Comparing Decision-Making and Control for Learning a Virtual Environment: Backseat Drivers Learn Where They are Going

Jonathan Z. Bakdash, Sally A. Linkenauger, and Dennis Proffitt
Department of Psychology, University of Virginia
{jzb3e, sal3g, drp} @ virginia.edu

A considerable amount of research has been conducted on the role interactivity, active versus passive navigation, for learning the spatial layout of a virtual environment (VE). However, active navigation is not unitary. It has two distinct components: decision-making and control. In the present work we investigated which main component of active navigation was critical for acquiring spatial knowledge of a virtual city. We found that spatial knowledge was comparable when the VE was learned with active navigation or decision-making in the absence of control, but was much worse when only control was present. These results suggest decision-making, not control, is the critical component for learning a VE.

Key Words: navigation, frames of reference, spatial updating, virtual environments

INTRODUCTION

There has been a substantial amount of research on spatial learning in virtual environments (VEs) comparing active versus passive navigation (Péruch & Wilson, 2004; Sun, Chan, & Campos, 2004; P. Wilson & Péruch, 2002; P. N. Wilson, 1999). In a typical study on this issue, one participant has active control over navigation whereas the other participant only observes the outcome of the actions. This situation is analogous to a car driver and a passive passenger. Missing from these studies is an appreciation that active navigation is not unitary; it has two distinct components: decision-making and control. For example, imagine that the driver of a car has a passenger who is using a map and directing the driver where and when to turn. The driver has control over the car, but the passenger (or “backseat driver”) is the one making the decisions about where to go. Which one of these individuals learns the environment better? The growing use of GPS navigation systems makes examining the largely unknown distinction between decision-making and control an important question.

Decision-making and control differ in a number of ways. Decisions are processes directed towards achieving an underlying distal goal, such as traveling from home to work. Control consists of more immediate, proximal goals, like avoiding obstacles and staying on the road. This distinction can be illustrated with semi-autonomous robots, where a human operator specifies a goal, such as having the robot travel to a distant location. However, with time delay and other problems associated with remote operation it makes sense to have the robot use control heuristics to execute immediate actions in the environment, such as avoiding a rock blocking its path. These complex, immediate actions can be successfully implemented without a representation of spatial layout through the use of visual control heuristics. For example, a car can be stopped in time to avoid hitting an object by keeping the rate of expansion of the object’s visual angle under a certain value (Lee, 1976). Thus, decision-making requires a representation of the environments global spatial layout, whereas control can be achieved using local information and control heuristics that do not need global spatial information.

Empirical and anecdotal evidence suggest that a backseat driver may acquire better spatial knowledge than a driver who only has control. Burnett and Lee (2005) used a virtual driving simulator to compare spatial knowledge for participants that used a paper map to plan their routes versus the use of a verbal navigation system. They found participants that used the navigation system drew less detailed maps, with fewer landmarks. Brooks (2007) commented that after using a GPS system for a few weeks, he was “… quickly shedding all vestiges of geographical knowledge.” Given the increasing popularity of GPS navigation systems (Rothman, 2007), the role of decision-making in acquiring spatial knowledge is a notable applied issue.

In the current research, we separated decision-making and control to assess their contribution in acquiring spatial knowledge of a virtual city environment. Three conditions were employed: active (decision-making and control), decider (decision-making alone), and controller (control alone). Visual information between the decider and controller conditions was comparable: see the procedure for details. Participants spent 20 minutes learning the VE which was displayed on a large screen projector. Then, they transferred to a head-mounted (HMD) virtual reality which displayed the same VE they had just learned. Using the HMD, spatial knowledge was measured by having participants point at unseen target locations, which is called spatial updating (Montello, Richardson, Hegarty, & Provenza, 1999; Rieser, 1989). Finally, participants placed the target locations on a blank map.

METHOD

Participants

Sixty-seven (34 male, 33 female) University of Virginia students and members of the Charlottesville community participated in this experiment. Participants were either paid $20 or received course credit for their participation.
Equipment and Virtual Environment

Alice99 was used to create and render the virtual city for this experiment in a first-person perspective, see Figure 1. This is the same virtual city that was used in an earlier navigation study (Bakdash, Augustyn, & Proffitt, 2006)). The VE consisted of streets laid out in an irregular grid and measured about 150 meters by 200 meters in size. Five target objects (gazebo, tank, school, helicopter, and humvee) were situated in the environment so that only one was visible simultaneously.

Figure 1. Example screenshot of the virtual city from the same viewpoint seen by participants.

