Tyler Thrash*,[1]
Postdoctoral Researcher
Chair of Cognitive Science
ETH Zurich
Clausiusstrasse 59
RZ E 22.2, Zurich
Switzerland 8092

Mubbasir Kapadia[1]
Computer Science Department
Rutgers University

Mehdi Moussaid[1]
Max Planck Institute for Human Development

Christophe Wilhelm
Disney Research Zurich

Dirk Helbing
Chair of Sociology, in particular of modeling and simulation
ETH Zurich

Robert W. Sumner
Disney Research Zurich

Christoph Hölscher
Chair of Cognitive Science
ETH Zurich

Presence, *Vol. 24, No. 4, Fall 2015,* 322–334
doi:10.1162/PRES_a_00237

Evaluation of Control Interfaces for Desktop Virtual Environments

Abstract

Tracking and analyzing the movement trajectories of individuals and groups is an important problem with applications in crowd management and the development of transportation systems. However, real-world tracking is limited due to the size of the trackable area and the precision with which a person can be tracked. Experiments in virtual environments have many advantages, including practically unlimited sizes and the precise measurement of spatial behavior. However, the generalizability of research using virtual environments to real-world scenarios is often limited by the translation of participants' movements to those of their avatars. We compared human movement patterns in virtual environments with different control interfaces: a handheld joystick, a mouse-and-keyboard setup, and a keyboard-only setup. With each of these controls, participants completed several movement-related tasks of varying difficulty in a limited amount of time. Questionnaires indicated that participants preferred the mouse-and-keyboard setup over the other two setups. Standard performance measures suggested that the joystick underperformed in a variety of tasks. Movement trajectories in the final task indicated that each of the control setups produced somewhat realistic behavior, despite some apparent differences from real-world trajectories. Overall, the results indicated that, given limited resources, mouse-and-keyboard setups consistently outperform joysticks and produce realistic movement patterns.

1 Introduction

Human movement provides the means by which people make decisions and interact with each other socially. At its most basic level, human movement may include such actions as pressing a button, stating a preference, or even directing gaze toward an interesting visual pattern. At larger scales, humans' capacity for movement has allowed for the transference of culture and language, the development of complex transportation systems, and the spread of our species around the world. Understanding a behavior relies on an understanding of its underlying movements, and thus, any science of behavior relies on researchers' abilities to track these movements.

Tracking human movement at a larger scale has allowed researchers to investigate a variety of topics at both individual (e.g., Hodgson et al., 2015) and collective (e.g., Johansson, Helbing, & Shukla, 2007; Robin, Antonini, Bierlaire, & Cruz, 2009) levels of spatial behavior. However, tracking

*Correspondence to tyler.thrash@gess.ethz.ch.
[1] The first three authors contributed equally to this work.

movement through a real environment poses several challenges such as the size of the trackable area and the precision with which a person can be tracked (Hodgson et al., 2015). Tracking solutions are typically limited to a small number of individuals in controlled setups (e.g., Moussaïd et al., 2009) and require supervision and manual labor. It is expensive to conduct multiple variations of tracking experiments and impractical to collect data for situations such as stress or panic.

In contrast, virtual environments allow researchers to precisely track users as they move through a practically unlimited space. However, one potential limitation of virtual environment studies is the generalizability of the obtained results to real-world scenarios (Loomis, Blascovich, & Beall, 1999). This generalizability is primarily limited by the realism of the graphics and the translation of a person's movements into those of his/her avatar in the virtual environment.

Previous research has found that type of control scheme affects participants' abilities to maneuver through different virtual environments (e.g., Lapointe, Savard, & Vinson, 2011; Reicke, Bodenheimer, McNamara, Williams, Peng, & Feuereissen, 2010; Ruddle, Volkova, & Bülthoff, 2013), but few studies have compared the use of control interfaces for desktop virtual environments across a variety of spatial tasks. For example, Ruddle and colleagues (2013) found that moving through a virtual environment using a treadmill was easier than using a joystick whether coupled with a desktop or head-mounted display. However, this study assessed the efficiency of movement along one particular route through one particular virtual environment (Ruddle et al., 2013).

In addition, across the literature, different control interfaces are selected for comparison depending on the specific application being considered. For example, Teixeira, Vilar, Duarte, Rebelo, and da Silva (2012) compared user performance with a joystick to user performance with a Nintendo® Wii Balance Board in order to generalize to commercial gaming applications. In contrast, the present study compares control setups in several movement-related tasks in order to generalize to interactive, multiuser scenarios.

Specifically, we compared joystick, mouse-and-keyboard, and keyboard-only setups in terms of user preference, the time required to complete various tasks, and the frequency with which avatars collided with obstacles. We also compared movement trajectories in the virtual environment to real-world trajectories for a similar task. Performance data were collected in a variety of virtual environments during different types of tasks. Participants were asked to complete a step-by-step tutorial in an irregularly shaped room, a maneuvering task in a circular corridor with and without stationary obstacles, and a maneuvering task in a straight corridor with another participant's avatar. Our results indicate (1) that the joystick underperforms relative to the other control setups in terms of user preferences and objective performance measures and (2) that trajectories in virtual environments are similar to those in real environments in systematic ways.

2 Method

The present experiment investigated preference and performance differences for three control setups that are commonly employed in research using desktop virtual reality. Users performed movement-related tasks in several different virtual environments.

2.1 Participants

There were 69 participants who were recruited for participation using the University Registration Center for Study Participants (http://www.uast.uzh.ch). All of the participants were between 19 and 36 years of age. Compensation depended on specific performance criteria. Reward points were given for completing particular tasks during the experiment and were removed for collisions in the virtual environments. Point totals were then normalized across participants and converted to monetary values between 25 and 55 CHF with a mean compensation of approximately 40 CHF. Data were collected in the ETH Decision Science Lab (DeSciL) which independently approved the experimental procedures according to its human participant regulations.

