arXiv:2003.03133v1 [cs.HC] 6 Mar 2020

DeFINE: Delayed Feedback based Immersive Navigation Environment for Studying

Goal-Directed Human Navigation

Kshitij Tiwari and Ville Kyrki

Aalto University, Finland

Allen Cheung

The University of Queensland, Australia

Naohide Yamamoto

Queensland University of Technology, Australia

Author Note

Kshitij Tiwari and Ville Kyrki, Department of Electrical Engineering and Automation, Aalto University, Finland; Allen Cheung, Queensland Brain Institute, The University of Queensland, Australia; Naohide Yamamoto, School of Psychology and Counselling and Institute of Health and Biomedical Innovation, Queensland University of Technology, Australia.

Correspondence concerning this article should be addressed to Kshitij Tiwari, Department of Electrical Engineering and Automation, Aalto University, Finland. E-mail: kshitij.tiwari@aalto.fi

Abstract

With the advent of consumer-grade products for presenting an immersive virtual environment (VE), there is a growing interest in utilizing VEs for testing human navigation behavior. However, preparing a VE still requires a high level of technical expertise in computer graphics and virtual reality, posing a significant hurdle to embracing the emerging technology. To address this issue, this paper presents Delayed Feedback based Immersive Navigation Environment (DeFINE), a framework that allows for easy creation and administration of navigation tasks within customizable VEs via intuitive graphical user interfaces and simple settings files. Importantly, DeFINE has a built-in capability to provide performance feedback to participants during an experiment, a feature that is critically missing in other similar frameworks. To demonstrate the usability of DeFINE from both experimentalists' and participants' perspectives, a case study was conducted in which participants navigated to a hidden goal location with feedback that differentially weighted speed and accuracy of their responses. In addition, the participants evaluated DeFINE in terms of its ease of use, required workload, and proneness to induce cybersickness. Results showed that the participants' navigation performance was affected differently by the types of feedback they received, and they rated DeFINE highly in the evaluations, validating DeFINE's architecture for investigating human navigation in VEs. With its rich out-of-the-box functionality and great customizability due to open-source licensing, DeFINE makes VEs significantly more accessible to many researchers.

Keywords: Virtual reality, Software, Speed, Accuracy, Closed-loop

DeFINE: Delayed Feedback based Immersive Navigation Environment for Studying

Goal-Directed Human Navigation

Behavioral researchers are increasingly becoming interested in understanding the underlying mechanisms for goal-directed navigation in humans (Cornwell, Johnson, Holroyd, Carver, & Grillon, 2008; Pezzulo, van der Meer, Lansink, & Pennartz, 2014; Spiers & Maguire, 2006). Whilst it would be ideal to gather data on navigation behavior in a well-controlled real-world setting, not every researcher has access to such a setup. However, with the advances in virtual reality (VR), it is becoming easier to set up economical game-like environments to test hypotheses about computational principles which could explain human navigational behaviors.

When it comes to rodents and bats, a rapidly growing body of in-vivo neuronal response studies provides deeper insights for simple goal-directed navigation tasks (Ainge & Langston, 2012; Geva-Sagiv, Las, Yovel, & Ulanovsky, 2015; Hollup, Kjelstrup, Hoff, Moser, & Moser, 2001; Koene, Gorchetchnikov, Cannon, & Hasselmo, 2003; Yartsev & Ulanovsky, 2013). However, well-defined computational models that can unify neuronal responses and animal behaviors are largely missing. In contrast, for humans, research on neuronal responses during navigation is developing but still limited (Ekstrom et al., 2003; Ekstrom, Spiers, Bohbot, & Rosenbaum, 2018; Jacobs et al., 2013), creating a large role for behavioral investigations to play in interrogating neural computation for navigation (X. Chen, He, Kelly, Fiete, & McNamara, 2015; Harris & Wolbers, 2012; Mokrisova et al., 2016; Yamamoto, Philbeck, et al., 2014). Thus, using VR for navigation research is not just promising but also crucial for advancing our understanding of neural mechanisms of human navigation. Furthermore, the use of VR in navigation research should not be limited to humans; applying it to animal research can provide new insights into the relationship between spatial navigation behavior and neuronal response patterns (Acharya, Aghajan, Vuong, Moore, & Mehta, 2016; G. Chen, King, Lu, Cacucci, & Burgess, 2018; Doeller, Barry, & Burgess, 2010; Ravassard et al., 2013). It also has potential to facilitate cross-species comparisons between humans and animals by making it easy to test different species in the same or equivalent environments (Commins et al., 2019; Machado et al., 2019).

Previous attempts have been made at developing easy-to-use VR frameworks for researchers who are novice when it comes to graphics designing or using game engines like Unity3D. Brookes, Warburton, Alghadier, Mon-Williams, and Mushtaq (2019) developed the Unity Experiment Framework (UXF) that allows researchers to set up rich three-dimensional (3D) visual scenes and log response to visual stimuli. Similarly, Vasser et al. (2017) presented a framework called VREX, which allows for easy creation of indoor environments with single or multiple connected rooms populated with furniture. Another notable line of recent work is development of virtual environments (VEs) that are meant to translate standard maze environments in animal research (e.g., T-maze and Morris water maze; Morris, Garrud, Rawlins, & O'Keefe, 1982; Packard & McGaugh, 1996) to human navigation studies (Commins et al., 2019; Machado et al., 2019; Wiener et al., 2019)

These frameworks solely depend on the stimulus–response causal relationship. That is, participants are presented with a stimulus with which they carry out a navigation response (e.g., walking to a goal location indicated by visual cues), and during and at the completion of the response, they do not receive any feedback on their performance. Although such a research design is appropriate and even required for investigating certain aspects of navigation behavior (J. Loomis & Knapp, 2003; Philbeck & Loomis, 1997), it makes it impossible to examine how participants modulate their subsequent response to the stimulus using feedback (Brand, 2008).

Importantly, when goal-directed navigation is performed in real-world settings, navigators often receive feedback with which they can adjust their behavior. For example, when they walk in the dark to reach a door and fail to touch its knob, the lack of tactile sensation serves as feedback and informs them that they still have a few more steps to go. In this instance, the human navigation behavior should be characterized as consisting of a closed stimulus–response–feedback loop, instead of an open stimulus–response loop as in the foregoing frameworks. Indeed, studies have been conducted to capture human navigation by the stimulus–response–feedback loop. For instance, Carton, Nitsch, Meinzer, and Wollherr (2016) have found evidence that humans tend to adapt their trajectory planning horizons when they detect potential collision situations. In this case, impending collisions constitute a

stimulus that triggers a response, which is witnessed in the form of a shortened planning horizon. The successful avoidance of collision or a failure thereof functions as feedback that can be used to tune the length of planning horizons in the future. In this manner, experiments on human navigation come in a variety of designs, utilizing both stimulus–response and stimulus–response–feedback loops. To fully accommodate this diversity of the experiments, there is a need for a new VR framework that allows for incorporation of feedback into the experimental design. To this end, we developed the Delayed Feedback based Immersive Navigation Environment (DeFINE).

Delayed Feedback based Immersive Navigation Environment

DeFINE is based on the Unity3D game engine, and hence, relies heavily on C# as a programming language. All the low-level implementation is already taken care of to minimize the workload of the end-users who can simply use the modular components to either modify the existing settings or customize them as required by the experimental design. DeFINE aims specifically to provide an easy-to-use experimental environment that is based on the stimulus–response–feedback architecture, which can be used to study goal-directed spatial navigation in humans. In order to reduce the burden of researchers when it comes to setting up an experiment in DeFINE, we provide short video demonstrations that succinctly explain how to use the basic functionality of DeFINE "out-of-the-box" (https://youtu.be/OVYiSHygye0) and also how to change various elements of DeFINE to suit the researchers' particular needs using the Unity software (https://youtu.be/smIp5n9kyAM). A detailed user manual is also available (https://gitlab.com/aalto-qut/environment/blob/master/user_manual.pdf). DeFINE is being released open-source under the MIT license and is freely available to download via Gitlab at https://gitlab.com/aalto-qut/environment.

Currently, DeFINE can be integrated into Unity tasks built for Windows personal computers. It is assumed that DeFINE will be used with a head-mounted display (HMD) such as HTC Vive and Oculus Rift. For example, DeFINE is designed to utilize the HMD's motion tracking sensors for implementing various methods of participants' locomotion within VEs

(see the locomotion methods section below for details). In addition to the HMD worn by participants, DeFINE simultaneously presents a VE to a desktop display so that experimentalists can monitor the progress of an experiment. Further details about hardware and software requirements for DeFINE are available in the user manual.

