Spatial updating in real and virtual environments - contribution and interaction of visual and vestibular cues

Bernhard E. Riecke*
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Markus von der Heyde†
SCC, Bauhaus-Universität Weimar, Germany

Heinrich H. Bülthoff‡
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

Abstract

INTRODUCTION: When we move through the environment, the self-to-surround relations constantly change. Nevertheless, we perceive the world as stable. A process that is critical to this perceived stability is "spatial updating", which automatically updates our egocentric mental spatial representation of the surround according to our current self-motion. According to the prevailing opinion, vestibular and proprioceptive cues are absolutely required for spatial updating. Here, we challenge this notion by varying visual and vestibular contributions independently in a high-fidelity VR setup. METHODS: In a learning phase, participants learned the positions of twelve targets attached to the walls of a 5x5m room. In the testing phase, participants saw either the real room or a photo-realistic copy presented via a head-mounted display (HMD). Vestibular cues were applied using a motion platform. Participants' task was to point "as accurately and quickly as possible" to four targets announced consecutively via headphones after rotations around the vertical axis into different positions. RESULTS: Automatic spatial updating was observed whenever useful visual information was available: Paticipants had no problem mentally updating their orientation in space, irrespective of turning angle. Performance, quantified as response time, configuration error, and pointing error, was best in the real world condition. However, when the field of view was limited via cardboard blinders to match that of the HMD (40x30°), performance decreased and was comparable to the HMD condition. Presenting turning information only visually (through the HMD) hardly altered those results. In both the real world and HMD conditions, spatial updating was obligatory in the sense that it was significantly more difficult to ignore ego-turns (i.e., "point as if not having turned") than to update them as usual. CONCLUSION: The rapid pointing paradigm proved to be a useful tool for quantifying spatial updating. We conclude that, at least for the limited turning angles used (<60 deg), the Virtual Reality simulation of ego-rotation was as effective and convincing (i.e., hard to ignore) as its real world counterpart, even when only visual information was presented. This has relevant implications for the design of motion simulators for, e.g., architecture walkthroughs.

CR Categories: H.1.2 [Models and Principles]: User/Machine Systems—Human factors, Human information processing H.5.1 [Information Interfaces and Presentation, (e.g. HCI]: Multimedia Information Systems—Artificial, augmented, and virtual realities H.5.2 [Information Interfaces and Presentation, (e.g. HCI]: User Interfaces—Input devices and strategies, Interaction styles J.4 [Social and Behavioral Sciences]: Psychology

Keywords: spatial updating, Virtual Reality, spatial orientation, motion simulation, human factors, psychophysics

*e-mail: bernhard.riecke@tuebingen.mpg.de
†e-mail: markus.von.der.heyde@scc.uni-weimar.de
‡e-mail: heinrich.buelthoff@tuebingen.mpg.de

1 Introduction

When we move through the environment, the self-to-surround relations constantly change in a rather complex manner. Nevertheless, we perceive the world as stable. Even though this is normally taken for granted, it is one of the most useful and remarkable properties of our spatial representations of the surround - that it is automatically updated according to our current self-motion, thus creating an expectation of what we will perceive. If this current percept matches the expectation, the surround is perceived as stable.

This automatic and seemingly effortless transformation process is typically referred to as "automatic spatial updating" [Farrell and Robertson 1998; Farrell and Thomson 1998; Klatzky et al. 1998; Presson and Montello 1994; Rieser et al. 1982]. It is this process that allows us to locomote, for example, 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. Under most circumstances, spatial updating is even "obligatory" in the sense of being largely beyond conscious control and hard to suppress [Farrell and Robertson 1998; May and Klatzky 2000]. This difference between "automatic" and "obligatory" spatial updating is illustrated in Figure 1.

Vestibular and kinesthetic motion cues proved to be sufficient to enable automatic spatial updating during rotations as well as translations in blindfolded participants [Easton and Sholl 1995; Farrell and Robertson 1998; May and Klatzky 2000]. Conversely, spatial updating is typically impaired when proprioceptive and vestibular cues in particular are missing [Klatzky et al. 1998; May and Klatzky 2000; Presson and Montello 1994; Rieser 1989; Simons and Wang 1998; Wang and Simons 1999; Wang and Spelke 2000; Wraga et al. 2004]. Qualitative errors seem to occur most often when kinesthetic and/or vestibular cues about ego-turns are missing. Klatzky et al. [1998] and May and 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. Such results lead to the prevailing opinion that vestibular and proprioceptive cues are absolutely required for triggering the spatial updating automatism.