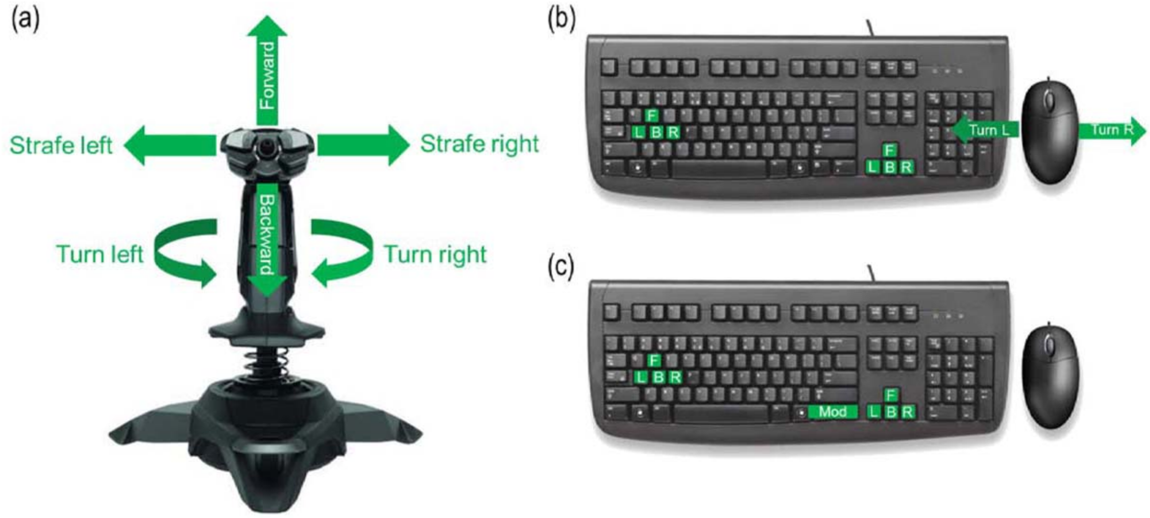

Figure 1. *Control mappings for the three control setups: (a) joystick, (b) mouse-and-keyboard, and (c) keyboard-only. "F" represents forward movement, "B" represents backward movement, "L" represents leftward movement, and "R" represents rightward movement.*

2.2 Materials

Participants were seated in a room containing 36 cubicles, each containing a desktop computer. Each of these computers was a Dell Optiplex 980 running Windows 7 Enterprise SP1 X64. Each computer was connected to a Dell 1909W monitor with a 19-inch diagonal and a resolution of 1400 $\times$ 900 pixels. Three other peripheral devices were also connected: a keyboard, a mouse, and a Saitek Cyborg V.1 joystick.

Three control setups were compared for the present study: keyboard-only, mouse-and-keyboard, and joystick. Several variables were held constant across the three control setups. Each control setup had three degrees of freedom (i.e., forward/backward, strafe left/strafe right, and turn left/turn right). Also, all of the control mappings (i.e., the functions relating button presses to movement through the virtual environment) used the same maximum velocity (i.e., 1.3 m/s for forward translations and 0.6 m/s for backward and sideways translations). However, control mappings necessarily varied from one control setup to another in some respects because of differences in the construction of the controls themselves. See Figure 1 for the mappings of the different control setups.

The experimental application was developed using the Unity3D game engine (Unity Technologies), and ADAPT (Shoulson, Marshak, Kapadia, & Badler, 2014) was used for animating the virtual avatars being controlled by the human subjects. The avatars were 1.85 m tall with a horizontal field of view of approximately 135 degrees and a collision radius of 0.25 m. The application frame rate was at least 30 frames per second, and the network latency was approximately 67 msec. The experimental application placed participants into three different virtual environments: an irregular training room, a circular corridor, and a straight corridor. See Figure 2 for overhead views and screenshots of each of these virtual environments.

2.3 Procedure

The experiment was conducted over the course of 12 separate sessions. Within each session, five to six participants completed the experimental phases in the same order for each of the three control setups. In order, the experimental phases included a step-by-step tutorial (in the irregular training room), Task A (in the circular corridor without obstacles), Task B (in the circular corridor with obstacles), and Task C (in the straight corridor with

