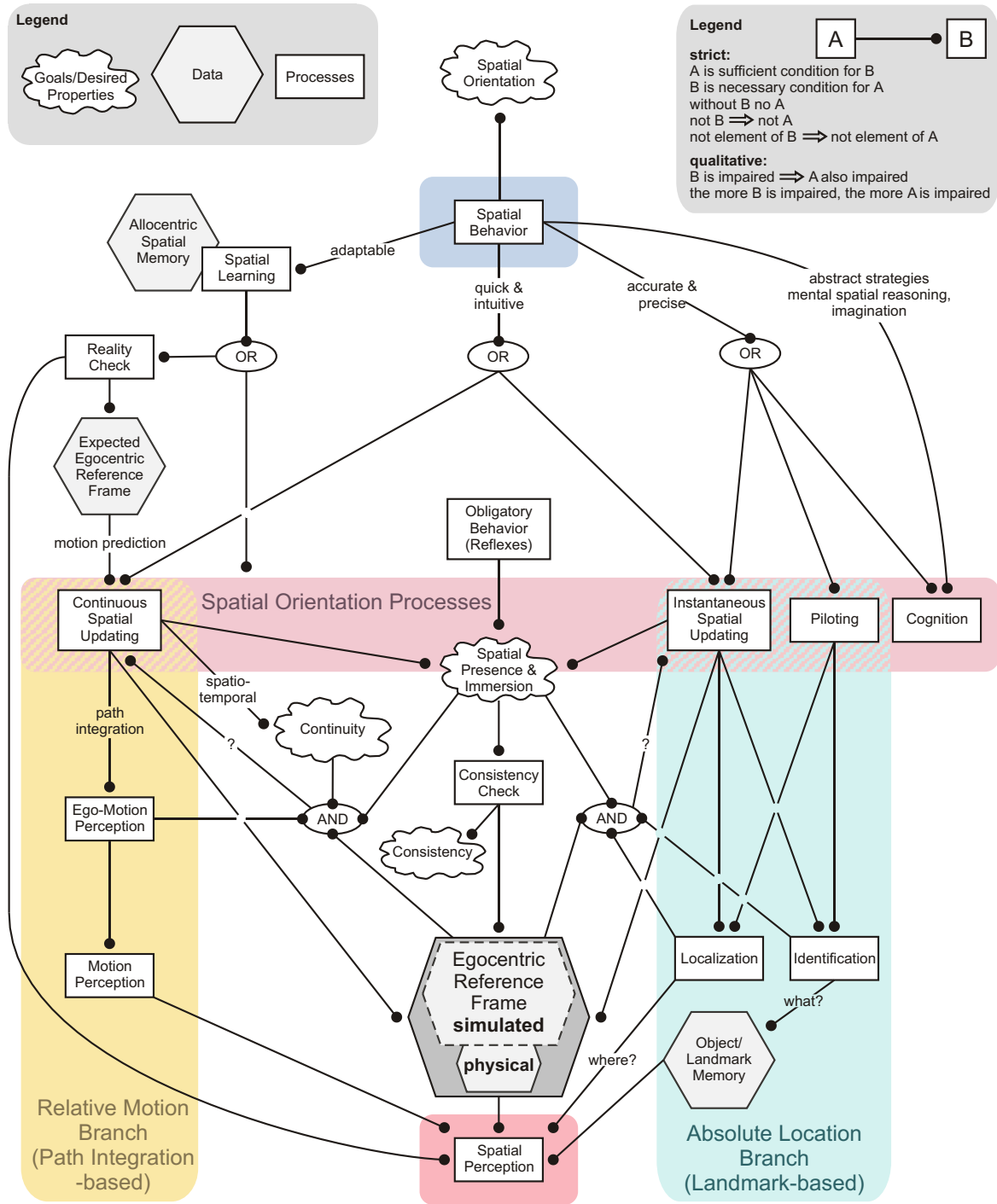
Figure 49: Conceptual framework applied to the UPDATE and CONTROL trials of the spatial updating experiments under full cue conditions. The whole framework is operational, allowing for the observed excellent spatial updating performance.

updating must be somewhat impaired if the FOV is limited. Even with the reduced FOV, however, overall performance was still excellent, indicating that automatic spatial updating was operational. Switching to an HMD-based VR simulation (block C of Experiment REAL WORLD VERSUS VR) instead of viewing the real surround (block B) while leaving the FOV unchanged did not impair performance systematically or significantly[16]. Furthermore, comparing block A and B of Experiment SIMULATION PARAMETERS showed that switching to a video-projection instead of using an HMD did not change performance either. Hence, we can conclude that our VR-based visual simulation allowed for the same excellent automatic spatial updating performance as an equivalent view onto the real surround. This was found irrespective of HMD or projection screen usage. This validates our approach of using Virtual Reality technology for high-level psychophysical experiments and suggests the transferability to the real world.

In the context of our framework, this result is rather interesting, as any VR simulation involves two potentially conflicting *egocentric reference frames*: On the one hand, it introduces the intended *egocentric reference frame* of the VR simulation (i.e., the virtual environment, see Figure 49). On the other hand, the physical reference frame of the VR simulator is still somehow present and cannot be removed completely. Hence, if the physical reference frame of the VR setup is too obvious and cannot be ignored easily, the two reference frames are in clear conflict, which reduces the *consistency* of the intended *egocentric reference frame*. This in turn reduces *spatial presence & immersion* and eventually impairs quick and intuitive *spatial behavior*, as is elaborated upon in subsections 12.1 and 17.1. Conversely, the observed quick and intuitive pointing responses under HMD as well as projection screen conditions suggest that *spatial presence & immersion* was not critically impaired. Consequently, the physical reference frame of the VR setup was apparently not strong or dominant enough to introduce any major conflict with the intended reference frame of the simulation. For the HMD condition, this might be explained by both the photorealistic rendering of a consistent, landmark-rich scene and the HMD blocking all visual cues of the physical surround. For the projection screen conditions, this suggests several things. First, the "window-onto-the-simulated-world" metaphor worked apparently quite well (with and without blinders). Second, the physical reference frame of the motion platform and video projection setup was weak enough to be easily dominated by the simulated visual and vestibular cues. Finally, this confirms our design of the projection setup and our approach of mounting the full video projection setup onto a motion platform. In this manner, participants knew that they could be moved physically. This might be an important difference to the earth-fixed half-cylindrical projection screen used in the experiments described in part II. There, participants might also have experienced a cognitive conflict ("I see a motion of the scene, but I know that I am stationary because I'm sitting in a room that cannot be moved"), whereas they did not on the motion platform ("I see a motion of the scene, and I might as well be moving physically, because I know that I can be moved").

From the full cue UPDATE conditions described in this subsection, however, we can not yet disambiguate between the contribution of *continuous* and *instantaneous spatial updating*. This would only be possible when considering conditions that render one or the other process impossible. On the one hand, removing all landmarks in the blindfolded or optic flow conditions allowed us to eliminate *instantaneous spatial updating* and investigate *continuous spatial updating* in isolation, as will be discussed in subsection 17.2.2. On the other hand, the "teleport" condition effectively eliminated all motion information, thus disabling *continuous spatial updating*. This allowed the study of the *instantaneous spatial updating* process in isolation, and will be discussed in subsection 17.2.3.
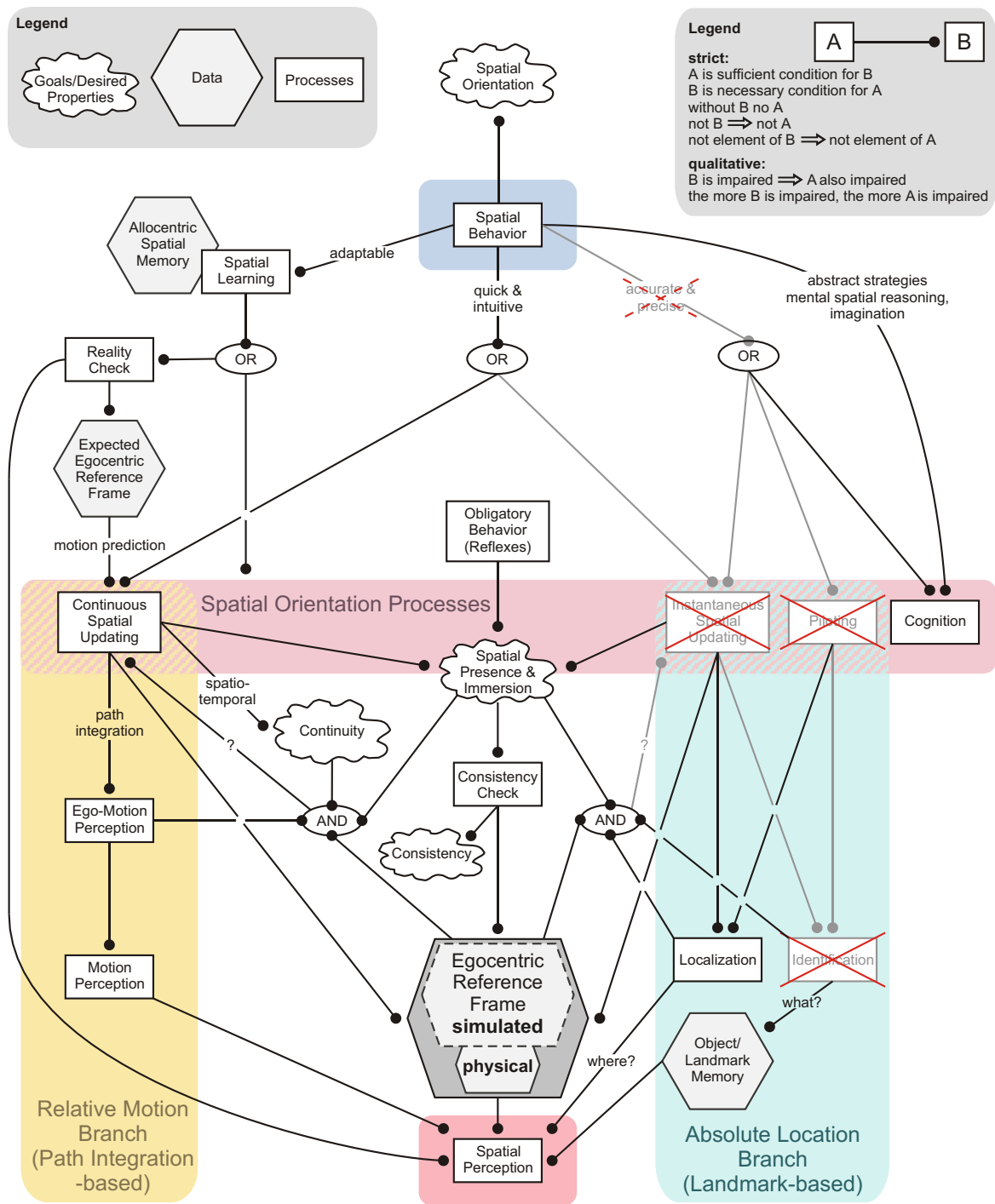
Figure 50: Conceptual framework applied to the UPDATE and CONTROL trials of the spatial updating experiments without useful landmarks (OPTIC FLOW conditions (C&D) of Experiment LANDMARKS VERSUS OPTIC FLOW and blindfolded condition (F) of Experiment REAL WORLD VERSUS VR). The lack of known and identifyable landmarks prevents *instantaneous spatial updating, piloting*, as well as landmark-based *cognition*, which in turn prevents *accurate & precise spatial behavior*.

17.2.2 Conditions without useful landmarks

In the OPTIC FLOW conditions of Experiment LANDMARKS VERSUS OPTIC FLOW, all landmarks were removed during the motion and pointing phase and replaced by a simple optic flow pattern. Even though participants performed well above chance, the optic flow turned out to be insufficient for automatic as well as obligatory spatial updating. Nevertheless, the optic flow did have some specific effect on the mental spatial representation of the participants and could not simply be ignored completely. Additional vestibular motion cues improved UPDATE performance somewhat by reducing the configuration error, but overall performance was still far from automatic spatial updating and significantly worse than in the conditions with landmarks. Using vestibular motion cues alone in block F (blindfolded motions) of Experiment REAL WORLD VERSUS VR showed a similar tendency, but the difference to the visual conditions was less pronounced. It seems like vestibular rotation cues alone might to some extent be sufficient for automatic spatial updating. However, the smooth vestibular motion cues were apparently insufficient to trigger obligatory spatial updating, and it remains to be investigated whether higher accelerations or more jerky motions might render the vestibular motion cues salient enough to enable obligatory spatial updating.

In the context of the framework, the lack of any usable landmarks renders both *instantaneous spatial updating* and *piloting* impossible, as was elaborated upon in subsection 17.1.2 and 17.1.3. Consequently, participants had to resort to either *cognition* or *continuous spatial updating* to solve the task (see Figure 50). As the observed response times were still rather low, however, we would argue that *continuous spatial updating* was intact and played the dominant role, whereas the contribution of *cognition* was, if present, low.

One might expect that the visual switching between the market place scene and the optic flow stimulus should disrupt the spatiotemporal *continuity* and the *consistency* of the *egocentric reference frames*. According to the framework, however, intact *continuous spatial updating* implies both spatiotemporal *continuity* and *spatial presence & immersion* (which in turn implies the *consistency* of the *egocentric reference frames*). How can this apparent contradiction be resolved? Verbal responses from the participants indicate that they did not interpret the optic flow stimulus as a separate scene, but rather some kind of overlay or fog that blocked the visibility of the market place scene. That is, the optic flow scene was apparently not sufficient to constitute a separate *egocentric reference frame* that could have conflicted with the reference frame of the market place. This would explain why neither spatiotemporal *continuity* nor *consistency* was severely affected by the visual switching between the market place scene and the optic flow stimulus.

17.2.3 Condition without motion information ("jump" or "teleport" condition)

In the jump or teleport condition of Experiment SIMULATION PARAMETERS (block I), participants were presented with a view of a new orientation without any motion in between, much like in a slide show. Unexpectedly, the lack of any motion cues did not impair CONTROL or UPDATE performance at all, compared to smooth motions in block K. Merely displaying an image of a new orientation was even sufficient to trigger obligatory, reflex-like spatial updating. That is, the visual cues were powerful enough to almost immediately align the egocentric mental reference frame with the view presented, even when participants were explicitly asked to try to suppress this updating in the IGNORE condition. This finding was unexpected and is to our knowledge unprecedented in the literature. Moreover, this result conflicts with the prevailing opinion that vestibular cues are required for proper updating of ego-turns (see section 12.2.4). Several factors might be responsible for this discrepancy, primarily the immersiveness of our visualization setup and the abundance of natural landmarks in a highly trained, consistent environment.

---

[16]The slight but insignificant increase in response time might have been caused by latencies in the VR simulation.
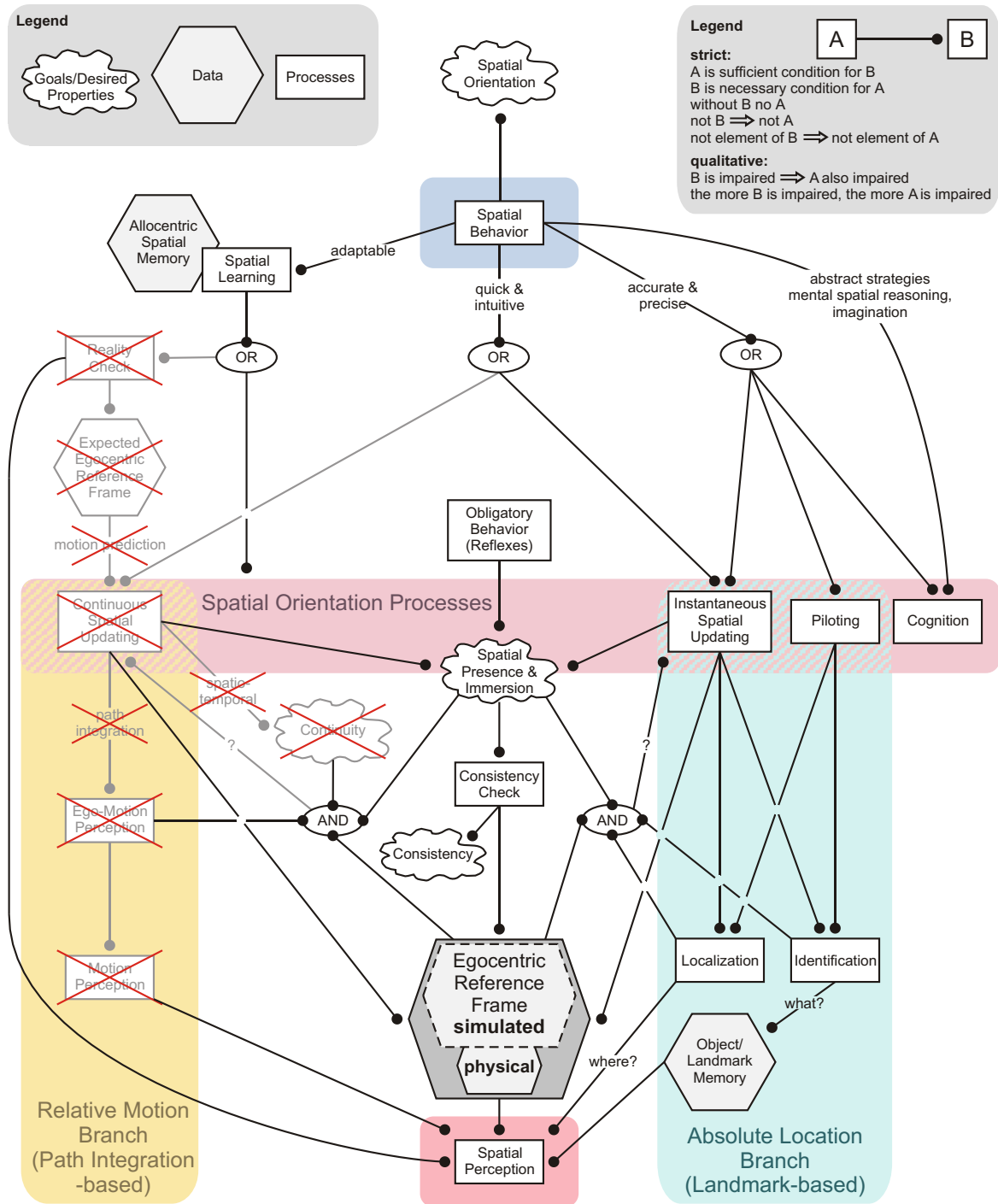
Figure 51: Conceptual framework applied to the UPDATE and CONTROL trials of the teleport condition of Experiment SIMULATION PARAMETERS (block I). The lack of any *motion perception* and spatiotemporal *continuity* disables the whole relative motion-based left branch.