The main capabilities and options of DeFINE are detailed below in the following order: (1) the generic experimental structure, (2) time- and accuracy-based feedback, (3) the graphical user interface (GUI), (4) DeFINE's diverse suite of locomotion methods, (5) static and dynamic goals, (6) performance leader-board, and (7) intra-VR surveys.

Experiment Structure

Human behavioral experiments are often defined by a *trial–block–session* architecture which allows the experimentalists to repeat a task multiple times to acquire requisite data (Figure 1). Just like the UXF (Brookes et al., 2019), DeFINE adopts this architecture. A *trial* is an instance of the experiment where the participant is presented with a stimulus and their response is recorded. At the end of the trial, the participant receives feedback. To clarify that the feedback is given after the response is made, as opposed to during the trial as the response unfolds, this feedback is referred to as "delayed" in DeFINE. Trials are often repeated multiple times for various purposes (e.g., to measure variability of responses to the same stimulus, to decipher the learning effects over trials, or to train the participants on the task), constituting a *block*. At the end of the block, the participant is assumed to be familiar with the environment and to have formulated a behavior of choice for the task at hand. In order to evaluate the quality of this behavior, the experimentalist may choose to make some modifications to the environment before proceeding with the next block. The experiment can consist of a single or multiple blocks, and when there are multiple blocks, a single iteration of the task over the blocks is called a *session*.

DeFINE has primarily two levels of abstraction as shown by the two columns (demarcated by orange dotted lines) in Figure 2. All modules mentioned on the left are those implemented at the low-level abstraction that come pre-programmed with DeFINE while those on the right are implemented at a higher level for easy and quick modifications. Such an
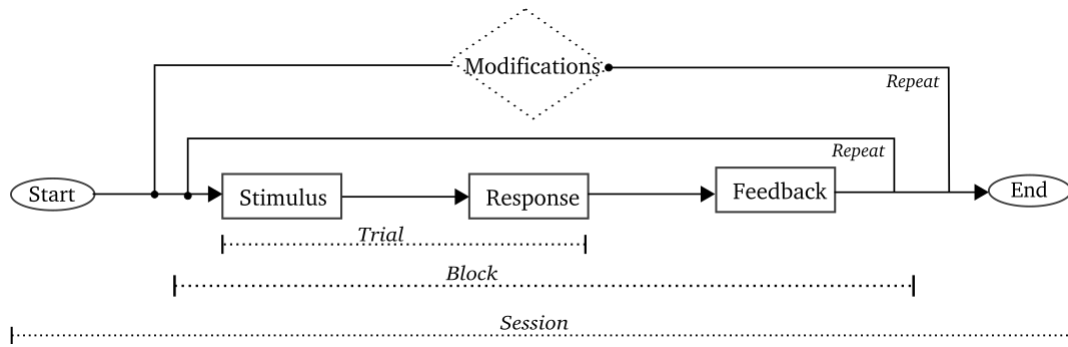
Figure 1. Structure of typical human spatial navigation experiments. Participants are presented with a *stimulus*, to which they provide a *response* and are given *feedback* after the *trial* has been fully executed. This *feedback* is then used to modify performance in the next trial. Usually, multiple trials are carried out under the same condition, which constitute a *block*. Once the participants are fully familiar with the task and the condition, they progress to the next block which may involve modifications of the condition, and their navigation performance is evaluated again over multiple trials. A single iteration of the task is called a *session*.

arrangement streamlines customization of an experiment to suit different research studies because the modules at the low level of abstraction are common across many experiments and thus, the experimentalists can rely on DeFINE's built-in functionality, which in turn allows them to focus their effort on customizing high-level modules that tend to be unique to each experiment. All modules shown by dotted black lines represent optional modules that can be utilized as seen fit by the experimentalist. In keeping with the trial–block–session architecture as described in Figure 1, all modules clustered under the green polygon ("Begin Session") represent the session, those under the red polygon ("Begin Block") represent the blocks, and those under the purple polygon ("Begin Trial") represent the trials. All other modules (under either column of abstraction) are preset before the session starts.

The flow of the experiment starts from the top of Figure 2. DeFINE takes care of initializing and rendering the immersive environment and setting up the session, block and trials, but the experimentalist is required to configure the VE (e.g., setting its size) and specify the experimental design (e.g., the number of trials per block) by modifying the settings files in order to make DeFINE behave in the way they need for their experiments. Details of the
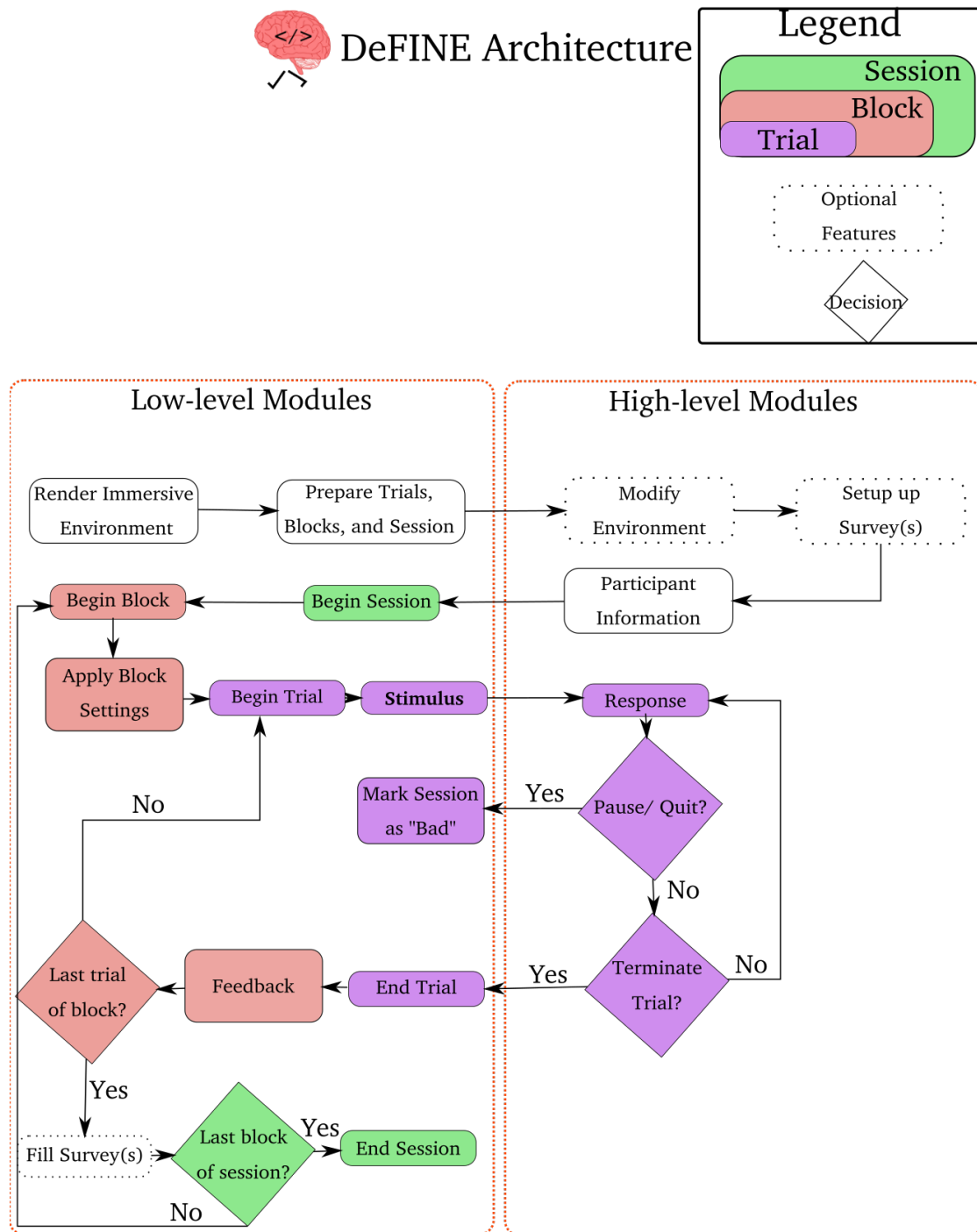
Figure 2. The architecture of the Delayed Feedback based Immersive Navigation Environment (DeFINE). The left panel shows the low-level functionality provided by the environment while the high-level experimentalist defined functionalities are shown on the right. The surveys/questionnaires are optional, and hence, shown using dashed lines. Modules shown in purple are parsed during a *trial*, those in red during a *block* and those in green during a *session*.

