Cognitive Factors Can Influence Self-Motion Perception (Vection) in Virtual Reality

BERNHARD E. RIECKE, JÖRG SCHULTE-PELKUM, MARIOS N. AVRAAMIDES,
MARKUS VON DER HEYDE, and HEINRICH H. BÜLTHOFF
Max Planck Institute for Biological Cybernetics

Research on self-motion perception and simulation has traditionally focused on the contribution of physical stimulus properties ("bottom-up factors") using abstract stimuli. Here, we demonstrate that cognitive ("top-down") mechanisms like ecological relevance and presence evoked by a virtual environment can also enhance visually induced self-motion illusions (vection). In two experiments, naive observers were asked to rate presence and the onset, intensity, and convincingness of circular vection induced by different rotating visual stimuli presented on a curved projection screen (FOV: 54°×45°). Globally consistent stimuli depicting a natural 3D scene proved more effective in inducing vection and presence than inconsistent (scrambled) or unnatural (upside-down) stimuli with similar physical stimulus properties. Correlation analyses suggest a direct relationship between spatial presence and vection. We propose that the coherent pictorial depth cues and the spatial reference frame evoked by the naturalistic environment increased the believability of the visual stimulus, such that it was more easily accepted as a stable "scene" with respect to which visual motion is more likely to be judged as self-motion than object motion. This work extends our understanding of mechanisms underlying self-motion perception and might thus help to improve the effectiveness and believability of virtual reality applications.

Categories and Subject Descriptors: 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

General Terms: Experimentation, Human factors, Measurement

Additional Key Words and Phrases: Ego–motion simulation, psychophysics, spatial orientation, spatial presence, vection, virtual reality

1. INTRODUCTION

When exposed to a visual stimulus that depicts a translation or a rotation with constant velocity, most observers experience a compelling illusion of self-motion in the direction opposite to that of the moving stimulus [Dichgans and Brandt 1978; Hettinger 2002]. This illusory self-movement occurs in natural settings when a stimulus occupying a large part of the visual field moves relative to a stationary observer. This compelling illusion was first described by Mach [1875] and was later termed "vection"

[Fischer and Kornmüller 1930; Tschermak 1931]. For example, when a train is leaving the railway station moving in the forward direction, a passenger looking out of the window of a motionless train in the adjacent track may suddenly experience a strong sense of backward self-movement.

The fact that the illusion of self-motion is sometimes very compelling is demonstrated in a study by Lepecq et al. [1993]. In this study, participants pointed at the memorized locations of targets before and after a period of forward vection. Results revealed that when pointing at the targets after being exposed to vection, participants had modified their responses as if they had really moved (i.e., they overestimated the eccentricity of laterally placed targets). In carefully designed studies, vection can even be indistinguishable from actual self-motion [Brandt et al. 1973b].

When a large visual stimulus starts to move, there is a mismatch between the visual stimulus, suggesting self-motion and the lack of any concurrent vestibularly sensed acceleration signal indicating the motion onset and increase in velocity. This visuovestibular conflict is thought to be responsible for the fact that it typically takes several seconds from the onset of visual motion before self-motion is perceived (reported values in the literature vary between 2 and 30 s). The phenomenon of vection has been traditionally linked to the vestibular system's limitation of only being able to register accelerations and decelerations and not velocity per se [Zacharias and Young 1981]. That is, as soon as a constant velocity is reached and no more accelerations are present, there is no longer any concurrent visuo-vestibular mismatch (apart from the lack of centrifugal and Coriolis forces for rotations). If there would have been an initial physical acceleration and a corresponding vestibular signal, that signal would have decayed over time, which might explain the possibility for experiencing compelling illusory self-motion if enough time has passed since the initial visual acceleration.

A number of studies have elucidated several factors relating to the stimulus and the experimental setting that can moderate the onset time, duration, and intensity of vection. Most, if not all, of these factors are bottom-up parameters (i.e., physical stimulus properties). For example, it has been shown that the speed of illusory self-motion increases linearly with increasing stimulus movement velocity up to 120°/s [Brandt et al. 1973b]. Furthermore, the solid angle (field of view) subtended by the moving visual stimulus shows a strong influence on vection, with full-field stimulation yielding the most compelling vection and the lowest vection onset time [Brandt et al. 1973b; Dichgans and Brandt 1978]. The perceived depth structure of the stimulus has been shown to affect vection as well: When the visual stimulus is comprised of two parts (either superimposed or spatially separate), the part that is *perceived* as more distant primarily determines vection—even when it is, in fact, physically closer [Howard and Heckmann 1989; Ohmi et al. 1987]. However, later studies have provided evidence that the motion of the foreground can also affect vection. For example, when the foreground remains stationary relative to the background [Howard and Howard 1994] or moves slowly to the opposite direction [Nakamura and Shimojo 1999] vection is facilitated. Another factor known to influence vection is the pattern of eye movement: When viewing a moving visual stimulus, the eyes normally follow the stimulus smoothly (optokinetic nystagmus). When participants fixate a stationary fixation cross that is presented on top of the moving stimulus, the resulting suppression of the optokinetic nystagmus significantly reduces the onset latency of vection, compared to a control condition where the eyes follow the motion of the stimulus in a natural way [Becker et al. 2002; Brandt et al. 1973b; Fushiki et al. 2000].

The fact that the majority of studies on vection have focused on bottom-up parameters implies that relatively little work has been carried out to examine how cognitive (higher-level or top-down) processes like the semantic interpretation of the moving stimulus might affect vection. The possibility that cognitive factors might affect the probability of experiencing vection or at least modulate its onset latency and strength has been largely neglected, apart from a few noteworthy exceptions:

Lepecq et al. [1995] investigated the influence of cognitive factors for translational vection in 7- and 11-year-old children. Latency of reported self-motion was reduced when participants were seated on a

movable chair as compared to a chair that could not move. This cognitive influence could, however, not be reproduced for adults in a circular vection experiment [Riecke et al. 2005d; Schulte-Pelkum et al. 2004]. In a translational vection experiment, Palmisano and Chan [2004] demonstrated that asking participants to focus on the *onset of self-motion* biases them to report an earlier vection onset, compared to the situation where they were asked to report the *offset of object motion* while watching the identical stimulus. Kitazaki and Sato [2003] demonstrated an attentional modulation of vection: When viewing two overlaid patterns of colored dots moving either upward or downward, vection was reported more often in accordance to the nonattended stimulus.

Apart from these few studies that explicitly addressed the influence of cognitive manipulation on vection, there are only anecdotal reports proposing that the knowledge that physical motion is possible, in principle, enhances vection or enables vection in situations where vection would otherwise not (or only in few cases) occur. This was done by seating blindfolded participants on a computer-controlled rotating chair for auditorily induced vection [Larsson et al. 2004] and by having observers stand in a movable booth while viewing a translational vection stimulus subtending only a small (7.5° diameter) visual angle [Andersen and Braunstein 1985].

In the context of VR simulations and the current experiment, a study by van der Steen and Brockhoff [2000] is of particular interest: The authors used a naturalistic stimulus presented on a large (142° ×110°) dome projection setup to induce forward or leftward rotational (yaw) vection. Participants were seated on a Stewart motion platform that added small physical motions 5 s after the onset of the visual motions. The latency until the moving visual stimulus induced saturated vection were exceptionally small (2.7 s for translations and 3.0 s for rotations), compared to experiments using nonnaturalistic stimuli where values ranged from 10 s [Brandt et al. 1973b] to 20–30 s [Howard and Howard 1994]. This lead van der Steen and Brockhoff [2000] to conjecture that saturated vection is induced more easily if the visual scene represents a natural environment. Note, however, that this conjecture was not directly investigated. Furthermore, the small number of observers in the study by van der Steen and Brockhoff [2000] ($N = 4$ for circular vection and $N = 5$ for translational vection; one of the participants was one of the authors) and the multitude of differences to other vection experiments both in terms of the visual stimuli, experimental setup, and response measures did not allow for any conclusive answers yet.

The present study was designed to test the conjecture of van der Steen and Brockhoff by investigating the influence of naturalism of the visual stimulus in *one* study using a within-subject design and carefully controlled visual stimuli. That is, we posit that cognitive or higher-level factors could, indeed, play an important role for perceiving vection. One such possible role might derive from our inherent assumption of a stable environment [Dichgans and Brandt 1978]: Under normal circumstances, it is only rarely the case that a large, distant portion of our surroundings physically move. In particular, natural surroundings ("scenes") are normally earth-stationary. Hence, if we see relative motion between ourselves and large, distant parts of our surroundings (or even whole scenes), we are more inclined to attribute the relative movement to ourselves instead of the surroundings. We would argue that such scenes can serve as "frames of reference" with respect to which perceived relative motion is more likely to be attributed to self-motion than object motion. Conversely, individual objects are not necessarily earth-fixed. That is, if we see individual objects or groups of objects move with respect to us, it seems ecologically plausible to assume that the perceived relative motion is due to the objects moving rather than oneself.

Note that most classic vection stimuli are abstract geometric patterns that seem to fall into the latter category. Here, we propose that the vection-inducing potential of visual stimuli might be enhanced if they depict a naturalistic scene instead, which is more likely expected to be earth-stationary and might thus more easily be accepted as a stable reference frame (in addition to, e.g., the reference frame defined by gravity). If VR is used instead of optokinetic drums, increasing the naturalism and realism of visual stimuli becomes relatively easy. One of the goals of VR is, indeed, to create the illusion of

being and acting in a computer-mediated world. Ideally, this mediated world should replace or at least dominate the real, physical world. In most hitherto existing VR simulations, however, representations of both the real environment (e.g., the physical VR setup and the surrounding laboratory room) and simulated environment (the computer-generated environment, which typically differs from the real environment) coexist and possibly interfere [Riecke and von der Heyde 2002; Riecke et al. 2006a]. The degree of perceived conflict between the two representations or reference frames of the real and virtual environment is closely related to the concept of spatial presence and immersion [IJsselsteijn 2004; Riecke and von der Heyde 2002; Sadowski and Stanney 2002]. Here, we propose that vection in a simulated environment might be enhanced if participants feel more immersed and spatially present in that environment and might thus more readily expect the virtual environment to be stable, just like corresponding real-world scenes are expected to be stable. This hypothesis has recently been confirmed for auditorily induced circular vection [Larsson et al. 2004; Riecke et al. 2005d]: Both vection and presence were increased when participants were presented with rotating sounds that normally stem from earth-stationary objects ("acoustic landmarks" such as the sound of a fountain), as compared to artificial sounds (e.g., pink noise) or sounds that normally originate from moving objects (e.g., the sound of footsteps or a driving vehicle).