In the context of our framework, the teleport condition allowed us to disentangle the contribution of *continuous* and *instantaneous spatial updating*: In most natural situations, they operate in parallel and we cannot distinguish between the individual contributions. The teleport condition, however, allowed us to effectively eliminate all motion information, thus rendering *motion perception* useless. Furthermore, the lack of spatiotemporal *continuity* prevents *continuous spatial updating*. Taken together, the teleport condition consequently disabled the whole relative motion-based left branch (*ego-motion perception* and *continuous spatial updating* in particular, see Figure 51).

As the observed response times were quite small, the framework would predict that the participants' responses were mainly based on *instantaneous spatial updating*, as both *piloting* and *cognition* do not allow for the observed quick and intuitive behavior. Following the logical connections, intact *instantaneous spatial updating* implies high *spatial presence & immersion*, and consequently also a consistent primary *egocentric reference frame* of the simulation. This implication is rather interesting, as one might expect that the spatiotemporal discontinuity introduced by the teleport should disrupt spatial presence & immersion by, e.g., causing an intermediate loss of presence or a so-called "break in presence" (BIP) (Slater, 2002). If presence was disrupted at all, it must have been rather short and/or weak, as the teleporting did not impair performance significantly, compared to smooth motions. That is, our VR simulation and projection setup allowed participants to feel present in the new orientation almost immediately after the jump, without much additional delay.

Comparing response times between UPDATE and CONTROL trials provides a first rough estimate for the time needed to adopt a new reference frame. For the jump condition, UPDATE trials were only about 30ms slower than CONTROL trials, which is about the same value as for the smooth motion conditions (see Figure 33). Further experiments are needed, however, to test whether 30ms might really be enough time to adopt a new reference frame. Even though only orientation changes in a well-known environment were investigated in the current experiments, this value seems rather low. The current results can, however, serve as a starting point guiding later experiments that tackle this question explicitly.

17.2.4 Implications of *instantaneous spatial updating*

We have introduced the notion of *instantaneous spatial updating* and claimed that it is the mechanism that allowed for both *quick & intuitive* and *accurate & precise* spatial behavior in the teleport condition, without any motion cues whatsoever. So how did *instantaneous spatial updating* affect performance in the conditions without teleport? Whenever there was useful visual landmark information available, *instantaneous spatial updating* most likely worked and might have recalibrated or even overridden the apparently slower and less accurate *continuous spatial updating*. That is, *continuous spatial updating* might just serve as a backup or control mechanism that operates in the background and generates an *expectation* of what one is about to perceive. Only when the two spatial updating mechanisms clearly disagree or when one of them fails (due to the lack of sufficient sensory input, for example) will the other one take over and ensure that we are not lost. Hence, having two spatial updating process running in parallel can serve as a mutual backup system and ensure quick and robust spatial orientation even in conditions when sensory information is sparse and noisy.

Furthermore, comparing the output of the two systems can serve as an alert system: If the expectation generated by the *continuous spatial updating* matches the perceived ego-position (generated by the *instantaneous spatial updating*), everything is fine and no further attention is needed. If they clearly disagree, however, this conflict will most likely come to consciousness, thus allowing for more flexible handling of complex or unexpected situations. Through evolution and life-long exposure to spatial stimuli, we have been trained to expect consistency between the two processes. Within the last centuries, however, more and more situations emerged where we are exposed to a prolonged disagreement between the two processes and the different sensory modalities. Earlier examples include

	parameter/available cues	clear influence	influence only in IGNORE cond.	no clear influence	experiment number & block
provides mainly static (display) information	FOV (blinders vs. unlimited vision)	x			6 A vs. B, 7 E vs. F
	landmarks vs. optic flow	x			8 A vs. C & B vs. D
	vis. & vest. vs. no vis. & vest. (blindf.)	x			6 C vs. F
	vis. & vest. vs. static vis. & vest.	x			6 C vs. E
	vis. & static vest. vs. static vis. & vest.	x			6 D vs. E
	vis. & static vest. vs. no vis. & vest.	x			6 D vs. F
	static vis. & vest. vs. no vis. & vest.		x		6 E vs. F
	presentation device (screen vs. HMD)		x		7 A vs. B
	real world vs. VR			x	6 B vs. C
provides both static & motion information	turning angle		x		7 F vs. J
	turning angle & turn velocity		x		7 J vs. K
	turning angle & teleport		x		7 I vs. J
	gain factors & turning angle (screen)		x		7 C vs. F, C vs. H
	gain factors & turning angle (blinders)			x	7 B vs. E, B vs. G
provides mainly dynamic (motion) information	turn velocity			x	7 F vs. K
	gain factors			x	7 E vs. G, F vs. H, B vs. J
	optic flow with vs. without vest. cues	x			8 C vs. D (config. err.)
	landmarks with vs. without vest. cues			x	8 A vs. B
	landmarks with vs. without vest. cues			x	6 C vs. D (CONTROL marginally significant)
	teleport vs. continuous motions			x	7 I vs. K

Table 23: Overview of variables affecting static (display) and/or dynamic (motion) information with respect to their influence on spatial updating performance. The clustering of the conditions around the diagonal shows clearly the dominant influence of static (display) information over dynamic (motion) information. This suggests that instantaneous spatial updating dominated continuous spatial updating.

riding in a coach or a ship without vision of the external (stable) world. More recent examples include watching movies and navigating virtual environments. In any case, prolonged conflict is often associated with motion discomfort and disorientation, especially if the perceived vertical is affected (Bles et al., 1998).

Coming back to the main question, does the notion of *instantaneous spatial updating* help explain the outcome of the experimental conditions with both smooth motion information *and* landmark information? If *instantaneous spatial updating* is really powerful enough to recalibrate or dominate *continuous spatial updating* (at least as long as there is no obvious conflict between the two updating systems), only differences in the available *static* information should have a strong effect on performance. *Motion* variables, on the other hand, should hardly affect performance.

Even thought the experiments were not designed to test these hypotheses, the results agree nicely with them (see Table 23): Motion variables like rotational velocity, gain factors between visual and vestibular motion, and the presence or absence of vestibular motion cues showed indeed no clear effect as long as useful visual landmark information was provided. Varying the FOV, on the other hand, varied mainly the amount of statically available visual information and did indeed affect performance consistently. See Table 23 for the complete set of parameters investigated.

Last but not least, the turning angle affected both the available static (display) and dynamic (motion) information, and is thus expected to take an intermediate role, showing at least some effect. The only consistent effect on performance was in the IGNORE condition, where larger turns were significantly harder to ignore than smaller ones. The underlying reasons remain unclear. *Instantaneous spatial updating* might force participants in the IGNORE condition to briefly adopt the new (to-be-ignored) reference frame. Following the mental rotation hypothesis, they would then try to mentally back-rotate to the previous orientation. If one assumes a limited mental rotational velocity, this could explain why larger turns took longer to ignore. It remains unclear, however, how the mental rotation hypothesis could explain why larger turns increased also the configuration error, absolute pointing error, absolute ego-orientation error and ego-orientation error against turning direction.

One might argue that participants in the blocks with smaller rotation ranges were aware of that and somehow used it to, e.g., create an expectation of the following possible orientation. In this manner, they might have been able to more easily ignore the next view as they could anticipate better how it might look. If this argument were true, one would expect a turning angle effect only in the conditions where participants *knew* about the possible amplitude of the next turn, that is, only when comparing blocks of smaller turning angles with blocks of larger turning angles. The data supports this view, as larger turns were only harder to ignore when blocks with larger or smaller turns were compared. Within-block correlation analyses, however, did not reveal any consistent turning angle effect for the IGNORE condition. For the UPDATE trials, only the OPTIC FLOW condition of Experiment LANDMARKS VERSUS OPTIC FLOW revealed a clear influence of turning angle on spatial updating performance: All performance measures showed a tendency towards larger turns being more difficult to update. The positive correlations between turning angle and both absolute pointing error and absolute ego-orientation error indicate the expected path integration errors.

17.2.5 Conclusions

One reason for the good overall performance in the spatial updating conditions that included landmarks might be that all participants knew the physical surround quite well: This is especially true in the case of the Tübingen market place, where participants had spent countless hours in the physical surround, thus having built up a detailed *object/landmark memory* as well as *allocentric spatial memory*. That is, no further *spatial learning* was required during the experiment, as the scene (including all landmarks) was already well represented from real world experience and the landmarks training

phase before the actual experiment. Even in the teleport condition, where the relative motion-based left branch of the framework was completely dysfunctional, a view of a new orientation was sufficient to probe participants' spatial memory and align the *egocentric reference frame* accordingly. We suspect that *instantaneous spatial updating* would not work as well if the scene presented was less realistic, consistent and/or not as well-known. Further experiments are needed, however, to test what features of a scene or object configuration are essential for allowing good *instantaneous spatial updating* and consequently also teleport performance, and whether jumps to new positions are as easily updated as jumps to new orientations. Nevertheless, the jump condition probably illustrates best the potential power and dominance of good visual cues. That is, the importance of an effective visual displays can hardly be overemphasized. More experiments are planned and currently being performed in order to determine what features of a visual display are critical, and how the vestibular and auditory cues contribute (Schulte-Pelkum, Riecke, von der Heyde, & Bülthoff, 2002, 2003; Schulte-Pelkum, Riecke, & von der Heyde, 2003).

17.3   Spatial updating experiments - IGNORE condition

In the IGNORE conditions, participants were instructed to ignore the next motion and respond as if they were still in the original orientation. That is, participants were indirectly asked to consciously establish a secondary egocentric reference frame from the current view before the to-be-ignored turn, and respond only with respect to that imagined reference frame. During the motion, the task was essentially to keep the current imagined reference frame no matter what happens, which implies decoupling this imagined reference from spatial perception to prevent it from being updated.

If there was no sensory input indicating any motion or change in orientation (e.g., if the screen was black), the IGNORE task should be easiest. Even though we have not run that condition, we think that the task should be rather simple – the reader may try him/herself. At least, performance should be as good as for the CONTROL trials without landmarks (Experiment REAL WORLD VERSUS VR, block F (blindfolded) and Experiment LANDMARKS VERSUS OPTIC FLOW, condition D (just optic flow)). Any additional sensory input potentially decreases performance if and only if it has a mandatory spatial updating effect on the primary egocentric reference frame. IGNORE trials can consequently be used to test the "power" of the distracting spatial cues.

The IGNORE task should consequently be harder if the available static as well as dynamic spatial cues are stronger in indicating an orientation different from the imagined one. This is the essence of obligatory spatial updating as defined earlier: The mental spatial reference frame is updated to and aligned with a different orientation even when participants explicitly try to consciously suppress this updating and keep their mental reference frame at rest.

The experimental results can be split into two groups: Obligatory spatial updating was always observed when participants were presented with useful visual cues (landmarks) while pointing. On the other hand, vestibular turn cues alone, optic flow alone, and a combination of both were clearly incapable of inducing obligatory spatial updating. In addition, we found several parameters that rendered the spatial updating process more or less obligatory (see also Table 23):

Turns presented via HMD were consistently easier to ignore than turns presented through blinders, whether viewing the real world or a virtual replica displayed on the projection screen. Comparing real word with Virtual Reality performance revealed a slight but insignificant tendency towards the VR simulation being easier to ignore. As already mentioned in subsection 17.2.4, larger turns were harder to ignore than smaller turns. This response pattern, however, was only apparent when blocks with larger or smaller turns were compared. Within-block correlation analyses did not reveal any consistent turning angle effect for the IGNORE condition. Comparing smooth motions with jumps

to new orientations (teleport condition) showed that jumps were as hard to ignore as the smooth motions.

So how can the IGNORE trials be described in the logical framework? If the mental reference frame was under complete conscious control, the IGNORE task should be as easy as the CONTROL task[17]. This was never the case, however, not even when only vestibular cues or optic flow indicated any motion (block F of Experiment REAL WORLD VERSUS VR and condition D of Experiment LANDMARKS VERSUS OPTIC FLOW, respectively). Conversely, this indicates that motion cues from the vestibular sense or from optic flow alone already had some updating effect on the mental spatial representation, even against the explicit conscious decision. These cues were, however, not sufficient for enabling UPDATE performance that was superior to IGNORE performance, indicating no or incomplete obligatory spatial updating. Nevertheless, our mental spatial representation seems to be directly influenced by vestibular as well as visual motion cues. Furthermore, the jump condition of Experiment SIMULATION PARAMETERS (block I) demonstrated that even static visual landmark cues alone can be sufficient to severely affect our mental spatial representation by triggering obligatory spatial updating. Even though we are to some degree able to adopt a secondary reference frame deliberately, this imagined reference frame is strongly influenced by vestibular and especially visual perception. Coming back to our framework, this phenomenon could be incorporated by adding an additional, cognitively imagined *egocentric reference frame* that cannot be completely decoupled from spatial perception (see Figure 52). Consequently, we have three *egocentric reference frames* now: The physical reference frame of the physical surround, the reference frame of the VR simulation, and the additional imagined reference frame in the IGNORE conditions. The IGNORE condition should be easy if participants were able to respond only according to their imagined *egocentric reference frame*, and if this imagined reference frame was completely decoupled from *spatial perception*, *spatial updating*, and the simulated and physical *reference frames* (see Figure 52 (a)).
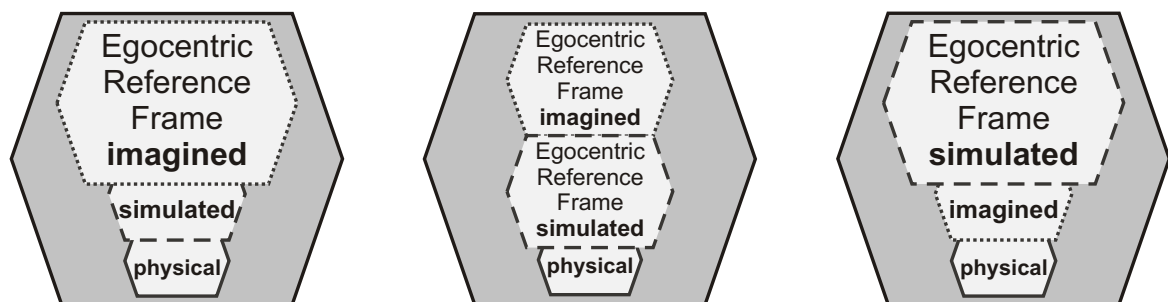
(a) IGNORE condition if the imagined reference frame was dominant and completely decoupled from spatial perception.

(b) IGNORE condition if the imagined reference frame was not strong enough to clearly dominate the simultated reference frame.

(c) IGNORE condition if the imagined reference frame was dominated by the simulated reference frame of the virtual environment.

Figure 52: Schematic illustration of the three egocentric reference frames involved in the IGNORE trials of the spatial updating experiments.

The experiments with only vestibular motion cues or visual motion cues from optic flow, however, indicated clearly that *spatial perception* and *motion perception* in particular do have an effect that cannot be completely suppressed. That is, the relative motion-based left part of the framework alone

---

[17]In both the IGNORE and CONTROL task, participants' task was essentially to point as if still being in the previous orientation.

has been proven to affect the imagined mental spatial representation, even without any perceivable landmarks. This is depicted in Figure 52 (b).

How about the influence of the absolute location-based right part of the framework? This was most directly addressed in the teleport condition, which provided only static visual cues, without any motion cues that could enable *motion perception*. Participants in that condition were incapable of ignoring the presented view and responding as if still being in the previous orientation. Presumably, the static visual cues triggered the *localization* and *identification* of items stored in *object/landmark memory*, and allowed for an *instantaneous spatial updating* process that almost immediately aligned the egocentric reference frame of the simulation with the view presented. In the UPDATE conditions, this allowed participants to almost immediately respond according to the new orientation. In the IGNORE condition, however, participants were apparently unable to not respond according to the new orientation. That is, they could not decouple their imagined reference frame from the updating of the reference frame of the VR simulation (see Figure 52 (c)).

17.4   Application of the framework to the literature

In this subsection, we will extend the scope of our framework by trying to apply it to different experiments and observations from the literature. On the one hand, this serves as an additional test of the framework using examples for which it was not designed. We would consider the framework to be useful if it helps in explaining or re-interpreting the main conclusions and especially some of the unanswered questions. Furthermore, it might suggest possible extensions of the framework that could be incorporated at a later stage. On the other hand, analyzing both our results and results from the literature (which sometimes disagree) in one unifying framework can enable a deeper understanding of the underlying mechanisms and (dis)similarities. Such comparisons might eventually help in developing a "big picture" of how spatial orientation works, especially in critical situations where some spatial information or sensory cues are missing or in conflict.

17.4.1   Object and scene recognition: Comparing physical observer and object motions

Simons & Wang (1998) investigated layout change detection for object-arrays presented on a table. Change detection across views was impaired when the array of objects rotated relative to the stationary observer ("object array rotations"), but not when the observer moved around the stationary display ("viewpoint changes")[18]. Further experiments confirmed that observers can update an egocentric representation of a scene when physically moving to a different viewing position. Such automatic spatial updating, however, did not occur during object array rotations even when additional visual or motor information for the magnitude of the orientation change were provided (Wang & Simons, 1999). Later experiments on single object change detection supported the authors' conclusion that extra-retinal information from physical observer-motions (and not visual background information) was responsible for the observed difference between observer motion and object rotation (Simons et al., 2002).