Practice and learning phases. For the active and controller conditions, participants used a Saitek Cyborg EVO joystick to control their movement through the VE. Heading direction was adjusted by moving the joystick and the throttle of the joystick controlled walking speed. In the decision-making condition the experimenter used the joystick to control movement, following verbal directions from the participant.

The VE was rendered at 640 by 480 at 60 frames per second using a Dell Dimension 8250 computer equipped with a GeForce Ti 4200 graphics card. The virtual city was displayed on a DA-LITE screen using a Sharp Notevision 6 projector, with an image size of 109.22 centimeters (width) by 147.32 centimeters (height). Participants sat about 5.28 meters from the screen.

Testing phase. During the testing phase, a Virtual Research V8 HMD was used to view the same VE rendered from a first-person perspective with Alice99. Using a Dell Precision 360 computer equipped with a GeForce 4 MX420 and GeForce 4 MX200, the HMD displayed stereo images at 640 by 480 at 60 frames per second with a horizontal field of view of 48°. Head movements were registered to update the VE seen through the HMD using an Intersense IS-900 motion tracking system. Participants rotated in place and used a tracked wand to point at target locations that were out of sight. Angular pointing error was measured as a function of the deviation of the center of mass from the target being pointed at, not taking into a possible difference in elevation. Higher pointing error indicates less accurate responses to target locations, see Figure 2 for a hypothetical example of pointing error. Pointing accuracy was assessed from the angle formed between a perfect pointing response and the actual pointing response.

Figure 2. Hypothetical example of pointing error from above. This view was never seen by participants. The red dot represents the standing location at the helicopter. A perfect pointing response (0°) to the gazebo is shown with the green line. The yellow dotted line shows a pointing error of 25°.

At the end of the study, participants used a program to place target locations on a blank map displayed on a small computer monitor. The map program recorded the locations of target placements in (x, y) pixel coordinates.

Design

There were three conditions in this study:
1) Active: decision-making and control.
2) Decider: decision-making only (i.e. “backseat driver”), participants instructed the experimenter where to go.
3) Controller: participants only had joystick control, the experimenter instructed participants where to go.

Each participant in the controller condition received comparable visual information to a participant in the decider condition. This was done by having the experimenter watch a playback of a decision-making participant while verbally instructing the control participant where to travel. Thus, controllers and deciders were yoked. Since the primary focus in the present work was to ascertain which component of active navigation was important for learning a VE, no passive viewing condition was presented.

A between-subjects design was used for the three conditions (active, decider, and controller). For the learning phase, the starting target location was randomized, except for the controller condition which was matched to the decider condition. During the testing phase, the order of targets to be pointed to was randomized. Since constantly jumping around to different target locations after each trial would be disconcerting, the randomization was conditional on having all of the pointing responses for a given location that was being stood at occur consecutively.
**Procedure**

Participants were randomly assigned to groups, with the limitation that there had to be a matched participant in the decider condition available for each controller participant. In the active condition there were (11 male, 9 female), decider (10 male, 10 female), and controller (10 male, 10 female).

To assess for spatial and navigation abilities, participants first completed the Santa Barbara Sense of Direction scale (SBSOD) (Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002). Also, level of video game experience was measured by having participants report their experience with first-person shooters (e.g. Halo, Unreal Tournament) using a one to seven point likert scale and the average number of hours they spend per week playing such games. After completing these questionnaires, participants were told to find and learn the locations of the five targets in the virtual city environment and that they would have 20 minutes to do so. Finally, participants were instructed their knowledge of the virtual city would be assessed by having them stand at target locations and pointing at the ones that were out of sight. The real-world example given to participants was asking them to point to the Rotunda (a salient, well-known landmark at the University of Virginia), which was not visible from the experiment room. The total experiment took approximately one hour to complete.

**Practice phase.**

**Active condition.** First, participants were instructed on using the joystick and then they practiced moving for about one minute, in a VE created for this purpose. The practice VE was similar in appearance to the actual city environment, but was much smaller and did not contain any target locations. Control proficiency was ascertained by having participants walk around a city block, in the practice VE, in under 45 seconds. Two participants were unable to circle the block in the allotted time on the first try. They were given additional practice with the controls and were then able to pass the criterion set for control proficiency.

**Controller and decider conditions.** In the controller condition, the procedure was the same as the active condition with one addition. The experimenter gave verbal directions (e.g. turn left at the next intersection) in the practice VE to ensure the participant would understand where to go. Two participants needed extra practice to attain control proficiency. Contrary to the other two conditions, participants in the decider condition did not receive any training or practice with the joystick controls. However, they did receive a brief training session giving the experimenter verbal directions about where to go in the practice VE (e.g. keep going straight, stop and turn around).