Even though Wann and Rushton [1994] have stressed the importance of an ecological context and a naturalistic optic array for studying self-motion perception more than 10 years ago, this hypothesis has never been directly investigated for visual stimuli. The goal of the present study is to determine whether vection can be modulated by the nature of the stimulus, depending on whether it depicts a natural scene that provides a compelling reference frame and allows for presence or not. The existence of such cognitive contributions would be of considerable theoretical interest, as it challenges the prevailing opinion that the self-motion illusion is mediated solely by physical stimulus parameters, irrespective of any higher cognitive contributions and how the stimulus is perceived. It could also be important in terms of applications, as it might be employed to increase the effectiveness and convincingness of self-motion simulations: Physically moving the observer on a motion platform is rather costly, labor-intensive, and requires a large setup and extensive safety measures. Thus, if cognitive or top-down mechanisms could help to improve the simulation in terms of effectiveness for the given task, this would be quite beneficial, especially because these factors can often be manipulated with relatively simple and cost-effective means.

2. EXPERIMENT 1—INFLUENCE OF GLOBAL SCENE CONSISTENCY ON VECTION AND PRESENCE

2.1 Introduction and Hypotheses

Experiment 1 compared two different types of stimuli in terms of their potential to induce vection: A photorealistic image of a natural scene and scrambled versions of it.[1] Various scrambled version of the stimulus were created by scrambling image parts either in a mosaiclike manner or by slicing the original image horizontally and randomly reassembling it (see Figure 1). Questionnaires administered after the experiment assessed the extent of experienced presence for each experimental stimulus. Three hypotheses were examined here:

1. Global scene consistency, coherent pictorial depth cues, and presence (cognitive or higher-level contributions): The purpose of the scene scrambling was to disrupt global scene consistency[2] while only

---

[1]Some of the data has previously been presented at conferences [Riecke et al. 2005a, and d].

[2]Global scene consistency refers here to the coherence of a scene layout that is consistent with our natural environment, where, e.g., houses are not floating in mid-air and a marketplace consists of houses not jumbled-up or upside-down, but arranged meaningfully around an open area.

Fig. 1. Top: 360° roundshot of the Tübingen Market Place. Middle: 54° × 45° view of the four stimuli used in one session: Original image and 2, 8, and 32 slices. Bottom: 54° × 45° view of the four stimuli used in the other session: Original image and 2 × 2, 8 × 8, and 32 × 32 mosaics per 45° FOV.

slightly changing overall image statistics (bottom-up contributions, see also hypothesis 3 below). Only the globally consistent scene implies the possibility of natural locomotion (walking or driving) and spatial presence (i.e., the subjective experience of "being there" in one place or environment, even when one is physically situated in another [Witmer and Singer 1998]) in a natural-looking— albeit simulated—scene. That is, only the intact (globally consistent) scene provides one with a globally consistent reference frame, which can be used to orient oneself. Conversely, any kind of scene scrambling should reduce believability and spatial presence in the virtual environment. This would predict that spatial presence and vection are highest in the globally consistent (unscram- bled) scene and reduced for all of the scrambled stimuli. Furthermore, the various scrambled stimuli should not differ from each other in either the presence ratings or the vection measures. (i.e., A > B = C = D & a > b = c = d; see Figure 1).

Note that only the naturalistic scene stimuli (A and a) comprise a globally consistent pictorial depth structure and consistent absolute and relative size cues. This might further enhance the perceptual realism and presence/immersions. It is known that a stimulus that is perceived as being further away dominates vection [Howard and Heckmann 1989; Nakamura and Shimojo 1999; Ohmi et al. 1987]. It is conceivable that the consistent pictorial depth structure and size cues might increase the overall perceived depth of the visual stimulus, such that it appears to extend beyond the projection screen and to be further away than the scrambled stimuli. Furthermore, increased perceived stimulus distance in known to increase the perceived self-rotation speed [Wist et al. 1975]. Both of these factors might also enhance vection for the globally consistent stimulus.

2. Level of stimulus degradation (severity of scene scrambling) (mixed-level contributions): Hypoth- esis (1) predicts a *binary* result depending on whether the visual stimulus is globally consistent (stimuli a and A)—and might thus serve as a stable frame of reference including coherent picto- rial depth cues—or not (stimuli b–d and B–D). It is, however, also conceivable that the level of

stimulus degradation (i.e., the severity of scene scrambling) might *gradually* decrease presence and/or vection. This might be attributed to both higher-level contributions (e.g., the ability to mentally reconstruct the original (intact) image or parts thereof) and lower-level mechanism (e.g., the local neighborhood relationships between adjacent or nearby pixels in the image). This would predict that vection and/or presence might decrease gradually as the severity of scrambling (number of mosaics/slices per solid angle) is increased (i.e., B > C > D, b > c > d). The least vection would then be expected for the most severe stimulus degradation (conditions C, D, c, d), where the original image can hardly be mentally reconstructed and the local neighborhood relationships between nearby pixels is largely altered. Furthermore, the horizontally sliced stimuli cut the stimulus into fewer pieces than the mosaiclike scrambled stimulus, which would predict overall lower vection and/or presence for the mosaiclike scrambled stimuli (B < b, C< c, D < d).

3. Number of vertical high-contrast edges (bottom-up contributions): As mentioned above, the scene scrambling in general affected physical stimulus properties or so-called bottom-up factors. One specific effect of the mosaiclike scrambling is the addition of vertical high-contrast edges that increased the higher spatial-frequency content of the stimulus— both of which are bottom-up factors that are known to increase perceived stimulus speed [Distler 2003] and vection [Dichgans and Brandt 1978]. Hence, if these bottom-up factors dominate over cognitive contributions (global scene consistency, presence, frame of reference, and pictorial depth structure) and the level of stimulus degradation, the mosaiclike scrambled stimuli would be expected to increase vection, compared to the horizontally scrambled stimuli, which did not contain such additional vertical edges and thus had a horizontal spatial frequency spectrum quite similar to the globally consistent stimulus (i.e., B > b, C > c, D > d).

2.2 Methods

Twelve naive participants (five of whom were female) participated in Experiment 1 in exchange for monetary compensation. Participants' age ranged from 17 to 34 years (mean: 23.8a, SD: 5.0a). All participants had stereo vision and normal or corrected-to-normal vision.

2.2.1 *Stimuli and Apparatus.* Participants were comfortably seated at a distance of 1.8 m from a curved projection screen on which the rotating visual stimuli were displayed nonstereoscopically (cf. Figure 2). The image was projected using a JVC D-ILA[3] DLA-SX21S video projector with 1400×1050 pixel resolution at an update rate of 60 Hz (V-sync on). The projection screen had a curvature radius of 2 m and the simulated FOV was set to $54° \times 45°$ to match the physical FOV. Stimuli comprised of several variations of $360°$ roundshots of the Tübingen market place (4096×1024 pixel, see Figure 1) wrapped around a virtual cylinder, which rotated around the vertical axis of the participant thus producing the perception of circular vection. The horizontally scrambled stimuli were created by slicing the $360°$ roundshot image horizontally, randomly reordering the slices, and adding a random horizontal offset (between 0 and $360°$) for each slice (see Figure 1 a–d). Mosaiclike scrambled stimuli were created by subdividing the original image into individual mosaiclike patches, which were subsequently reshuffled (see Figure 1, A–D). Note that the edges introduced by the various scrambling procedures moved together with the rest of the image at the same angular speed.

The top and the side walls of the cabin were covered with black curtains to reduce visual cues and the physical reference frame of the room. Furthermore, spatial auditory cues were masked by the sound of several layers of flowing water played through active noise-canceling headphones that participants

---

[3]The Digital Direct Drive Image Light Amplifier (D-ILA) developed by JVC is a special LCoS (Liquid Chrystal on Silicon) technology and can be thought of as a reflective type of LCD. Colors are presented nonsequentially, thus avoiding rainbow artifacts for moving stimuli.
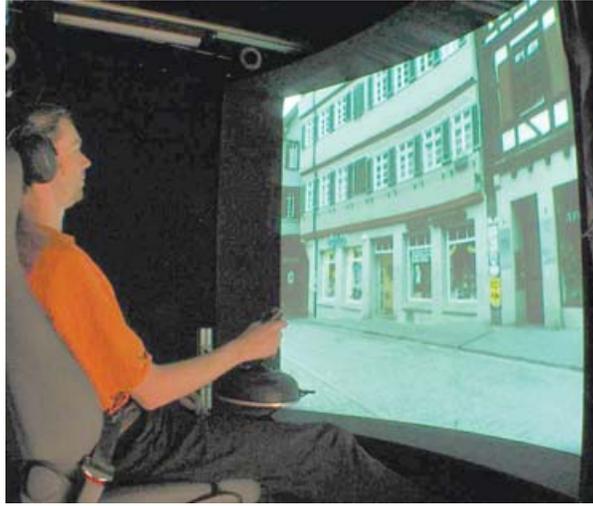
Fig. 2.   Participant seated in front of the curved projection screen displaying a view of the Tübingen market place.

wore throughout the experiment. Responses were collected using a Microsoft force-feedback 2 joystick that was mounted at a comfortable distance in front of the participants:

2.2.2   *Procedure.*   Experimental trials were initiated by participants pressing a designated button on the joystick, upon which the static image started rotating clockwise or counterclockwise around the vertical axis with constant acceleration for 3 s. Maximum rotational velocities were 20 and 40°/s. The assignment of trials to the two velocity levels and rotation directions was randomized within the experimental session. The maximum duration of constant velocity rotation was 60 s, after which the stimulus decelerated at a constant rate for 6 s and came to a stop. Visual accelerations were 6.7 and 13.3°/s^2 for maximum velocities of 20 and 40°/s, respectively. Values for decelerations were one-half of that.

Participants were instructed to pull the joystick in the direction of their perceived self-motion as soon as it was sensed. The time interval between the onset of stimulus rotation and the first deflection of the joystick indicated the vection onset time and was the primary dependent measure. Participants were also asked to deflect the joystick more the stronger the perceived self-motion became; this allowed recording the time course of vection intensity. Participants were instructed that zero-vection intensity (no joystick deflection) refers to no perceived self-motion at all, while maximum-vection intensity (full joystick deflection) denotes *saturated vection*, i.e., the case where all visual motion is attributed to self-motion and the visual stimulus is perceived as earth-stationary. Hence, full joystick deflection denotes saturated vection independent of the stimulus velocity. Note that we explicitly asked participants to rate the vection intensity and not the perceived self-motion velocity [Dichgans and Brandt 1978] or displacement angles [Becker et al. 2002], as the naturalistic stimulus used in the present study could have been used to assess the angle turned and, hence, indirectly, the self-motion velocity.