How can these results be understood in the context of our logical framework? During physical locomotion around the table, *continuous spatial updating* should be intact and lead to an *expected egocentric reference frame* of the display. The *reality check* can then be used to compare the *expected egocentric reference frame* with the perceived one, similar to snapshot matching. If they match, no change occurred; if one object changed, this should easily be detectable. Hence, the change detection task should have been easy as long as the *expected egocentric reference frame* matches the currently

---

[18]Vision of the objects and table was always blocked during the table and observer motion.

perceived one apart from the one object changed. This should be the case in both conditions where the world remained stable, i.e., where the table was not moved relative to the room.

In the condition where the table rotated (unknown to the participants), the *expected egocentric reference frame* (assuming a stable world) does not match the perceived one. Hence, a simple snapshot matching mechanism does not suffice for change detection, as all objects are in an unexpected position, not just the one that was physically moved on the table. Therefore, the framework predicts that change detection performance should be good when only the observer moves and impaired when the world is unstable (object rotation), whether the observer moves or not. This prediction agrees perfectly with the experimental outcome.

In another experiment by Wang & Simons (1999), participants rotated the table themselves. Interestingly enough, even though participants knew and actively controlled how far the objects moved, they were nevertheless unable to update the objects on the table as well as when they moved themselves around the table. A similar advantage of ego-motions over object motions was found whenever object arrays had to be rotated as a whole (Wraga et al., 1999b, 2003). This advantage extends even to the case of *imagined* self or object motions (Wraga et al., 1999a, 2000), at least for rotations around the earth-vertical axes – which seem ecologically most common and relevant for humans (Carpenter & Proffit, 2001). This confirms our assumption that the automated *continuous spatial updating* requires ("$\Longrightarrow$") *ego-motion perception*.

We have seen that *continuous spatial updating* alone can explain the observed data pattern. But did *instantaneous spatial updating* nevertheless have any relevant influence, too? As visibility of the room surrounding the table did not improve performance, *instantaneous spatial updating* was most likely not operating for the objects presented on the table. This seems to conflict with the results from the spatial updating experiments presented in part III of this paper, where a new view of the same scene was capable of triggering *instantaneous spatial updating*.

One of the differences between the experiments might already explain this apparent contradiction: Simons & Wang used objects presented on a table, whereas we used target objects that were part of the surrounding scene and constituted the scene geometry. This might be an important factor, as scene geometry is known to be updated robustly and even recovered after disorientation (Wang & Spelke, 2000, 2002). The surrounding scene was probably updated in both cases via *instantaneous spatial updating*. In our experiments, the target objects were consequently also updated and part of one consistent egocentric reference frame. In the table experiments, however, the objects on the table might not have been linked closely enough to the surrounding room to be instantaneously updated with the room. That is, the objects on the table and the surrounding room might not have formed one coherent scene or egocentric reference frame that was updated as a whole. The "object array rotation" trials where the table with the objects did indeed rotate relative to the room might have contributed to this lack of coherency and holistic updating. Hence, care should be taken that the target objects are sufficiently embedded into one coherent scene if *instantaneous spatial updating* is to be studied or relied upon.

There is further evidence suggesting that only *continuous spatial updating* and none of the other spatial orientation processes played any major role in the table experiments: Control experiments by Simons & Wang (1998) showed that change detection performance was considerably impaired when participants were disoriented (by spinning them around in a wheelchair) before the test phase. One purpose of *instantaneous spatial updating* is to re-orient oneself after getting lost. It should consequently not be affected by disorientation. The same is true for *piloting* and *cognition*. *Continuous spatial updating*, on the other hand, is based on path integration and is therefore rendered useless by any kind of disorientation. This is consistent with our reasoning that *continuous spatial updating* is responsible for the good change detection performance in the non-disoriented conditions, and neither *instantaneous spatial updating*, *piloting*, nor *cognition* played any major role.

17.4.2   Object and scene recognition: Visually simulated observer versus object motions

In object and scene recognition studies, the response latency and error typically increases with the angular distance between the test view and the studied view (Christou & Bülthoff, 1999; Diwadkar & McNamara, 1997; Edelman & Bülthoff, 1992; Roskos-Ewoldsen et al., 1998; Shelton & McNamara, 1997). This orientation or view dependency can be reduced or even removed when automatic spatial updating induced by physical observer movements connects the novel view to the studied view (Simons & Wang, 1998; Wang & Simons, 1999; Simons et al., 2002). Inspired by those findings, Christou and colleagues tested whether visual information *alone* can also achieve spatial updating and view-invariance (Christou et al., 1999; Christou & Bülthoff, 1999; Christou & Bülthoff, 2000; Christou & Bülthoff, 2003). In Virtual Reality experiments similar to the real world table experiments by Simons and Wang, Christou et al. showed that visual information indicating the current observer position can indeed improve performance. The performance increase was the same no matter if photorealistic visual background information or a compass-like arrow indicated the current observer position. Interestingly, the performance benefit was basically due to an unspecific overall decrease in response time. That is, object and scene recognition remained view-dependent, resulting in the well-known inverted U-shaped profile (if proportion of errors or response time are plotted over turning angle).

Can those results be understood in the context of our framework? The authors claim that the visual background cues facilitated spatial updating (e.g., Christou & Bülthoff, 2000). In the following, we will use the logical connections of the framework to question this claim and argue that mainly higher cognitive processes and not automated spatial updating was facilitated by the additional visual cues indicating the current viewer position.

As participants were seated behind a solid desk that definitely did not move, and the virtual scene was presented on a computer monitor subtending a rather small physical FOV, it is unlikely that participants had any percept of ego-motion. This lack of *ego-motion perception* already excludes *continuous spatial updating* as a possible spatial orientation mechanism.

If *instantaneous spatial updating* was responsible for the response time decrease in the condition where visual background information was available, performance should drop considerably in the condition where only a compass-like arrow in front of a black background indicated the current ego-orientation (thus providing only a very weak simulated *egocentric reference frame*). As the visual room reference and the pointer reduced response time equally, *instantaneous spatial updating* can presumably be excluded as a possible explanation.

There are at least two reasons why *instantaneous spatial updating* might not have contributed in updating the objects: On the one hand, *instantaneous spatial updating* could have been disrupted by the lack of *spatial presence & immersion* due to the conflict between the simulated (moving) and physical (stable) *egocentric reference frame* (see subsection 17.1.1). On the other hand, even if *instantaneous spatial updating* did work and participants did update to the new location in the simulated room, it could be that the objects presented on the table were not sufficiently connected to the (properly updated) room to be updated themselves. That is, they might not have formed one consistent scene or egocentric reference frame that was updated holistically. As in the the studies by Simons & Wang (1998), this lack of consistency and holistic updating might have been amplified by the trials in which the object or table did indeed rotate relative to the room (object motion condition).

Non-automated landmark-based spatial orientation processes (*Piloting*) could have helped to determine the simulated ego-position from the visual background information or compass. *Piloting*, however, cannot tell how an object should look like from the new perspective. *Piloting* as well as *continuous* and *instantaneous spatial updating* are apparently not fully operational. The framework consequently predicts that *cognition* must have been the main process responsible for the observed

effects. Apparently, knowing the new viewer position and consequently also the turning angle decreased the overall computational demand. One conceivable cognitive strategy is mental rotation, which indeed predicts the observed inverted-U shaped response characteristics and the lack of view-independence. Hence, we conclude that not automatic spatial updating but higher cognitive processes (most likely deliberate, non-automated mental rotation) was responsible for the observed data pattern.

17.4.3   Spatial updating in nested environments

Wang & Brockmole (2003) investigated whether spatial updating occurred automatically for two nested environments simultaneously. Participants were asked to turn either with respect to the room (local environment) or the surrounding campus buildings (global environment), and then point to targets from both the updated environment (relative to which they turned) or the other, non-updated environment. The data revealed an asymmetry between the local and global environment: When moving relative to the global environment, the local one was automatically updated, but when moving with respect to the local environment, the global one was *not* updated automatically.

This is an example where our framework cannot be applied, as it does not yet include nested or hierarchical spatial representations. One possible extension would be to include nested *egocentric reference frames*. According to Wang & Brockmole (2003), the logical connections should be that an automatic spatial updating of the larger (global) egocentric reference frame logically implies ("$\Longrightarrow$") an updating of the smaller (local) egocentric reference frame. We suspect, however, that such a logical implication would only hold true if both levels of the nested environment were represented egocentrically. This was also suggested by Wang & Spelke (2002) for when both levels are behaviorally relevant. Such an egocentric representation and updating explains why pointing performance in Experiment LANDMARKS VERSUS OPTIC FLOW (A & B) was so good, even though most participants completely failed in the map-drawing task. Apparently, an allocentric representation of the surround in form of, e.g., a "cognitive map" is not required for most everyday behavior (Wang & Spelke, 2002).

17.4.4   Contribution of physical motion cues for rotations in VR

To our knowledge, Wraga et al. (2003) were among the first to extend spatial updating work beyond the typically studied nonvisual (blindfolded or imagined) conditions to include visual landmark cues presented in VR. As their methodology is rather similar to our experiment, but their results are not, we would like to discuss them here in more detail.

Participants in the Wraga et al. (2003) study wore a headtracked HMD and saw a simple square or circular room with four or five target objects. The participants' task was to name or point to previously-learned target objects after real or visually simulated self-motions. In Experiments 1 and 2, spatial updating performance was quantified in terms of response time and number of errors for verbal responses ("where is object X?"). Experiment 1 compared physical ego-rotations to simulated room (display) rotations, and revealed an advantage for ego-rotations in both response time and error scores. Experiment 2 compared active and passive ego-motions and found only minor (insignificant) benefits from actively rotating. Experiments 3 and 4 were similar to experiments 1 and 2, respectively, but used five irregularly spaced targets in a circular room instead of a square room and a compass-like pointer instead of verbal responses. This resulted in a comparable advantage for self-rotations, but no difference between active and passive motions were found. The considerable overall increase in response time might be due to the complicated pointing metaphor used[19].

---

[19]The response times in experiment 3 and 4 were typically longer than 8s, which is more than five times larger than in experiments 1 and 2 with verbal responses. Compared to the literature at large and the experiments presented in part

In Experiment 1 of Wraga et al. (2003), removing proprioceptive and vestibular cues from physical self-rotations degraded participants' performance considerably. In the experiments described in part III, however, it did not, even when the presentation device was an HMD (Experiment REAL WORLD VERSUS VR, block C vs. D). How can this apparent contradiction be explained?

In our experiments, participants were always rotated passively, which reduces proprioceptive cues to a minimum. Wraga et al. found no clear difference between active and passive rotations, however, suggesting that neither active control nor proprioceptive cues are absolutely required for spatial updating tasks.

Another difference between the experiments is the layout of the scene and the targets. Wraga et al. (2003) used a highly symmetrical simulated room where the target objects were the only salient landmarks. In our studies, however, the virtual environment provided an asymmetric room geometry and an abundance of salient landmarks other than the target objects. As room geometry is known to be an important factor for robust spatial orientation (see, e.g., Wang & Spelke, 2002), this might be a critical difference.

Can our framework assist in further pinpointing the underlying reasons for the different results? In our studies, the presence or absence of physical motions hardly influenced performance as long as the landmarks were visible. As a lack of vestibular motion cues implies a reduced *ego-motion perception* and consequently also a somewhat impaired *continuous spatial updating* performance, we concluded that *instantaneous spatial updating* jumped in and compensated for the lack of vestibular motion cues.

In the Wraga et al. (2003) study, however, the lack of sensory cues from physical motions did reduce participants' performance. This suggests that *instantaneous spatial updating* was not able to compensate for the impaired *continuous spatial updating*. As the landmarks were always visually presented, their *localization* and *identification* should be unaffected by the lack of physical motion. Hence, the framework predicts that the *egocentric reference frame* and/or *spatial presence & immersion* should have been impaired, as these are the other two necessary prerequisites for *instantaneous spatial updating*. Both possibilities are conceivable:

First, the lack of physical motions might have increased the conflict between the egocentric reference frame of the physical surround (indicating no motion) and the visually simulated reference frame (indicating motion). This reduced *consistency* would then be detected by the *consistency check*, leading to a reduced *spatial presence & immersion*.

Second, the *egocentric reference frame* of the virtual room itself might already have been insufficient for *instantaneous spatial updating*. As discussed in subsections 17.4.1 and 17.4.2, this could be due to the target objects not constituting one coherent reference frame that can be automatically updated as a whole.

Both explanations are rather speculative, though, and need to be tested thoroughly. Nevertheless, the comparison with our results and especially the application of our framework allowed us to deduce testable hypotheses that can inspire both the scientific discussion and future experiments.

III, these response times and pointing errors are exceptionally large, which is unfortunately not discussed in the paper. We suspect that this response time increase is mainly due to the cumbersome handling of the pointing device and only partly due to the different target and room layout: The virtual pointer had a horizontal default position and was controlled using a dial. Hence, pointing took longer if the desired (goal) pointer position was further away from the default (starting) position. This procedure unfortunately both increases overall response time and confounds spatial updating performance with relative target location, which might in turn produce artifacts and increase the response variance unnecessarily.

17.4.5 Disorientation in VR

In most Virtual Reality and multi-media applications involving simulated rotations of the observer, users are easily disoriented after only a few simulated motions, even when an abundance of landmarks is available (see Loomis et al. (1999), Péruch & Gaunet (1998), Richardson, Montello, & Hegarty (1999), Ruddle et al. (1997) and also subsection 12.1). In comparable real world situations, however, people are not as easily disoriented. Does our framework help in determining which critical factors might be missing in those VR applications?

Piloting should generally be intact as long as a sufficient amount of salient and useful landmarks is available. *Cognition* should also be operational and comparable to the real world. Consequently, VR performance should potentially be *accurate & precise* and may involve higher cognitive processes. This is in agreement with the literature and our own findings in Experiment LANDMARKS, section 7: As long as reliable landmarks were available, participants successfully used both piloting (e.g., snapshot view matching) and mental spatial reasoning for homing. As both *piloting* and *cognition* are most likely intact, our framework consequently predicts that *spatial updating* must be considerably impaired. This makes sense, as disorientation can be paraphrased as a lack of robust, quick, and intuitive spatial orientation, which is exactly what automatic spatial updating is good for. So why is spatial updating not operating properly in most VR applications?

Continuous spatial updating is in most cases already disabled due to the lack of convincing ego-motion simulation. As a long tradition of vection studies showed, merely presenting a moving visual stimulus is indeed not immediately (if at all) accepted as self-motion. Furthermore, most VR displays do not allow for accurate path integration of rotations and tend to produce rather large artifacts, especially for HMDs (see, e.g., Bakker et al., 1999, 2001; Schulte-Pelkum et al., 2002, 2003).

In terms of *instantaneous spatial updating*, most virtual environments provide a sufficient amount of landmarks that can be *localized* and *identified* and often constitute a consistent simulated *egocentric reference frame*. Hence, only one necessary prerequisite for *instantaneous spatial updating* seems to be missing: *Spatial presence & immersion*. As *spatial presence & immersion* is also a necessary prerequisite for *continuous spatial updating*, it seems to play a central role for overcoming the disorientation problem associated with VR.

Unfortunately, though, little is known on how to reliably achieve high *spatial presence & immersion* without having to simulate all sensory information as realistically as possible. Such a brute force approach (simulate everything that is computationally and financially possible) might work for a few specialized applications, but is bound to fail for most natural, dynamic, or complex situations.

Only recently has *spatial* presence been investigated independently from the general feeling of presence, which is often enough only vaguely defined (Regenbrecht & Schubert, 2002; Regenbrecht, 1999; Schubert et al., 2001). Regenbrecht & Schubert (2002) for example used introspective questionnaires and found that spatial presence should be enhanced by the mental representation of possible actions. They conclude that: "Designers of virtual environments who aim at creating a high sense of presence, be it in games or in architectural simulations (Regenbrecht, 1999), can conclude from these results that, to create high spatial presence, they must allow the users to choose their own point of view in the VE and to navigate their virtual body, give them possibilities to interact with objects in the VE, and enact simple interactions with virtual characters or other real users. However, it should be noted that what counts are the users' representations of these interactions, not their objective availability per se."

We agree that the knowledge about possible interaction can enhance the subjective feeling of spatial presence. Whether following such rules-of-thumb is enough to elicit not only the subjective feeling of spatial presence, but also to enable spatial orientation that is as robust and effortless as in the real world is debatable. It is, for example, conceivable that one's subjective impression of spatial

presence is quite high, but that one's mental spatial representation of the surround is not automatically updated during self-motions, which is responsible for robust and effortless spatial orientation. In our framework, this is reflected in *spatial presence & immersion* being only one of several necessary prerequisites for *spatial updating* and consequently also *quick & intuitive spatial behavior*.

We believe that subjective evaluations are an important contribution and should always accompany spatial presence research. They do not, however, allow one to reliably predict whether a subjective feeling is reflected in the ability to perform appropriate actions. If we aim at stepping beyond mere theory and applying our research to help humans when confronted with mediated environments like virtual environments, subjective methods need to be complemented by behavioral measures. A theoretical background can then help by deducing hypothesis and predictions and thus assist in designing and understanding behavioral experiments. This was one motivation for us to devise the framework presented in section 16. We hope that we have shown that our framework can indeed be useful in connecting theory and experiments and already satisfies most of the desired properties for a comprehensive framework as stated in the introduction (see part IV on page 123): "First, it might allow for a coherent representation of the experimental paradigms and results. Second, it could help to structure and clarify our reasoning and discussions. Maybe most important, it might allow for a deeper understanding of the underlying processes and mutual dependencies. Last but not least, it could suggest novel experiments and experimental paradigms, allow for testable predictions, and stimulate the scientific discussion."

18 Implications and final conclusions

We hope that we have shown that a framework based on logical propositions can indeed assist in analyzing spatial situations and experimental results. In particular, it was helpful in structuring and clarifying our reasoning and understanding the implications if certain processes are impaired or goals/assumptions are not met.

We argued that both the relative motion-based left branch and the absolute location-based right branch of the framework have a mandatory, irrepressible influence on the imagined egocentric reference frame. This influence is most likely mediated by *continuous* and *instantaneous spatial updating*, respectively. In our experiments, however, only the absolute location-based right branch of the framework was capable of triggering obligatory spatial updating, indicating the power and potential dominance of visual cues.