In this study, we questioned this notion by performing a spatial updating experiment where different combinations of visual and vestibular cues were compared. In order to be able to independently control vestibular and visual cues, a Virtual Reality setup was used including a motion simulator (6 degree of freedom (DOF) motion

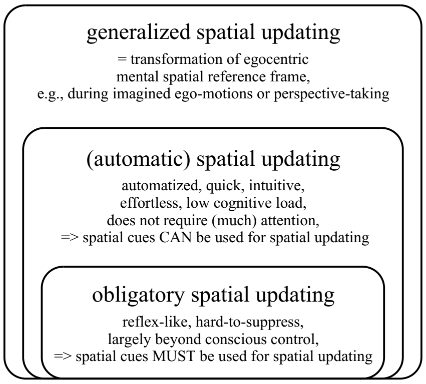

Figure 1: Connection between generalized, automatic, and obligatory spatial updating. At the most general level, **generalized spatial updating** refers to all spatial transformations of our egocentric mental spatial representation. This includes mental perspective-taking or consciously updating our egocentric representation 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 representation. Due to this automaticity, both the cognitive load and attentional demands are minimal, if not zero [e.g., Rieser 1989]. **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.

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.

If visual cues alone would prove to be inferior to combined visuo-vestibular cues, all Virtual Reality (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 effortless navigation, as they do not sufficiently enable spatial updating. There is indeed a number of studies showing that spatial orientation abilities largely deteriorate when non-visual sensory modalities are excluded, reduced, or only insufficiently simulated [Chance et al. 1998; Bakker et al. 1999; May and Klatzky 2000; Péruch and Gaunet 1998; Sholl 1989; Simons and Wang 1998; Wang and Simons 1999; Wraga et al. 2004]. We suspected, however, that this apparent insufficiency of visual cues might in fact be largely due to insufficiencies of the visual simulation, namely a lack of a naturalistic scene presented with sufficient detail, resolution, and field of view. To test this hypothesis, a detailed photorealistic replica of a real scene was used in the current study.

2 Methods

Twelve naive participants completed the experiment, with ages ranging from 19 to 33 years (mean: 26.3).

2.1 Stimuli and apparatus

Scenery and visualization The pointing stimuli consisted of twelve target objects (the numbers from 1 to 12, arranged in a

Figure 2: **Top:** A photorealistic virtual replica of the real Motion-Lab was created from a 360° roundshot (4096 x 1024 pixel) of the real room. **Bottom:** Participant pointing with the position-tracked pointer and wearing a position-tracked head-mounted display (HMD) and active noise cancellation headphones.

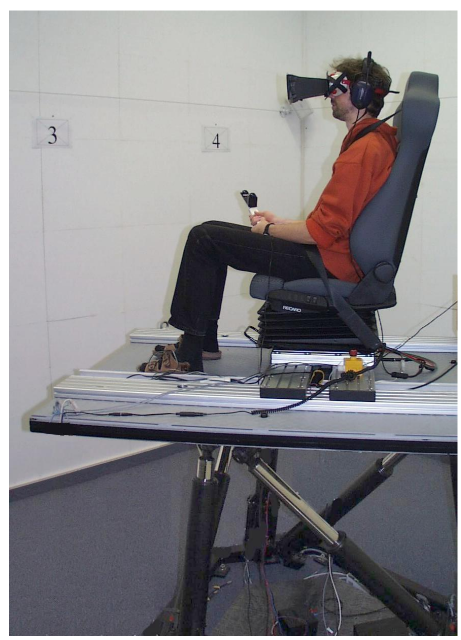

Figure 3: Experimental setup displaying a participant seated on the motion platform and wearing headphones and purpose-designed blinders (vision delimiting cardboard goggles) reducing the FOV to that of the HMD (40°x30°). Note the targets on the wall.

clock face manner) attached to the walls of the Motion-Lab at eye height (see Fig. 2 and 3). Participants saw either the real room or a photorealistic virtual replica of it (see Fig. 2) presented through a position-tracked head-mounted display (HMD Kaiser ProView XL50, see Fig. 2). The HMD had a resolution of 1024 × 768 pixel and subtended a physical field of view (FOV) of 40° × 30°.