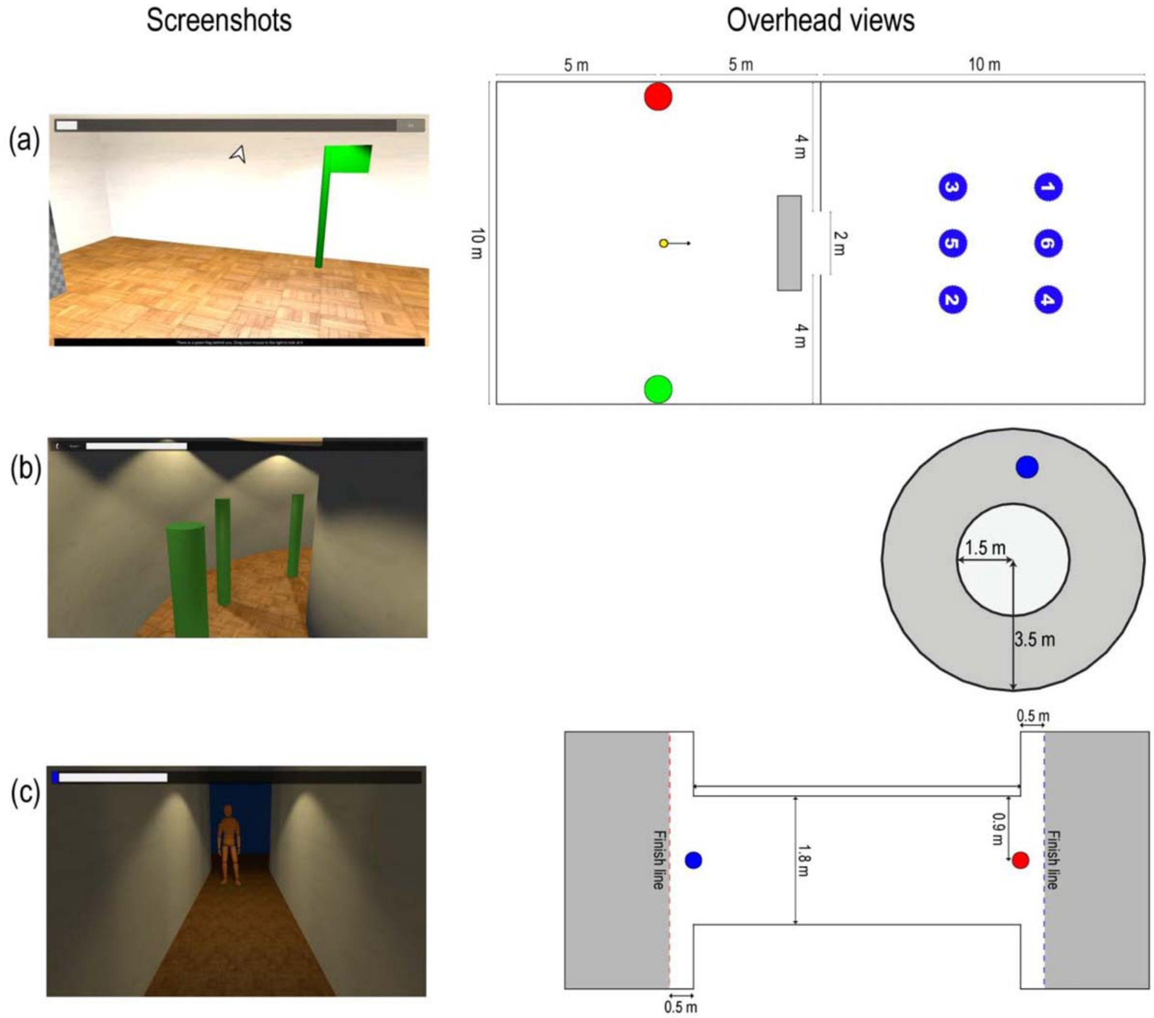

Figure 2. *Screenshots and overhead views of the virtual environments used in (a) the step-by-step tutorial, (b) Tasks A and B, and (c) Task C.*

another person's avatar). Brief instructions were presented on the computer screen before the step-by-step tutorial and between each of the other phases.

During the step-by-step tutorial, each pair of opposing controls was enabled separately in order for participants to complete several simple navigation tasks. For example, one task instructed participants to "turn left to view the red flag." Once participants used each of the six controls within a particular control setup, they were asked to move into a room containing a blue flag. Participants were then asked to walk toward the blue flag with all of the controls for a particular setup enabled. After reaching the blue flag, another blue flag would appear at a different location. This procedure was repeated until the participants had reached blue flags in six different, randomized but predetermined, locations. The flag locations were the same for every participant.

During Task A, participants were asked to move around a circular corridor in a randomly determined direction—clockwise or counterclockwise—until they obtained 50 "points." Each participant started Task A with 30 points, gained 10 points for each completed revolution around the circular corridor, and lost 1 point for each collision with a wall. These point values were defined in such a way as to avoid values below zero and to ensure that each participant would complete at least two trials. Participants were instructed that more points would result in more compensation at the end of the experiment.

Task B was similar to Task A except that the circular corridor then contained green, stationary obstacles. Each obstacle was 1.5 m tall in the virtual environment, and its location was randomly chosen from a set of 40 locations that were evenly distributed around the corridor. Task B began with 5 obstacles; this number increased at a rate of 1 obstacle per completed revolution with a maximum of 7. Pilot testing suggested that the task became too difficult for participants with more than 7 obstacles. For each collision with an obstacle, participants lost a point.

Once a participant obtained at least 50 points in Task B, he/she was paired with another participant, and his/her avatar was placed at one end of the virtual, straight corridor facing the other participant at the other end in order to begin Task C. Each participant was then asked to move past the other participant without colliding. As in Task A and Task B, any collision resulted in a 1-point penalty. Completing each trial also resulted in a 10-point reward. Once a participant obtained at least 50 points in Task C, he/she continued to perform similar trials in the straight corridor with another person who also reached the same performance criterion. This final phase continued until every participant within a session had completed at least 2 testing trials or a time limit of 10 minutes was reached.

After training with each of the control setups for up to 10 minutes, the participants completed a short survey. The critical question on this survey was "Which of the control setups from today's experiment did you prefer to use?" Other questions on the survey were related to gaming experience and were used to recruit participants for another set of experiments.

2.4 Design and Analysis

The primary independent variable "control setup" had three conditions: keyboard-only, mouse-and-keyboard, and joystick. The order of control setups was counterbalanced across sessions so that every possible permutation was used for exactly two sessions. Because of an isolated experimenter error that resulted in extreme lag, the data were discarded for one control setup (i.e., the joystick) in one session. Answers to the survey question about preferred control setup and performances in each of the experimental phases were compared across the three setups. The proportion of participants who reported a preference for each of the control setups was compared to chance using a chi-squared goodness-of-fit test. Performances in the step-by-step tutorial, Task A, and Task B phases were evaluated using the mean time required to complete each phase. Performances in the Task B phase were also evaluated using the mean number of collisions with any wall or obstacle.

Performances in Task C were evaluated by comparing movement trajectories in the virtual straight corridor and two pseudo-random simulations to analogous data previously collected in a real-world study (Moussaïd et al., 2009). These analyses were conducted in several different ways in order to prevent the over-interpretation of any particular finding. Inferential statistics are provided where possible.

2.4.1 Bayes' Information Criterion. This set of analyses allowed us to assess the similarity between distributions of real-world trajectories and distributions of virtual trajectories resulting from each of the control setups and to compare these similarities across control setups using established standards (see Kass & Raftery, 1995). Virtual and real-world trajectories that included fewer than 50 or more than 100 data points were eliminated from these analyses. One trial had fewer than 50 data points because it ended prematurely. Trajectories with more than 100 data points (i.e., 6.5% of trials) most likely represented trials in which participants ignored the instructions. For example, at least one participant was observed moving back and forth through the hallway without heading toward the goal. Distributions for the remaining trajectories were calculated by aggregating over locations that were recorded while participants moved through the virtual/real-world environment. The virtual and real-world corridors were discretized into 100×20 equally sized spatial grid cells. Each of these grid cells represented an approximately 10 cm $\times$ 10 cm space. We then counted the frequency of data