The rotation of the stimulus stopped automatically if maximum joystick deflection was sustained for 10 s (otherwise it continued for 60 s) to reduce the potential occurrence of motion sickness. Finally, at the end of each trial participants were asked to provide a "convincingness rating" of perceived self-motion by moving a lever next to the joystick to select 1 of the 11 possible steps of a 0–100% rating scale. The value of 0% corresponded to "no perceived motion at all" (i.e., perception of a rotating stimulus and a stationary self) and that of 100% to "very convincing sense of vection" (i.e., perception of a stationary

stimulus and a rotating self). After each session, participants completed a 14-item Igroup Presence Questionnaire (IPQ, [Schubert et al. 2001]) for each of the four scenes that were presented in the experimental session. In total, we obtained eight sets of questionnaires from each participant.

During each of the two sessions, participants performed 2 blocks containing 16 trials each. In one session, the mosaiclike scrambled stimuli were presented, in the other session, the horizontally scrambled stimuli were used. The presentation order of the two sessions was counterbalanced across participants. The experiment followed a 2 (session: mosaic, slices) × 4 (scrambling severity: intact, 2, 8, 32 mosaics/slices) × 2 (rotation velocity: 20, 40°/s) × 2 (turning direction) within-subject factorial design with two repetitions per condition. A pause of 15 s was inserted between trials to reduce potential motion aftereffects. In order to familiarize participants with the setup, a practice block containing four trials (one for each scrambling severity level) preceded the experimental blocks. Furthermore, because none of the participants had experienced vection in the laboratory before, they were exposed, prior to beginning the practice block, to a vection stimulus until they reported a strong sense of ego-motion.

Participants were instructed to watch the stimuli "as relaxed and naturally" as possible. They were also told not to suppress the optokinetic reflex (OKR) and neither to stare through the screen nor to fixate on a static point on the screen, but to concentrate on the image in the central part of the projection screen. We did not use any fixation point, even though it is known that a fixation point reduces vection onset times [Becker et al. 2002; Fushiki et al. 2000]. The main reason was that from an applied perspective for ego-motion simulation, it is more relevant to investigate how one can induce vection under natural viewing conditions, i.e., without a fixation point. Furthermore, this also reduced the perceived flicker and ghost images because of the 60 Hz projection: Even the moderate rotation velocities of 20 and 40°/s can already produce strong flicker and ghost images if the eyes fixate one point and do not follow the image motion. For example, a single vertical line translating sideways is seen as multiple flickering lines as it moves across the fixation point.

2.3   Results

The data for vection onset time, convincingness, and vection intensity are summarized in Figure 3 for reference. Repeated-measures ANOVAs were performed for the three dependent variables using session, scrambling severity, and rotation velocity as within-subject factors. Furthermore, correlation analyses between vection measures and the presence questionnaire data were performed.

2.3.1   *Vection Onset Time.*   The three-way ANOVA for vection onset time revealed two significant main effects. First, participants were faster reporting the onset of vection when the stimuli rotated at 40°/s than at 20°/s, $F(1, 10) = 23.9$, $p < 0.001$. Second, vection onset times varied depending on scrambling severity, $F(3, 30) = 6.23$, $p < 0.01$. More specifically, participants indicated the onset of vection faster with the intact stimuli than any of the scrambled stimuli, all pairwise $p's < 0.05$. No differences for vection onset times were obtained among the 2, 8, and 32 slices/mosaics stimuli. Neither the effect of session nor any of the interactions were significant.

2.3.2   *Vection Intensity.*   As in the vection onset time analysis, the only significant statistics for vection intensity were the main effects for rotation velocity and scrambling intensity ($F(1, 10) = 42.0$, $p < 0.001$ and $F(3, 30) = 8.29$, $p < 0.001$, respectively). Participants indicated stronger vection for stimuli rotating at 40 than at 20°/s. Furthermore, vection was rated as stronger for the intact stimulus than any of the 2, 8, or 32 slices/mosaics, all pairwise $p's < 0.05$. Intensity ratings did not differ significantly among the 2, 8, and 32 slices/mosaics. Again, neither the effect of session nor any of the interactions were significant.
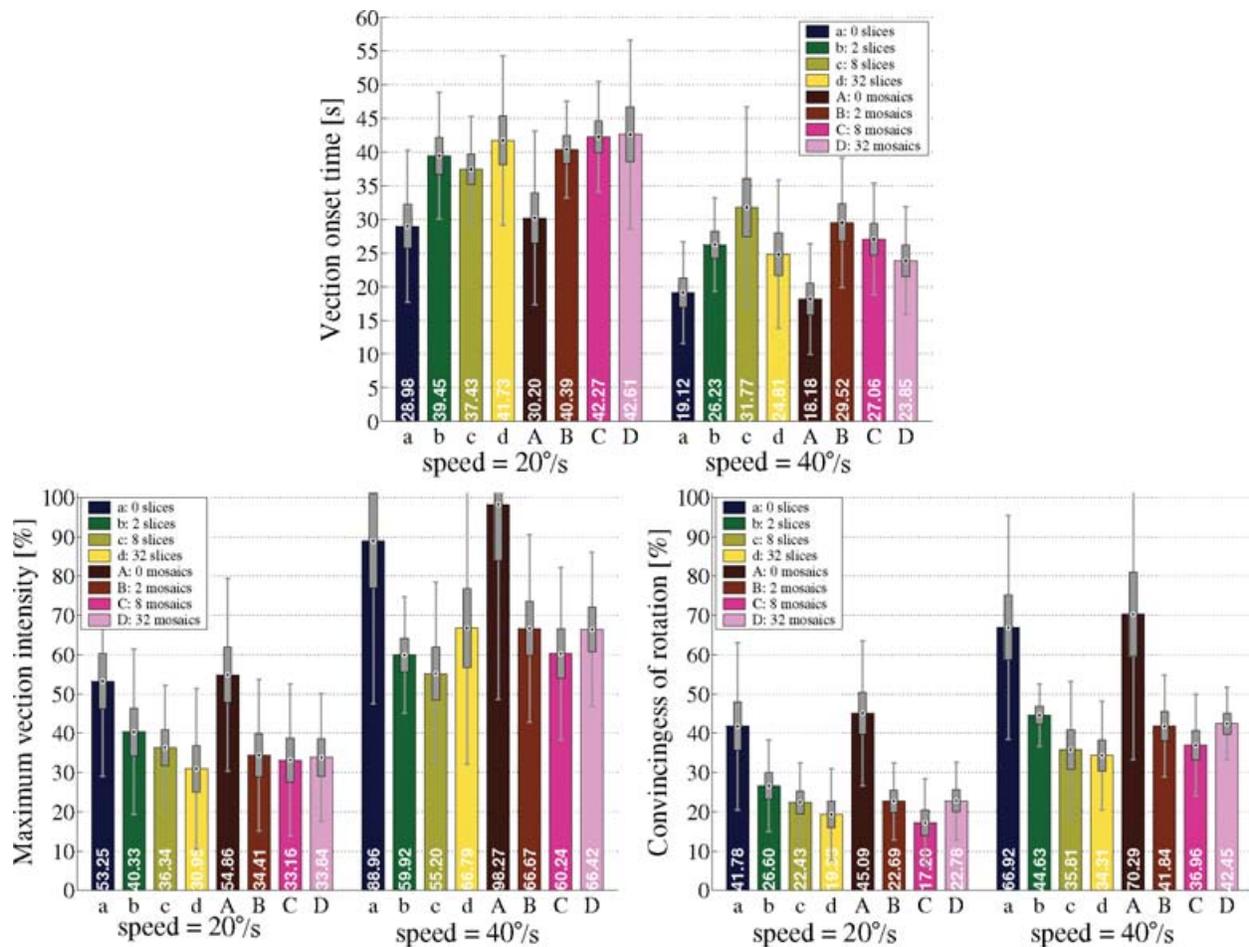
Fig. 3. *Top:* Plotted are mean vection onset times for each of the 16 experimental conditions of Experiment 1. The left and right group of eight bars represent the low and high velocity conditions (20 and 40°/s, respectively). Boxes and whiskers depict one standard error of the mean and one standard deviation, respectively. The eight visual stimulus conditions are explained in the figure legend at the top. For the sake of clarity in all three data plots, overall between-subject differences in vection responses were removed using the following normalization procedure: Each data point per participant was divided by the ratio between the mean performance of that participant across all conditions and the mean of all participants across all conditions. Note that this procedure was only applied for the data plotting and not for any of the statistical tests. *Bottom left:* Perceived vection intensity, quantified as the maximum joystick deflection reached. *Bottom right:* Convincingness ratings.

2.3.3 *Convincingness of Vection.* The analyses for the convincingness ratings revealed effects that paralleled those of the other two measures. Participants rated as more convincing the illusory self-movement produced by stimuli rotating at 40 than at 20°/s ($F(1, 10) = 23.7$, $p = 0.001$). Moreover, they rated vection as being more convincing for the globally consistent stimulus than any of the other stimuli ($F(3, 30) = 41.4$, $p < 0.001$; all pairwise $p's < 0.001$). There was also a significant difference between the 2 and the 8 slices/mosaics ($t(11) = -4.16$, $p < 0.01$). None of the remaining pairwise tests reached significance. Neither session nor any of the interactions were significant.

2.3.4 *Gender Effects.* Both the onset latency and intensity of vection showed no systematic influence of gender ($F(1, 10) = 0.025$, $p = 0.88$ and $F(1, 10) = 0.57$, $p = 0.47$, respectively). The rated convincingness of vection, however, showed significantly higher ratings for females (mean: 50.5%, SD: 25.3%) than males (mean: 27.3%, SD: 19.4%), $F(1, 10) = 6.72$, $p = 0.027$. Note that this finding is not consistent with findings by Darlington and Smith [1998] and Kennedy et al. [1996], who reported significantly longer vection onset latencies for males (intensity and convincingness of vection were not assessed in those studies). The gender effect observed in the current study should, however, be treated with caution because of the relatively small number of participants.