Conversely, *cognition* (i.e., consciously trying to control one's imagined egocentric reference frame and act accordingly) was clearly unable to suppress or override *continuous* and *instantaneous spatial updating*. This indicates the power and reflexiveness of the spatial updating processes, and highlights their importance for effortless spatial orientation.

Hence, any attempt to simulate self-motions that does not enable these reflexive, highly automatized spatial updating processes is bound to decrease spatial orientation performance and unnecessarily increase cognitive load. So, how can we enable obligatory spatial updating in virtual environments? As obligatory spatial updating is reflex-like, it occurs by default as long as it is not prevented. The same is true for spatial presence: It is the "default" in real world situations. Thus, any situation where spatial presence is impaired must have prevented it somehow. Therefore, our question should be rephrased: Instead of asking "how we can produce spatial updating or cause spatial presence?", we should ask "what exactly prevents spatial updating or spatial presence?". Following the logic of our framework, any item implied by, for example, *continuous spatial updating* ($A \Rightarrow B$) can also potentially prevent it ($A \Rightarrow B \iff \neg B \Rightarrow \neg A$) and is therefore critical. The potential of such an analysis is one of the strengths of a model using logical propositions, and was one reason for us not to use the commonly employed information flow analysis. So what can we conclude from following the logical implications of our framework?

Continuous spatial updating $\Rightarrow$ ego-motion perception Following the logical arrows, we see that *continuous spatial updating* requires *ego-motion perception*. From the vection literature, though, we know that merely presenting a rotating visual stimulus will not immediately evoke the percept of ego-motion. Many parameters have been found to influence vection, including the FOV, the amount of foveal and peripheral vision, and the spatial frequency content (see, e.g., Dichgans & Brandt, 1978; Warren & Wertheim, 1990; Wertheim, 1994b, 1994a). Nevertheless, it typically takes several seconds until a rotating visual stimulus can be interpreted as self-rotation. We are not aware of any study demonstrating immediate vection onset. As *continuous spatial updating* is based on integrating the perceived ego-motion, and any vection onset delay potentially introduces considerable path integration errors, just presenting a rotating visual stimulus (without any additional motion cues) might already prevent *continuous spatial updating*. To tackle this issue, we are currently designing experiments to investigate whether any combination of visual, auditory and vibrational cues can yield an immediate vection onset [20]. As many VR displays do not allow for accurate path integration of rotations and tend to produce rather large artifacts, especially for HMDs (e.g., Bakker et al., 1999, 2001; Kearns et al., 2002; Péruch et al., 1997), we are currently also investigating the influence of

---

[20]This research is part of the recently started EU-project "POEMS" (Perceptually Oriented Ego-Motion Simulation, see `www.POEMS-project.info`).

various display parameters on such systematic path integration errors (Schulte-Pelkum, Riecke, von der Heyde, & Bülthoff, 2002, 2003; Schulte-Pelkum, Riecke, & von der Heyde, 2003).

Continuous spatial updating ⇒ continuity   *Continuous spatial updating* further requires spatiotemporal *continuity* of the perceived stimulus and the environment in general. Any violation of that *continuity* can consequently prevent *continuous spatial updating*. Violations typical for many VR applications include discontinuous jumps of the observer position, suddenly appearing or disappearing objects, and jerky motions due to, e.g., a low or unsteady frame rate or multiple images. Care should therefore be taken to avoid these and other discontinuities.

Instantaneous spatial updating ⇒ localization ∧ identification   Both the *localization* and *identification* of known landmarks are a necessary prerequisite for *instantaneous spatial updating*. This is easy to achieve with current VR technology, especially when using photo-based texturing. Care should, however, be taken to use salient and unique landmarks and to avoid the often used repetitive and streamlined scenes. Sufficient familiarity with the scene and objects can easily be achieved through appropriate exposure or training phases or simply by building virtual mock-ups of real scenes that the users are already familiar with.

Spatial updating ⇒ egocentric reference frame   *Spatial updating* requires, of course, the existence of a *egocentric reference frame*, or else there would be nothing to update. Furthermore, relevant objects need to be embedded into a consistent scene, thus constituting a coherent *egocentric reference frame* that can be updated as a whole (see subsection 17.4.1 and 17.4.2). Consequently, repetitive or featureless scenes with simple, symmetrical layout and objects floating in mid-air should be avoided.

Spatial updating ⇒ spatial presence & immersion   Both *spatial updating* processes require and implicate *spatial presence & immersion*. This connects our results to current presence research and highlights the importance of interdisciplinary research. Considering *spatial presence & immersion* as an important factor and being able to reliably quantify it is an important research endeavor beyond the mere theoretical interest in the subjective feeling of "being there". As *obligatory behavior (reflexes)* imply *spatial presence & immersion* and are experimentally well accessible, such research can assist in ensuring that a potential lack of obligatory *spatial updating* is not due to a lack of *spatial presence & immersion*.

Spatial presence & immersion ⇒ consistency check ⇒ consistency ∧ egocentric reference frame   *Spatial presence & immersion* can only occur if the *consistency check* agrees on one *consistent egocentric reference frame*. This is a critical prerequisite that is probably not met in most VR setups (see section 17 for examples). Often enough, the physical VR setup provides too strong of an egocentric reference frame to allow for immersion, thus preventing spatial presence (see Figure 53 (a)). Examples include desktop VR and probably most VR setups where the FOV is small and the physical room provides a strong reference frame (e.g., if it is cluttered and not darkened, or if background noise provides strong spatial cues). Consequently, the ultimate goal here is to make the VR interface vanish and become "invisible" to the user, in a way that it can really open up a "window onto the virtual world" (see Figure 53 (c)). That is, the VR setup should capture or require as little "interface attention" as possible, because any need to cognitively suppress the physical surround increases cognitive load and interferes with *spatial presence & immersion*. On the other hand, the simulated egocentric reference frame should be as strong as possible, such that it can easily dominate the physical counterpart.
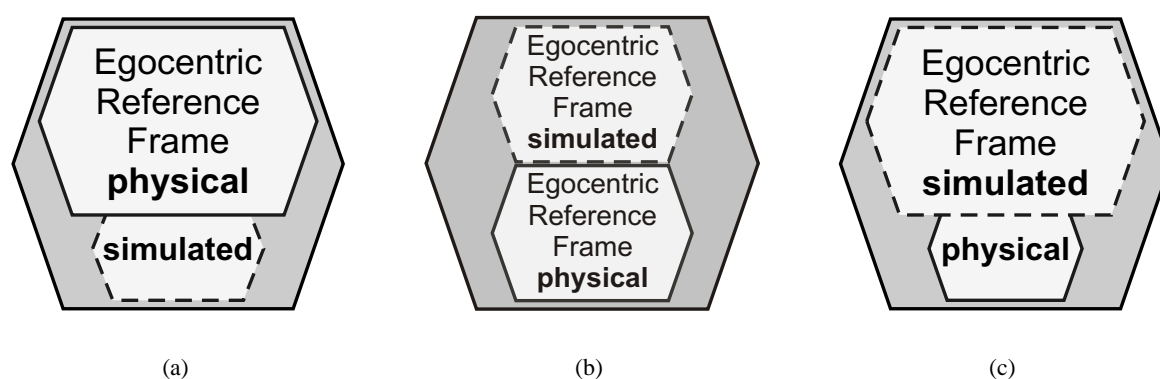
Figure 53: Schematic illustration of the reference frame conflict observed in many VR applications. (a) The physical reference frame of the VR setup is particularly strong and dominates the weaker simulated reference frame, thus preventing *spatial presence & immersion*. (b) The simulated reference frame of the virtual environment is stronger than in (a), but not strong enough to dominate the physical reference frame, which results in an impaired *spatial presence & immersion*. (c) Only if the physical reference frame vanishes perceptually or is clearly dominated by the simulated reference frame can we achieve intact *spatial presence & immersion*.

Thus, enabling (i.e., not preventing) *spatial presence & immersion* requires a human-centered approach of ergonomically designing the virtual environment and VR interfaces to suit the humans, instead of asking the humans to adapt to the VR setup. As it is currently impossible to make the interface vanish physically, one should instead focus on making it vanish *perceptually*. We know from good actors, magicians, or movies that one's attention can be guided, and we can be made to "want to believe". This knowledge could be used to achieve our goal more elegantly, without having to simulate everything with perfect realism. (See Best (1994) for an amusing introduction to that issue.) Many factors are probably relevant, including interesting, challenging, and capturing tasks, the feeling of excitement, flow, involvement, and achievement, and a clear goal or purpose of the human-machine interaction. Obviously, this challenges VR-designers and requires a truly interdisciplinary effort.

We demonstrated that spatial updating experiments like the ones presented in this thesis, especially when embedded in a conceptual framework, are a successful paradigm for investigating and quantifying the contribution and interaction of different sensory cues and modalities for spatial orientation. In addition to assessing this spatial cognition aspect, the human factors issues involved can be tackled by determining the relevant simulation and display parameters necessary for good spatial orientation. This demonstrates the truly interdisciplinary and comprehensive character of this piece of research.

Summary We demonstrated that visual path integration without any vestibular or kinesthetic cues can be sufficient for elementary navigation tasks like rotations, translations, and triangle completion. When a consistent scene with reliable, salient landmarks is used, visual cues alone can even be sufficient to trigger reflex-like obligatory spatial updating. Devising a comprehensive framework based on logical propositions allowed for a deeper understanding of the underlying mechanisms in both our experiments and experiments from the literature. In particular, it enabled us to clearly distinguish between the well-known continuous spatial updating and teleport-like "instantaneous spatial updating".

19   Epilogue

So what happened in our initial example of walking around in darkness (section 3 on page 1)? As long as we do not see, hear, or otherwise perceive any landmarks, *continuous spatial updating* is probably the dominant spatial orientation mechanism that prevents us from getting lost. The further we walk, the larger the uncertainty about our exact location becomes. We might get nervous and try to recapitulate and memorize the walked path, visualize it in a bird's eye view, or count steps and turns to compensate for the increasing path integration error. Without any exact information like landmarks, even such *cognitive spatial orientation processes* often do not help much. As soon as we perceive any landmark, however, the situation changes completely. If we are familiar with the environment, a brief glimpse of a known reference point might be enough to trigger *instantaneous spatial updating*, and we are promptly and automatically reoriented. If we are rather unfamiliar with the surround, a landmark might not be sufficient for *instantaneous spatial updating*, but might allow for a rough position fixing via *piloting* - if we smell coffee or hear the coffee machine, for example, we can deduce that we must be close to the kitchen.

A Additional data plots for reference

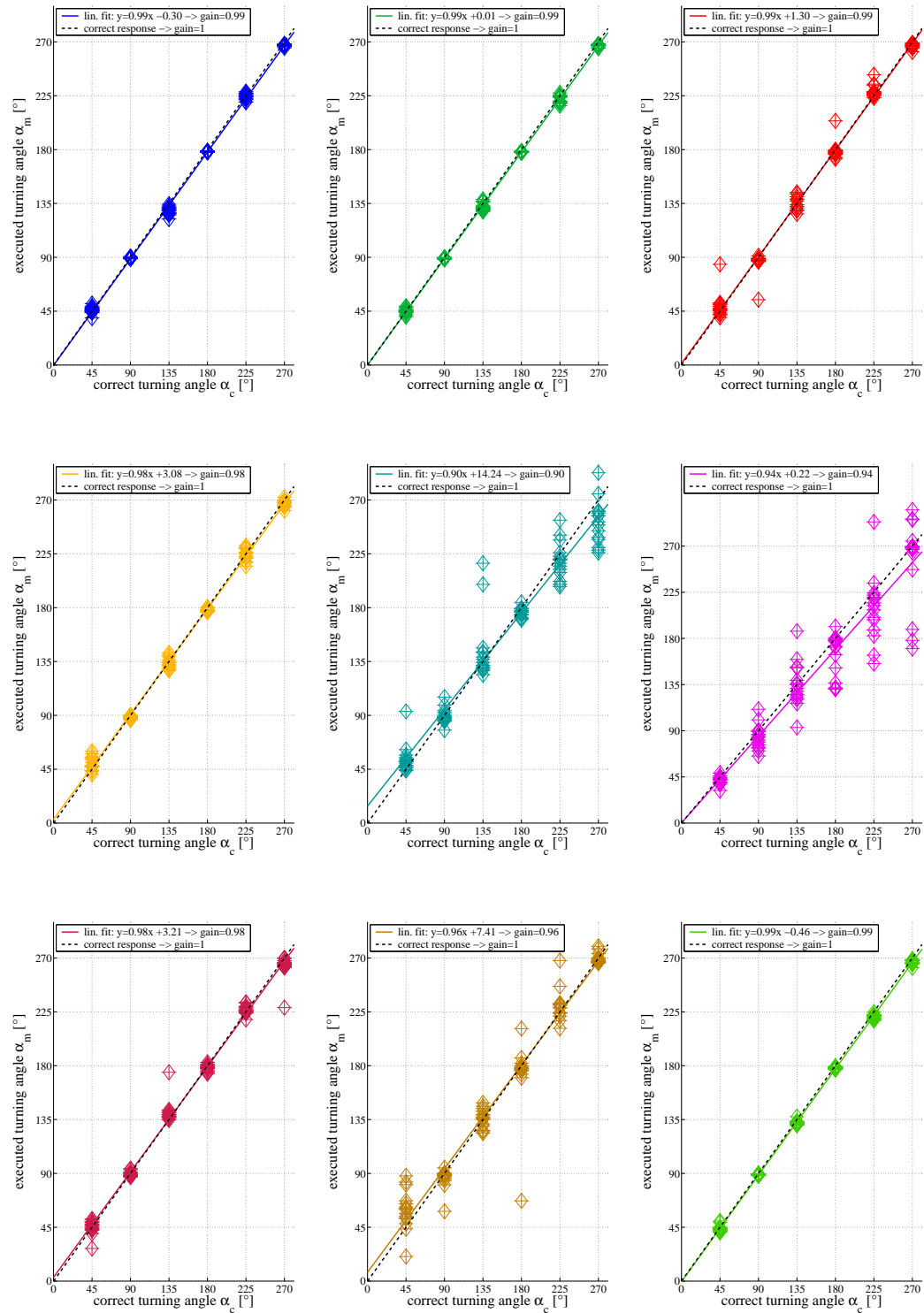## A.1 Overview figures for Experiment TURN&GO

Figure 54: Visual turn execution performance for all nine participants of Experiment TURN&GO. Note the high accuracy and precision, indicating only minimal systematic errors and variability, respectively.
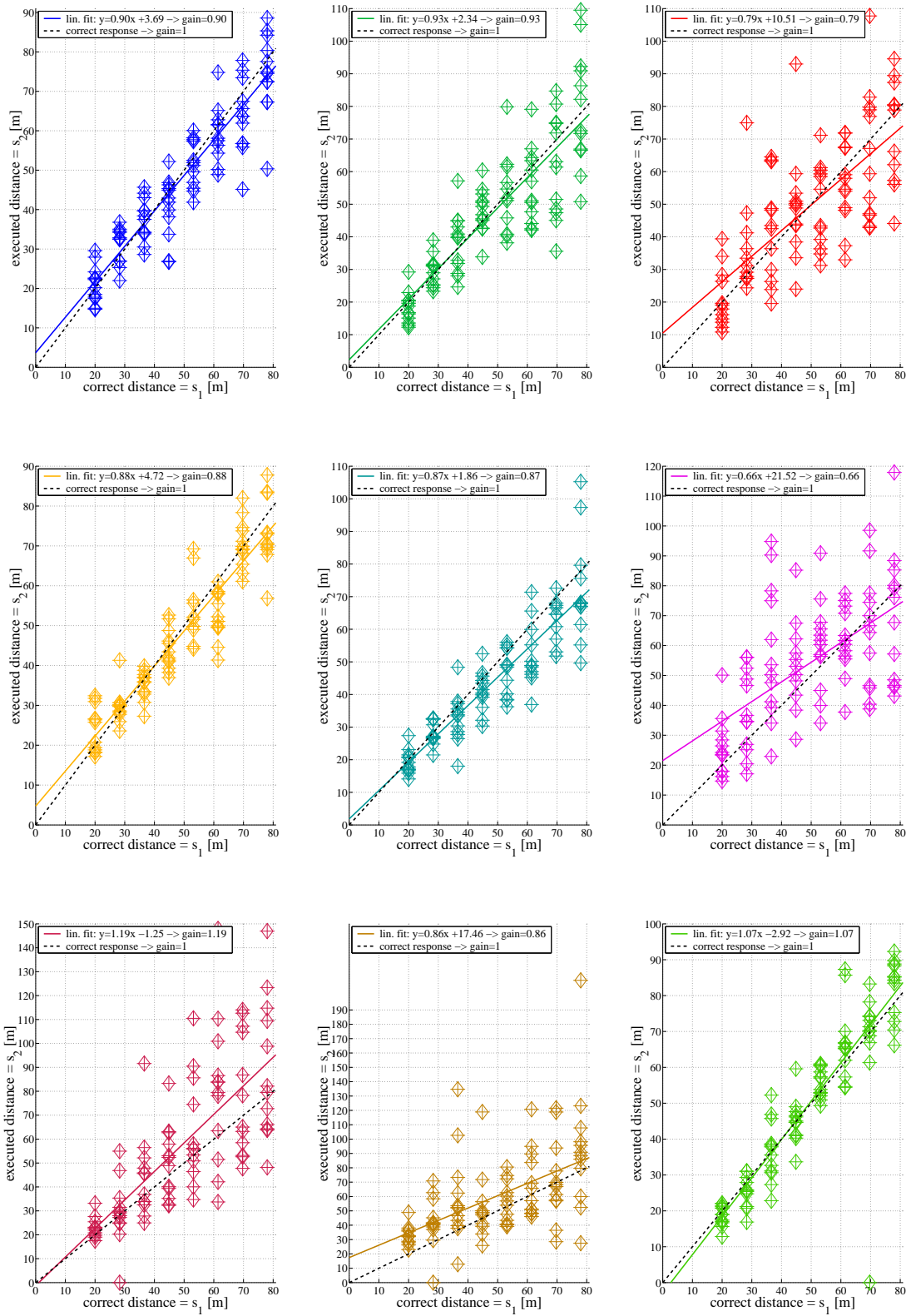
Figure 55: Visual distance reproduction performance for all nine participants of Experiment TURN&GO. Even though the within-subject variability is rather high, the mean systematic errors are only moderate.