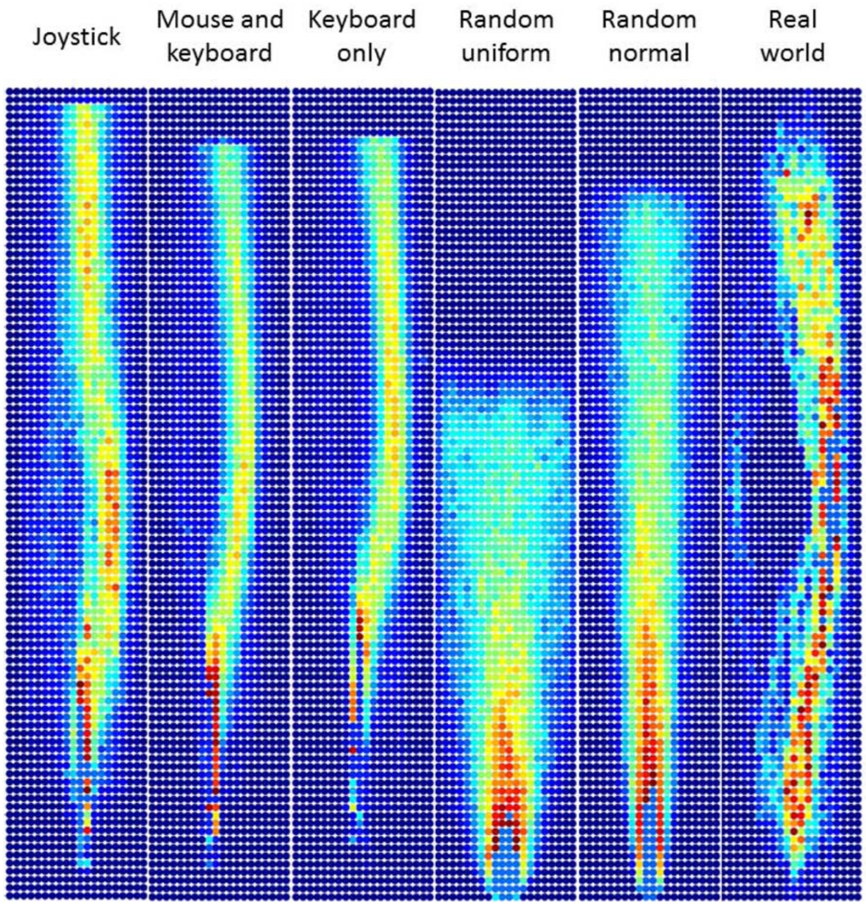

Figure 3. *Density maps representing movement trajectories from each of the three experimental conditions, two pseudorandom simulations, and the real-world experiment. The trajectories begin at the bottom of these maps and finish at the top of these maps.*

points within each of these cells. See Figure 3 for density maps representing the aggregated trajectories. We then divided the frequencies for each grid cell by the sum of observed data points in the corresponding condition. We also defined outliers as values beyond three standard deviations of the mean value across grid cells and replaced each of them with the mean. Defining and replacing outliers using the median gives largely the same pattern of results.

Movement trajectories were also simulated using a few basic assumptions. Each data point of each trajectory was assumed to be $1/100^{\text{th}}$ of the length of the corridor in order to produce a similar number of data points as for the real-world and virtual-environment trajectories. Each of 100 random data points was produced for each trajectory by generating a random vector extending from the previous data point and constrained so that the agent never moved backwards or outside of the virtual corridor. All of the points for each trajectory were generated from either uniform or normal distributions. For the uniform distribution simulation, the probability of moving in any forward or sideways direction was the same as the probability of moving in any other forward or sideways direction. The normal distribution simulation was additionally constrained so that movement perpendicular to the length of the hallway corresponded to a randomly generated value that was three standard deviations from the mean of the distribution. For both uniform and normal distribution simulations, one thousand trajectories were generated and aggregated using the same method as for the empirical data described earlier.

We then computed the error variance for the difference between two probabilities: the probability of someone moving within each specified grid cell in the virtual environment and the probability of someone moving within each specified grid cell in the real environment:

$$\widehat{\sigma_e^2} = \frac{\Sigma(pi_v - pi_r)^2}{n_i - 1}. \quad (1)$$

Here, $\widehat{\sigma_e^2}$ represents the estimated error variance, pi_v represents the probability of someone moving within a specified grid cell in the virtual environment in one of the experimental conditions v, pi_r represents the probability of someone moving within a specified grid cell in the real world, and n_i represents the number of grid cells over which the error was summed.

Estimated error variance was then converted to Bayes' information criterion (BIC; Lewandowsky & Farrell, 2010) using the following equation:

$$BIC_v = \left[n_i \times \ln\left(\widehat{\sigma_e^2}\right)\right] + [k \times \ln(n_i)]. \quad (2)$$

Here, k represents the number of free parameters in the model used to predict the data. For this particular statistical application, we consider the virtual and random trajectory data as models predicting the real-world data. Thus, there are zero free parameters for this case, and the calculation essentially becomes

$$BIC_v = \left[n_i \times \ln\left(\widehat{\sigma_e^2}\right)\right]. \quad (3)$$

In order to compare the fit of the different conditions to the real-world data using established standards (see Kass & Raftery, 1995), we then calculated the difference between *BIC*s for each pair of conditions:

$$\ln(B) = BIC_1 - BIC_2. \quad (4)$$

Here, $\ln(B)$ represents the natural logarithm of the Bayes' factor. Low values of $\ln(B)$ indicate evidence favoring the first model. See Table 1 for the standards established by Kass and Raftery (1995) for interpreting the magnitude of this term.

Table 1. *Established Standards for Interpreting $-\ln(B)$ as Strength of Evidence for One Model Over Another (Kass & Raftery, 1995)*

$-\ln(B)$	Strength of evidence
0 to 1	Not worth more than a bare mention
1 to 3	Positive
3 to 5	Strong
> 5	Very strong

In order to help determine whether the magnitude of each effect was driven by the large number of grid cells used to compute the trajectory distributions, we also calculated the size of n needed to produce very strong effects:

$$n_{req} = \frac{-5}{\ln\left(\widehat{\sigma_{e_1}^2}\right) - \ln\left(\widehat{\sigma_{e_2}^2}\right)}. \quad (5)$$

Here, n_{req} refers to the number of data points required to produce very strong evidence toward the first model.