2.3.5 *Questionnaires.* Results from the presence questionnaire by Schubert et al. [2001] are summarized in Figure 4. In our sample, the IPQ showed high reliability ($\alpha = 0.91$). To examine the structure and constituent elements of the presence questionnaire, we analyzed similarities and correlations between the responses to the different questions of the IPQ using a factor analysis. The factor analysis revealed a two-dimensional structure of presence: Factor 1 contained items about realism of the simulated scene and spatial presence (e.g., sense of acting in the virtual environment), while factor 2 contained items that addressed attentional aspects or involvement (e.g., awareness of real surroundings of the simulator versus the simulated environment). Factor 1 and 2 correspond to the bottom right and middle plots, respectively, in Figure 4.

Participants also completed a simulator sickness questionnaire (SSQ) before and after each session [Kennedy et al. 1992]. As expected, the simulator sickness ratings were somewhat higher after the experiment (0.336 ± 0.049 SE versus 0.173 ± 0.048 on a 0–3 point rating scale), but all participants felt comfortable finishing the experiment. An additional presence susceptibility questionnaire (unpublished), which is supposed to measure a person's general susceptibility to presence, did not show any clear results or correlations with any of the vection measures. Therefore, only the results form the IPQ presence questionnaire will be discussed in the following.

Mean presence scores obtained with the IPQ were computed for each level of the scrambled scenes (see Figure 4, top left plot). A repeated-measures ANOVA with session (horizontally versus mosaiclike sliced) and number of slices (unsliced, 2, 8, and 32) showed a significant effect only for the number of slices ($F(3, 18) = 21.5$, $p = 0.001$). A post hoc analysis showed that only the presence ratings of the intact market scene differed significantly from all other levels (Bonferroni-corrected $p = 0.003$), but no significant differences between the 2, 8, and 32 slices were found (see Figure 4). That is, two slices were enough to impair presence significantly, and no further decrease in presence was observed for the 8 and 32 slices. Mean presence scores for each of the four original subscales of the IPQ (realism, presence, space, and involvement) and also of the compound scales that were merged according to the factor analysis (factor 1 = "spatial presence" and factor 2 = "attention/involvement") are shown in Figure 4.

In the next step, we investigated how presence in the simulated scene related to the different aspects of the self-motion illusion by performing correlation analyses between the mean presence scores and the three measurands from the vection experiment (vection onset time, vection intensity, and convincingness). Table I shows the paired-samples correlations (r) and the corresponding p values for the overall presence score and the two factor values from the factor analysis, which were interpreted as "spatial presence" (factor 1) and "attention/involvement" (factor 2).

For the overall presence score, we found significant correlations between presence ratings and convincingness ratings (7 out of 8 correlations for the different visualization conditions reached significance, overall correlation of $r = 0.772$, $p = 0.003$; see Table I). Vection onset time correlated negatively with presence (3 out of 8, overall correlation of $r = -0.616$, $p = 0.033$), which means that higher presence was associated with shorter onset times, in these cases. No significant correlations were observed between presence ratings and vection intensity. The more detailed analysis revealed significant negative
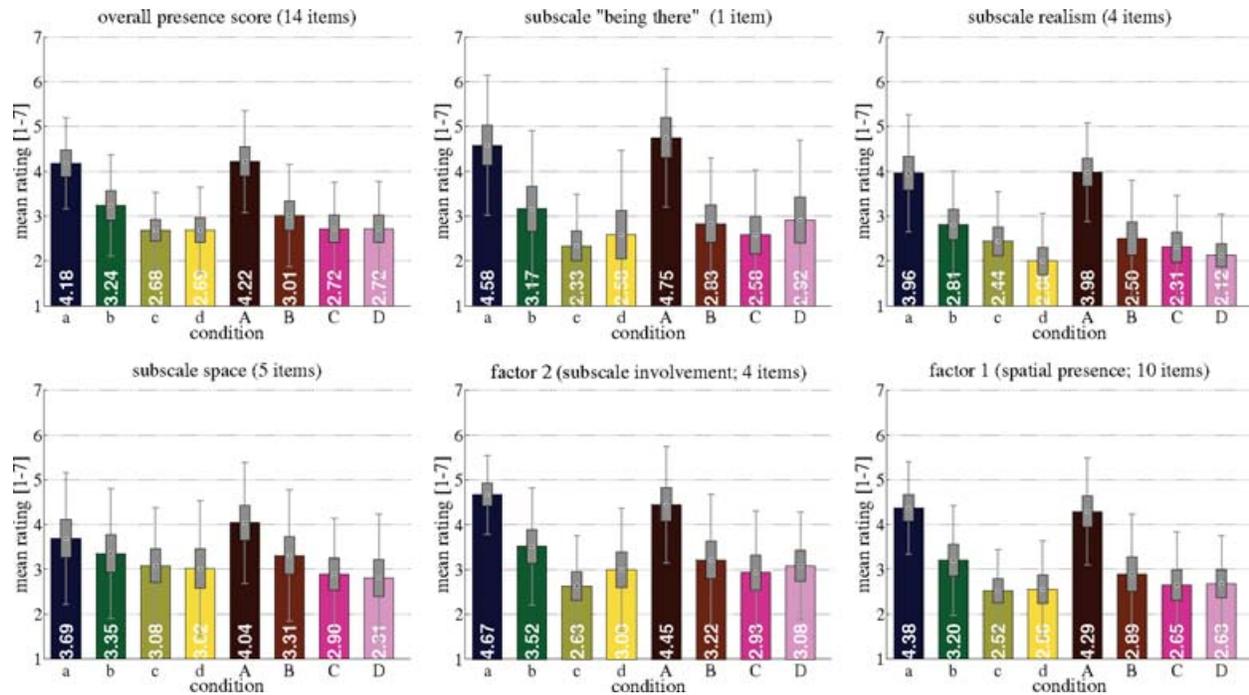
Fig. 4.   Presence ratings for the eight different visual stimuli used in Experiment 1. The top left plot shows the mean overall presence score over all 14 items of the IPQ. The four plots next to that show the data that were split up according to the four original subscales as described by Schubert el al. [2001]: "Being there," realism, space, and involvement. The remaining two plots on the bottom right show the mean presence ratings according to the results of the factor analysis. The involvement subscale coincides with factor 2 of the factor analysis, and the remaining three subscales (10 items) constitute factor 1, which can be interpreted as the spatial presence aspect. It can be seen that only the consistent scenes induced high spatial presence and high attentional involvement. Note the qualitatively similar pattern of results for all scales: Only the intact scenes (a, A) yielded high presence ratings, while all scene scrambling reduced presence consistently.

correlations of the involvement scale (factor 2) with vection onset time (4 out of 8, overall correlation of $r = -0.620$, $p = 0.031$), while the compound spatial presence scale (factor 1) correlated positively with the convincingness ratings (4 out of 8, overall correlation of $r = 0.579$, $p = 0.049$). Vection intensity correlated positively with the involvement scale for three of the eight visualization conditions, but did not reach significance in the overall correlation ($r = 0.469$, $p = 0.124$). Furthermore, vection intensity showed no clear correlation with the compound spatial presence scale (0 out of 8, $r = -0.229$, $p = 0.474$). It should be pointed out that given the small sample size ($N = 12$), these correlations are quite substantial. Interestingly, there was an asymmetry in the correlation results: Factor 1 (spatial presence) correlated mainly with convincingness ratings, whereas factor 2 (involvement) correlated mainly with vection onset time (see Table I).

2.4   Discussion

Both the vection and presence data showed virtually no difference for the different scrambling severities. That is, the level of stimulus degradation did seem to not *gradually* decrease vection and/or presence as predicted by hypothesis 2 ["Level of stimulus degradation (severity of scene scrambling) (mixed-level contributions)"], but instead resembled more a *binary, one-step* decrease as predicted by hypothesis 1 ["Global scene consistency, coherent pictorial depth cues, and presence (cognitive or higher-level

Table I. Paired-Samples Correlations between Vection Measures and Presence Ratings[a]

		Overall Presence Score			Factor 1 (Spatial Presence)			Factor 2 (Involvement)		
		Onset Time	Convin-cingness	Vection Intensity	Onset Time	Convin-cingness	Vection Intensity	Onset Time	Convin-cingness	Vection Intensity
Horizontal										
Presence rating unsliced	r	−0.289	**0.665***	0.209	−0.154	0.731**	0.146	−0.427	0.323	0.247
	p	0.36	**0.018**	0.515	0.634	0.007	0.65	0.166	0.306	0.44
Presence rating 2 slices	r	**−0.675***	0.892**	0.302	−0.462	0.853**	0.218	**−0.868***	**0.631***	0.611*
	p	**0.016**	**0.001**	0.208	0.131	0.0001	0.497	**0.0002**	**0.028**	**0.035**
Presence rating 8 slices	r	**−0.790***	0.625*	0.313	−0.571	0.602*	0.038	**−0.802***	0.369	0.653*
	p	**0.002**	**0.03**	0.321	0.051	0.038	0.908	**0.002**	0.238	**0.021**
Presence rating 32 slices	r	−0.319	0.302	−0.004	−0.051	0.262	−0.286	**−0.62***	0.202	0.506
	p	0.312	0.34	0.991	0.875	0.411	0.368	0.031	0.529	0.093
Scrambled										
Presence rating unsliced	r	0.086	**0.776***	−0.1	0.246	**0.802***	−0.197	−0.292	0.519	0.140
	p	0.791	**0.003**	0.757	0.44	**0.002**	0.54	0.357	0.084	0.664
Presence rating 2 slices	r	−0.448	**0.596***	0.364	−0.24	0.513	0.222	**−0.687***	0.467	0.492
	p	0.144	**0.041**	0.245	0.453	0.088	0.487	**0.014**	0.126	0.104
Presence rating 8 slices	r	**−0.728***	**0.753***	0.385	−0.564	**0.728***	0.219	**−0.778***	0.461	0.600*
	p	**0.007**	**0.005**	0.216	0.056	**0.007**	0.493	**0.003**	0.132	**0.039**
Presence rating 32 slices	r	−0.086	**0.646***	−0.057	0.102	0.518	−0.241	−0.445	**0.695***	0.302
	p	0.79	**0.02**	0.869	0.752	0.084	0.451	0.181	0.012	0.339
Overall correlation	r	**−0.616***	**0.772***	0.227	0.041	**0.579***	−0.229	**−0.620***	0.307	0.469
	p	**0.033**	**0.003**	0.478	0.90	**0.049**	0.474	**0.031**	0.332	0.124

[a]N = 12. Note that convincingness ratings correlated highly with the factor 1 (spatial presence), but not with factor 2 (involvement). Conversely, vection onset time was negatively correlated with factor 2 (involvement), but not with factor 1 (spatial presence). Vection intensity was only moderately correlated with factor 2 (involvement), but not at all with factor 1 (spatial presence).

contributions)"]: Vection and presence were highest for the globally consistent (intact) stimulus and almost equally reduced for all scrambling levels.