settings files are described later. If the experimentalist wants the questionnaires to be filled within the environment, these questionnaires need to be created online, and the respective links to the questionnaires need to be added to the environment settings. Setting up the questionnaires is followed by inputting the participant's information and selecting preset settings that are appropriate for the experiment. Once the participant information has been entered, the experimentalist can start the experiment session that DeFINE will generate. If there are block-specific settings, those are applied by DeFINE before starting the first trial. The trials follow the stimulus–response–feedback structure, by showing the participant the stimulus in the environment, to which the participant responds, and after the experimentalist-defined end condition has been met, the trial ends and feedback is shown to the participant. A new trial is started automatically after the feedback has been given, until the program reaches the end of the block. At the end of the block, the participant will be shown the questionnaires, if any were specified during setup. During the trials the experimentalist can take down notes, or manually mark the session as bad to indicate that it should not be taken into account for data analysis. At any point of the session the experimentalist can abort the session, which will also mark the session as bad in the stored session notes. If there are more blocks remaining in the session, the next block is started after DeFINE has applied its specific settings to the environment. After the final block of trials has been completed, and the questionnaires following it have been answered, the session ends and the participant and experimentalist are returned to the startup screen. For each of the trials executed with DeFINE , the participant's trajectory during the trial, the status and changes of environment variables, the score obtained at the end of the trial (see the next section for details), trial start time, trial end time, total time taken for the trial, and straight-line distance to the goal when the participant ends the response are logged.

Feedback

As opposed to other closed-loop systems where real-time feedback may be made available to participants as they carry out a response, DeFINE provides the delayed feedback at the end of each trial. This is because continuous feedback will most likely go against the

purpose of typical goal-directed navigation experiments. Usually, these experiments are to test participants' ability to estimate their location relative to the goal by using sensory and other spatial cues in the environment (J. M. Loomis & Philbeck, 2008). If the participants were given external feedback about whether they were moving in the right direction for every step they took, this feedback would essentially be a non-spatial cue that would directly aid them in their location estimation. In extreme cases, with continuous feedback, the participants could perform the task by moving in an arbitrary direction without processing the sensory and spatial cues and seeing if that would result in positive feedback. Such a strategy would lead them to take myopic unstructured paths to a goal, causing non-optimal navigation performance. Thus, by default, DeFINE is designed to give performance feedback only after the trial is completed. However, it is possible for experimentalists to modify DeFINE's source code in Unity and have it provide real-time feedback, if they so choose.

Feedback on a goal-directed navigation behavior can be given in a number of different forms, but DeFINE adopts a reward/cost function that evaluates participants' performance and provides feedback as gains and losses of scores. It has been shown that feedback of this type is very effective in affecting participants' behavior and decision making under a variety of conditions (Brand, 2008; Hossain & List, 2012; Yechiam & Hochman, 2014). The reward and cost in the context of navigation can also be defined in various ways, and it is up to experimentalists' discretion how they formulate the reward/cost function in DeFINE, but one straightforward method would be to define them by using speed and accuracy of navigation behavior. That is, the quicker the participants are in performing a trial, the greater the reward (or the smaller the cost); and the more accurate they are in reaching the goal, the greater the reward (or the smaller the cost). By default, DeFINE implements a reward/cost function of this form. Specifically:

$$R = \beta_1 \exp(-\alpha_1 t) + \beta_2 \exp(-\alpha_2 d) \tag{1}$$

In Equation (1), t refers to the time taken for the navigation towards the goal, and d refers to the residual straight-line distance to the goal from the location at which the participant ended the trial. β_1 and β_2 are weights used for combining the time and distance into the decaying reward function, which penalizes both the time taken and the residual
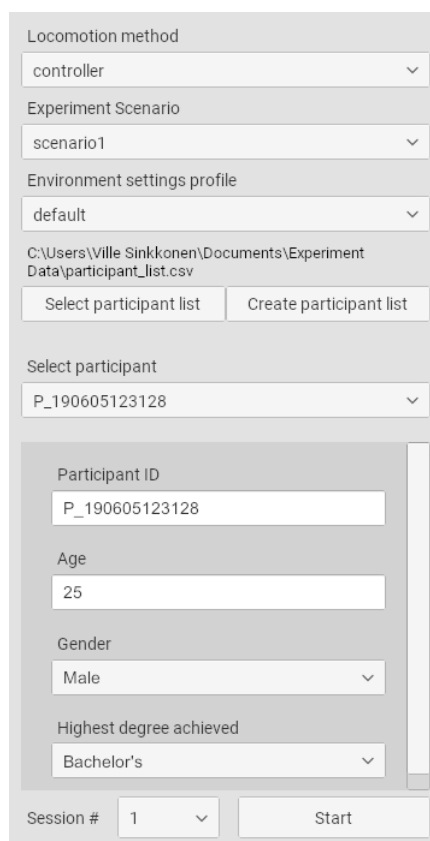
distance to the goal (i.e., rewarding shorter times and smaller distances with higher scores).

α_1 and α_2 are factors for scaling the effects of time and distance. If experimentalists choose to

use this function in their experiments, they can assign values of their choice to these

parameters simply by specifying them in a settings file (details shown later). If they are to

calculate the reward/cost scores using their own equation, they can do so by modifying a

relevant section of DeFINE's codes in Unity. It should further be noted that, by changing the

relevant codes and implementing their own equation, the experimentalists can use *any* kinds of

feedback that do not take the form of a cost/reward function. For example, it is possible to

simply present how far away participants are from the goal at the end of a trial.

Graphical User Interface (GUI)

Utilizing the GUI of UXF (Brookes et al., 2019), DeFINE allows the experimentalist to

log participant information including, but not limited to, name (or participant identification),

age, gender, and educational qualification (Figure 3a). Should other personal particulars be

required as per the experimental design, they can be easily added to the framework by

appropriately modifying the settings files. As an extension to UXF's original GUI,

DeFINE allows the experimentalist to quickly set up the environment of choice with the

desired locomotion method (see the next section for details about the locomotion methods).

This unique feature can also be scaled and automated to handle multiple combinations of

environments and locomotion methods via DeFINE's auto-pilot mode. In this mode, the

experimentalist can provide DeFINE with preset instructions so that it loads specific

combinations of the environment and the locomotion method in a specified order. This way, a

sequence of participants can be tested automatically, doing away with the need to individually

set up an appropriate combination of the environment and the locomotion method for every

participant. For example, if an experimental design requires that each participant be shown a

different environment, a sequence of environments can be explicitly listed in the settings files

which will then be autonomously parsed when executing the auto-pilot mode. Similarly, if

each participant is to do trials with a different locomotion method, explicit

participant-locomotion method combinations can be listed in the settings files in a similar

fashion.

As a significant extension to the predecessor, UXF, DeFINE also provides

functionalities to study the role of lighting conditions and auditory cues in spatial navigation.

At any point during the trial, the experimentalist can toggle the sound and lights of the VE on

or off by clicking on the dedicated buttons on the user interface, shown in Figure 3b. The

change of the status of these environmental variables are logged along with the information

about participants' performance in a navigation task (e.g., their position within the

environment at a given time point).

(a) Drop-down menus to specify experimental
parameters.

(b) The buttons for toggling sound and lights.

Figure 3. DeFINE's intuitive GUI for the experimentalist to set up an experiment.

Locomotion Methods

In order to provide a locomotion suite for participants to perform goal-directed

navigation in VR, DeFINE comes equipped with a variety of locomotion methods.

Teleoperated locomotion. In order to allow teleoperated locomotion, DeFINE is compatible with both a keyboard based and the VR controller based teleoperated methods. A typical use case may involve head direction sensors in an HMD being used to update participants' headings in the VE while the keyboard or the joystick controller being used to linearly traverse at preset velocity. The necessary key-bindings and further details are available in the user manual

(https://gitlab.com/aalto-qut/environment/blob/master/user_manual.pdf).

Arm-swing locomotion. Arm-swing locomotion is an implementation of walking-in-place locomotion. In this method, the participants walk in place, including swinging their arms in a manner consistent with their pace of walking. It has been shown that such arm swings are effective in having participants experience a naturalistic sense of locomotion without actually moving in real space (Kunz, Creem-Regehr, & Thompson, 2009; Yamamoto, Mach, Philbeck, & Van Pelt, 2018). This locomotion method uses the physical movement of the controller(s), held by the participant, to determine forward speed in the VE. The movement speed is calculated from the positional difference of the tracked controller(s) between two consecutive frames. This calculation can be done either by requiring movement of two controllers (typically, one in each hand) or by using one controller (or either of the two controllers) that moves more than a given threshold amount between the frames. When the use of the two controllers is required, the forward speed in the environment is set to be zero, unless both of the controllers exceed the threshold value.