Vestibular stimuli and apparatus For vestibular stimulation, participants were seated on a six degree of freedom Stewart motion platform (Motionbase Maxcue, see Fig. 3 and von der Heyde [2000]). For the 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 and Sholl 1995; May and Klatzky 2000; May 1996; Presson and Montello 1994; Rieser 1989]).

Vibrations Additional broad-frequency vibrations were applied to the participants's seat and floor plate 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.

Auditory stimuli Instructions during the experiment were given by a computer-generated voice and were presented via active noise canceling headphones (Sennheiser HMEC 300).

2.2 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. 2 and 3). After each pointing, participants raised the pointer to an upright position, 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. [2004]).

This rapid 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 (e.g., Riecke et al. [2002]). 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. Via triangulation, we can then backtrack where participants thought they were.

2.3 General procedure

After a training phase, each participant completed a test phase consisting of six blocks of different cue combinations (see Table 1), split into two sessions. In a repeated-measures, within-subject design, four typical spatial updating conditions were used in each block of this experiment. 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.

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 rotated to a different orientation, but asked beforehand to *ignore* that motion and "respond as if you had not moved". 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 than to UPDATE. 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.

Due to limitations of the platform turning range, the maximum heading deviation from straight ahead (12 o'clock) was ± 57°. Three different turning angles were used (±9.5°, ±19°, ±28.5° in the CONTROL conditions and ±19°, ±38°, ±57° in the other ones). Movement time was always set to seven seconds, resulting in peak angular velocities of 5.4, 10.9, and 16.3°/s. 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 or CONTROL trial).

2. **Motion phase**, which always lasted 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.

3. **Pointing phase**, consisting of four repetitions of auditory target announcement (e.g., "Object 9"), subsequent pointing, and raising pointer to upright (default) position.

In the test phase, each participant was presented with six stimulus conditions (blocks A-F, ca. 15 min. each) in pseudo-balanced order, with different degrees of visual and vestibular information available (see Table 1 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). Block D was the only one where participants were not turned physically, and asked to just use visual information. In blocks A-C, the amplitudes of the visual and vestibular (physical) turns were equal.

The pointing data ware analyzed in terms of five dependent variables, revealing different aspects of spatial updating (see below).

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 (no motion)	yes
Block E: "HMD constVis. + vest. cues"	40° x 30°	no (static image)	yes	yes
Block F: "Blindfolded just vest. cues"	—	no (blindfolded)	yes	no

Table 1: Summary of the six different stimuli (cue combinations).

As pointing data are inherently directional (circular) data, we used circular statistics for computing the dependent variables (see, e.g., Batschelet [1981] for an introduction).

1. **Response time**: How easy and intuitive (fast) is the access to our spatial knowledge?

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.

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 participants somehow misperceive their orientation by 10°, this might already explain up to 10° of their absolute pointing error.

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.

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 4 for the five dependent variables. Figures 4 clearly show the typical response pattern for automatic as well as obligatory spatial updating: UPDATE performance is comparable to CONTROL performance (implying automatic spatial updating), whereas IGNORE performance is considerably worse (implying obligatory spatial updating). 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.

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, 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 cues.

The CONTROL data are summarized in Figure 5, the corresponding t-tests are compiled in Table 2. Different questions guided the choice of cue combinations and will be discussed in the following subsections by comparing CONTROL performance between adjacent blocks.

Influence of FOV (block A vs. B) Comparing real world performance with unrestricted vision (block A) versus constrained FOV (block B, see Figure 5) reveals a clear performance decrease for limiting the FOV to 40°×30° even in the rather simple baseline (CONTROL) task.

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 5 (left) reveals a small but insignificant response time increase of approximately 90ms for using the HMD in block C (cf. Table 2). 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.

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 2). Hence, vestibular cues seem to play only a minor role for the simple CONTROL trials.

Influence of (missing) useful visual cues (block 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 2). 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 orientation ranges, these ego-orientation errors would most likely be considerably larger.