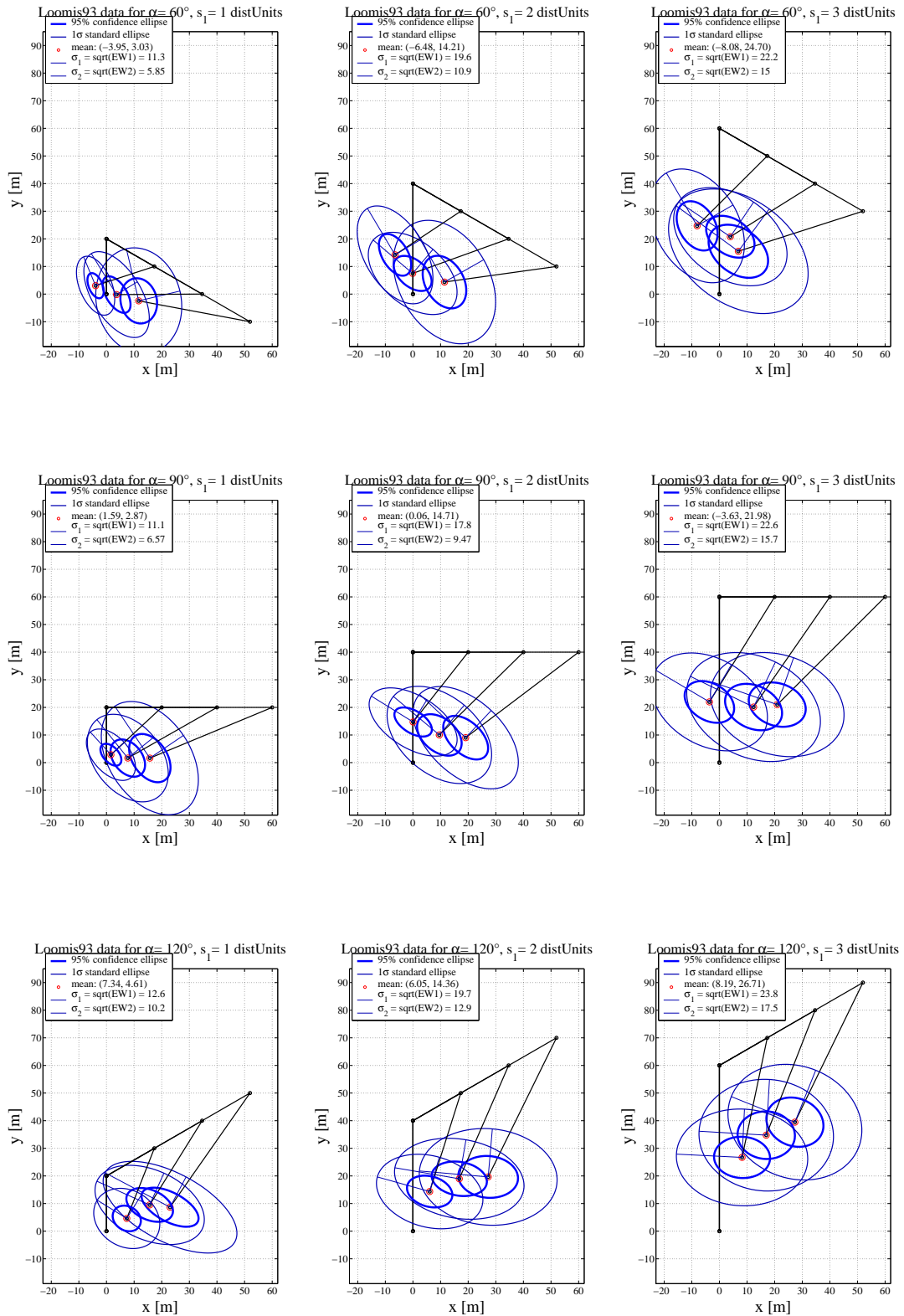## A.2   Overview figures for Loomis et al. (1993)

Figure 56: Homing performance for blindfolded walking in the study by Loomis et al. (1993), plotted as in Figure 6 and 9. The data is scaled to match the triangles used in our experiments. Note the considerable variability and systematic errors, especially the general distance undershoot.

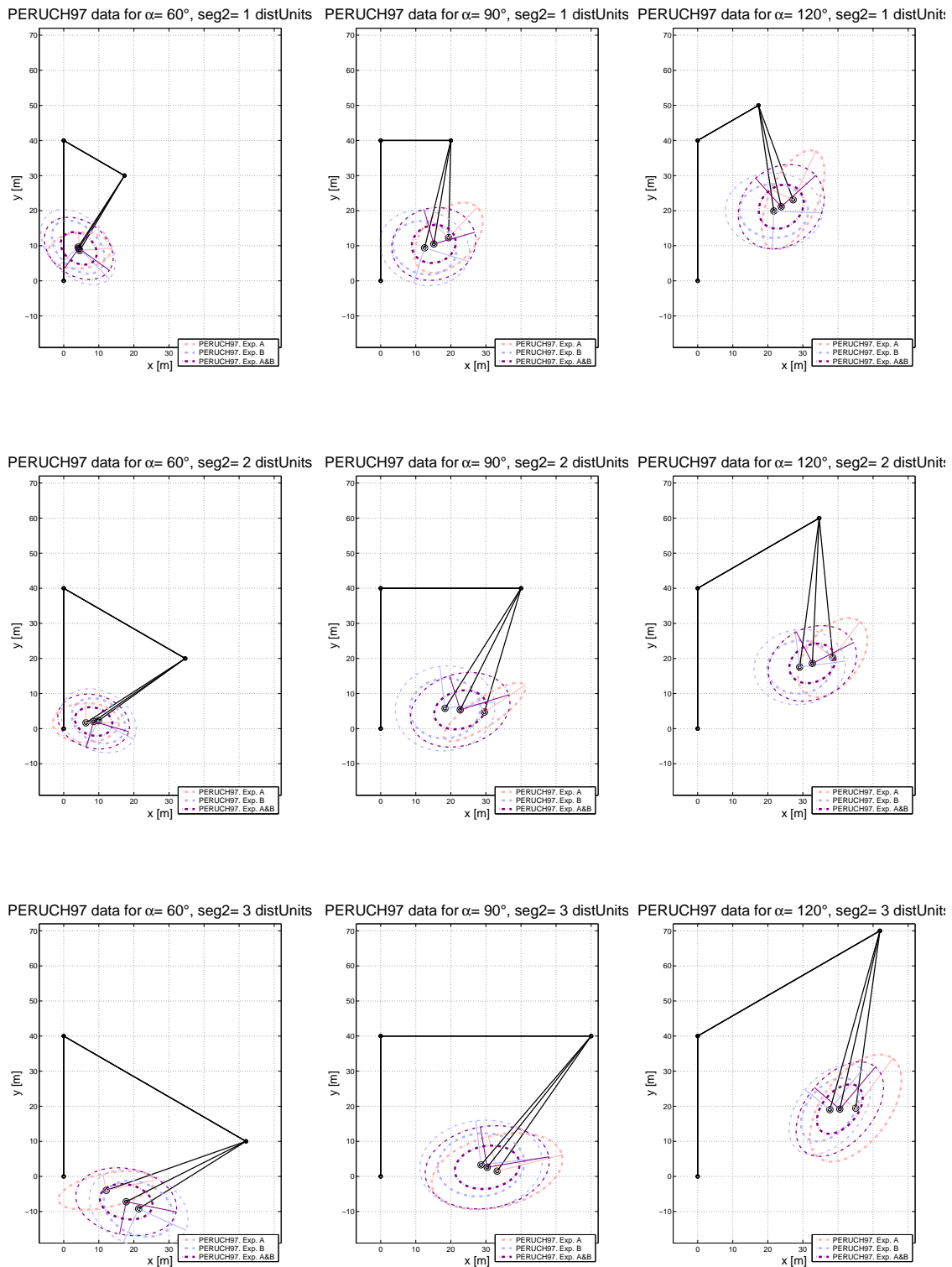## A.3   Overview figures for Péruch et al. (1997)

Figure 57: Visual homing performance in the study by Péruch et al. (1997), plotted as in Figure 6 and 9. Experiment B was identical to Experiment A apart from a difference in the instructions: Before executing the homing response at goal 2, participants were asked to turn back until they could see goal 1. The data is scaled to match the triangles used in our experiments. Note the large systematic errors and low stimulus response.

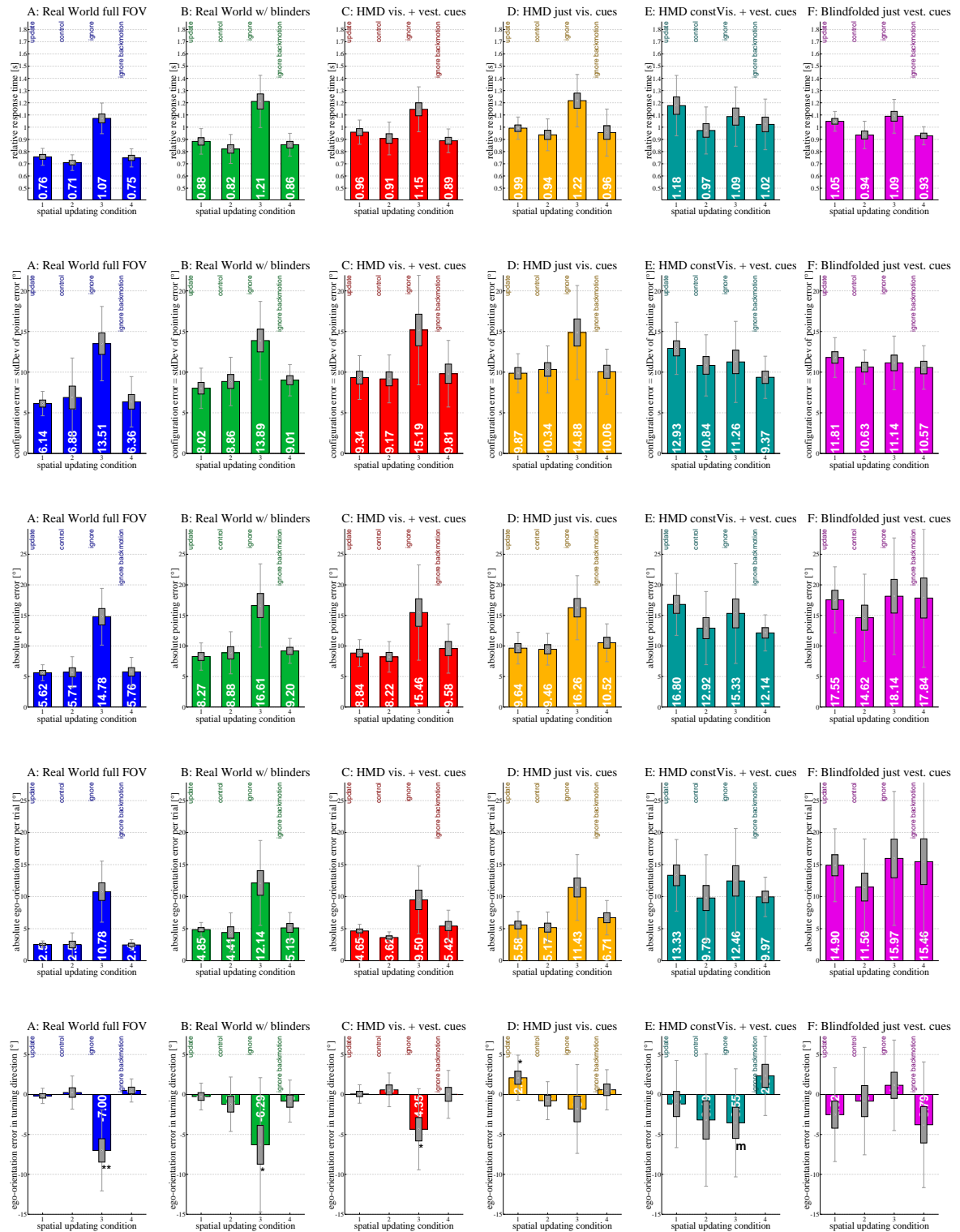## A.4 Overview figures for Experiment REAL WORLD VERSUS VR

Figure 58: Compilation of all dependent variables for Experiment REAL WORLD VERSUS VR, grouped by block (stimulus combination). Note the typical response pattern for obligatory and automatic spatial updating in all conditions with useful visual cues: UPDATE performance is comparable to CONTROL performance, whereas IGNORE performance is considerably worse.
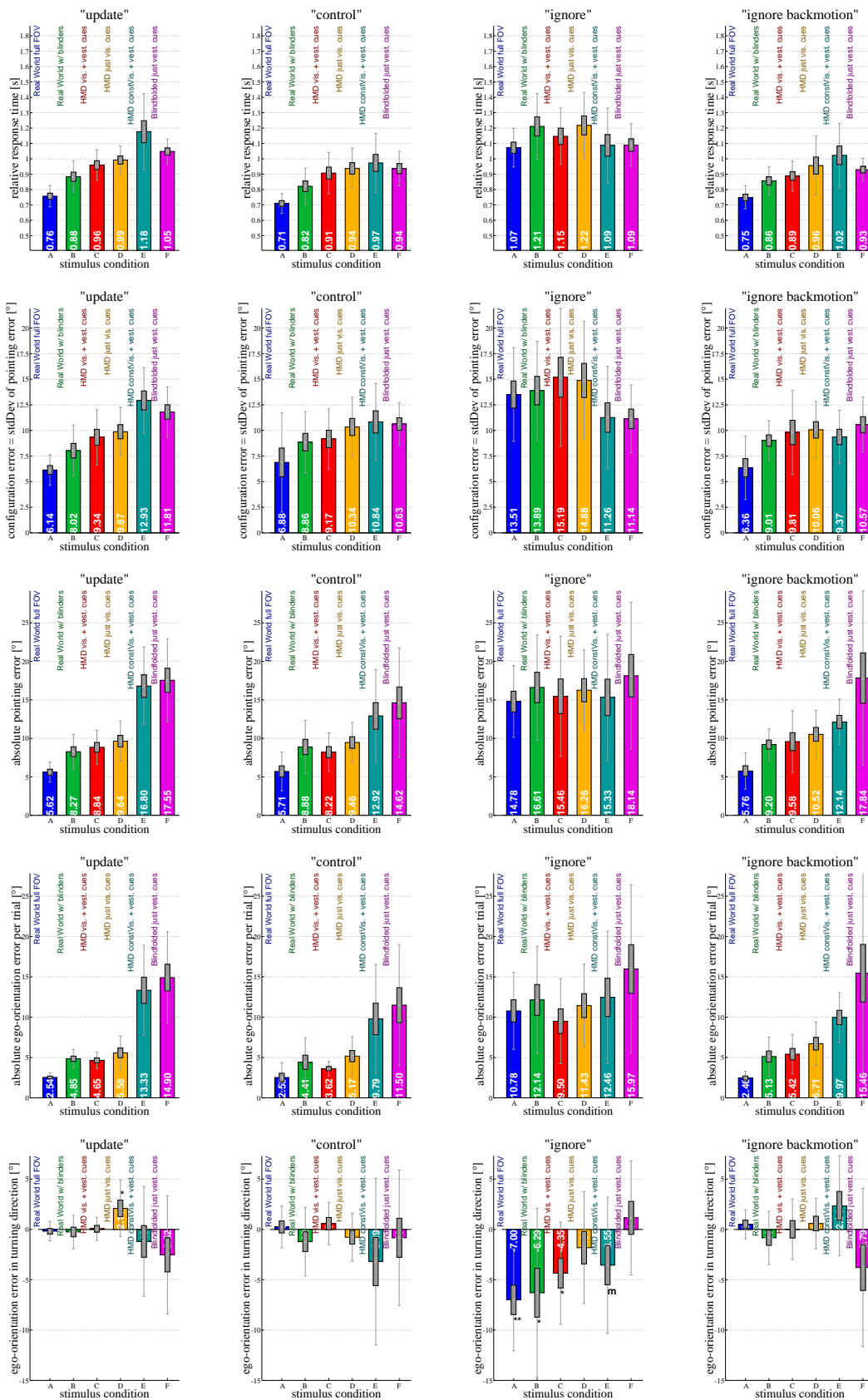
Figure 59: Compilation of all dependent variables for Experiment REAL WORLD VERSUS VR, grouped by spatial updating condition.