Head-bob locomotion. Head-bob locomotion is another implementation of walking-in-place locomotion. In order to move forward in the VE, the participants need to walk in place, and as they do, their head, and in particular, the HMD, bobs slightly vertically. This locomotion method uses this vertical *bobbing* to determine the forward velocity. DeFINE tracks the vertical direction of the bobbing and its starting position. Once the direction changes, the participant is considered to have stepped, if the vertical height difference between two successive flexion points exceeds a threshold value specified in the settings of the locomotion method. The detected physical step is then translated into a step in the VE so that the participants walk forward in the VE at a preset velocity (to be set in a

settings file by the experimentalist). Due to the fact that the HMD is in front of the participant's face, turning their head up or down causes the HMD to move vertically. In order to avoid reading these vertical movements as steps of the participant, DeFINE also tracks the participant's rotational head movements about the pitch axis and ignores any "bobs" that are accompanied with the rotational head movements that exceed a specified threshold value.

Physical walking. Physical walking is the only locomotion method in which the participant is expected to physically move around in the real world. The movement of the participant is tracked by using an HMD's motion tracking sensors and the participant's position in the VE is updated accordingly. Owing to the limited size of the physical area in which the participant's movement can be tracked (which is typically up to 10×10 m), the size of the VE is going to be limited. To alleviate this limitation, sometimes modified physical walking methods such as redirected walking are adapted (Nilsson et al., 2018; Paludan et al., 2016). In these methods, the rotations and translations of the participant are slightly altered between physical and virtual worlds in order to steer the participant away from the edges of the available physical area. However, DeFINE does not utilize these methods because they can induce disruption to mental and neural spatial representations as well as to navigation behavior by causing a mismatch between intended (and physically carried out) movements and consequent virtual movements (Du, Mou, & Zhang, 2018; Tcheang, Bülthoff, & Burgess, 2011).

In DeFINE, a visible grid barrier, shown in Figure 4, is displayed in the HMD when participants approach the limits of a configured area in which they can safely move around. The grid barrier serves two purposes. First and foremost, it prevents the participants from going out of the physical safe area, ensuring their safety. It is advisable that a navigation task in DeFINE be well confined to an area smaller than the safe area so that the participants will never encounter the barriers in the first place. If they do view the barrier, it essentially functions as an extra landmark that informs about the boundary of an environment, which can induce significant bias in their navigation behavior (Bird, Capponi, King, Doeller, & Burgess, 2010; Cheung, 2014; Mou & Zhou, 2013).

Second, the barrier makes it possible to extend a navigable virtual space beyond the

physical safe area, in case it is necessary for an experiment. To do this, participants hold the trigger button in the VR controller, which locks the VE in place. While the VE is locked, the participants' physical rotation in real space is not reflected in their virtual heading. Thus, the participants appear to keep facing the same direction in the VE, despite physically turning to face away from the edge of the safe area. The grid barrier remains visible and in correct orientation in respect to the physical safe area, allowing the participants to reorient themselves before continuing. In order to minimize the motion sickness caused by the VE remaining still during the participants' physical rotation, DeFINE blurs the VE during the rotation. Unlike similar approaches used in the literature (Williams et al., 2007), DeFINE does not require the participants to rotate a fixed amount as long as they steer clear of the physical boundary.

Although this method of extending the virtual space can be practical, it must be used with caution because by physically rotating in the locked VE, the participants will most likely be forced to go through a mental process of dissociating real and virtual spaces once and realigning them after the physical rotation. It is very probable that this process will have significant impact on the participants' mental and neural spatial representations, and in turn, on their subsequent navigation behavior.
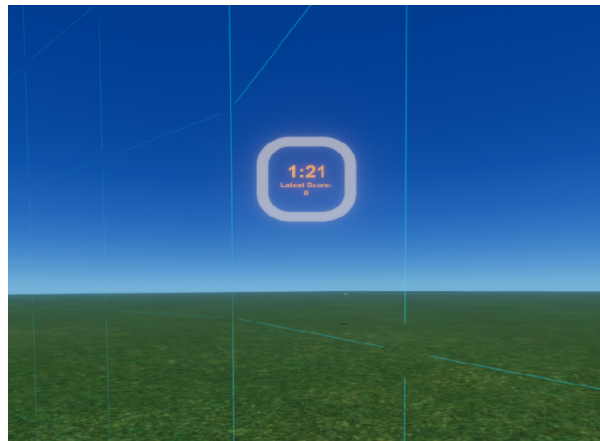
Figure 4. The visible blue grid marking the edge of a configured safe area.

Teleportation. Teleportation locomotion differs from all of the other locomotion methods in that participants never move through space, but instead teleport directly to their desired location some distance away. Before and after teleporting, the direction of the participants' body and head remain unchanged relative to the environment. In DeFINE, the

participants teleport by holding the trigger of a VR controller, which brings up the teleportation marker, as seen in Figure 5. Then the participants place the marker on the desired teleportation target location by aiming the controller at the location. Once the participants release the trigger, they are teleported to the marked location, given that the location is on the horizontal X-Z plane and clear of all collision regions around the objects in the environment. A valid target location is indicated by the blue color of the marker, whereas invalid locations turn the marker red. Although teleportation is not a naturalistic method of locomotion, its use is increasingly common in VEs including those for spatial navigation research (Cherep et al., 2020; Kelly, Ostrander, Lim, Cherep, & Gilbert, 2020).
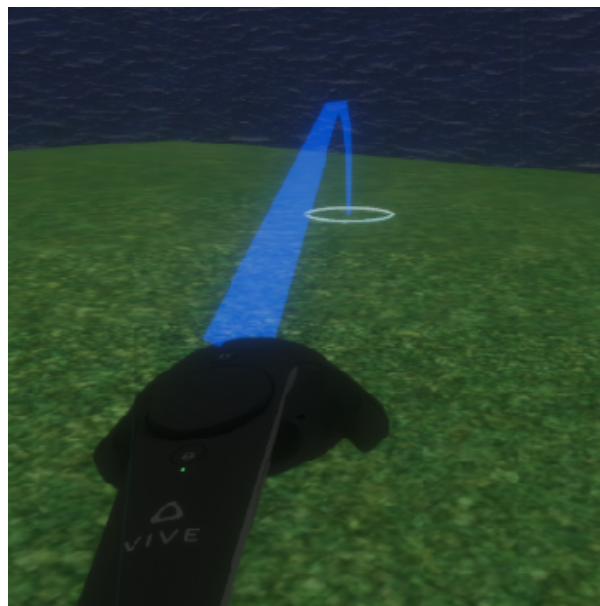
Figure 5. Teleportation locomotion with the teleportation marker visible.

Goal Demarcation

To accommodate a variety of experiments, DeFINE offers two possible ways of demarcating the goal location for a navigation task: firstly, presenting static objects at goal locations (e.g., arrows, exclamation marks, or other similar objects) and secondly, showing dynamic objects like a buzzing fly that can give imprecise indication of the goal location. An example of the dynamic goal markers is available in the case study section below.