2.4.2 Theil's Uncertainty Coefficient (Symmetrized).

This set of analyses was also used to compare distributions of trajectory data and was based on a similar aggregation procedure as that used to calculate the BIC except for one major difference. We used this additional measure of distribution similarity to prevent overreliance on any one particular measure. Also, the relationship between measures based on Bayes' theorem and those based on Shannon entropy is largely unknown.

Because Theil's uncertainty coefficient (Theil's U) is an extension of information entropy measures, probabilities of zero cannot be incorporated easily. Thus, in order to account for grid cells that did not contain any observations, the probability of a participant moving through each grid cell was subtracted from one, essentially inverting the trajectory distributions for all virtual and real-world data. Although this transformation prevents the interpretation of the obtained Theil's U's in terms of information bits, it should not affect the comparison of Theil's U's across control setups.

Theil's U requires the calculation of entropy for single distributions and conditional entropies for pairs of distributions. The entropy for single distributions represents the uncertainty associated with each set of trajectories separately (see Equation 6), whereas conditional entropy for pairs of distributions represents the uncertainty associated with one set of trajectories given that another set of trajectories is known (see Equation 7):

$$H(D) = -\Sigma[p_{i_D} \times \ln(p_{i_D})]]. \tag{6}$$

Here, $H(D)$ represents the entropy (H) associated with any distribution (D), and p_{i_D} represents the probability (p) of an observation (i) with respect to that distribution (D). The distribution D may refer to the aggregated set of trajectories from a virtual environment condition (v) or from the real-world data (r).

$$H(D_1|D_2) = -\Sigma\left[p_{i_{D_1}, i_{D_2}} \times \ln\left(p_{i_{D_1}|i_{D_2}}\right)\right]. \tag{7}$$

Here, $H(D_1|D_2)$ represents the conditional entropy of one distribution (D_1) given that another distribution (D_2) is known. Also, $p_{i_{D_2}, i_{D_2}}$ represents the product of the probability of each observation i from D_1 and the probability of the corresponding observation i from D_2; $p_{i_{D_2}|i_{D_2}}$ represents the conditional probability of each observation i from D_1 given that the corresponding observation i from D_2 is known.

The original Theil's U can then be calculated using Equations 6 and 7:

$$U(D_1|D_2) = \frac{H(D_1) - H(D_1|D_2)}{H(D_1)}. \tag{8}$$

Here, $U(D_1|D_2)$ represents the number of bits in distribution D_1 that are predictable given distribution D_2. For our application to the present data, we used a symmetrized version of Theil's U in order to measure the association between distributions D_1 and D_2:

$$S(D_1, D_2) = \frac{[H(D_1) \times U(D_1|D_2)] + [H(D_2) \times U(D_2|D_1)]}{H(D_1) + H(D_2)}. \tag{9}$$

The symmetrized Theil's U was used to compare aggregated trajectories from the different control conditions in terms of their association with the aggregated trajectories from the real-world data.

2.4.3 Minimum Predicted Distance. Following Olivier, Marin, Crétual, and Pettre (2012), we used this measure in order to analyze the time course of motion adaptation during Task C. Along each observed data point of the task, we calculated the smallest distance between the two participants that would occur if they continued at the same velocity. The resulting distribution of minimum predicted distances for each trial was then standardized so that every data point fell between time 0 and time 100. Here, a data point at time 0 represents the beginning position of the avatar and a data point at time 100 represents the point at which the two participants passed each other. These standardized distributions of minimum predicted distances were then averaged for each control setup. Correlations were then used in order to determine the degree of association between the standardized and aggregated distributions of minimum predicted distance for each of the control setups and the real-world data.

2.4.4 Other Measures for Trajectory Comparison. Across control conditions, we also compared variability in the locations along the length of the corridor at which the two participants passed each other (i.e., "crossing points"), the distance between participants as they passed each other, and number of two-person collisions. This set of statistical comparisons must be interpreted cautiously because the data could not be meaningfully aggregated for each participant without dependencies in the data. For example, each two-person collision would be counted twice if we would have analyzed data aggregated for individual participants. Thus, each trial was considered independently.

3 Results

A two-tailed, chi-squared test was used to evaluate user preferences. The proportion of participants who preferred each control setup (24 for keyboard-only, 35 for mouse-and-keyboard, and 7 for joystick) significantly differed from the equal proportions expected by chance,

$x^2(2) = 18.09$, $p < .001$. One participant did not complete this survey item because of simulator sickness during the experiment (see Vinson, Lapointe, Parush, & Roberts, 2012), and two participants were eliminated from this analysis because they chose two of the three control setups. Regardless, this effect was clearly driven by the low number of participants who reported a preference for the joystick.

Two-tailed, between-subjects, one-way ANOVAs were used in order to evaluate differences among the three control setups in the step-by-step tutorial, Task A, and Task B in terms of mean time required to complete each of these phases. The same type of test was also used in order to evaluate differences in the number of collisions exhibited by users in Task B. Because all participants progressed through these phases in the same order with a 10-minute time limit overall, a few participants were included in the analysis of the earlier phases but not in the later phases. Also, some participants progressed more quickly with one control setup than another. Because of these slight imbalances in the design, between-subjects ANOVAs were used to analyze a factor that was essentially manipulated within-subjects. Note that this approach is relatively conservative because between-subjects comparisons have lower power than within-subjects comparisons for the same sets of data (Keppel & Wickens, 2004).

There were no significant differences among the three control setups in terms of completion time for the step-by-step tutorial, $F(2,201) = 1.30$, $MSE = 6257.4$, $p = .276$ (see Figure 4). However, there were significant differences among the control setups in terms of completion time for both Task A, $F(2,200) = 9.75$, $MSE = 6046.6$, $p < .001$, and Task B, $F(2,192) = 11.69$, $MSE = 6112.5$, $p < .001$ (see Figure 4). Participants required more time to complete Task A and Task B with the joystick than with the other types of controls. There was also a significant difference among the three control setups in terms of number of collisions for Task B, $F(2,192) = 29.06$, $MSE = 550.2$, $p < .001$ (see Figure 5).