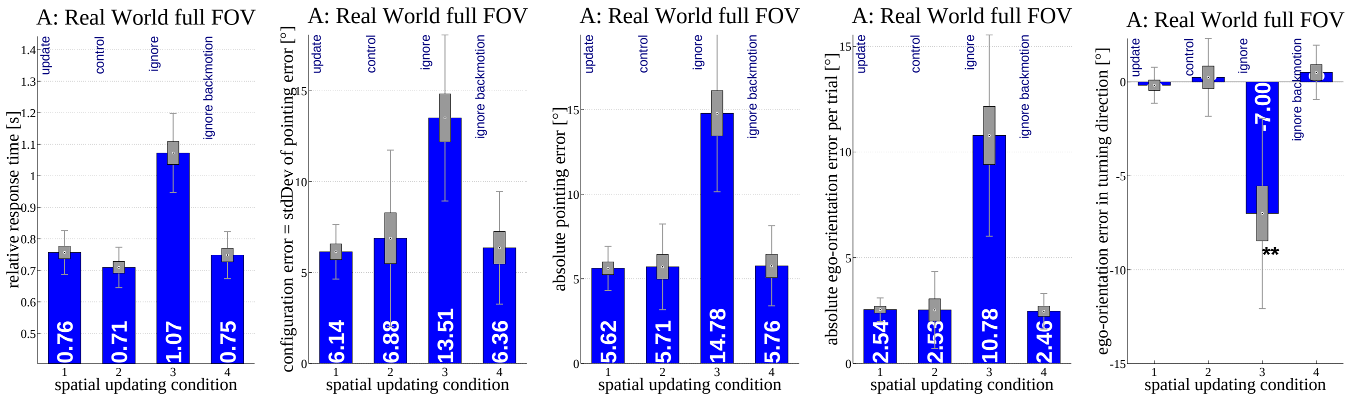

Figure 4: Pointing performance showing the typical response pattern for spatial updating: UPDATE performance is comparable to CONTROL performance, whereas IGNORE performance is considerably worse. Performance in block A (Real World full FOV) is plotted for the five dependent variables, each for the four 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 the right plot indicate whether the mean differs significantly from zero (on a 5%, 0.5% or 0.05% significance level, using a two-tailed t-test).

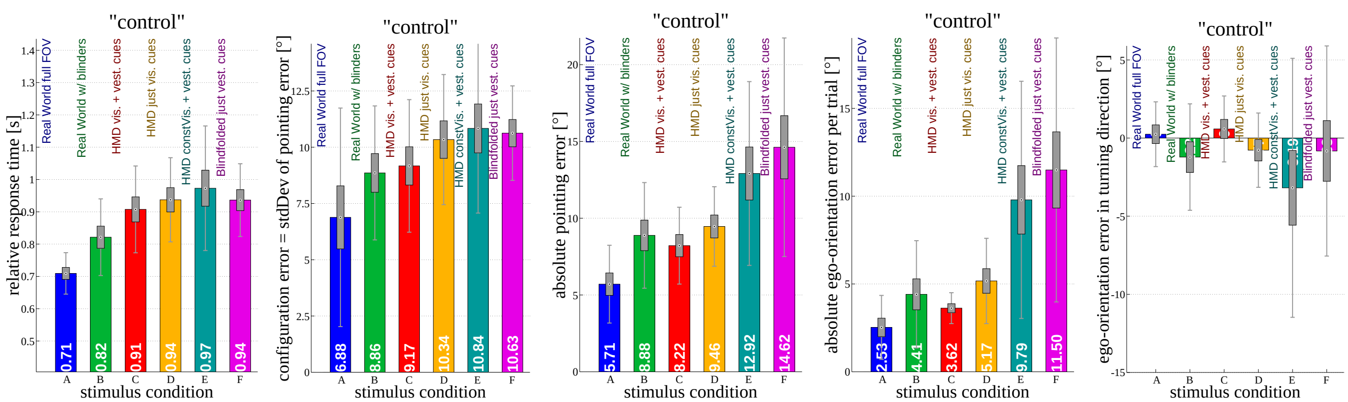

Figure 5: Baseline spatial updating performance plotted for the five dependent variables, each for the six different cue combinations. Note the FOV effect even for the simple baseline task (block A vs. B).

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 demonstrates the generalizability to comparable real world situations.

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 (see Figure 6). 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 there-and-back to the same orientation.