This result was somewhat unexpected, as a pretest using the same stimuli, but nonnaive observers from our own lab showed a gradual vection decrease for increasing scrambling severity. From the current data, it remains unclear whether the lack of a significant influence of scrambling severity for the naive participants might just be a result of the insufficient statistical power. To say the least, global scene consistency seems to have played the dominant role in facilitating vection in the current study, and any global inconsistency reduced vection as well as spatial presence and involvement considerably. Even though it might seem plausible and to be expected that a naturalistic scene induces stronger vection than more abstract stimuli like the scrambled stimuli used here, we are not aware of any previous publication explicitly investigating this issue.

Previous research has shown that adding vertical high contrast edges facilitates circular vection [Dichgans and Brandt 1978]. This bottom-up mechanism would predict that the mosaiclike scrambled stimuli should be more powerful in inducing vection than the horizontally scrambled stimuli, which contained fewer vertical high-contrast edges (hypothesis 3). Furthermore, we know from previous literature that increasing contrast and spatial frequency of a moving stimulus leads to higher perceived velocity [Distler 2003]. This would predict that the additional vertical high-contrast edges contained in the mosaiclike scrambled stimuli should increase the perceived velocity of the stimulus and thus indirectly enhance vection as compared to the intact stimulus and horizontal scrambled stimuli.

This study showed, however, no such vection-facilitating effect of the additional vertical edges at all. Instead, vection was actually *reduced* compared to the intact stimulus. This suggests the data cannot be explained by low-level, bottom-up processes alone, and that the bottom-up contributions (more vertical contrast edges in the mosaiclike scrambled stimulus) were dominated by cognitive or top-down processes (consistent reference frame and depth structure for the intact market scene). This is corroborated by the fact that the additional vertical contrast edges in the mosaiclike scrambled stimuli did not increase vection compared to the horizontally scrambled stimuli (which did not have any

more vertical contrast edges than the intact stimulus). This lack of differences between the mosaiclike and horizontally scrambled stimuli was rather surprising, as previous work showed a clear vection-facilitating effect if the number of vertical high-contrast elements was increased [Dichgans and Brandt 1978].

How can this apparent contradiction between the current study and the study by Dichgans and Brandt [1978] be explained? There are a number of differences in the experimental procedures and, in particular, the vection-inducing stimuli that might have contributed. Most importantly, Dichgans and Brandt [1978] used abstract geometrical patterns of alternating vertical black and white stripes, whereas the current study used reshuffled horizontal or mosaiclike patches of a roundshot photograph of a natural scene. That is, the individual vertical stripes in the study by Dichgans and Brandt contained no substructure or texture like the patches used in the current study. The resulting differences in contrast and the spatial frequency spectrum of the stimuli might be critical and would need to be explicitly investigated to yield a conclusive answer. Apart from such lower-level differences, there are also higher-level differences that might have contributed. For example, the mosaiclike and horizontal patches used in the current study might have been perceived as more natural than the abstract geometric pattern used by Dichgans and Brandt [1978]. One way to investigate the contribution of naturalness or realism might be to degrade the realism of a natural stimulus gradually not by scrambling but by using different rendering methods known from computer graphics (such as nonphotorealistic rendering, different shading or lighting methods, or wire-frame rendering).

In summary, Experiment 1 suggests that cognitive or higher-level mechanism do, in fact, contribute to the perception of self-motion induced by moving visual stimuli. We propose that the classic view that lower-level factors are the only factors contributing to vection should be extended to include higher-level and cognitive factors. Cognitive factors that might have contributed to this effect include global consistency of the scene, consistency of the pictorial depth cues and the resulting increase in perceived depth, believability of the stimulus (e.g., naturalism of scene layout, lighting, and size cues), presence and involvement in the simulated scene, and/or the affordance (the implied possibility) of moving through the scene. This will be elaborated upon in more detail in the general discussion (Section 4).

3.   EXPERIMENT 2—INFLUENCE OF SCENE INVERSION ON VECTION AND PRESENCE

The previous experiment demonstrated that scene scrambling reduced vection and presence in all dependent variables assessed. We argued that this effect was, to a large degree, caused by cognitive factors, such as the interpretation of the intact scene as a stable reference frame with respect to which a displayed motion is more likely to be interpreted as self-motion than object motion. Even though the image scrambling introduced only relatively small changes in the image statistics (mainly additional high-contrast edges, which increased higher spatial frequencies in the stimulus and changed the local relationships between close-by pixels), it cannot be excluded that these minor changes might have affected vection via bottom-up factors.

Experiment 2 was designed to investigate potential influences of cognitive factors on vection and/or presence *without* changing the overall image statistics (i.e., lower-level factors) of the stimulus at all. This was achieved by comparing the intact market place stimulus with an upside-down version of the same stimulus.

The reasoning is as follows: Placing the visual stimulus upside-down should decrease the naturalness of the stimulus and reduce the implied possibility of naturally walking or driving through it, which might, in turn, reduce spatial presence and potentially also vection (either directly or mediated by the reduction in presence), while leaving the pictorial depth cues unchanged. Note that all of these factors are cognitive or higher-level factors and that inverting the scene does neither change global image
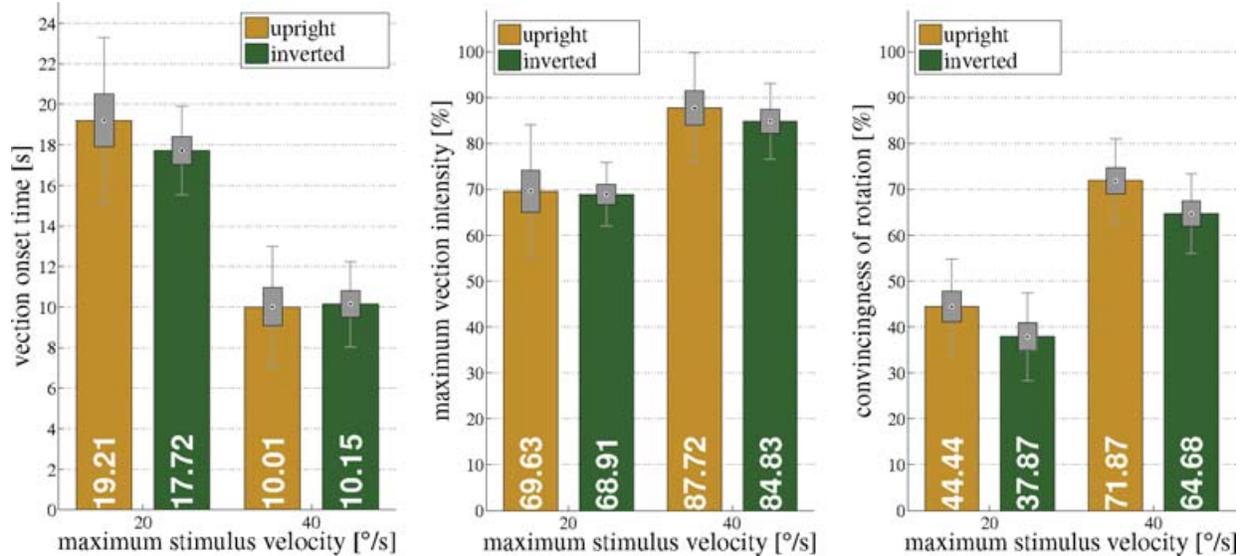
Fig. 5. *Left:* Mean vection onset times for each of the four experimental conditions, averaged over the ten participants and plotted as in Figure 3. *Middle:* Perceived vection intensity, quantified as the maximum joystick deflection reached. *Right:* Convincingness ratings. Note that scene inversion yielded slightly less convincing self-motions (right plot), but showed otherwise no clear effect. Increasing the stimulus velocity enhanced vection in all three dependent variables.

statistics[4] nor introduce additional high-contrast edges, as the scrambling in previous experiment. Hence, any observed effects would argue in favor of at least some cognitive or top-down contribution to vection.

3.1 Methods

Ten naive participants (gender-balanced) completed the experiment in a single session containing 24 trials. Participants were between 17 and 27 years old (mean: 22.8a, SD: 2.7a). The 24 trials consisted of a factorial combination of 2 visual stimuli (normal and upside-down) $\times$ 2 velocities (left and right) $\times$ 2 turning directions (20 and 40°/s) $\times$ 3 repetitions. Apart from this, the experimental procedures were identical to the first experiment.

3.2 Results and Discussion

3.2.1 *Vection Measures.* The vection data are displayed in Figure 5 for reference. Repeated-measures ANOVAs were performed for the three dependent vection measures using a 2 (scene: upright, upside-down) $\times$ 2 (rotation velocity: 20, 40°/s) design. For all three dependent variables, rotation velocity yielded highly significant main effects ($F(1, 9) = 14.3$, $p = 0.004$ for vection onset time, $F(1, 9) = 38.6$, $p < 0.001$ for convincingness ratings, and $F(1, 9) = 12.5$, $p = 0.006$ for vection intensity). The factor visual stimulus yielded significant effects only for the convincingness ratings ($F(1, 9) = 13.2$, $p = 0.005$). That is, inverting the visual scene reduced the rated convincingness of the illusory self-motion by about 15% for the lower stimulus velocity and 10% for the faster stimulus velocity (see Figure 5), but had no effect on either vection onset time or vection intensity. None of the interactions reached significance.