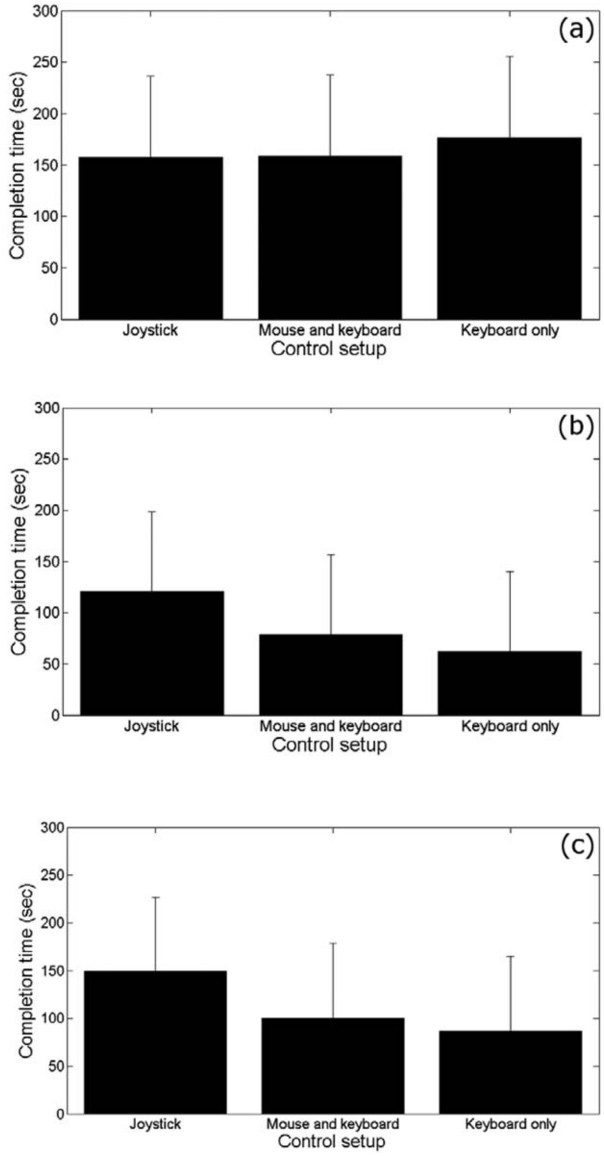

Figure 4. *Mean completion time for each of the three control setups in (a) the step-by-step tutorial, (b) Task A, and (c) Task B. Error bars represent the magnitude of MSE for the between-subjects ANOVA.*

3.1 Bayes' information criterion. For the trajectory data, we first interpreted the magnitude of $-\ln(B)$ for the comparison of every pair of control setups in terms of "realism." Recall that $-\ln(B)$ accounts for similarities between the control setups and the real-world data. There was very strong evidence for the joystick over both the mouse-and-keyboard setup, $-\ln(B) = 1484$, and the keyboard-only setup, $-\ln(B) = 1023$. There was also very strong evidence for the keyboard-only setup over the mouse-and-keyboard setup, $-\ln(B) = 131$. The very large values obtained for

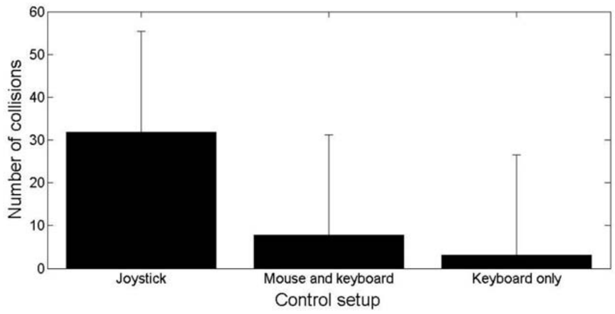

Figure 5. *Mean number of collisions for each of the three control setups in Task B. Error bars represent the magnitude of MSE for the between-subjects ANOVA.*

each of these comparisons may be attributable to the large number of data points being fit (i.e., $n = 2000$). In order to evaluate the extent to which the magnitude of $-\ln(B)$ was attributable to the number of data points being fit, we next calculated the size of n_{req} needed to obtain very strong evidence (i.e., $-\ln(B) > 5$) for each of the previously given comparisons (see Equation 5). For very strong evidence for the joystick over the mouse-and-keyboard setup, $n_{req} = 7$. For very strong evidence for the joystick over the keyboard-only setup, $n_{req} = 10$. For very strong evidence for the mouse-and-keyboard setup over the keyboard-only setup, $n_{req} = 22$.

Next, we interpreted the magnitude of $-\ln(B)$ for the comparison of every control setup to the uniform and normal distribution simulations. There was very strong evidence for each of the different control setups over the uniform distribution simulation. For the comparison of the joystick and the uniform distribution simulation, $-\ln(B) = 1527$, $n_{req} = 7$. For the comparison of the mouse-and-keyboard setup and the uniform distribution simulation, $-\ln(B) = 43$, $n_{req} = 232$. For the comparison of the keyboard-only setup and the uniform distribution simulation, $-\ln(B) = 504$, $n_{req} = 20$. There was also very strong evidence for the joystick and keyboard-only setups over the normal distribution simulation but not for the mouse-and-keyboard setup. For the comparison of the joystick and the normal distribution simulation, $-\ln(B) = 1252$, $n_{req} = 8$. For the comparison of the keyboard-only setup and the normal distribution simulation, $-\ln(B) = 228$, $n_{req} = 44$. For the comparison of the normal distribution simulation and the mouse-and-keyboard setup, $-\ln(B) = 233$, $n_{req} = 43$.

3.2 Theil's uncertainty coefficient (symmetrized). As an attempt to (dis)confirm the results of the Bayesian analyses, we then compared the three control setups in terms of Theil's U. For the joystick, $U = 3.82 \text{ x } 10^{-3}$. For the mouse-and-keyboard setup, $U = 3.81 \text{ x } 10^{-3}$. For the keyboard-only setup, $U = 3.81 \text{ x } 10^{-3}$. Although differences between Theil's U's for different control conditions are very small and should not be over-interpreted, they slightly favor the joystick over the mouse-and-keyboard and keyboard-only setups.