Leader-board

Gamification of learning has shown increased participant engagement be it for online programs (Looyestyn et al., 2017) or education (Barata, Gama, Jorge, & Gonçalves, 2013). Thus, as an optional feature, DeFINE is also equipped with a leader-board which provides a ranking based on scores obtained using Equation (1) or other equivalent equations implemented by experimentalists (Figure 6). DeFINE keeps track of the scores and displays ten best scores in the leader-board. A new high-score is indicated with red font in the leader-board, while a score that was not high enough to get to the leader-board is shown at the bottom to illustrate the difference between the latest score and pre-existing scores. If participants are to carry out some practice or training trials first, it may not be appropriate to compare their scores against the pre-existing scores before they become fully familiarized with an experimental task. In that case, it is possible to show a provisional ranking which is not integrated with the leader-board. For clarity, this is labeled with a red *Practice* tag in the board. Once the practice phase is finished, the actual scores of the participants are integrated into the leader-board that includes their own previous high scores.
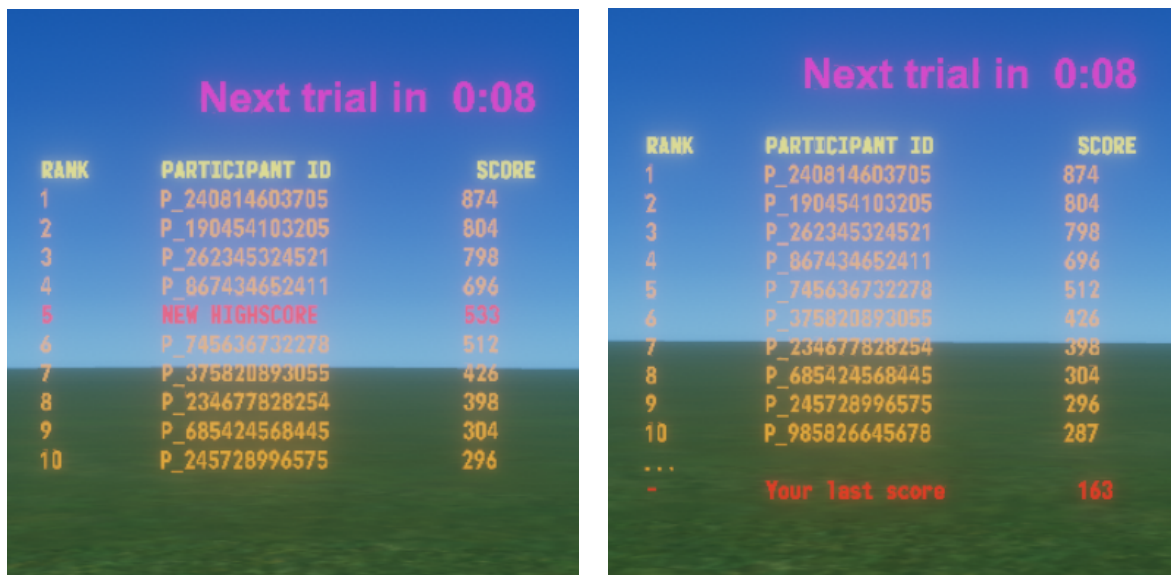
(a) A new high-score is shown in red in the leader-board.

(b) The latest score, which does not make it to the top ten, is shown at the bottom.

Figure 6. Leader-boards inform participants about their performance in a given trial relative to their own and other participants' previous trials.

While having a leader-board can motivate participants, it can also cause the conditions of an experiment to be different between the participants. As earlier participants obtain their place in the leader-board, they keep replacing lower scores on it. As such, it gets systematically more difficult for later participants to score high enough to make it to the top-ten scores of all time. Having a leader-board that is seemingly unreachable might provide a different motivation to the later participants than having an easily reachable one would. In order to ensure that each participant can have an equal experiment condition, DeFINE offers two options. First, because the leader-board is an optional feature, experimentalists can choose to remove it entirely. Second, they can use a fake leader-board that behaves like a normal leader-board during the session of one participant, except that the changes to the board are not in fact stored to a log file. Once the next participant begins a session, the board reverts to its original condition, giving subsequent participants the same competitive challenge.

Surveys

Often in behavioral studies, the experimentalists would like participants to fill surveys for quality assurance or other related purposes. Some of the most commonly used surveys for such studies using VR include the simulator sickness questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993) and the NASA task load index (NASA TLX; Hart, 2006). The SSQ studies the onset of simulator sickness symptoms like nausea or headache owing to being immersed in VR. It contains 27 questions and the participants answer each of them using a scale ranging from none (0) to severe (3). The NASA TLX is a survey for evaluating the workload of a given task utilizing six questions. Administering these and any other surveys has been made conveniently possible in DeFINE (Figure 7). The surveys are visible in an HMD to the participants and also on a desktop display to the experimentalist. While questions that have preset choices can be answered directly by the participant using the VR controller(s), questions that require free-form responses are to be typed in by the experimentalist on behalf of the participant.

While DeFINE's survey system allows the experimentalists to administer surveys while keeping participants immersed in VR, other systems typically require participants to take off

an HMD to answer surveys. Thus, if an experiment involves multiple sessions, each of which contains surveys, participants need to be re-immersed in VR every time they remove the HMD and put it back on (Schatz, Sackl, Timmerer, & Gardlo, 2017). This can be very cumbersome and make the participants feel uncomfortable, possibly inducing cybersickness. An alternative could be that the experimentalist orally asks the questions and fill in the surveys on behalf of the participants, but this can feel intrusive to the participants and reduce the sense of immersion in VR because the participants have to directly communicate with the experimentalist who does not belong to the virtual world (Bowman, Gabbard, & Hix, 2002). DeFINE remedies these issues by displaying the surveys in the HMD. To our knowledge, only Regal, Schatz, Schrammel, and Suette (2018) implemented a similar system previously, but it does not support the recent versions of the Unity game engine released post 2017.
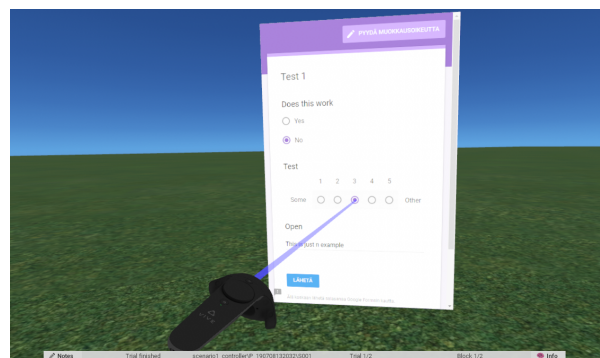
Figure 7. Filling in surveys whilst being immersed in VR using DeFINE.

Case Study

Be it rodents or humans, there is one behavioral effect in navigation that may be ubiquitous—i.e., the speed-accuracy trade-off (SAT) (Bogacz, Hu, Holmes, & Cohen, 2010; Heitz, 2014). Hasty decisions often lead to sub-optimal choices, whereas accurate decisions are futile if they take too much time. For example, when faced with a threat, a sensible course of action would be to hasten and escape as fast as possible. On the other hand, some tasks like walking on scaffolding require accuracy but the time taken to accomplish the task cannot be ignored either. Thus, making an optimal choice requires cognitive mechanisms to determine the appropriate balance between speed and accuracy, which then dictates the decision making process. This is referred to as the SAT and there is renewed interest in developing

computational models of SAT (Heitz, 2014; Heitz & Schall, 2012). While focus in this endeavor is on the neural mechanisms in general, understanding the impact of SAT on navigation behavior is also crucial as it affects the decisions made while navigating to a goal. To further illustrate the benefit of using DeFINE to investigate such behavioral characteristics, a simple goal-directed navigation scenario is presented.

The aim of this study was to investigate whether the SAT leads participants to adapt their navigation behavior through the feedback from previous trials. On the basis of previous studies demonstrating the existence of SAT in humans (Bogacz et al., 2010), we hypothesized that providing scores as a measure of navigation performance, which differentially rewarded speed and accuracy in different conditions, would affect participants' navigation behavior. More specifically, it was predicted that given a sufficient number of trials, the participants should be able to make a trade-off between speed and accuracy of their navigation behavior in a way that increases the feedback scores in each condition. In addition, this case study demonstrates that DeFINE does not induce cybersickness in participants who did not already report any such symptoms prior to immersion in VR.

Method

Participants.   Twenty-four participants (15 males, 8 females, and 1 other) took part in this study. Twenty-three of them were students of Aalto University, and one was from the vocational university Metropolia, Finland. The mean age of the participants was 24.8 ± 2.9 years. The participants' educational background ranged from having graduated from high school to having a master's degree. All participants gave written informed consent to participate in the study and received a movie ticket in return for their participation. The protocol of the study was approved by the Aalto University Research Ethics Committee.