A.5   Overview figures for Experiment SIMULATION PARAMETERS

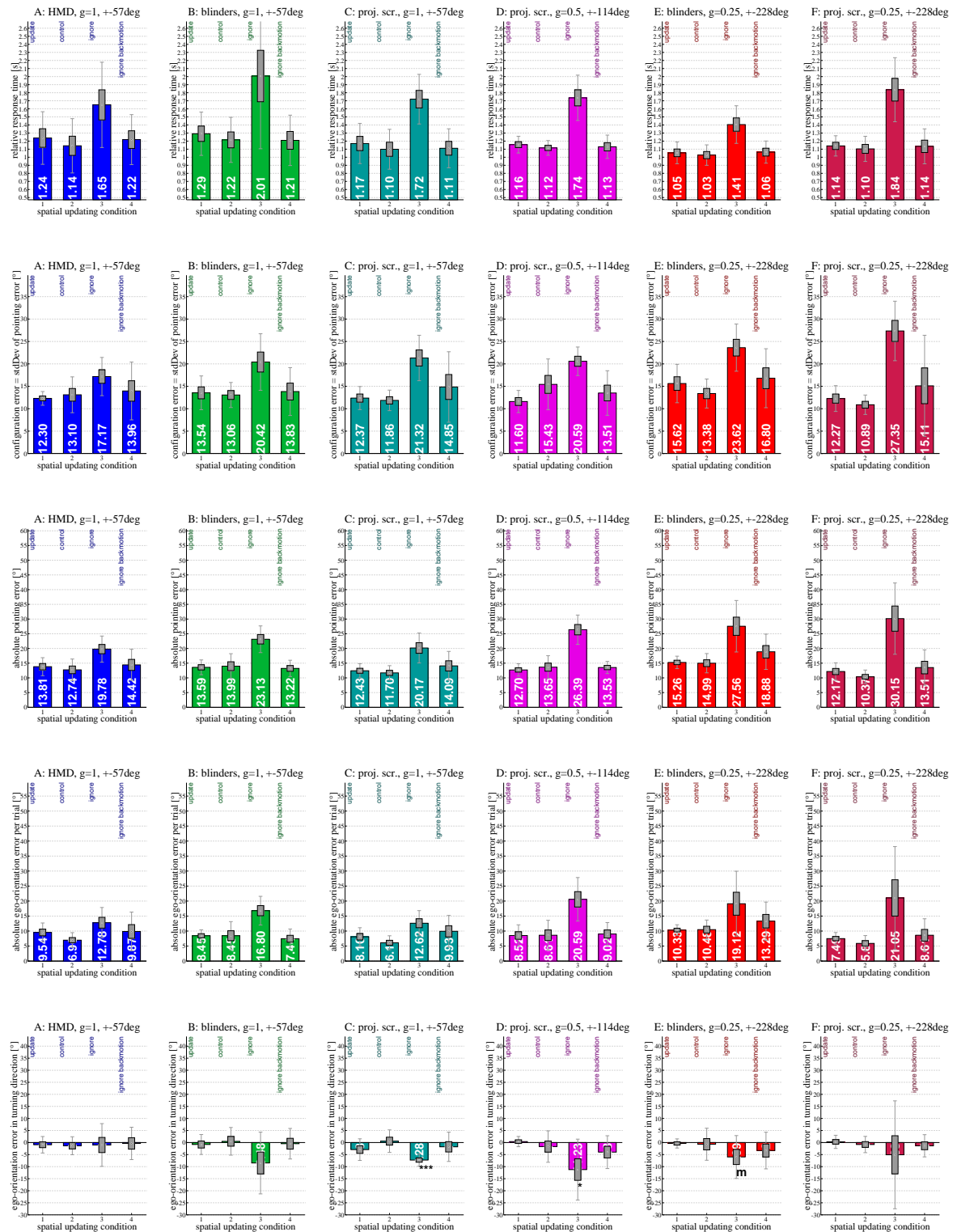**Figure 60:** Compilation of all dependent variables for Experiment SIMULATION PARAMETERS, for block A-F. Note the typical response pattern for obligatory and automatic spatial updating in all conditions: UPDATE performance is comparable to CONTROL performance, whereas IGNORE performance is considerably worse.
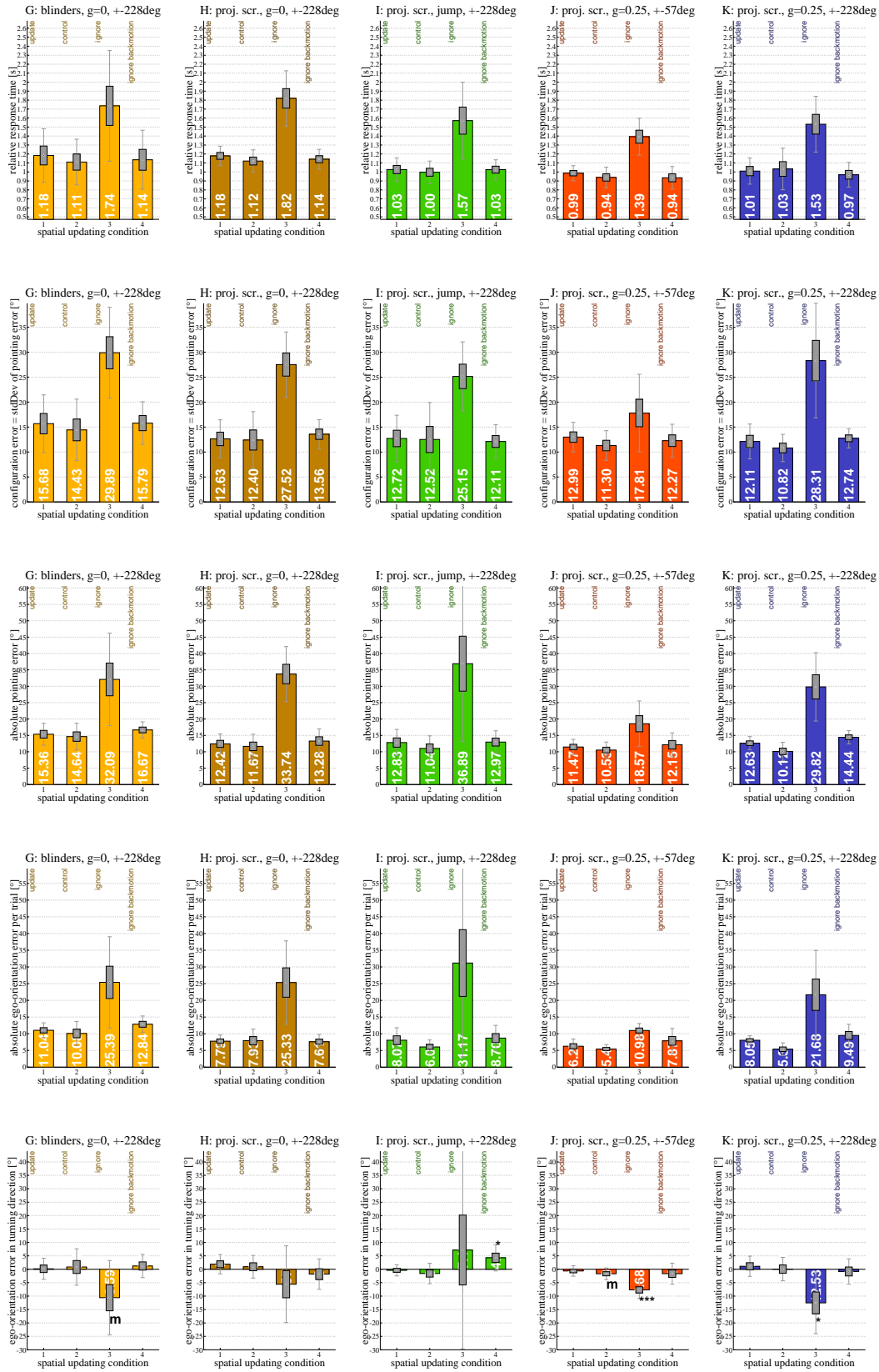
Figure 61: Compilation of all dependent variables for Experiment SIMULATION PARAMETERS, for block H-K.
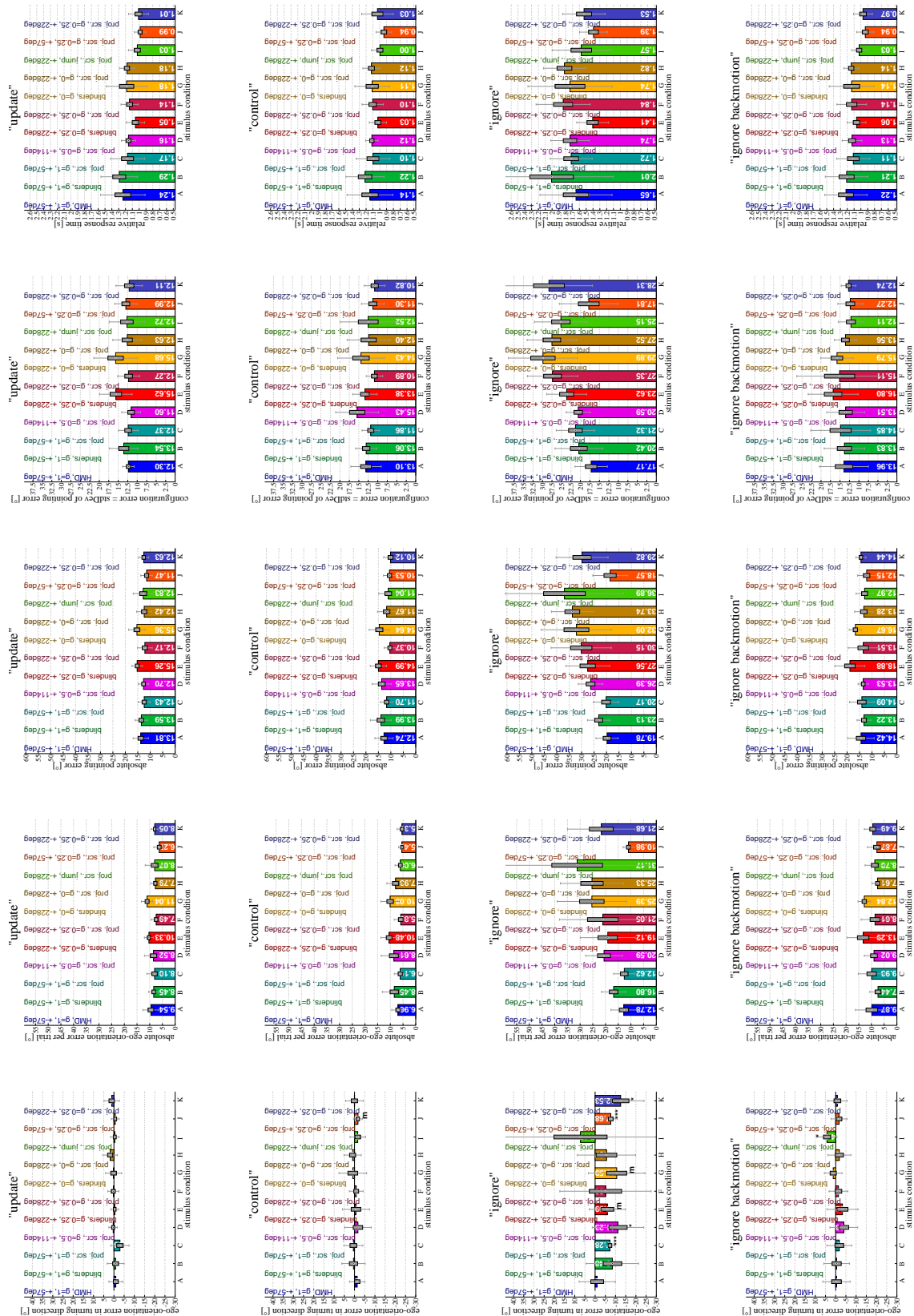
Figure 62: Compilation of all dependent variables for Experiment SIMULATION PARAMETERS, grouped by spatial updating condition.

A.6 Overview figures for Experiment LANDMARKS VERSUS OPTIC FLOW

Figure 63: Compilation of all dependent variables for Experiment LANDMARKS VERSUS OPTIC FLOW, grouped by stimulus combination. Note that only the LANDMARKS conditions show the typical response pattern for obligatory and automatic spatial updating.

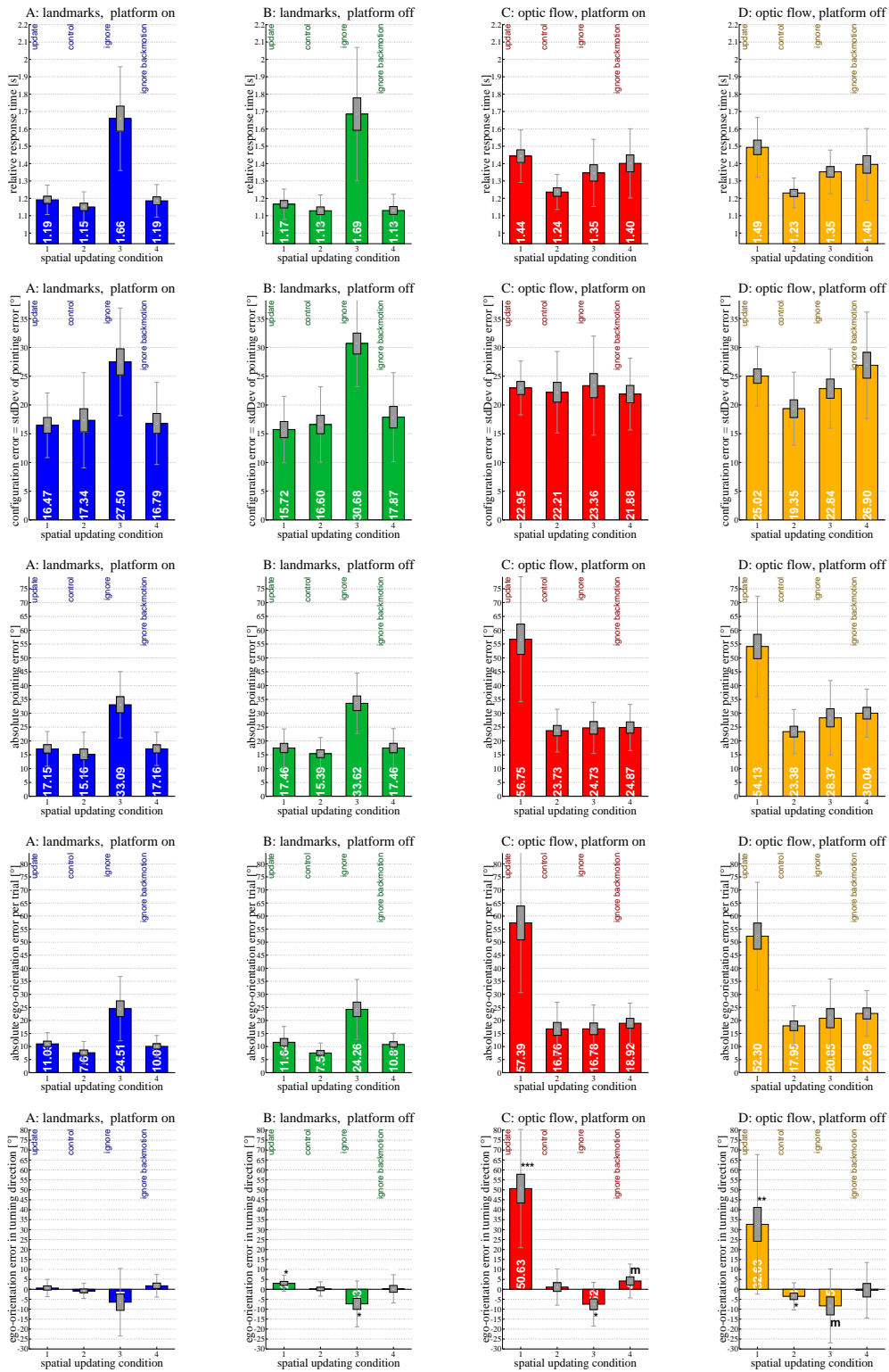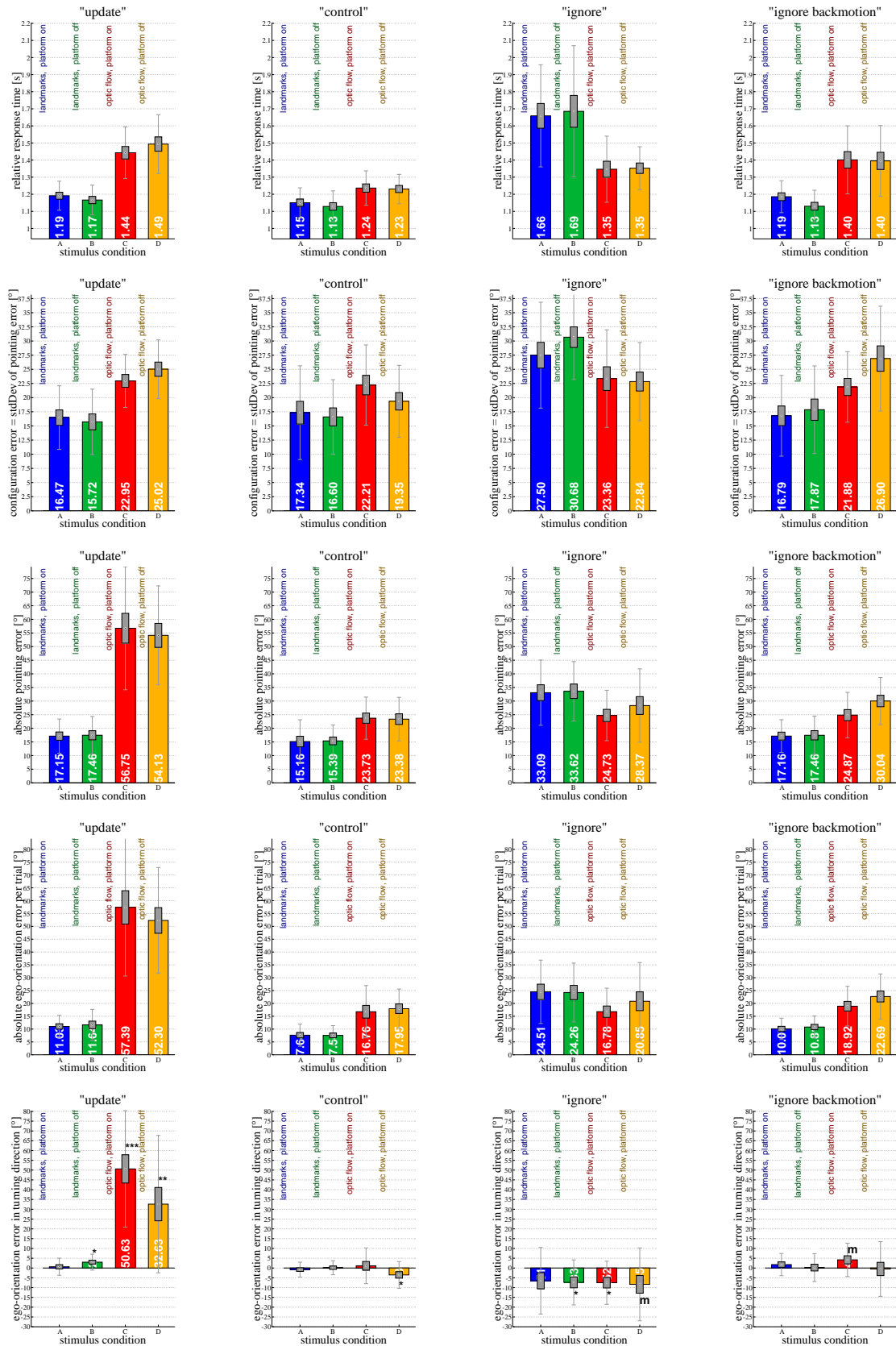**Figure 64:** Compilation of all dependent variables for Experiment LANDMARKS VERSUS OPTIC FLOW, grouped by spatial updating condition. Note that overall performance decrease in the UPDATE condition when landmarks are missing.