The literature on blindfolded spatial updating suggests a slight response time increase of approximately 100ms for UPDATE trials [Farrell and Robertson 1998; May 2000], and a considerable increase in pointing error (e.g., from 15° to 24° in the study by Farrell and 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.

Conditions with useful visual information (block 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 6 (left)). This difference was significant for the two real world conditions (blocks A & B, $p < 0.05$), but only marginally significant for the

		response time		configuration error		absolute pointing error		absolute ego-orientation error		
		t(11)	p	t(11)	p	t(11)	p	t(11)	p	p
CONTROL										
influence of FOV	A vs. B	-2.65	0.022*	-1.35	0.2	-3.46	0.0054*	-2.13	0.057m	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	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	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	0.36
const. vis. vs. no vis. cues	E vs. F	0.49	0.63	0.159	0.88	-0.679	0.51	-0.704	0.5	0.54
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*	0.98
UPDATE-CONTROL										
influence of FOV	A vs. B	-0.549	0.59	0.0658	0.95	0.927	0.37	-0.569	0.58	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	0.34
influence of vest. turn cues	C vs. D	-0.0606	0.95	0.926	0.37	0.745	0.47	0.915	0.38	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	0.78
const. vis. vs. no vis. cues	E vs. F	0.957	0.36	0.787	0.45	0.433	0.67	0.0564	0.96	0.37
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	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	0.71
real world vs. VR	B vs. C	1.93	0.079m	0.00579	1	0.839	0.42	1.2	0.26	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	0.81
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*	0.61
const. vis. vs. no vis. cues	E vs. F	-2.6	0.025*	-0.694	0.5	-0.808	0.44	-0.618	0.55	0.035*
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	0.049*

Table 2: Tabular overview of the paired two-tailed t-tests for the different comparisons. 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 ($p < 0.1$).

two HMD conditions (blocks C & D, $p < 0.1$). The effect size is less than the 100ms expected from the literature, indicating that updating to new orientations is almost as easy as updating there-and-back to the same orientation. This, in turn, suggests that spatial updating using landmark-rich visual cues is rather easy and automatic. 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.

Conditions without useful visual information (block 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 and Robertson 1998; May 2000], corroborating the hypothesis that blindfolded spatial updating to new locations is not as quick and easy as for there-and-back motions.

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, 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. The data 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). 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.

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 consistent scene. Blindfolded spatial updating showed response time differences (UPDATE - CONTROL) similar to those observed in the literature (e.g., Farrell and Robertson [1998]), validating our methodology. Additional conflicting visual information was rather hard to ignore and increased response times further. Maybe the most rel-

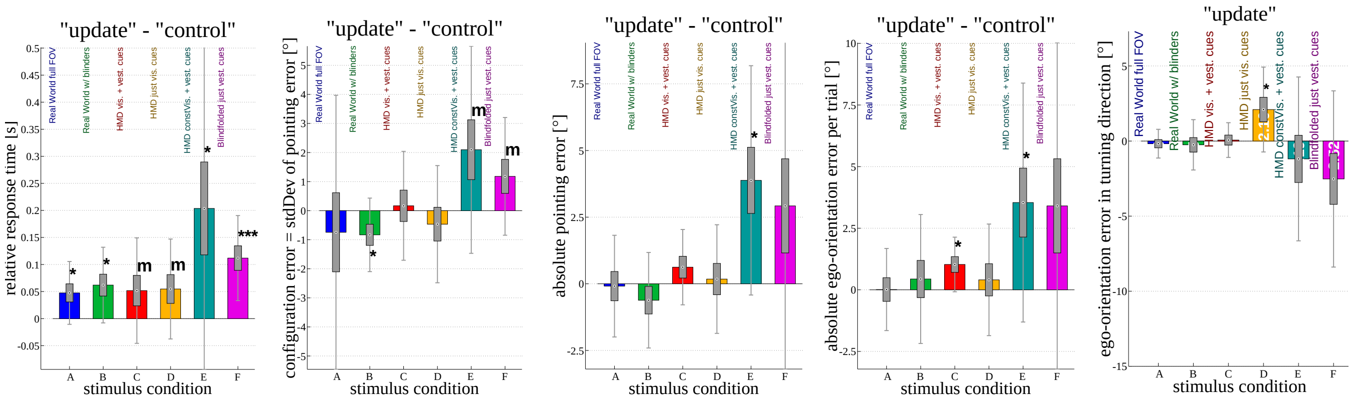

Figure 6: Automatic spatial updating performance, quantified as the difference between UPDATE and CONTROL performance (except right plot). If updating to new orientations is harder than for baseline there-and-back motions, UPDATE performance should be worse than CONTROL performance, resulting in a positive 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).

evant 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 could find in the literature[1]. 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: Participants in the study by Wraga et al. [2004] needed for example 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. [2004] 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 [Farrell and Robertson 1998; May 2000]) might not be visible any more.