Design and materials.   Participants were asked to navigate from start to goal positions within a virtual room of 10×10 m with non-repeating textures as shown in Figure 8. The walls were 4-m tall. The participants navigated using the controller teleoperation method. The participants' eye height in the virtual room was set at 1.36 m, which approximately corresponded to their actual eye height while seated in a real room. The participants first went

Figure 8. The default experimentalist view of the environment containing walls and floor with non-repeating textures. A countdown timer and score window helps participants keep track of their performance. Light and sound can be toggled by the experimentalist. The firefly serves as a noisy visual cue to guide the participants to the goal. The light/audio toggle switches (top right) and experimentalist button descriptors (top left) are only visible to the experimentalist on the desktop running DeFINE, and are not made visible to the participant in the HMD. Except for this difference, the participant's and experimentalist's views are the same. The arrows specifying X, Y, and Z axes in the lower-left corner are shown in this figure only; they appear neither in the participant's nor experimentalist's view. See text for further details.

through a training block containing 15 trials in which the room walls were visible. Subsequently, they performed the same navigation task in a testing block in which the walls were removed and the floor texture was extended to the horizon. The testing block also consisted of 15 trials. The starting and goal positions were fixed across trials as well as across the blocks. Relative to the center of the room, in a left-handed coordinate system, the starting position was at (4.5 m, 4.5 m) and the goal position was at (−3 m, −1 m). At the starting position, the participants directly faced the hidden goal location with an orientation of −135° about the vertical Y-axis. As a dynamic goal marker, a firefly buzzed around the goal position in such a way that its randomly fluctuating flying trajectory had its center directly over the goal position. Specifically, in each frame, the fly's position along the horizontal X-Z plane was randomly sampled within the radius of 0.75 m from the goal for training and 1.5 m for testing trials. The height of the fly along the Y-axis was randomly sampled in the range of 0.75–1.25 m. To make the fly move smoothly, its position was incremented with a step size of 5 mm. For a graphical presentation of X, Y, and Z axes, see Figure 8.

In this manner, the fly represented a noisy visual cue to guide the participants to the goal position. That is, the exact goal position was never revealed to the participants, and instead they were told that the goal was somewhere inside the area delimited by the fly's trajectory. Hence, the goal position was provided imprecisely to the participants via the noisy visual cue, and also the feedback score.

The participants were assigned to one of two groups, both of which received scores for their performance that depended on both the navigation speed and accuracy, but with different weights. *Time group* received feedback that put more importance on speed while the feedback in *accuracy group* was weighted in favor of being as close as possible to the goal. The feedback provided to the participants was computed using Equation (1) with the constants shown in Table 1. The score was presented to the participants at the end of each trial.

The participants were informed that they would be graded according to the time elapsed and residual distance to the actual goal. However, they were not told about the existence of the two groups or which group they belonged to. Instead, the participants of both groups were told that the scores obtained for the trials would be based on both speed and accuracy. In order to

Table 1

Constants of the reward function for two participant groups.

Constant	Time group value	Accuracy group value
α_1	-0.05	0.2
α_2	0.2	1
β_1	-2	0.5
β_2	6.2	3.4

make the scores easier for the participants to follow, they were scaled up by a factor of 300 and their minimum value was set at 0 (i.e., no negative scores).

DeFINE was run on a high performance Windows 10 personal computer with an Intel core i5 processor, 32 GB of RAM and a Nvidia GTX 1070 graphics card. The HTC Vive Pro with a wireless extension was used for a VR HMD. DeFINE was used for presenting the VE and recording data for each frame (approximately 90 frames per second). Specifically, the following log files were created for each participant per session:

Environment settings.   This file specified the size of the virtual room, specifics of the fly's trajectory (buzzing speed, minimum and maximum height of flight, buzzing radius) and whether to remove or retain the bounding walls during the testing phase. Additionally, the link to survey forms (if applicable) was added here.

Locomotion settings.   This log file specified the locomotion method used and its presets like traversal gain and rotation speed.

Scenario settings.   This log file specified how many trials were to be presented in training and testing blocks, the longest-possible duration of a trial (with the maximum of 120 s) and the start and goal locations.

Participant particulars.   The participant information as collected via the GUI (i.e., ID, age, gender and highest qualification achieved) was recorded.

Movement logs.   This log file recorded the participant's X and Z positions and rotation about the Y axis with time stamps. Owing to flexibility provided by DeFINE to toggle lights and sounds even during a trial, the status of these parameters were also logged every frame in

these logs. A new log was created per trial along with trial numbers.

Trial results.   This log file assimilated the component wise and cumulative rewards per participant along with distance covered and time elapsed during the trial.

Notes.   The experimentalist's notes during the experiments were recorded in this file. For instance, if some participants felt dizzy and opted for early termination, the particular session can be marked as bad and further details can be stored as notes for later use.

Procedure.   Participants sat in the middle of the room that was clear of any obstacles. At the outset of the experiment, the participants were asked to fill the SSQ to log their state of health before being immersed in VR. Their age, gender, and the highest qualification achieved were also recorded using DeFINE's GUI (Figure 3a). The experimentalist then put the HMD on the participant's head (over the spectacles, as and when necessary) and handed the participant hand-held VR controllers.

As soon as the participant had verbally confirmed to be ready, the training block was started by the experimentalist. The participant began a trial by leaving the start position, using the controller teleoperation method to navigate, and ended it by pressing a key on a VR controller when they thought they had reached the goal. The goal was positioned diagonally across the other side of the room and remained unchanged across trials. The participant then received a score from the trial in a leader-board (Figure 6). The fake leader-board feature was used so that all participants performed the navigation task with the same competitive challenge. The leader-board was displayed for 10 seconds (or until the participants pressed the "End Trial" key on the VR controller), and the room was automatically shown from the start position again for the next trial thereby resetting the scene to the exact same configuration for each trial. The participant completed the trials at their own pace, until reaching the end of the block. The participants were allowed to have a short break between trials. When necessary, the participants were able to skip a trial by pressing a controller key. At the end of the training block, the participant filled the NASA TLX in the DeFINE's form system (i.e., without taking off the HMD) using a 7-point Likert scale. Upon having filled the form, the participant started the testing block at their own input, prompted on the HMD. Once again the participant performed the trials at their own pace, until filling the NASA TLX again at the end of the

testing block. Filling the form completed the VR part of the experiment.

After taking off the HMD and the controllers, the participant filled the SSQ again to evaluate their simulation sickness after the exposure to the immersive VR. In addition, the participants were invited to provide feedback about the experiment and DeFINE by indicating the degree of agreement with each of the following five statements in a 5-point Likert scale: "Instructions were easy to understand"; "I understood what the score depended on"; "moving in the VE was easy"; "the walls in practice phase were helpful"; and "filling a form in the VE was easy".

Results

Two participants from each group misunderstood task instructions and simply chased the fly rather than navigating towards the goal it indicated. Due to this behavior, data from these participants were excluded from analysis. The data presented in this section represent the results of the remaining 10 participants per group, accounting for 20 participants in total. In addition, in 0.7% of all trials, participants accidentally pressed the button to end the trial immediately after it had begun. These trials were also discarded for the analysis presented herewith.

For each trial, the total elapsed time, the residual distance to the goal, and the score were derived from log files. For each dependent measure, data points that were more than 3*SD* away from each participant's mean of each block (training or testing) were defined as outliers and removed from analysis. This resulted in removal of 1.17% of trials on average.

Effect of feedback on navigation behavior. Table 2 shows descriptive statistics of all three dependent measures as a function of participant groups and blocks. Overall, participants in the time group performed trials more quickly and less accurately in the testing block than in the training block, suggesting the presence of the SAT. Participants in the accuracy group carried out trials more slowly in the test block than in the training block, which is also suggestive of the SAT. However, these participants decreased their accuracy and worsened their scores in the testing block, making it unclear whether they successfully utilized the feedback weighted in favor of accurate performance.

Table 2

Means and standard deviations of the total elapsed time, the residual distance to the goal, and the score as a function of participant groups and blocks.

	Time Group		Accuracy Group	
	Training	Test	Training	Test
Time (s)	11.13 (5.93)	7.81 (4.44)	10.97 (5.25)	13.33 (10.77)
Distance (m)	0.43 (0.18)	0.67 (0.32)	0.63 (0.34)	0.80 (0.36)
Score	664.97 (239.02)	718.87 (255.85)	624.77 (141.36)	542.41 (160.26)

Note. Standard deviations are shown in parentheses.

Figure 9 shows the elapsed time and residual distance in each trial, providing a more detailed picture of participants' performance. In terms of speed, the two groups performed similarly in the training block, but they differed in the testing block. Specifically, the time group maintained approximately the same speed throughout the block, performing the trials consistently quicker than the accuracy group. This pattern suggests that the feedback scores affected participants' navigation differently in the two groups. On the other hand, the effects of the scores were less clear on the accuracy of performance. Participants in the accuracy group showed no visible improvement of accuracy in later trials, even though they received scores that rewarded accurate performance.