**Learning phase.** For the active and decider conditions, 20 minutes were given to free-explore the virtual city and participants were told to use whatever strategy they wanted to learn the locations of the five targets. The 20 minute learning time was sufficient for every target location to be visited at least once. In the active condition, participants had both decision-making and joystick control, whereas in the decider condition only the ability to choose where to go was available.

Controllers used the joystick but were instructed where to go by the experimenter, who was following the trajectory of a decider. This made visual information between controllers and deciders comparable.

Trajectory data from each participant in the decider condition was recorded. The experimenter then played back the path each decider participants followed on a Dell Latitude C610 laptop giving verbal directions to controller participants so they followed the same path. The experimenter sat next to each controller participant, but in order to keep participants attention on the projected screen and controlling their movement, the laptop screen was tilted away from their viewpoint.

**Testing phase.** The test phase was identical for all three conditions.

**Pointing error.** Using VR, participants stood at each of the five target locations and pointed at the other four unseen locations using the wand. For pointing, participants were asked to imagine the virtual pointer that extended from at the end of the wand go straight through buildings and other objects to directly hit the target they were pointing at. Pointing responses were indicated by pressing a button on the bottom of the wand which also changed the color of the virtual pointer. Participants were asked to hold the wand steady and quickly click the button for pointing. However, if there was more than a 2° disparity between the initial button press and release they were told to repeat their pointing response.

**Map construction task.** Participants were told to place all five of the target locations on the blank map displayed on a computer screen. They were further instructed that the orientation and scale of the constructed map did not matter; targets should be placed wherever they thought they belonged relative to each other.

**RESULTS**

Due to motion sickness, five participants were unable to complete the experiment. Also, one participant appeared intoxicated and was not capable of finishing the study. Finally, a malfunction with the VR equipment prevented completion of the study for one participant. All seven participants were excluded from analysis and replaced.

Participants in the active and decider conditions had comparable pointing error and map construction accuracy; whereas participants in the controller condition had higher pointing error and constructed less accurate maps, see Figures 3 and 4. These results indicate that having the ability to make decisions about where to go, not control, is the critical component for learning the spatial layout of a virtual environment. Furthermore, visual experience by participants in the decider and controller conditions was comparable.

**SBSOD and video game experience**

Since experience with first-person shooters and hours spent playing these games were reliably correlated ($r = .69$), a composite video game experience variable was created by averaging $z$-scores from these two questions. The SBSOD and the composite video game experience variable were assessed
using separate two-way ANOVAs with 3 (condition) by 2 (sex) both specified as between-subjects factors.

No main effect was found for sex on the SBSOD, \( p = .10 \). By condition, no differences were found for the SBSOD, \( p > .62 \). Also, no main effect of sex was present for the composite measure of videogame experience, \( p = .77 \). However, using the composite videogame experience variable, males (\( M = 1.09, SE = .32 \)) were higher than females (\( M = -1.16, SE = .13 \)), \( p < .001 \), \( F(1, 54) = 40.72, p < .001, \eta^2_p = .43 \). No condition by sex interactions were observed for the SBSOD or the video game experience, \( ps > .5 \). These results indicate equality in the SBSOD and video game experience across conditions.

**Testing phase: Pointing error**

Angular pointing error was analyzed using absolute values, which provides an overall measure accuracy. Absolute error was assessed using a 3 (condition) x 2 (sex) x 5 (target stood) x 4 (target pointed) repeated measures ANOVA. The first two factors specified as between-subjects variables and the last two factors as between-subjects variables. The most important finding was a main effect of condition, \( F(2, 54) = 4.51, p < .001, \eta^2_p = .14 \), show above in Figure 3. Planned comparisons indicated that pointing error was significantly larger for the controller condition than the active and decider conditions, \( ps < .02 \), and there was no reliable difference between the active and decider conditions, \( p = .82 \).

In addition, there was a significant main effect of sex, \( F(1, 54) = 33.09, p < .001, \eta^2_p = .38 \), males had lower pointing error (\( M = 19.67^\circ, SE = 3.33^\circ \)) than females (\( M = 46.93^\circ, SE = 3.55^\circ \)). No interaction was observed between sex and condition, \( p = .96 \).