References

Alfano, P. L. & Michel, G. F. (1990). Restricting the field of view: Perceptual and performance effects. *Perceptual and Motor Skills*, **70**, 35–45.

Amorim, M. A. & Stucchi, N. (1997). Viewer- and object-centered mental explorations of an imagined environment are not equivalent. *Cognit. Brain Res.*, **5**(3), 229–239.

Arthur, K. W. (2000). *Effects of Field of View on Performance with Head-Mounted Displays*. Ph.D. thesis, Department of Computer Science, University of North Carolina, Chapel Hill. Available: ftp://ftp.cs.unc.edu/pub/publications/techreports/00-019.pdf.

Bakker, N. H., Werkhoven, P. J., & Passenier, P. O. (1999). The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. *Presence - Teleoperators and Virtual Environments*, **8**(1), 36–53.

Bakker, N. H., Werkhoven, P. J., & Passenier, P. O. (2001). Calibrating Visual Path Integration in VEs. *Presence - Teleoperators and Virtual Environments*, **10**(2), 216–224.

Batschelet, E. (1981). *Circular statistics in biology*. London: Acad. Pr.

Beall, A. C. & Loomis, J. M. (1997). Optic flow and visual analysis of the base-to-final turn. *Int. J. Aviat. Psychol.*, **7**(3), 201–223.

Berger, D., von der Heyde, M., & Bülthoff, H. H. (2002). Attention to visual or vestibular cue appears not to change the weights in the sensor fusion process during body yaw-rotation perception. In H. H. Bülthoff, K. G. Gegenfurtner, H. A. Mallot, & R. Ulrich (Eds.), *Beiträge zur 5. Tübinger Wahrnehmungskonferenz*, p. 186. Knirsch Verlag, Kirchentellinsfurt, Germany. Available: www.kyb.tuebingen.mpg.de//publication.html?publ=1318.

Berthoz, A. (1997). Parietal and hippocampal contribution to topokinetic and topographic memory. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.*, **352**(1360), 1437–1448.

Best, K. (1994). *The Idiots' Guide to Virtual World Design*. Little Star Press. Available: http://www.hitl.washington.edu/scivw/scivw-ftp/pubs/IdiotsGuidetoVR/best.html.

Bles, W., Bos, J. E., de Graaf, B., Groen, E., & Wertheim, A. H. (1998). Motion sickness: Only one provocative conflict? *Brain Res. Bull.*, **47**(5), 481–487.

Bülthoff, H. H., Riecke, B. E., & van Veen, H. A. H. C. (2000). Do we really need vestibular and proprioceptive cues for homing. *Invest. Ophthalmol. Vis. Sci. (ARVO)*, **41**(4), 225B225. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=111.

Bülthoff, H. H., Wallraven, C., & Graf, A. B. A. (2002). View-based dynamic object recognition based on human perception. In *ICPR*, pp. 768–776. IEEE CS Press.

Braitenberg, V. (1984). *Vehicles*. Cambridge, MA: MIT Press.

Bremmer, F., Klam, F., Duhamel, J. R., Ben Hamed, S., & Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *European Journal of Neuroscience*, **16**(8), 1569–1586.

Bremmer, F. & Lappe, M. (1999). The use of optical velocities for distance discrimination and reproduction during visually simulated self motion. *Exp. Brain Res.*, **127**(1), 33–42.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., Zilles, K., & Fink, G. R. (2001). Polymodal motion processing in posterior parietal and promotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, **29**(1), 287–296.

Brockmole, J. R. & Wang, R. F. (2002). Switching between environmental representations in memory. *Cognition*, **83**(3), 295–316.

Bülthoff, H. H. & van Veen, H. A. H. C. (2001). Vision and Action in Virtual Environments: Modern Psychophysics in Spatial Cognition Research. In L. R. Harris & M. Jenkin (Eds.), *Vision and Attention* chap. 12. New York: Springer.

Carpenter, M. & Proffit, D. R. (2001). Comparing viewer and array mental rotations in different planes. *Memory & Cognition*, **29**(3), 441–448.

Chance, S. S., Gaunet, F., Beall, A. C., & Loomis, J. M. (1998). Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence - Teleoperators and Virtual Environments*, **7**(2), 168–178.

Cheung, B. S. K., Howard, I. P., & Money, K. E. (1991). Visually-induced sickness in normal and bilaterally labyrinthine-defective subjects. *Aviat. Space Environ. Med.*, **62**(6), 527–531.

Christou, C. G. & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Mem. Cogn.*, **27**(6), 996–1007.

Christou, C. G. & Bülthoff, H. H. (2000). Spatial updating is facilitated by purely visual cues in a virtual environment. *Invest. Ophthalmol. Vis. Sci.*, **41**(4), 3858B956.

Christou, C. G. & Bülthoff, H. H. (1998). Using Realistic Virtual Environments in the Study of Spatial Encoding. In C. F. et al. (Ed.), *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, Vol. 1849 of *Lecture notes in computer science: Lecture notes in artificial intelligence* (pp. 317 – 332). Berlin Heidelberg: Springer.

Christou, C. G. & Bülthoff, H. H. (2003). Environment-centered reference frames in shape recognition. (submitted).

Christou, C. & Bülthoff, H. (1999). The perception of spatial layout in a virtual world. Tech. rep. 75, Max-Planck Institut für biologische Kybernetik. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1540.

Christou, C., Tjan, B., & Bülthoff, H. (1999). Viewpoint information provided by familiar environment facilitates object identification. Tech. rep. 68, Max-Planck Institut für biologische Kybernetik. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1536.

Cobb, S. V. G., Nichols, S., Ramsey, A., & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence - Teleoperators and Virtual Environments*, **8**(2), 169–186.

Creem, S. H. & Proffitt, D. R. (2000). Egocentric measures of spatial updating: Is there an advantage for action? Poster presented at the 41 st Meeting of the Psychonomic Society, New Orleans, LA.

Creem, S. H., Wraga, M., & Proffitt, D. R. (2001). Imagining physically impossible self-rotations: geometry is more important than gravity. *Cognition*, **81**(1), 41–64.

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*. HarperCollins.

Darken, R. P., Allard, T., & Achille, L. B. (1998). Spatial orientation and wayfinding in large-scale virtual spaces: An introduction. *Presence - Teleoperators and Virtual Environments*, **7**(2), 101–107.

Darken, R. P. & Sibert, J. L. (1996). Navigating Large Virtual Spaces. *International Journal of Human-Computer Interaction*, **8**(1), 49–71.

Dichgans, J. & Brandt, T. (1978). Visual-Vestibular Interaction: Effects on Self-Motion Perception and Postural Control. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Perception*, Vol. VIII of *Handbook of Sensory Physiology* (pp. 756–804). Berlin Heidelberg: Springer.

Diwadkar, V. A. & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science*, **8**(4), 302–307.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors*, **43**(1), 129–146.

Duchon, A., Bud, M., Warren, W. H., & Tarr, M. J. (1999). The role of Optic Flow in Human Path Integration. In *Proceedings of the 40th Annual Meeting of the Psychonomic Society*, p. 48.

Easton, R. D. & Sholl, M. J. (1995). Object-array structure, frames of reference, and retrieval of spatial knowledge. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **21**(2), 483–500.

Edelman, S. & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3-dimensional objects. *Vision Research*, **32**(12), 2385–2400.

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

Etienne, A. S., Maurer, R., & Séguinot, V. (1996). Path Integration in Mammals and its Interaction with Visual Landmarks. *The Journal of Experimental Biology*, **199**(1), 201–209.

Farrell, M. J. (1996). Topographical disorientation. *Neurocase*, **2**(6), 509–520.

Farrell, M. J. & Robertson, I. H. (1998). Mental rotation and the automatic updating of body-centered spatial relationships. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **24**(1), 227–233.

Farrell, M. J. & Robertson, I. H. (2000). The automatic updating of egocentric spatial relationships and its impairment due to right posterior cortical lesions. *Neuropsychologia*, **38**(5), 585–595.

Farrell, M. J. & Thomson, J. A. (1998). Automatic spatial updating during locomotion without vision. *Q. J. Exp. Psychol. Sect A-Hum. Exp. Psychol.*, **51**(3), 637–654.

Fischer, M. H. & Kornmüller, A. E. (1930). Optokinetisch ausgelöste Bewegungswahrnehmung und optokinetischer Nystagmus [Optokinetically induced motion perception and optokinetic nystagmus]. *Journal für Psychologie und Neurologie*, 273–308.

Flach, J. M. (1990). Control with an eye for perception: Precursors to an active psychophysics. *Ecol. Psychol.*, **2**(2), 83.

Franz, M. O., Schölkopf, B., Mallot, H. A., & Bülthoff, H. (1998). Where did I take that snapshot? Scene-based homing by image matching. *Biol. Cybern.*, **79**(3), 191–202.

Freyd, J. J. & Finke, R. A. (1984). Representational momentum. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **10**, 126–132.

Fujita, N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (1993). The Encoding-Error Model of Pathway Completion without Vision. *Geographical Analysis*, **25**(4), 295–314.

Gallistel, C. R. (1990). *The organization of learning*. Learning, development, and conceptual change. Cambridge, MA, USA: MIT Press.

Goldin, S. E. & Thorndyke, P. W. (1982). Simulating Navigation for Spatial Knowledge Acquisition. *Hum. Factors*, **24**(4), 457–471.

Goldstein, E. B. (1996). *Sensation and perception* (4th edition). Brooks/Cole.

Golledge, R. G. (Ed.). (1999). *Wayfinding Behavior: Cognitive mapping and other spatial processes*. Baltimore: Johns Hopkins.

Gouteux, S. & Spelke, E. S. (2001). Children's use of geometry and landmarks to reorient in an open space. *Cognition*, **81**(2), 119–148.

Guedry, F. E., Rupert, A. R., & Reschke, M. F. (1998). Motion sickness and development of synergy within the spatial orientation system. A hypothetical unifying concept. *Brain Research Bulletin*, **47**(5), 475–480.

Haber, L., Haber, R. N., Penningroth, S., & Novak et al., K. (1993). Comparison of nine methods of indicating the direction to objects: Data from blind adults. *Perception*, **22**(1), 35–47.

Hendrix, C. & Barfield, W. (1996a). Presence within virtual environments as a function of visual display parameters. *Presence - Teleoperators and Virtual Environments*, **5**(3), 274–289.

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

Hettinger, L. J., Nelson, W. T., & Haas, M. W. (1996). Target detection performance in helmet-mounted and conventional dome displays. *International Journal of Aviation Psychology*, **6**(4), 321–334.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in Cognitive Maps. *Cognitive Psychology*, **13**(2), 149–206.

Hollins, M. & Kelley, E. K. (1988). Spatial updating in blind and sighted people. *Percept. Psychophys.*, **43**(4), 380–388.

Howarth, P. A. & Costello, P. J. (1997). The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, **18**(2), 107–116.

Hubbard, T. L. & Bharucha, J. J. (1988). Judged displacements in apparent vertical and horizontal motion. *Perception and Psychophysics*, **44**(3), 211–221.

Hunt, E. & Waller, D. (1999). Orientation and Wayfinding: A Review. Available: depts.washington.edu/huntlab/vr/pubs/huntreview.pdf.

IJsselsteijn, W., de Ridder, H., Freeman, J., Avons, S. E., & Bouwhuis, D. (2001). Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence. *Presence - Teleoperators and Virtual Environments*, **10**(3), 298–311.

IJsselsteijn, W. (2002). Elements of a multi-level theory of presence: Phenomenology, mental processing, and neural correlates. In F. Gouveia (Ed.), *PRESENCE 2002*, pp. 245–259. Universidare Fernando Pessoa, Porto, Portugal.

ISA (1998). *Intelligenz-Struktur-Analyse*. Frankfurt: Sweets & Zeitlinger B. V., Sweets Test Services. ITB Institut für Test- und Begabungsforschung GmbH, Bonn.

Ivanenko, Y. P., Viaud-Delmon, I., Siegler, I., Israël, I., & Berthoz, A. (1998). The vestibulo-ocular reflex and angular displacement perception in darkness in humans: adaptation to a virtual environment. *Neurosci. Lett.*, **241**(2-3), 167–170.

Kappe, B., van Erp, J., & Korteling, J. E. H. (1999). Effects of head-slaved and peripheral displays on lane-keeping performance and spatial orientation. *Human Factors*, **41**(3), 453–466.

Kearns, M. J., Warren, W. H., Duchon, A. P., & Tarr, M. J. (2002). Path integration from optic flow and body senses in a homing task. *Perception*, **31**(3), 349–374.

Kennedy, R. S., Lanham, D. S., Drexler, J. M., Massey, C. J., & Lilienthal, M. G. (1997). A comparison of cybersickness incidences, symptom profiles, measurement techniques, and suggestions for further research. *Presence: Teleoperator & Virtual Environment*, **6**(6), 638–644.

Klatzky, R. L. (1999). Path completion after haptic exploration without vision: Implications for haptic spatial representations. *Perception & Psychophysics*, **61**(2), 220–235.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychol. Sci.*, **9**(4), 293–298.

Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (1997). Encoding spatial reresentations through nonvisually guided locomotion: Test of human path integration. In D. Medin (Ed.), *The psychology of learning and motivation*, Vol. 37 (pp. 41–84). San Diego, CA: Acad. Press.

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1990). Acquisition of route and survey knowledge in the absence of vision. *J. Mot. Behav.*, **22**(1), 19–43.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*. Cambridge, MA: MIT Press: A Bradford Book.

Kozhevnikov, M. & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, **8**(3), 439–453.

Lathan, C. E., Wall, C. W., & Harris, L. R. (1995). Human eye-movement response to z-axis linear acceleration - the effect of varying the phase-relationships between visual and vestibular inputs. *Exp. Brain Res.*, **103**(2), 256–266.

Lehnung, M., Haaland, V. O., Pohl, J., & Leplow, B. (2001). Compass-versus finger-pointing tasks: The influence of different methods of assessment on age-related orientation performance in children. *Journal of Environmental Psychology*, **21**(3), 283–289.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence - Teleoperators and Virtual Environments*, **10**(3), 282–297.

Loomis, J. M. & Beall, A. C. (1998). Visually controlled locomotion: Its dependence on optic flow, three-dimensional space perception, and cognition. *Ecol. Psychol.*, **10**(3-4), 271–285.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behav. Res. Methods Instr. Comput.*, **31**(4), 557–564.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1993). Nonvisual navigation by blind and sighted: assessment of path integration ability. *J. Exp. Psychol. Gen.*, **122**(1), 73–91.

Loomis, J. M., Da Silva, J. A., Philbeck, J. W., & Fukusima, S. S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science*, **5**(3), 72–77.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human Navigation by Path Integration. In R. G. Golledge (Ed.), *Wayfinding Behavior: Cognitive mapping and other spatial processes* (pp. 125–151). Baltimore: Johns Hopkins.

Loomis, J. M., Klatzky, R. L., & Lederman, S. J. (1991). Similarity of tactual and visual picture recognition with limited field of view. *Perception*, **20**(2), 167–177.

Mach, E. (1922). *Die Analyse der Empfindungen [The analysis of sensations]*. Jena: Gustav Fischer.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998a). Knowing where and getting there: A human navigation network. *Science*, **280**(1), 921–924.

Maguire, E. A., Frith, C. D., Burgess, N., Donnett, J. G., & O'Keefe, J. (1998b). Knowing where things are: Parahippocampal involvement in encoding object locations in virtual large-scale space. *Journal of Cognitive Neuroscience*, **10**(1), 61–76.

Marlinsky, V. V. (1999a). Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man - II. Estimations of rotations about the earth-vertical axis. *Neuroscience*, **90**(2), 395–401.

Marlinsky, V. V. (1999b). Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man - III. Route inference. *Neuroscience*, **90**(2), 403–411.

Maurer, R. & Séguinot, V. (1995). What is modelling for? – A critical review of the models of path integration. *Journal of Theoretical Biology*, **175**(4), 457–475.

May, M. & Klatzky, R. L. (2000). Path integration while ignoring irrelevant movement. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **26**(1), 169–186.

May, M. (1996). Cognitive and embodied modes of spatial imagery. *Psychologische Beiträge*, **38**, 418–434.

May, M. (2000). *Kognition im Umraum*. Studien zur Kognitionswissenschaft. Wiesbaden: DUV: Kognitionswissenschaft.

May, M., Péruch, P., & Savoyant, A. (1995). Navigating in a virtual environment with map-acquired knowledge: Encoding and alignment effects. *Ecol. Psychol.*, **7**(1), 21–36.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., Gothard, K., Jung, M. W., Knierim, J. J., Kudrimoti, H., Qin, Y., Skaggs, W. E., Suster, M., & Weaver, K. L. (1996). Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *Journal of Experimental Biology*, **199**, 173–185.

Mergner, T. & Becker, W. (1990). Perception of horizontal self-rotation: Multisensory and cognitive aspects. In R. Warren & A. H. Wertheim (Eds.), *Perception & Control of Self-Motion* (pp. 219–263). New Jersey, London: Erlbaum.

Mergner, T. & Glasauer, S. (1999). A simple model of vestibular canal-otolith signal fusion. *Ann.NY Acad.Sci.*, **871**, 430–434.

Mergner, T. & Rosemeier, T. (1998). Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions - a conceptual model. *Brain Res. Rev.*, **28**(1-2), 118–135.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience*, **6**, 414–417.