To statistically examine whether feedback scores led participants to make a trade-off between speed and accuracy in their navigation performance, we conducted a detailed analysis of how the scores changed as the participants went through trials. Given that the scores were derived by giving differential weights to speed and accuracy, trial-by-trial changes in the scores would constitute a useful index of whether and how participants differentially weighted speed and accuracy in the course of the experiment. To this end, the gain or loss in the score was calculated for each trial by subtracting the score in an immediately preceding trial from the score of a current trial, and this difference score was analyzed by a mixed analysis of variance (ANOVA) in which block (training and testing) was a within-participant factor and group (time and accuracy) was a between-participant factor. Mean difference scores are
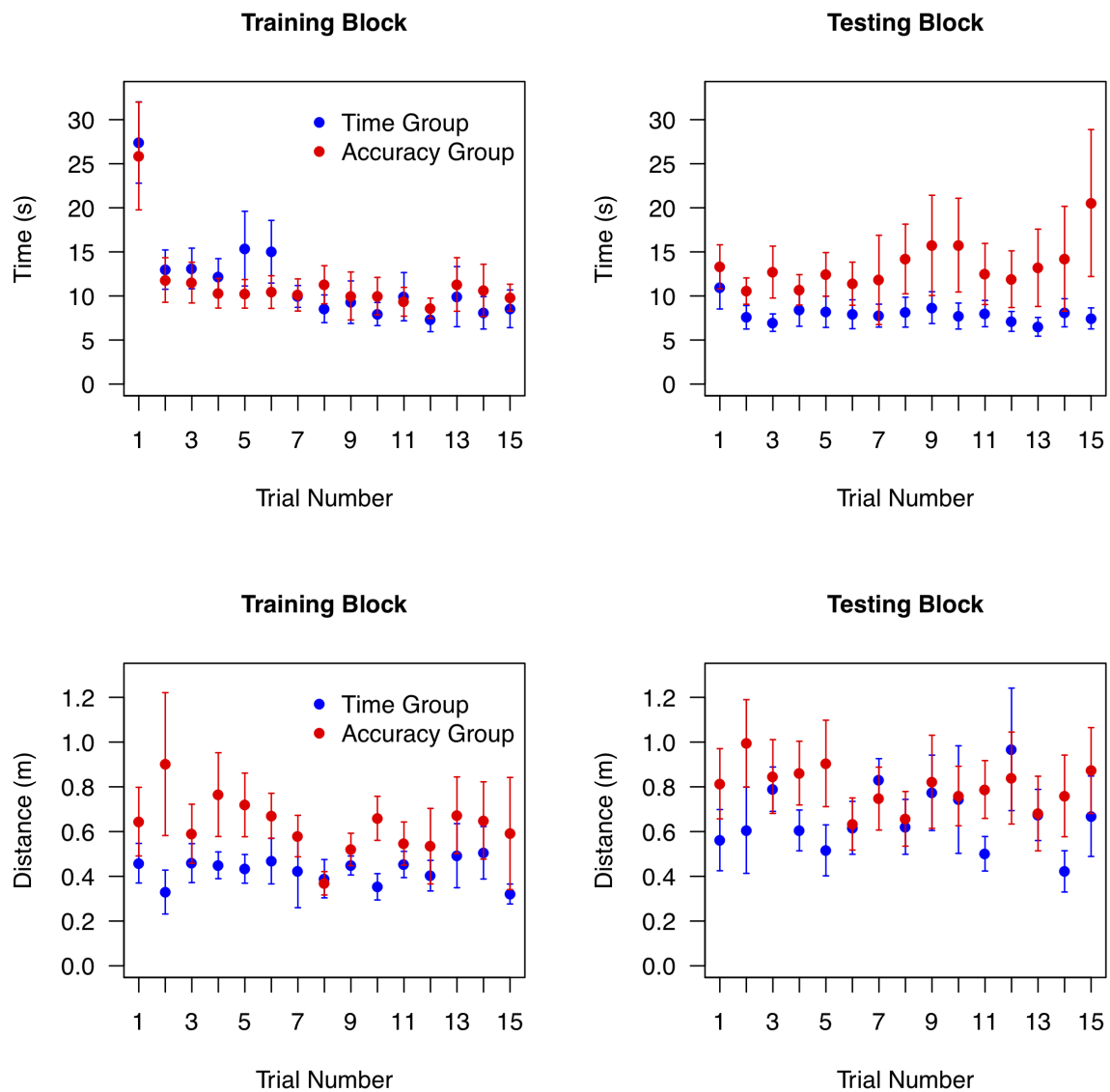
Figure 9. Mean times participants spent for performing each trial (top row) and mean residual distances to the goal at the end of each trial (bottom row) as a function of participant group, block, and trial number. Error bars represent ±1 standard error of the mean (*SEM*).

plotted for each trial in Figure 10. As in the observations made above, positive changes of the scores were most evident in the time group, particularly in the training block—participants in this group had a large gain in the second trial, and they also made a clear increase midway through the block. On the other hand, the scores remained largely unchanged in the testing block of the time group as well as in both blocks of the accuracy group. These patterns of data were reflected in the ANOVA that yielded the significant interaction between block and group,

$F(1, 18) = 4.56, p = 0.047, \eta_G^2 = 0.090$, and the significant main effect of block,

$F(1, 18) = 11.60, p = 0.003, \eta_G^2 = 0.20$. The main effect of group also approached

significance, $F(1, 18) = 4.23, p = 0.054, \eta_G^2 = 0.13$.

Taken together, the elapsed time, residual distance, and score suggest that the time and

accuracy groups performed the navigation task differently, even though they were tested in the

same manner except the parameters of the feedback function.
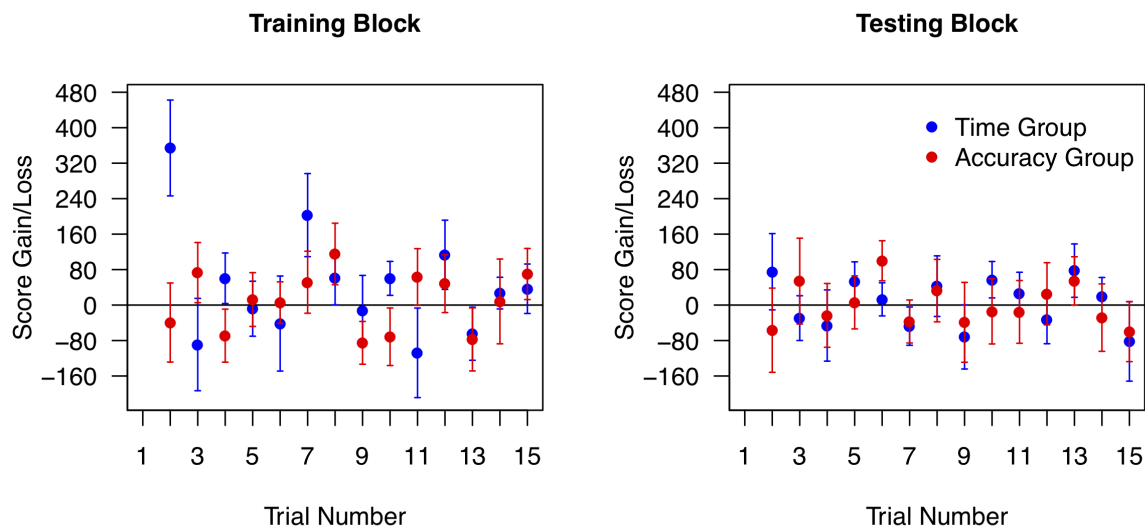
Figure 10. Mean gain or loss of a feedback score relative to an immediately preceding trial as a function of participant group, block, and trial number. Error bars represent ± 1 *SEM*.

Simulation sickness questionnaire.   Responses to the SSQ are summarized in

Table 3. As shown in the table, participants scored very low not only before but also after

exposure to the VE. Because the scores were very low overall, we used total raw scores for

analysis, instead of deriving weighted scores for each sub-scale of the SSQ (Kennedy et al.,

1993). The total raw SSQ scores were analyzed by a mixed ANOVA with exposure (before

and after) as a within-participant factor and group (time and accuracy) as a

between-participant factor. This ANOVA yielded no significant effects,

$F_s(1, 18) < 0.22, ps > 0.64, \eta_G^2 s < 0.010$, suggesting that the SSQ scores did not differ

between pre- and post-exposure to the VE as well as between the time and accuracy groups.

These results indicate that the use of DeFINE did not induce any major symptoms of

cybersickness.