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 7 and will be discussed in detail below.

[1]Response times for pointings after blindfolded rotations differ considerably, with values ranging from 1.6s [May 2000] and 1.7s [Farrell and Robertson 1998] over 1.8-3.2s [Rieser 1989] up to more than 3s [Creem and Proffitt 2000; Presson and Montello 1994]. A recent study on visually assisted spatial updating in VR reported even response times between 8 and 12s [Wraga et al. 2004].

Conditions with useful visual information (block A-D) Both real world conditions (blocks A&B) 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 largely increased, too. The considerable increase in configuration error indicates that the mental representation from the previous (to-be-remembered) orientation was less consistent and could not be remembered or accessed properly. The increase in absolute pointing error and absolute ego-orientation error can for the most part be explained by a direction-specific misperception of the correct ego-orientation in the opposite direction of the ignore motion (see Figure 7 (right)). That is, participants were apparently unable to correctly recall 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, which would predict a misperception *in* the direction of the motion, not *against* it. 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. 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 (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.

Conditions without useful visual information (block E & F) Figure 7 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

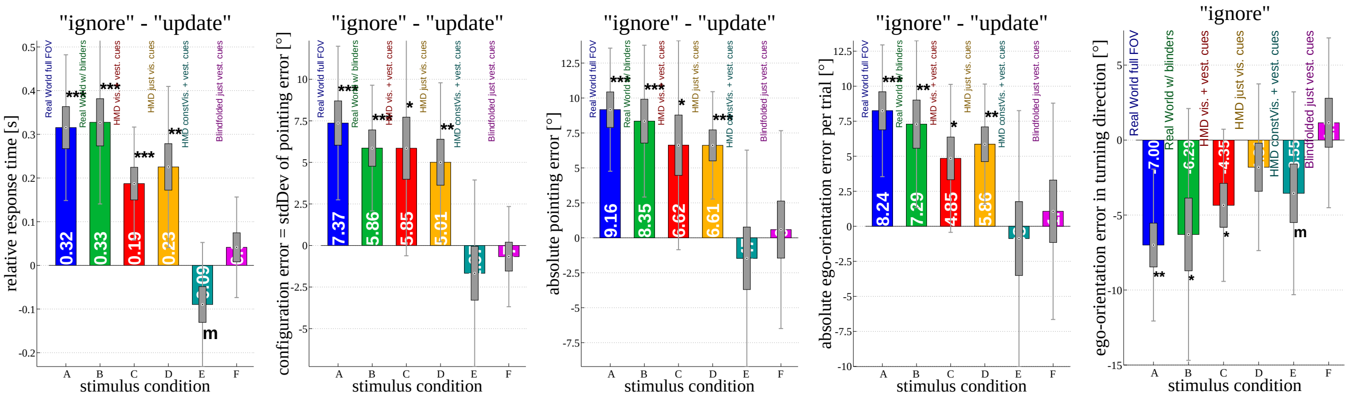

Figure 7: Obligatory spatial updating performance, quantified as the difference between IGNORE and UPDATE performance (except right plot). Values above zero indicate that ignoring is considerably harder than updating, implying obligatory spatial updating. This was the case for all conditions with useful visual cues (block A - D).

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 and 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-held pointing wand enabled considerably smaller response times than those observed in the literature (see subsection 3.2). 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 imagine 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 disorientation [Wang and 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 [Easton and Sholl 1995; Farrell and Robertson 1998; Farrell and Thomson 1998; May and Klatzky 2000; May 1996; Presson and Montello 1994; Rieser et al. 1982; Rieser 1989; Wang and 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.

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 almost impenetrable (i.e., obligatory) 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 knowledge 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.