There were also reliable main effects for target stood and target pointed, \( ps < .001 \). This reflects less accurate responses when standing at locations closer to the center of the VE, as there was nothing to bound responses. Similarly, greater pointing error was observed for pointing responses to target locations away from the edge of the virtual city. Higher order interactions were not meaningful and therefore were not reported.

Since pointing error is a cyclical measure, its components can be evaluated with unit vectors using circular statistics (Batschelet, 1981). Two circular measures can be derived:

1. **Constant error**: systematic directional bias.
2. **Variable error**: deviations independent of bias.

Both these components were found in our data. There was a high positive correlation between the absolute pointing error and variable pointing error, \( r(60) = .94, p < .001 \), one-tailed. This indicates minimal constant bias in the pointing data, making variable and absolute error nearly identical. Therefore, only absolute pointing error was reported.

**Testing phase: Map construction accuracy**

A bidimensional regression (BDR) (Friedman & Kohler, 2003; Tobler, 1994) was used to assess the overall accuracy of map construction. BDR fits a solution between two sets of (x, y) coordinates that minimizes the difference. This analysis is scale and rotation invariant and computes a measure of similarity, \( r \), which is the BDR equivalent of a correlation coefficient. Using the supplemental materials from (Friedman & Kohler, 2003), a four parameter Euclidean BDR was used to calculate map construction accuracy. This was done by computing the \( r \) value for each participant, by comparing their constructed map to the configuration of actual target locations. Since the distribution of the \( r \) values was negatively skewed, a Fisher Z-transform was applied to the data to make it more normal. The transformed values were evaluated in a two-way ANOVA, 3 (condition) x 2 (sex). A main effect of condition was found, \( F(2, 54) = 7.30, p < .001, \eta^2_p = .21 \), see Figure 4. Like pointing error, planned comparisons indicated less accurate maps for the controller condition compared to the active and decider conditions, \( ps < .01 \), and no difference between the active and decider conditions, \( p = .81 \).

In addition, there was a main effect of sex, \( F(1, 54) = 15.31, p < .001, \eta^2_p = .22 \), males constructing more accurate maps (\( r = .88, SE = .05 \)) than females (\( r = .64, SE = .05 \)).
Our findings demonstrate that the ability to decide where to go in a VE, not control, is the critical component for learning the environment. Despite the decider and control group experiencing similar visual information, the deciders acquired superior spatial knowledge of the virtual city. In fact, decision-making alone resulted in equivalent spatial knowledge for both decision-making and control (active condition).

We propose that the means of representing space through frames of reference differed between the decider and controller conditions. An egocentric reference frame is viewer-centered (self relative to object), whereas an exocentric (or allocentric) frame of reference is environment-centered (object to object) (Klatzky, 1998). Goal planning requires knowledge of both the observer’s current location and the spatial relationships between locations in the environment, requiring egocentric and exocentric reference frames. This egocentric representation of space creates a link between the body and the environment. However, action execution only requires an egocentric perspective, not a spatial representation of the environment. Making decisions is interactive, it has consequences, and therefore it creates a relationship between the body and the environment. Whereas, performing instructed actions is reactive, does not require a representation, and there are no consequences which means no link between the body and the environment. The notion that cognitive processes are related to bodily states and interactions with the environment is called embodied cognition (M. Wilson, 2002). In the present work, spatial encoding with control may have been deficient because control alone did not facilitate an embodied representation of the environment. Additionally, executing spatial verbal tasks have been shown to cause interference in spatial memory (Salthouse, 1974). Our results suggest that decision-making and control, perhaps via embodiment, affect the use and selection of reference frames which determines how the spatial layout of the virtual city is encoded.

The results of this work imply that using a GPS navigation system may impede acquiring spatial knowledge. This possibility is bolstered by the earlier mentioned virtual driving study from Burnett and Lee (2005). It is possible these findings extend to real-world navigation where the malfunction of a GPS system could be catastrophic in life or death situations, such as for law enforcement or the military. One potential way of ameliorating some of the deleterious effects of control may be to present navigation instructions visually in addition to or perhaps instead of verbally. In summary, we found that decision-making, not control, is the most important active navigation component for learning a VE.

ACKNOWLEDGEMENTS

This work was supported by grant W911QY-08-P-0030 from the U.S. Army Natick Soldier Research, Development, and Engineering Center to the third author. The authors would like to thank Jeanine Stefanucci and Jessi Witt for their contributions, Steve Jacquot for his amazing programming skills, and the many research assistants that helped with data collection.

REFERENCES


Wilson, P. N. (1999). Active exploration of a virtual environment does not promote orientation or memory for objects. Environment and Behavior, 31(6), 752.