---

[4]The image statistics for the upper and lower part of the image are, however, swapped because of the stimulus inversion and might thus contribute if vection and/or presence depend differently on the upper and lower part of the visual stimulus.
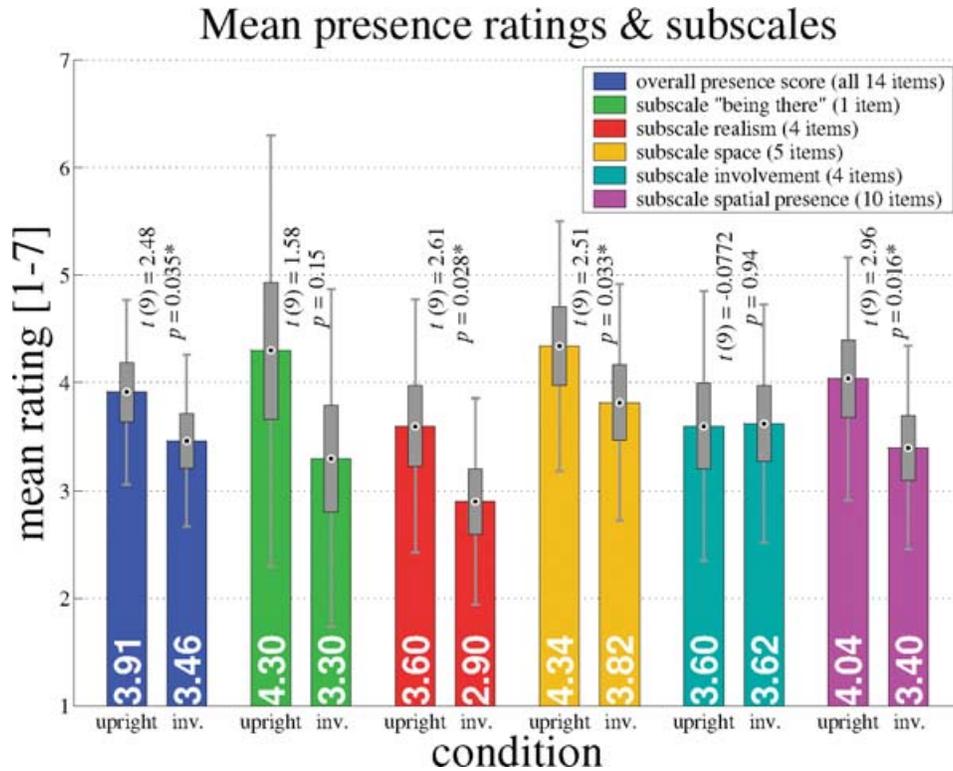
Fig. 6.   Presence ratings for the two different visual stimuli used in Experiment 2. Plotted are the overall presence score of all 14 items (left pair of bars), which was split up into the four subscales "being there," realism, space, and involvement (see legend at the top of the figure). The compound scale spatial presence refers to factor 1 of the factor analysis of Experiment 1. The left bar of each pair refers to the upright stimulus and the right one to the upside-down stimulus. Note the qualitatively similar pattern of results for the overall presence score and the three subscales "being there," realism, and space: There was always a tendency toward reduced presence for the upside-down scene. Only the involvement subscale did not show any clear influence of scene inversion.

3.2.2   *Gender Effects.*   As in Experiment 1, both vection onset latency and vection intensity showed no systematic influence of gender ($F(1, 8) = 2.19$, $p = 0.18$ and $F(1, 8) = 0.33$, $p = 0.58$, respectively). Experiment 1 showed significantly higher convincingness ratings for females. In the current experiment, the convincingness of vection was also rated slightly higher for females (mean: 59.8%, SD: 22.6%) than males (mean: 49.7%, SD: 20.6%). This difference did not reach significance, though ($F(1, 8) = 0.97$, $p = 0.35$). Because of the small number of participants in both Experiment 1 and 2, these results should, of course, be interpreted with caution.

3.2.3   *Presence Ratings.*   Presence ratings for both visualization conditions were obtained using the IPQ by Schubert et al. [2001], as before. The results for the overall presence ratings and the individual subscales are displayed in Figure 6. Results of paired *t*-tests between the upright and upside-down condition are indicated above the respective bars.

With a mean value of 3.91 (SE = 0.25), overall presence ratings for the upright scene were significantly higher than for the upside-down scene (3.46, SE = 0.22). All but the involvement subscale showed a similar tendency of higher ratings for the upright scene. This effect reached significance for the subscales

Table II. Table of Paired-Samples Correlations between Vection Measures and Presence Ratings[a]

		Onset Time		Convincingness		Vection Intensity	
Overall presence score	r	0.262		0.187		0.187	
(all 14 items)	p		0.233		0.303		0.303
Subscale "being there"	r	*0.472m*		0.276		0.093	
(1 item)	p		*0.083*		0.221		0.398
Subscale realism	r	0.224		−0.135		0.294	
(4 items)	p		0.267		0.355		0.205
Subscale space	r	*0.532m*		**0.555***		0.150	
(5 items)	p		0.057		**0.048**		0.340
Subscale involvement (factor 2)	r	−0.382		−0.199		−0.201	
(4 items)	p		0.138		0.291		0.289
Subscale spatial presence (factor 1)	r	*0.455m*		0.293		0.215	
(10 items)	p		*0.093*		0.206		0.276

[a]Note that only the correlation between convincingness and the space subscale of presence reached significance.

realism, space, and the compound scale spatial presence, which was identified as factor 1 in the factor analysis of the previous experiment (see Figure 6).

How do these results relate to the previous study? The factor analysis and subsequent correlation analysis performed for Experiment 1 showed a close positive correlation between spatial presence aspects of the presence questionnaire and the convincingness ratings. Furthermore, the involvement aspects of the questionnaire correlated negatively with vection onset time. Results of the current study might be interpreted in a similar manner. Inverting the stimulus affected both the spatial presence aspects of the questionnaire and the convincingness ratings (which were positively correlated in the previous study), but not the involvement subscale of the questionnaire and the corresponding vection measure onset time. The subsequent correlation analysis was intended to investigate if the correlation pattern observed for Experiment 1 would also hold for Experiment 2.

3.2.4 *Correlation Analysis Between Vection and Presence Measures.* As in Experiment 1, we investigated how the different aspects of presence related to the self-motion illusion using correlation analyses between the different presence and vection measures.[5] The resulting paired-samples correlations (r values) and corresponding p values are summarized in Table II.

Only the correlation between the convincingness ratings and the space subscale reached significance. In addition, vection onset time correlated marginally ($5\% < p < 10\%$) with the subscales "being there," space, and the compound scale spatial presence (which was identified as factor 1 in the factor analysis of Experiment 1). In sum, the strongest correlations were found for the space subscale of the presence questionnaire, with r values of 0.532 and 0.555 for the vection onset time and convincingness, respectively. That is, an increased spatial presence in the scene correlated with an increased convincingness of the self-motion illusion, but also surprisingly with a *later* vection onset. The latter result seems rather counterintuitive and we do not have a convincing explanation for it yet.

The overall correlational pattern in this study differs from the one observed for Experiment 1, where higher spatial presence (factor 1) correlated with increased convincingness ratings (but none of the other vection measures), whereas increased involvement (factor 2) correlated with earlier vection onset (but none of the other vection measures). In Experiment 2, there were similar trends, but they did not reach significance: Spatial presence was positively correlated with convincingness ($r = 0.293$, $p = 0.206$), and involvement was negatively correlated with onset time ($r = -0.382$, $p = 0.138$). Nonetheless,

---

[5]Because of the smaller number of participants and experimental conditions in Experiment 2, we refrained from performing another factor analysis.

the correlation analyses for the two experiments yielded somewhat different results, which cannot be fully reconciled or explained from the current data. Differences in the participant population might have contributed to the different results and a larger number of participants and a within-subject comparison would be needed to allow for more conclusive answers.

4. GENERAL DISCUSSION AND CONCLUSIONS

Previous studies have typically used abstract stimuli like black and white geometric patterns to induce vection. Here, we show that the illusion can be enhanced if a natural scene is used instead of more abstract patterns: Experiment 1 revealed that a visual stimulus depicting a natural, globally consistent scene including consistent pictorial depth cues can produce faster, stronger, and more convincing perception of illusory self-motion than more abstract, scrambled versions of the same stimulus. One possible explanation for why this happens is that natural scenes are less likely to be interpreted as moving because they are accepted as a stable reference frame [Dichgans and Brandt 1978]. This might also explain the data from Experiment 2, where presenting the natural scene upside-down reduce the convincingness of the self-motion illusion, even though it failed to affect vection onset time or intensity.

4.1 Presence and Potential Relation to Vection

Results from the presence questionnaires showed that the natural, globally consistent scene was associated with higher presence ratings than any of the scrambled or upside-down stimuli. This raises the possibility that presence and vection might be directly linked. It could be the case that vection was facilitated by the natural, upright scene stimulus because participants felt more present in it. Compatible with this hypothesis are the results from the various scrambled stimuli: Neither the presence ratings nor the vection onset time or intensity showed any consistent difference in the statistical tests. Two slices/mosaics were sufficient to reduce presence and impair vection as compared to the natural scene.

In this first experiment, the presence questionnaire showed a two-dimensional structure, namely spatial presence (factor 1) and attention/involvement (factor 2). Furthermore, we found a differential influence of these two factors: While spatial presence was closely related to the convincingness of the rotation illusion, involvement in the simulation was more closely related to the onset time of the illusion. This should be taken into consideration when attempting to improve VR simulations. Depending on task requirements, different aspects of presence might be relevant and should receive more attention or simulation effort.

The second experiment showed that mere scene inversion was sufficient to reduce the convincingness of illusory self-motion as well as rated spatial presence, even though the effect size was relatively small. Neither vection onset time, vection intensity, nor the rated involvement in the simulation were, however, affected by the stimulus inversion. It might be the case that scene inversion was just too subtle of a manipulation to show a significant effect. Alternatively, one might argue that the scene inversion did not affect vection onset latency and intensity because it did not affect the spatial depth structure and global scene consistency of the visual stimulus. This would suggest that the naturalism of the stimulus is more closely related to the overall convincingness and spatial presence evoked by the simulation, whereas the depth structure and global scene consistency would predominately affect the onset latency and intensity of vection, as well as the attention/involvement aspect of presence.

This idea is consistent with the finding from Experiment 1 that spatial aspects of presence correlated significantly with the convincingness ratings of vection, but not with vection onset time or intensity. The fact that the "involvement" factor correlated significantly with vection onset time and somewhat with intensity casts new light on the importance of attentional factors in vection research, as Kitazaki and Sato [2003] have also recently reported. In terms of presence research, this finding suggests that

attention might be conceptualized as one building block of presence which seems to be especially important for ego-motion simulation.

Even though the present study showed a clear correlation between vection and presence, further research is needed to determine if there might actually be a *causal* relation between presence and vection, and whether presence might also be affected by the perception of self-motion, as suggested by a recent study using the same VR setup [Riecke et al. 2004]: In that study, vection onset times were unexpectedly decreased when minor scratches were added to the projection screen. These hardly noticed scratches also enhanced vection in terms of both intensity and convincingness ratings. We are not aware of any theoretical reason why these imperfections in the simulation setup should increase presence in the simulated environment. If anything, one might rather expect a decrease in presence. Nevertheless, these minor modifications increased presence ratings significantly, which suggests that the presence increase might have been mediated by the increase in vection.