3.3 Minimum predicted distance. For the three control conditions and the real-world data, Figure 6 depicts changes in the mean minimum predicted distance over time from the beginning of each trial to the point at which the participants crossed each other. The correlations between mean minimum predicted distance for each of the three control conditions and the real-world data were all significantly above chance. For the joystick, $r = .49$. For the mouse-and-keyboard setup, $r = .53$. For the keyboard-only setup, $r = .54$. The results tend to favor the mouse-and-keyboard and keyboard-only setups over the joystick, but the differences between control conditions are relatively small.

3.4 Other measures for trajectory comparison. Variability in crossing points (vcp) was slightly lower for the mouse-and-keyboard setup than for the other two control setups. For the joystick, $vcp = 0.38$. For the mouse-and-keyboard setup, $vcp = 0.34$. For the keyboard-only setup, $vcp = 0.39$.

Using a single-factor, between-subjects ANOVA, we also found a significant difference among mean crossing distances for the joystick ($M = 0.73$, $SD = 0.21$), mouse-and-keyboard ($M = 0.73$, $SD = 0.17$), and keyboard-only ($M = 0.77$, $SD = 0.17$) setups, $F(2,2101) = 1342.64$, $MSE = 0.015$, $p < .0001$ (see Figure 7). There was more distance between pairs of participants in the keyboard-only condition than in the other two control conditions.

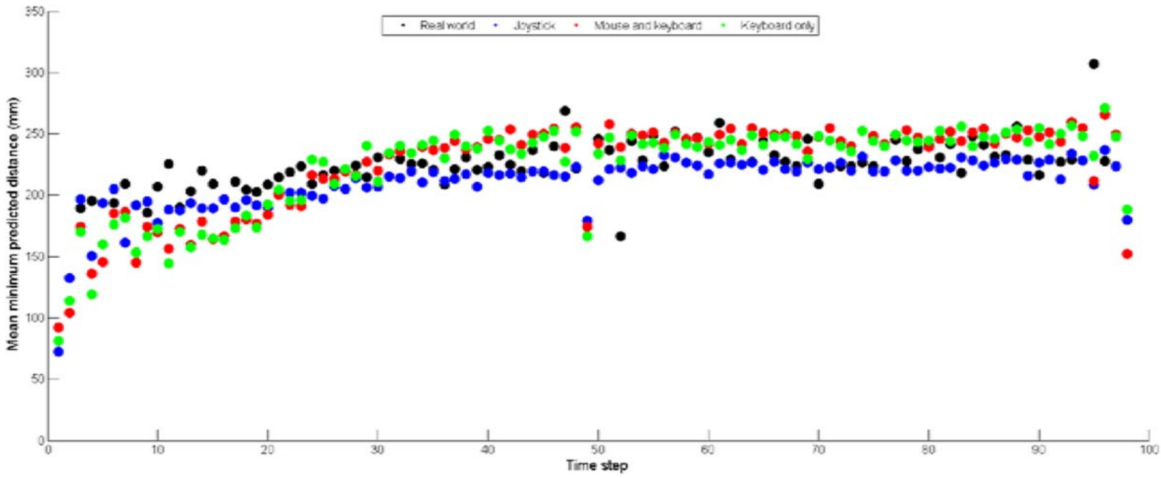

Figure 6. *Graph representing the mean minimum predicted distances for all three control conditions and the real-world data.*

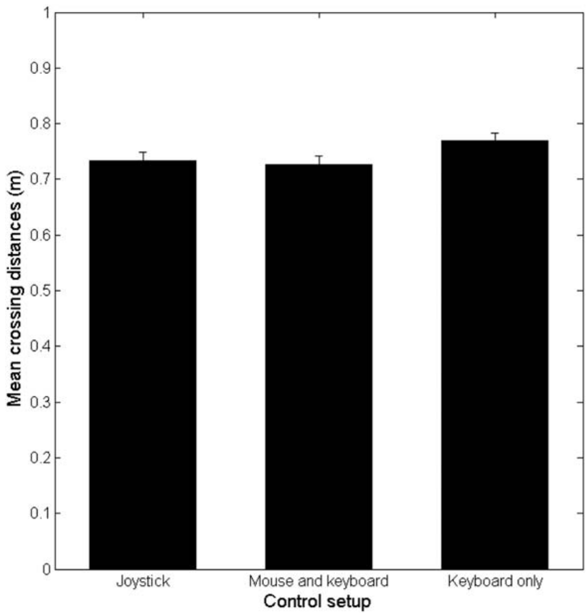

Figure 7. *Mean crossing distances for each of the three control setups in Task C. Error bars represent the magnitude of MSE for the between-subjects ANOVA.*

There were also a larger number of two-person collisions for the keyboard-only condition (66) than for the joystick (52) or mouse-and-keyboard (49) conditions.

4 Discussion

There are two sets of findings in the present work. First, simple control setups that required only a mouse and keyboard outperformed joysticks in terms of user preferences and standard performance measures for qualitatively different tasks. Second, movement trajectories appeared to fit real-world data relatively well, despite some apparent differences.

In the present study, users explicitly preferred the mouse-and-keyboard setup over the joystick. Participants also performed different spatial tasks more efficiently and with fewer collisions while using the mouse-and-keyboard setup relative to the other two conditions. One possible explanation for these performance advantages is that participants have had much more experience with desktop computers than with video games using a joystick interface (see Lapointe et al., 2011, for a similar argument). Future research should investigate the effects of extended practice on differences in performance across a variety of movement-related tasks and control setups.