Mittelstaedt, H. (2000). Triple-loop model of path control by head direction and place cells. *Biol. Cybern.*, **83**(3), 261–270.

Mittelstaedt, H. & Mittelstaedt, M.-L. (1982). Homing by path integration. In F. Papi & H. Wallraff (Eds.), *Avian navigation* (pp. 290–297). Berlin: Springer.

Mittelstaedt, M.-L. & Glasauer, S. (1991). Idiothetic Navigation in Gerbils and Humans. *Zool. Jb. Physiol.*, **95**, 427–435.

Mon-Williams, M. & Wann, J. P. (1998). Binocular virtual reality displays: When problems do and don't occur. *Human Factors*, **40**(1), 42–49.

Müller, M. & Wehner, R. (1988). Path integration in desert ants *Cataglyphis fortis. Proceedings of the National Academy of Sciences*, **85**, 5287–5290.

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., & McKinley, R. L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. *Human Factors*, **40**(3), 452–460.

O'Keefe, J. & Dostrovsky, J. (1971). The hippocampus as spatial map. Preliminary evidence from unit activity in the freely moving rat. *Brain Research*, **34**, 171–175.

O'Keefe, J. & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford, England: Clarendon.

Palmisano, S. & Gillam, B. (1998). Stimulus eccentricity and spatial frequency interact to determine circular vection. *Perception*, **27**(9), 1067–1077.

Poucet, B. (1993). Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological-Review*, **100**(2), 163–182.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychol. Bull.*, **86**(4), 777–803.

Presson, C. C. & Hazelrigg, M. D. (1984). Building Spatial Representations Through Primary and Secondary Learning. *Journal of Experimental Psychology-Learning Memory and Cognition*, **10**(4), 716–722.

Presson, C. C. & Montello, D. R. (1994). Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception*, **23**(12), 1447–1455.

Péruch, P. & Gaunet, F. (1998). Virtual environments as a promising tool for investigating human spatial cognition. *Cah. Psychol. Cogn.-Curr. Psychol. Cogn.*, **17**(4-5), 881–899.

Péruch, P., May, M., & Wartenberg, F. (1997). Homing in virtual environments: Effects of field of view and path layout. *Perception*, **26**(3), 301–311.

Regenbrecht, H. (1999). *Faktoren für Präsenz in virtueller Architektur [Factors for the sense of presence within virtual architecture]*. Unpublished doctoral dissertation, Bauhaus University, Weimar, Germany. Available: http://www.uni-weimar.de/ub/diss/Regenbrecht12012000.html.

Regenbrecht, H. & Schubert, T. (2002). Real and illusory interactions enhance presence in virtual environments. *Presence - Teleoperators and Virtual Environments*, **11**(4), 425–434.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communicaions* (pp. 303–318). New York: Wiley.

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Mem. Cogn.*, **27**(4), 741–750.

Riecke, B. E. & van Veen, H. A. H. C. (1999). Heimfinden in virtuellen Umgebungen. In H. Bülthoff, M. Fahle, K. Gegenfurtner, & H. Mallot (Eds.), *Beiträge der 2. Tübinger Wahrnehmungskonferenz*, p. 84 Max Planck Institute for Biological Cybernetics, Germany. Knirsch Verlag, Kirchentellinsfurt, Germany. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=308.

Riecke, B. E., van Veen, H. A. H. C., & Bülthoff, H. H. (1999). Is homing by optic flow possible? *J. Cogn. Neurosci.*, **1**, 76. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=309.

Riecke, B. E., van Veen, H. A. H. C., & Bülthoff, H. H. (2000a). Reicht optischer Fluß wirklich nicht zum Heimfinden? In H. Bülthoff, M. Fahle, K. Gegenfurtner, & H. Mallot (Eds.), *Beiträge der 3. Tübinger Wahrnehmungskonferenz*, p. 139 Max Planck Institute for Biological Cybernetics, Germany. Knirsch Verlag, Kirchentellinsfurt, Germany.

Riecke, B. E., van Veen, H. A. H. C., & Bülthoff, H. H. (2000b). Visual Homing is possible without Landmarks: A Path Integration Study in Virtual Reality. Tech. rep. 82, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. [Avaliable: ftp://ftp.kyb.tuebingen.mpg.de/pub/mpi-memos/pdf/TR-082.pdf]. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1203.

Riecke, B. E., van Veen, H. A. H. C., & 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. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1202.

Riecke, B. E., van Veen, H. A., & Bülthoff, H. H. (2000). Visual Homing is possible without Landmarks: A Path Integration Study in Virtual Reality. In M. von der Heyde & H. H. Bülthoff (Eds.), *Perception and Action in Virtual Environments* chap. 6, (pp. 97–134). Max Planck Institute for Biological Cybernetics, Germany: Cognitive and Computational Psychophysics Department.

Riecke, B. E. & von der Heyde, M. (2002). Qualitative Modeling of Spatial Orientation Processes using Logical Propositions: Interconnecting Spatial Presence, Spatial Updating, Piloting, and Spatial Cognition. Tech. rep. 100, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. Avaliable: http://www.kyb.tuebingen.mpg.de/publication.html?publ=2021.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2001a). How do we know where we are? Contribution and interaction of visual and vestibular cues for spatial updating in real and virtual environments. In H. H. Bülthoff, K. G. Gegenfurtner, H. A. Mallot, & R. Ulrich (Eds.), *Beiträge zur 4. Tübinger Wahrnehmungskonferenz*, p. 146 Max Planck Institute for Biological Cybernetics, Germany. Knirsch Verlag, Kirchentellinsfurt, Germany. Available: www.kyb.tuebingen.mpg.de/bu/poster/2001/b_riecke_twk2001.pdf.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2001b). How Real is Virtual Reality Really? Comparing Spatial Updating using Pointing Tasks in Real and Virtual Environments. *Journal of Vision*, **1**(3), 321a. http://journalofvision.org/1/3/321/.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2002a). Spatial updating experiments in Virtual Reality: What makes the world turn around in our head? In H. H. Bülthoff, K. G. Gegenfurtner, H. A. Mallot, & R. Ulrich (Eds.), *Beiträge zur 5. Tübinger Wahrnehmungskonferenz*, p. 162. Knirsch Verlag, Kirchentellinsfurt, Germany. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=632.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2002b). Teleporting works - Spatial updating experiments in Virtual Tübingen. In *OPAM*, Talk presented at the 10th annual meeting of OPAM, Kansas City, United States. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1952.

Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2003). Spatial updating in virtual environments: What are vestibular cues good for? *Journal of Vision*, **2**(7), 421a. http://journalofvision.org/2/7/421/.

Riecke, B. E. (1998). Untersuchung des menschlichen Navigationsverhaltens anhand von Heimfindeexperimenten in virtuellen Umgebungen. Master's thesis, Eberhard-Karls-Universität Tübingen, Fakultät für Physik. Available: www.kyb.tuebingen.mpg.de/bu/people/bernie/diplomaThesis.pdf.

Rieser, John, J., Guth, David, A., & Hill, Everett, W. (1982). Mental processes mediating independent travel: Implications for orientation and mobility. *Journal of Visual Impairment and Blindness*, **76**(6), 213–218.

Rieser, J. J. & Rider, E. A. (1991). young childrens spatial orientation with respect to multiple targets when walking without vision. *Developmental Psychology*, **27**(1), 97–107.

Rieser, J. J. (1989). Access to knowledge of spatial structure at novel points of observation. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **15**(6), 1157–1165.

Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, **15**(2), 173–188.

Rieser, J. J., Hill, E. W., Talor, C. R., Bradfield, A., & Rosen., S. (1992). Visual Experience, Visual Field Size, and the Development of Nonvisual Sensitivity to the Spatial Structure of Outdoor Neighborhoods Explored by Walking. *Journal of Experimental Psychology: General*, **121**(2), 210–221.

Roskos-Ewoldsen, B., McNamara, T. P., Shelton, A. L., & Carr, W. (1998). Mental representations of large and small spatial layouts are orientation dependent. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **24**(1), 215 – 226.

Ruddle, R. A. & Jones, D. M. (2001). Movement in cluttered virtual environments. *Presence - Teleoperators and Virtual Environments*, **10**(5), 511–524.

Ruddle, R. A., Payne, S. J., & Jones, D. M. (1997). Navigating Buildings in 'Desk-Top' Virtual Environments: Experimental Investigations Using Extended Navigational Experience. *J. Exp. Psychol.-Appl.*, **3**(2), 143–159.

Sadalla, E. K. & Montello, D. R. (1989). Remembering changes in direction. *Environ. Behav.*, **21**(3), 346–363.

Samsonovich, A. & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, **17**(15), 5900–5920.

Sauvé, J. P. (1989). *L'orientation spatiale: formalisation d'un modèle de mémorisation égocentrée et expérimentation chez l'homme.* Ph.D. thesis, Université d'Aix-Marseille II.

Schloerb, D. W. (1995). A Quantitative Measure of Telepresence. *Presence - Teleoperators and Virtual Environments*, **4**(1), 64–81.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence - Teleoperators and Virtual Environments*, **10**(3), 266–281.

Schubert, T. (2002). Five Theses on the Book Problem: Presence in Books, Film, and VR. In F. Gouveia (Ed.), *PRESENCE 2002*, pp. 53–58. Universidare Fernando Pessoa, Porto, Portugal.

Schulte-Pelkum, J., Riecke, B. E., & von der Heyde, M. (2003). Influence of display parameters on perceiving visually simulated ego-rotations - a systematic investigation. In *TWK*, (submitted). Available: www.kyb.tuebingen.mpg.de/publication.html?publ=2024.

Schulte-Pelkum, J., Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2002). Perceiving and controlling simulated ego-rotations by optic flow: Influence of field of view (FOV) and display devices on ego-motion perception. In *OPAM*, Poster presented at the 10th annual meeting of OPAM, Kansas City, United States. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=1960.

Schulte-Pelkum, J., Riecke, B. E., von der Heyde, M., & Bülthoff, H. H. (2003). Screen curvature does influence the perception of visually simulated ego-rotations. In *VSS*, (submitted to VSS 2003) Sarasota, Florida, United States. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=2025.

Schweigart, G., Mergner, T., Evdokimidis, I., Morand, S., & Becker, W. (1997). Gaze stabilization by optokinetic reflex (OKR) and vestibulo- ocular reflex (VOR) during active head rotation in man. *Vision Res.*, **37**(12), 1643–1652.

Seguinot, V., Maurer, R., & Etienne, A. S. (1993). Dead reckoning in a small mammal: the evaluation of distance. *J. Comp. Physiol. A.*, **173**, 103–113.

Shelton, A. L. & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychon. Bull. Rev.*, **4**(1), 102–106.

Shelton, A. L. & McNamara, T. P. (2001). Systems of Spatial Reference in Human Memory. *Cognitive Psychology*, **43**(4), 274–310.

Sholl, M. J. (1989). The relation between horizontality and rod-and-frame and vestibular navigational performance. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **15**(1), 110–125.

Simons, D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, **1**(7), 261–267.

Simons, D. J. & Wang, R. F. (1998). Perceiving real-world viewpoint changes. *Psychol. Sci.*, **9**(4), 315–320.

Simons, D. J., Wang, R. X. F., & Roddenberry, D. (2002). Object recognition is mediated by extraretinal information. *Perception & Psychophysics*, **64**(4), 521–530.

Slater, M. (2002). Presence and the sixth sense. *Presence - Teleoperators and Virtual Environments*, **11**(4), 435–439.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence-Teleoperators and Virtual Environments*, **7**(4), 327–351.

Stevens, S. S. & Greenbaum, H. B. (1966). Regression effect in psychophysical judgement. *Percept. Psychophys.*, **1**, 439–446.

Stumpf, H. & Fay, E. (1983). *Schlauchfiguren - Ein Test zur Beurteilung des räumlichen Vorstellungsvermögens*. Göttingen: Hogrefe.

Surdick, R. T., Davis, E. T., King, R. A., & Hodges, L. F. (1997). The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence - Teleoperators and Virtual Environments*, **6**(5), 513–531.

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

Taube, J. S., Muller, R. U., & Ranck, J. B. J. (1990b). Head direction cell recorded from the postsubiculum in freely moving rats. II. Effect of environmental manipulations. *Journal of Neuroscience*, **10**(2), 436–447.

Thornton, I. M. & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition*, **9**(1-2), 1–7.

Tolman, E. C. (1948). Cognitive maps in Rats and Men. *Psychological Review*, **55**, 189–208.

Trullier, O., Wiener, S. I., Berthoz, A., & Meyer, J. A. (1997). Biologically based artificial navigation systems: Review and prospects. *Prog. Neurobiol.*, **51**(5), 483–544.

van Veen, H. A. H. C., Riecke, B. E., & Bülthoff, H. H. (1999). Visual Homing to a Virtual Home. *Invest. Ophthalmol. Vis. Sci.*, **40**, 4200B3.

van Veen, H. A. H. C., Distler, H. K., Braun, S. J., & Bülthoff, H. H. (1998). Navigating through a virtual city: Using virtual reality technology to study human action and perception. *Future Generations Computer Systems*, **14**(3-4), 231–242.

von der Heyde, M. (2000). The Motion-Lab - A Virtual Reality Laboratory for Spatial Updating Experiments. Tech. rep. 86, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=633.

von der Heyde, M. & Riecke, B. E. (2001). How to cheat in motion simulation - comparing the engineering and fun ride approach to motion cueing. Tech. rep. 89, Max Planck Institute for Biological Cybernetics, Tübingen, Germany. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=635.

von der Heyde, M. (2001). *A Distributed Virtual Reality System for Spatial Updating: Concepts, Implementation, and Experiments*. Ph.D. thesis, Universität Bielefeld - Technische Fakultät.

Wallis, G. & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(8), 4800–4804.

Wallraven, C. & Bülthoff, H. H. (2001). View-based recognition under illumination changes using local features. In *CVPR 2001 - Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation (Proceedings-CD)*.

Wang, R. F. (1999). Representing a stable environment by egocentric updating and invariant representations. *Spatial Cognition and Computation*, **1**, 431–445.

Wang, R. F. & Brockmole, J. R. (2003). Human Updating in Nested Environments. Journal of Experimental Psychology - Learning Memory & Cognition (in press).

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

Wang, R. X. F. & Simons, D. J. (1999). Active and passive scene recognition across views. *Cognition*, **70**(2), 191–210.

Wang, R. X. F. & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, **77**(3), 215–250.

Warren, R. & Wertheim, A. H. (Eds.). (1990). *Perception & Control of Self-Motion*. New Jersey, London: Erlbaum.

Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nat. Neurosci.*, **4**(2), 213–216.

Wartenberg, F., May, M., & Péruch, P. (1998). Spatial Orientation in Virtual Environments: Background Considerations and Experiments. In C. Freska, C. Habel, & K. F. Wender (Eds.), *Spatial Cognition: an interdisciplinary appraoch to representing and processing spatial knowledge*, Vol. 1404 of *Lecture notes in computer science: Lecture notes in artificial intelligence* (pp. 469–489). Berlin, Heidelberg: Springer.

Wehner, R., Michel, B., & Antonsen, P. (1996). Visual navigation in insects: Coupling of egocentric and geocentric information. *Journal of Experimental Biology*, **199**(1), 129–140.

Wertheim, A. H. (1994a). Motion perception - rights, wrongs and further speculations - response. *Behav. Brain Sci.*, **17**(2), 340 – 348.

Wertheim, A. H. (1994b). Motion perception during self-motion - the direct versus inferential controversy revisited. *Behav. Brain Sci.*, **17**(2), 293 – 311.

Witmer, B. G. & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence - Teleoperators and Virtual Environments*, **7**(3), 225–240.

Wolpert, L. (1990). Field-of-view information for self-motion perception. In R. Warren & A. H. Wertheim (Eds.), *Perception & Control of Self-Motion* (pp. 101–126). New Jersey, London: Erlbaum.

Wraga, M., Creem, S. H., & Proffitt, D. R. (1999a). The influence of spatial reference frames on imagined object- and viewer rotations. *Acta Psychol.*, **102**(2-3), 247–264.

Wraga, M., Creem, S. H., & Proffitt, D. R. (1999b). Spatial updating of an irregularly shaped virtual array during self- and display rotations. *Invest. Ophthalmol. Vis. Sci.*, **40**(4), 1.

Wraga, M., Creem, S. H., & Proffitt, D. R. (2000). Updating displays after imagined object and viewer rotations. *J. Exp. Psychol.-Learn. Mem. Cogn.*, **26**(1), 151–168.

Wraga, M., Creem, S. H., & Proffitt, D. R. (2003). Spatial updating of virtual displays during self- and display rotation. (under review).

Yardley, L. & Higgins, M. (1998). Spatial updating during rotation: The role of vestibular information and mental activity. *J. Vestib. Res.-Equilib. Orientat.*, **8**(6), 435–442.

Acknowledgments

Thanks to

Prof. Bülthoff and Prof. Ruder for being my advisers and giving me the wonderful opportunity to do this research;

Markus von der Heyde and Douglas W. Cunningham for their friendship, discussions, and support in writing, technical issues, getting myself organized, and many outstanding gustatory experiences;

Claudia Holt for friendship, proofreading, and help with graphics;

Douglas W. Cunningham for helping me to learn how to write scientifically - something the university did unfortunately neither manage nor attempt;

The members of the AG Bülthoff for inspiring discussions, grill-parties, and the enjoyable and supportive atmosphere;

My participants for their patience and contributions to this work;

The authors of Loomis et al. (1993) and Péruch et al. (1997) for letting us use their raw data for the comparisons in subsection II.11.1;

The Max Planck Society, the Deutsche Forschungsgemeinschaft (SFB 550 "Erkennen, Lokalisieren, Handeln: neurokognitive Mechanismen und ihre Flexibilität"), and the European Community (IST-2001-39223, FET Proactive Initiative, project "POEMS" (Perceptually Oriented Ego-Motion Simulation, www.poems-project.info) for financial support;

and most importantly,

My family for their belief in me and for supporting me for all these years;

My beloved Iris Torchalla for her love and support.