3.4 Additional analyses

Learning effect Correlation analyses between performance and session number revealed no significant learning effect 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 $p's \geq 0.068$).

Turning angle effect If spatial updating was non-automatic and thus effortful and requiring considerable 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 change with turning angle for any of the dependent variables or 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.

4 Summary and conclusions

The rapid pointing paradigm allowed 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 shows that spatial updating can be reliably quantified with our rapid pointing paradigm.

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 demonstrates the generalizability of results obtained in this VR setup to comparable real world tasks.

References

BAKKER, N. H., WERKHOVEN, P. J., AND PASSENIER, P. O. 1999. The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. *Presence - Teleoperators and Virtual Environments 8*, 1 (Feb.), 36–53.

BATSCHELET, E. 1981. *Circular statistics in biology*. Acad. Pr., London.

CHANCE, S. S., GAUNET, F., BEALL, A. C., AND 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 (Apr.), 168–178.

CREEM, S. H., AND 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.

EASTON, R. D., AND SHOLL, M. J. 1995. Object-array structure, frames of reference, and retrieval of spatial knowledge. *J. Exp. Psychol.-Learn. Mem. Cogn. 21*, 2 (Mar.), 483–500.

FARRELL, M. J., AND ROBERTSON, I. H. 1998. Mental rotation and the automatic updating of body-centered spatial relationships. *J. Exp. Psychol.-Learn. Mem. Cogn. 24*, 1 (Jan.), 227–233.

FARRELL, M. J., AND THOMSON, J. A. 1998. Automatic spatial updating during locomotion without vision. *Q. J. Exp. Psychol. Sect A-Hum. Exp. Psychol. 51*, 3 (Aug.), 637–654.

KLATZKY, R. L., LOOMIS, J. M., BEALL, A. C., CHANCE, S. S., AND GOLLEDGE, R. G. 1998. Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychol. Sci. 9*, 4 (July), 293–298.

MAY, M., AND KLATZKY, R. L. 2000. Path integration while ignoring irrelevant movement. *J. Exp. Psychol.-Learn. Mem. Cogn. 26*, 1 (Jan.), 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. DUV: Kognitionswissenschaft, Wiesbaden.

PÉRUCH, P., AND GAUNET, F. 1998. Virtual environments as a promising tool for investigating human spatial cognition. *Cah. Psychol. Cogn.-Curr. Psychol. Cogn. 17*, 4-5 (aug-oct), 881–899.

PRESSON, C. C., AND MONTELLO, D. R. 1994. Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception 23*, 12, 1447–1455.

RIECKE, B. E., VAN VEEN, H. A. H. C., AND BÜLTHOFF, H. H. 2002. Visual Homing Is Possible Without Landmarks: A Path Integration Study in Virtual Reality. *Presence - Teleoperators and Virtual Environments 11*, 4 (Aug.).

RIESER, JOHN, J., GUTH, DAVID, A., AND HILL, EVERETT, W. 1982. Mental processes mediating independent travel: Implications for orientation and mobility. *Journal of Visual Impairment and Blindness 76*, 6 (jun.), 213–218.

RIESER, J. J. 1989. Access to knowledge of spatial structure at novel points of observation. *J. Exp. Psychol.-Learn. Mem. Cogn. 15*, 6 (Nov.), 1157–1165.

SHOLL, M. J. 1989. The relation between horizontality and rod-and-frame and vestibular navigational performance. *J. Exp. Psychol.-Learn. Mem. Cogn. 15*, 1 (Jan.), 110–125.

SIMONS, D. J., AND WANG, R. F. 1998. Perceiving real-world viewpoint changes. *Psychol. Sci. 9*, 4 (July), 315–320.

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, Dec. Available: www.kyb.tuebingen.mpg.de/publication.html?publ=633.

WANG, R. X. F., AND SIMONS, D. J. 1999. Active and passive scene recognition across views. *Cognition 70*, 2 (Mar.), 191–210.

WANG, R. X. F., AND SPELKE, E. S. 2000. Updating egocentric representations in human navigation. *Cognition 77*, 3 (DEC 15), 215–250.

WRAGA, M., CREEM, S. H., AND PROFFITT, D. R., 2004. Spatial updating of virtual displays during self- and display rotation. (under review).