4.2   Comparison of Vection Onset Latencies between Experiments

Comparing the absolute values of the different measurands between the two experiments shows considerable differences: Vection onset times for the globally consistent, upright stimulus were about 10 s lower in the second experiment, whereas vection intensity, convincingness, and presence ratings were roughly comparable. Further experiments using the same stimulus and display setup yielded vection onset times as low as 10 s [Riecke et al. 2005b; Schulte-Pelkum et al. 2004], which suggests that the participant population might be the determining factor. In addition, vection onset times seem to decline with increasing practice and exposure to vection-inducing stimuli (at least in our setup): In a control experiment (unpublished), vection onset times for circular vection dropped over the course of two consecutive sessions (25 min each) from an initial value of 16 s down to less than 8 s. Furthermore, Palmisano and Chan [2004] demonstrated that the specific instructions can also influence vection onset latencies.

Together, these observations emphasize the importance of within-subject designs and show the difficulty of comparing absolute vection onset latencies between studies, even when FOV, stimulus velocity, and the stimulus itself is matched. Thus, from the present experiments we can mainly conclude that decreasing naturalism, global scene consistency, and coherent depths cues has a detrimental influence on vection. Control studies (unpublished) showed that using abstract geometric stimuli (black and white checkerboard or vertically striped pattern) in the same VR setup further increased vection onset times and was, for some participants, even insufficient for inducing vection at all. This highlights the importance of using naturalistic and complex visual stimuli in VR, where the available FOV is typically much smaller than for the classic optokinetic drums, which allow full-field stimulation.

Vection onset times reported in circular vection studies using optokinetic drums vary considerably between studies and range from 2 to 30 s depending on a number of factors, including the velocity and spatial frequency content of the moving stimulus, FOV, stimulus eccentricity, eye movements, and perceived depth of the stimulus [Becker et al. 2002; Brandt et al. 1973b; Dichgans and Brandt 1978; Riecke et al. 2006b]. Vection onset latencies found in the present study range from 10 to 40 s and are thus clearly above the minimum values reported in the literature. A number of factors might have contributed, including the relatively small FOV, the lack of a fixation point in our study, and the different participant population.

Comparing not the absolute values of the dependent measures but their relative change between Experiments 1 and 2 shows a larger effect size for scene scrambling in Experiment 1 (25–50%, all three vection measures) than for scene inversion in Experiment 2 (10–15%, convincingness ratings only). Similarly, presence ratings show a somewhat smaller effect size for the scene inversion. One might argue that the change in the visual stimulus is indeed more subtle in the second experiment, where

the image statistics, depth structure, and global scene consistency of the stimulus remains unaltered, and only the orientation with respect to gravity is inverted and thus rendered less natural.

4.3   Potential Mechanisms Underlying the Observed Higher-Level Effect

Different mechanisms might have contributed to the observation that scene scrambling decreases vection. Three potential contributing mechanisms are discussed in the following and graphically represented in Figure 7 (a, b, and c). Note that the current study was not designed to disambiguate between these three mechanisms and the current data is, indeed, insufficient to do so.

1. *Presence and stable reference frame as factors mediating vection increase.* As a tentative first explanation of the higher-level contributions observed in the current study, we proposed above that the interpretation of the globally consistent, upright stimulus as a natural scene allowed for higher believability and presence in the simulated environment and provided observers with a more convincing, stable reference frame and primary rest frame with respect to which stimulus motion is being judged more easily as self-motion instead of object or image motion (see Figure 7, pathway (a)). The proposed mediating influence of presence for the self-motion illusion is in agreement with the "presence hypothesis" proposed by Prothero [1998] which states that "the sense of presence in the environment reflects the degree to which that environment influences the selected rest frame."

   Alternatively: *Increased depth and depth structure as factors mediating vection increase.* There is, however, an alternative hypothesis about the potential mechanism underlying the observed higher-level effect, as sketched in Figure 7, pathway (b) and (c): The consistent pictorial depth cues contained in the depiction of the natural scene might have increased the perceived depth and/or distance of the stimulus, since the spatial layout extending in depth was perceivable through an abundance of visual cues, including linear perspective, texture gradients and relative size (cf. Figure 1A and a). This means that the scene might have been perceived as being further away than the projection screen and the surrounding VR setup. In contrast, the scrambled stimuli was most likely interpreted as a 2D surface at the distance of the projection screen, since all meaningful information about spatial depth had been removed or made globally inconsistent by the scrambling procedure (cf. Figure 1, stimuli b–d and B–D). The latter is in agreement with participants' reports in a postexperimental interview.

2. *Increase in perceived vection velocity.* It is known from the literature that increasing perceived stimulus distance has a strong vection-enhancing effect: A study by Wist et al. [1975] showed that perceived speed of circular vection (which is often used as a measure of vection intensity) increased linearly with perceived distance of a stereoscopically presented stimulus, even though the angular velocity was unaltered. That is, the perceived velocity of circular vection is surprisingly not solely dependent on the angular velocity of the stimulus as one should expect, but also on the perceived distance of the stimulus. Similar increases of circular vection with increasing apparent distance were found irrespective of the method employed for altering perceived distance (Pulfrich effect, accommodative and fusional convergence). As increased physical stimulus speed (40 versus $20°$/s) enhanced vection in the current study, an increase in perceived distance thus might have indirectly enhanced vection here. The results of the current study might indicate that pictorial depth alone can also be sufficient for triggering this mechanism. Further experimentation is, however, necessary to clarify whether this explanation holds true.

3. *Perceived background motion.* The literature suggests another possibility how changes in perceived depth and/or distance of the visual stimulus might have affected vection in the current study: Howard and Heckmann [1989] demonstrated that a moving central display results in stronger vection if it is physically placed behind a stationary surround instead of before. Hence, the pictorial
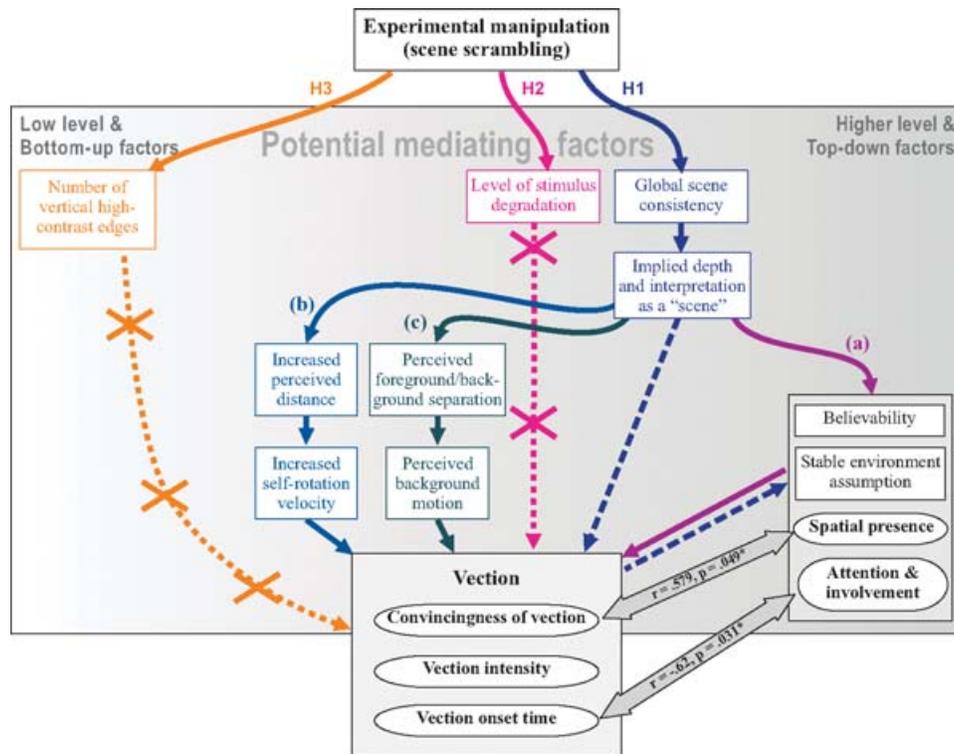
Fig. 7. Schematic illustration depicting how the experimental manipulation in Experiment 1 (scene scrambling; see top box) might have affected vection and presence via different mediating factors, ranging from lower-level (bottom-up) factors (left part of the graphic) to higher-level (cognitive/top-down) factors (right). The three different hypothesis outlined in the introduction are color-coded and labeled H1, H2, and H3. Only H1 (global scene consistency) was supported by the data and might have affected vection via three different mechanisms labeled (a), (b), and (c) (see text). Solid arrows indicate the most probable pathways; dashed and dotted arrows indicate less and least likely pathways, respectively. Vection and presence measures are displayed as oval boxes. Significant correlations between vection and presence measures are indicated by gray connecting arrows.

depth cues contained in the globally consistent stimulus might have resulted in a *perceived* foreground–background separation between the projection screen and surrounding setup (being perceived as foreground) and the projected globally consistent stimulus (being perceived as further away and thus as a background). In this manner, the stimulus motion might have been perceived as background motion, which, in turn, might have enhanced vection. If this was the case, a physical foreground–background depth separation would not be necessary to enhance vection, and pictorial depth cues would be sufficient. Note that in the current study there was no physical depth separation whatsoever between the moving stimulus and the projection setup. These results extend findings by Ohmi and colleagues [1987] who showed that when observers are presented with two similar competing visual stimuli (one moving and one stationary), circular vection occurs only when the moving display is *perceived* as being behind the stationary display (even though it might physically be closer).

4.4 Conclusions

Even though further experiments are required to disambiguate between these three potential contributing mechanisms, the current study supports the notion that cognitive and top-down mechanisms

do indeed affect the visually induced self-motion illusion, a phenomenon that was traditionally believed to be mainly bottom-up driven. A similar cognitive or top-down contribution was observed in a recent study on auditorily induced vection [Larsson et al. 2004]: Presenting sound sources that are normally associated with stationary objects (so-called "acoustic landmarks" like church bells) were more powerful in inducing circular vection than artificial sounds or sound typically generating from moving objects (e.g., foot steps). Hence, we propose that cognitive factors should be considered and further investigated both in self-motion simulation applications and in basic research–where they have been largely neglected apart from a few recent studies [Larsson et al. 2004; Lepecq et al. 1995; Palmisano and Chan 2004; Riecke et al. 2005d; Wright et al. 2005]. In this context, a recent study by Wright et al. [2005] is of particular interest, as it shows that visual motion of a photorealistic visual scene can dominate even conflicting inertial motion cues in the perception of self-motion.