Aggregated movement trajectories from the virtual environment conditions were also generally more similar to trajectories obtained in the real world than pseudo-random simulations. While this threshold for the veridicality of virtual environment data may be relatively low, these results suggest some systematic similarities between real-world and virtual-environment trajectories. For example, participants largely passed each other in the straight corridor on their right side (see Figure 3). For real-world data, Moussaid and colleagues (2009) interpreted this "right-side bias" as evidence for self-reinforcing behavioral conventions. People adjust their movements with the expected movements of other oncoming pedestrians in order to avoid collisions. As more and more pedestrians pass each other, a preference for either the right or left side develops. In order to investigate the emergence of such phenomena independently of aspects of the immediate environment such as bodily asymmetries or the direction of car-traffic (Moussaïd et al., 2009), researchers could experimentally vary the chosen passing side of confederates in a larger crowd of naïve volunteers. This variation could indicate whether the right-side bias emerges over the course of several interactions or is culturally defined beforehand.

Taken together, these findings suggest both opportunities and limitations for future research using these common control setups.

4.1 Limitations

One possible limitation of the present experiment is that the results reflect only one possible set of mappings between participants' manipulations of the controls and their avatars' movements. For example, it is possible that a slower maximum velocity of 0.7 m/s for forward translations would not result in a difference in the realism of movement trajectories. However, a maximum forward velocity of 1.3 m/s best reflects individuals' walking behavior in similar, real environments (see Moussaid et al., 2009).

Another possible limitation of the present experiment is the overall number of collisions by participants. For example, there were 167 two-person collisions in 2107 trials for Task C in the present experiment. Participants were penalized monetarily for colliding with walls, other stationary obstacles, or each other, but this penalty appears to be insufficient for preventing collisions that would probably not occur during movement through real environments. However, this does not necessarily preclude the comparison of different control setups in the same virtual environments. Indeed, the number of collisions in each experimental condition was useful in the present study for distinguishing between the different control setups.

4.2 Possible Applications

Despite these potential limitations, the present results clearly demonstrate (1) that mouse-and-keyboard setups outperform handheld joysticks in a variety of movement-related tasks and (2) that movement trajectories in virtual environments can reflect real-world trajectories in systematic ways. Fine-grained analyses of movement trajectories (such as those employed here) may be useful for evaluating local interactions and emergent crowd dynamics among multiple avatars in virtual environments (e.g., Olivier, Bruneau, Cirio, Pettŕe, 2014) or for translating a user's movements into those of a robot (i.e., teleoperation; e.g., Her, Hsu, Lan, & Karkoub, 2002; Mavrikios, Karabatsou, Pappas, & Chryssolouris, 2007). For the former example, the extent to which emergent crowd behaviors reflect real-world behaviors may depend in part on the veridicality of the crowd members' movements. With appropriate control training, researchers may be able to study crowd behavior in otherwise dangerous conditions such as during a stressful evacuation or at high crowd densities. For the latter example, precise control over a robot may be necessary for maneuvering through remote environments such as the sea floor or the surface of Mars.

Acknowledgments

The authors wish to thank Felix Thaler for his help with developing the virtual environments and experimental program, and Hany Abdelrahman and Iva Barišić for their help with creating the figures. Mehdi Moussaid was supported by a grant from the German Research Foundation (DFG) as part of the priority program on New Frameworks of Rationality (SPP 1516) to Ralph Hertwig and Thorsten Pachur (HE 2768/7-2).

References

Her, M. G., Hsu, K. S., Lan, T. S., & Karkoub, M. (2002). Haptic direct-drive robot control scheme in virtual reality. *Journal of Intelligent and Robotic Systems, 35*(3), 247–264.

Hodgson, E., Bachmann, E. R., Vincent, D., Zmuda, M., Waller, D., & Calusdian, J. (2015). WeaVR: A self-contained and wearable immersive virtual environment simulation system. *Behavior Research Methods, 47*, 296–307.

Johansson, A., Helbing, D., & Shukla, P. K. (2007). Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems, 10*, 271–288.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th edition). Upper Saddle River, NJ: Pearson.

Lapointe, J. F., Savard, P., & Vinson, N. G. (2011). A comparative study of four input devices for desktop virtual walkthroughs. *Computers in Human Behavior, 27*(6), 2186–2191.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Los Angeles: SAGE.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers, 31*(4), 557–564.

Mavrikios, D., Karabatsou, V., Pappas, M., & Chryssolouris, G. (2007). An efficient approach to human motion modeling for the verification of human-centric product design and manufacturing in virtual environments. *Robotics and Computer-Integrated Manufacturing, 23*(5), 533–543.

Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., & Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 1–8.

Olivier, A. H., Bruneau, J., Cirio, G., & Pettré, J. (2014). A virtual reality platform to study crowd behaviors. *Transportation Research Procedia, 2*, 114–122.

Olivier, A. H., Marin, A., Crétual, A., & Pettré, J. (2012). Minimal predicted distance: A common metric for collision avoidance during pairwise interactions between walkers. *Gait & Posture, 36*(3), 399–404.

Riecke, B. E., Bodenheimer, B., McNamara, T. P., Williams, B., Peng, P., & Feuereissen, D. (2010). Do we need to walk for effective virtual reality navigation? Physical rotations alone may suffice. In *Spatial Cognition VII* (pp. 234–247). Berlin Heidelberg: Springer.

Robin, T., Antonini, G., Bierlaire, M., & Cruz, J. (2009). Specification, estimation and validation of a pedestrian walking behavior model. *Transportation Research Part B: Methodological, 43* (1), 36–56.

Ruddle, R. A., Volkova, E., & Bülthoff, H. H. (2013). Learning to walk in virtual reality. *ACM Transactions on Applied Perception (TAP), 10* (2), 11:1–11:16.

Shoulson, A., Marshak, N., Kapadia, M., & Badler, N. I. (2014). ADAPT: The agent development and prototyping testbed. *IEEE Transactions on Visualization and Computer Graphics, 20*(7), 1035–1047.

Teixeira, L., Vilar, E., Duarte, E., Rebelo, F., & da Silva, F. M. (2012). Comparing two types of navigational interfaces for virtual reality. *Work: A Journal of Prevention, Assessment and Rehabilitation, 41*, 2195–2200.

Vinson, N. G., Lapointe, J. F., Parush, A., & Roberts, S. (2012). Cybersickness induced by desktop virtual reality. *Proceedings of the 2012 Graphics Interface Conference*, 1–7.