Extending our knowledge about cognitive contributions to self-motion perception could also be advantageous from a practical standpoint: Compared to other means of increasing the convincingness and effectiveness of self-motion simulations like increasing the visual field of view or using a motion platform, cognitive factors can often be manipulated rather easily and without much effort, such that they might be an important step toward a lean and elegant approach to effective ego-motion simulation. This could be achieved through enhancing overall believability and spatial presence in the simulated scene, for example by using a naturalistic stimulus and/or providing consistent multimodal stimuli (e.g., adding acoustic landmarks to the visual scene [Larsson et al. 2004; Riecke et al. 2005b]). The current study suggests that the effectiveness of motion simulations could also be improved by increasing the perceived distance of the presented scene with respect to the VR setup using, for example, pictorial depth cues. Furthermore, and in agreement with Wright et al. [2005], the current study suggests that the effectiveness of motion simulations could be improved by using photorealistic visual stimuli: Those stimuli provide an abundance of cues about pictorial depth, relative distances, visual orientation, etc., and might thus serve as a cognitive anchor such that the simulated visual scene is more easily interpreted and accepted as the stable reference frame, such that displayed visual motion is more likely to be perceived as self-motion rather than object motion.

REFERENCES

ANDERSEN, G. J. AND BRAUNSTEIN, M. L.   1985.   Induced self-motion in central vision. *Journal of Experimental Psychology-Human Perception and Performance 11*, 2, 122–132.

BECKER, W., RAAB, S., AND JÜRGENS, R.   2002.   Circular vection during voluntary suppression of optokinetic reflex. *Experimental Brain Research 144*, 4 (June), 554–557.

BRANDT, T., DICHGANS, J., AND KOENIG, E.   1973b.   Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research 16*, 476–491.

DARLINGTON, C. L. AND SMITH, P. F.   1998.   Further evidence for gender differences in circularvection. *Journal of Vestibular Research-Equilibrium & Orientation 8*, 2 (Mar.-Apr.), 151–153.

DICHGANS, J. AND BRANDT, T.   1978.   Visual-vestibular interaction: Effects on self-motion perception and postural control. In *Perception*, R. Held, H. W. Leibowitz, and H.-L. Teuber, Eds. Handbook of Sensory Physiology, vol. VIII. Springer New York, 756–804.

DISTLER, H. K.   2003.   Wahrnehmung in Virtuellen Welten. Ph.D. thesis, Justus-Liebig-Universität Gießen.

FISCHER, M. H. AND KORNMÜLLER, A. E.   1930.   Optokinetisch ausgelöste Bewegungswahrnehmung und optokinetischer Nystagmus [Optokinetically induced motion perception and optokinetic nystagmus]. *Journal für Psychologie und Neurologie*, 273–308.

FUSHIKI, H., TAKATA, S., AND WATANABE, Y.   2000.   Influence of fixation on circular vection. *Journal of Vestibular Research-Equilibrium & Orientation 10*, 3, 151–155.

HETTINGER, L. J.   2002.   Illusory self-motion in virtual environments. In *Handbook of Virtual Environments*, K. M. Stanney, Ed. Lawrence Erlbaum Assoc. Hillsdale, NJ, 471–492.

HOWARD, I. P. AND HECKMANN, T.   1989.   Circular vection as a function of the relative sizes, distances, and positions of 2 competing visual-displays. *Perception 18*, 5, 657–665.

HOWARD, I. P. AND HOWARD, A.   1994.   Vection—the contributions of absolute and relative visual motion. *Perception 23*, 7, 745–751.

IJSSELSTEIJN, W. A.   2004.   Presence in depth. Ph.D. thesis, Technische Universiteit Eindhoven, Netherland.

KENNEDY, R. S., LANE, N. E., LILIENTHAL, M. G., BERBAUM, K. S., AND LAWRENCE.   1992.   Profile analysis of simulator sickness symptoms: Application to virtual environment systems. *Presence—Teleoperators and Virtual Environment 1*, 3, 295–301.

KENNEDY, R. S., HETTINGER, L. J., HARM, D. L., ORDY, J. M., AND DUNLAP, W. P.   1996.   Psychophysical scaling of circular vection (cv) produced by optokinetic (okn) motion: Individual differences and effects of practice. *Journal of Vestibular Research-Equilibrium & Orientation 6*, 5 (Sep.-Oct.), 331–341.

KITAZAKI, M. AND SATO, T.   2003.   Attentional modulation of self-motion perception. *Perception 32*, 4, 475–484.

LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M.   2004.   Perception of self-motion and presence in auditory virtual environments. In *Proceedings of Seventh Annual Workshop Presence 2004*. 252–258. Available: www.kyb.mpg.de/publication.html?publ=2953.

LEPECQ, J. C., JOUEN, F., AND DUBON, D.   1993.   The effect of linear vection on manual aiming at memorized directions of stationary targets. *Perception 22*, 1, 49–60.

LEPECQ, J. C., GIANNOPULU, I., AND BAUDONNIERE, P. M.   1995.   Cognitive effects on visually induced body motion in children. *Perception 24*, 4, 435–449.

MACH, E.   1875.   *Grundlinien der Lehre von der Bewegungsempfindung*. Engelmann, Leipzig, Germany.

NAKAMURA, S. AND SHIMOJO, S.   1999.   Critical role of foreground stimuli in perceiving visually induced self-motion (vection). *Perception 28*, 7, 893–902.

OHMI, M., HOWARD, I. P., AND LANDOLT, J. P.   1987.   Circular vection as a function of foreground-background relationships. *Perception 16*, 1, 17–22.

PALMISANO, S. AND CHAN, A. Y. C.   2004.   Jitter and size effects on vection are immune to experimental instructions and demands. *Perception 33*, 8, 987–1000.

PROTHERO, J. D.   1998.   The role of rest frames in vection, presence and motion sickness. Ph.D. thesis, University of Washington. Available: www.hitl.washington.edu/publications/r-98-11/.

RIECKE, B. E. AND 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, MPI for Biological Cybernetics. Avaliable: www.kyb.mpg.de/publication.html?publ=2021.

RIECKE, B. E., SCHULTE-PELKUM, J., AVRAAMIDES, M. N., AND BÜLTHOFF, H. H.   2004.   Enhancing the visually induced self-motion illusion (vection) under natural viewing conditions in virtual reality. In *Proceedings of Seventh Annual Workshop Presence 2004*. 125–132. Available: www.kyb.mpg.de/publication.html?publ=2864.

RIECKE, B. E., SCHULTE-PELKUM, J., AVRAAMIDES, M. N., VON DER HEYDE, M., AND BÜLTHOFF, H. H.   2005a.   Scene consistency and spatial presence increase the sensation of self-motion in virtual reality. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)*. La Coruña, Spain. 111–118. Available: www.kyb.mpg.de/publication.html?publ=3489.

RIECKE, B. E., CUNNINGHAM, D. W., AND BÜLTHOFF, H. H.   2006a.   Spatial Updating in Virtual Reality: The Sufficiency of Visual Information. *Psychological Research*. (in press).

RIECKE, B. E., SCHULTE-PELKUM, J., AND CANIARD, F.   2006b.   Using the perceptually oriented approach to optimize spatial presence & ego-motion simulation. In *Handbook of Presence*. Lawrence Erlbaum, Assoc. Hillsdate, NJ, submitted. 49–57

RIECKE, B. E., SCHULTE-PELKUM, J., CANIARD, F., AND BÜLTHOFF, H. H.   2005b.   Influence of Auditory Cues on the visually-induced Self-Motion Illusion (Circular Vection) in Virtual Reality. In *Proceedings of Eigth Annual Workshop Presence 2005*.

RIECKE, B. E., SCHULTE-PELKUM, J., CANIARD, F., AND BÜLTHOFF, H. H.   2005c.   Towards lean and elegant self-motion simulation in virtual reality. In *Proceedings of IEEE VR2005*. Bonn, Germany. 131–138.

RIECKE, B. E., VÄSTFJÄLL, D., LARSSON, P., AND SCHULTE-PELKUM, J.   2005d.   Top-down and multi-modal influences on self-motion perception in virtual reality. In *Proceedings of HCI international 2005*. Las Vegas, NV.

SADOWSKI, W. AND STANNEY, K.   2002.   Presence in virtual environments. In *Handbook of Virtual Environments*, K. M. Stanney, Ed. Lawrence Erlbaum Assoc., Hillsdale, NJ 791–806.

SCHUBERT, T., FRIEDMANN, F., AND REGENBRECHT, H. 2001. The experience of presence: Factor analytic insights. *Presence—Teleoperators and Virtual Environments 10*, 3, 266–281.

SCHULTE-PELKUM, J., RIECKE, B. E., AND BÜLTHOFF, H. H. 2004. Vibrational cues enhance believability of ego-motion simulation. In *International Multisensory Research Forum (IMRF)*. Available: www.kyb.mpg.de/publication.html?publ=2766.

TSCHERMAK, A. 1931. Optischer raumsinn. In *Handbuch der Normalen und Pathologischen Physiologie*, A. Bethe, G. Bergmann, G. Embden, and A. Ellinger, Eds. Springer, Berlin, 834–1000.

VAN DER STEEN, F. A. M. AND BROCKHOFF, P. T. M. 2000. Induction and impairment of saturated yaw and surge vection. *Perception & Psychophysics 62*, 1 (Jan.), 89–99.

WANN, J. AND RUSHTON, S. 1994. The illusion of self-motion in virtual-reality environments. *Behavioral and Brain Sciences 17*, 2 (June), 338–340.

WIST, E. R., DIENER, H. C., DICHGANS, J., AND BRANDT, T. 1975. Perceived distance and perceived speed of self-motion—linear vs angular velocity. *Perception & Psychophysics 17*, 6, 549–554.

WITMER, B. G. AND SINGER, M. J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence—Teleoperators and Virtual Environments 7*, 3, 225–240.

WRIGHT, W. G., DIZIO, P., AND LACKNER, J. R. 2005. Vertical linear self-motion perception during visual and inertial motion: More than weighted summation of sensory inputs. *Journal of Vestibular Research 15*, 185–195.

ZACHARIAS, G. L. AND YOUNG, L. R. 1981. Influence of combined visual and vestibular cues on human perception and control of horizontal rotation. *Experimental Brain Research 41*, 159–171.