To ensure that the lack of significant effects in the ANOVA was not a mere consequence of having inconclusive data, a Bayes factor analysis was conducted to gauge the extent to which the data actually supported the null hypothesis that neither exposure nor group had an effect on the SSQ scores (Rouder, Morey, Speckman, & Province, 2012). When the null model was compared against the full model that included the main effects of exposure and group as well as the interaction between the two, it yielded a Bayes factor of 13.9. This constitutes positive evidence for the null hypothesis (Kass & Raftery, 1995), supporting the conclusion that DeFINE's VE (and which group participants were in) did not cause cybersickness above and beyond what the participants had prior to navigating in the VE.

NASA task load index.   Responses to each item of the NASA TLX ranged from one to seven, with smaller scores indicating lower task load. As shown in Table 4, participants generally indicated that doing the navigation task in DeFINE required medium workload. There was some variation of the scores between groups, blocks, and questions. For example, the scores of the temporal demand question suggest that the time group felt stronger time pressure than the accuracy group, which is consistent with the feedback function that put emphasis on speedy response in the time group. In addition, scores in the testing block tended to be higher than those in the training block, which corresponds to the fact that the task was made more difficult in the testing block. In line with these observations, a mixed ANOVA with block (training and test) and question (six questions of the NASA TLX) as within-participant factors and group (time and accuracy) as a between-participant factor yielded a significant interaction between question and group, $F(5,90) = 2.96, p = 0.035, \eta_G^2 = 0.049, \varepsilon = 0.66$ (this ANOVA was corrected for non-sphericity with the Greenhouse-Geisser method when appropriate). The interaction between question and block as well as the main effect of question were also significant, $F(5,90) = 3.31, p = 0.008, \eta_G^2 = 0.019$ and $F(5,90) = 7.14, p < 0.001, \eta_G^2 = 0.11, \varepsilon = 0.66$, respectively. The main effect of block was marginally significant, $F(1,18) = 3.58, p = 0.075, \eta_G^2 = 0.018$. On the other hand, the interaction between block and group and the main effect of group were not significant, $Fs(1,18) < 0.36, ps > 0.55, \eta_G^2s < 0.010$, suggesting that overall, the two groups tolerated the workload of using DeFINE in a similar way.

Table 3

Means and standard deviations of raw scores of the simulator sickness questionnaire (SSQ).

	Pre-exposure		Post-exposure	
	Time group	Accuracy group	Time group	Accuracy group
General discomfort	0.1 (0.32)	0.2 (0.63)	0.2 (0.42)	0.3 (0.48)
Fatigue	0.4 (0.52)	0.7 (0.95)	0.2 (0.42)	0.6 (0.70)
Boredom	0.3 (0.67)	0.1 (0.32)	0.3 (0.67)	0.1 (0.32)
Drowsiness	0.5 (0.71)	0.2 (0.42)	0 (0)	0.2 (0.63)
Headache	0.1 (0.32)	0 (0)	0.1 (0.32)	0.2 (0.42)
Eyestrain	0.5 (0.53)	0.5 (0.97)	0.7 (0.67)	0.5 (0.97)
Difficulty focusing	0.2 (0.42)	0.4 (0.70)	0.1 (0.32)	0.3 (0.48)
Salivation increase/decrease	0.1 (0.32)	0.2 (0.42)	0.1 (0.32)	0 (0)
Sweating	0 (0)	0.1 (0.32)	0.2 (0.42)	0.2 (0.63)
Nausea	0 (0)	0 (0)	0.2 (0.42)	0.1 (0.32)
Difficulty concentrating	0.3 (0.48)	0.3 (0.67)	0.1 (0.32)	0.3 (0.48)
Mental depression	0.2 (0.42)	0.2 (0.42)	0.2 (0.42)	0.2 (0.42)
Fullness of the head	0.3 (0.48)	0.4 (0.52)	0.3 (0.48)	0.2 (0.42)
Blurred vision	0.1 (0.32)	0.3 (0.48)	0.2 (0.42)	0.4 (0.52)
Dizziness with eyes open/closed	0.1 (0.32)	0.2 (0.63)	0.2 (0.42)	0.3 (0.48)
Vertigo	0 (0)	0 (0)	0.1 (0.32)	0.1 (0.32)
Visual flashbacks	0 (0)	0 (0)	0.3 (0.67)	0.1 (0.33)
Faintness	0.1 (0.32)	0 (0)	0 (0)	0.1 (0.32)
Breathing awareness	0.4 (0.52)	0 (0)	0.3 (0.48)	0.1 (0.32)
Stomach awareness	0 (0)	0.1 (0.32)	0.1 (0.32)	0.1 (0.32)
Loss of appetite	0 (0)	0 (0)	0 (0)	0.1 (0.32)
Increase of appetite	0.1 (0.32)	0.3 (0.48)	0.1 (0.32)	0.4 (0.52)
Desire to move bowels	0.1 (0.32)	0 (0)	0.1 (0.32)	0 (0)
Confusion	0 (0)	0.4 (0.70)	0.1 (0.32)	0.1 (0.32)
Burping	0 (0)	0 (0)	0 (0)	0 (0)
Vomiting	0 (0)	0 (0)	0 (0)	0.1 (0.32)
Others	0 (0)	0 (0)	0 (0)	0 (0)
Total	3.9 (3.51)	4.6 (5.66)	4.2 (3.74)	5.1 (4.12)

Note. Standard deviations are shown in parentheses. The possible range of the total score was from 0 to 81.

Table 4

Means and standard deviations of scores of the NASA task load index (NASA TLX).

	Time group		Accuracy group	
	Training	Test	Training	Test
Mental demand	3.0 (1.33)	3.7 (1.64)	2.3 (1.25)	3.0 (1.41)
Physical demand	1.7 (0.82)	2.2 (1.14)	1.7 (1.25)	2.7 (1.77)
Temporal demand	3.5 (1.35)	4.1 (1.79)	2.9 (1.60)	2.9 (1.85)
Effort	3.4 (1.58)	3.6 (1.71)	2.8 (1.14)	3.2 (1.32)
Performance	3.5 (1.43)	2.6 (1.35)	3.9 (1.60)	4.0 (1.70)
Frustration level	2.6 (1.71)	3.2 (1.62)	1.9 (1.20)	2.5 (1.43)
Total	17.7 (6.38)	19.4 (7.29)	15.5 (5.15)	18.3 (7.89)

Note. Standard deviations are shown in parentheses. The possible range of the total score was from 6 to 42.

Participant feedback on the experiment and DeFINE. Scores of the participant feedback survey at the end of the experiment are summarized in Table 5. Larger scores denote stronger agreement with the statements. Overall, participants gave high scores, indicating that DeFINE provided an easy-to-use interface for doing the navigation experiment. The scores were analyzed by a mixed ANOVA in which statement (five statements in the survey) was a within-participant factor and group (time and accuracy) was a between-participant factor. The main effect of statement was significant, $F(4,72) = 6.51, p < 0.001, \eta_G^2 = 0.20$, which suggests that scores were reliably lower in the statement about the usefulness of walls than in the other statements. The interaction between statement and group and the main effect of group were not significant, $F(4,72) = 1.73, p = 0.15, \eta_G^2 = 0.062$ and $F(1,18) = 0.18, p = 0.67, \eta_G^2 = 0.003$, respectively, suggesting that there was no overall difference between the groups in the way they responded to the feedback survey.

Discussion

The purpose of this case study was to demonstrate the usability of DeFINE by using it to investigate the phenomenon of the SAT in goal-directed navigation. Past studies of

Table 5

Means and standard deviations of scores of the participant feedback survey.

	Time group	Accuracy group
Clarity of instructions	4.2 (0.63)	4.1 (0.99)
Score interpretation	4.5 (0.71)	4.0 (1.05)
Ease of movement	4.3 (0.67)	4.2 (0.79)
Usefulness of walls	2.9 (1.37)	3.3 (0.82)
Ease of filling forms in DeFINE	3.4 (0.97)	4.2 (1.03)
Total	19.3 (2.63)	19.8 (2.62)

Note. Standard deviations are shown in parentheses. The possible range of the total score was from 5 to 25.

goal-directed navigation in small-scale space, whether they were conducted in real or VEs, tended to put emphasis on accuracy or precision of responses with little regard for how quickly participants carried out navigation tasks (e.g., X. Chen, McNamara, Kelly, & Wolbers, 2017; Chrastil & Warren, 2014; Harris & Wolbers, 2012; Yamamoto, Meléndez, & Menzies, 2014; Yamamoto, Philbeck, et al., 2014). However, it is important to consider the speed of the responses in evaluating their accuracy because there can be a trade-off relationship between them (Bogacz et al., 2010). DeFINE allows researchers to examine the speed and accuracy of navigation either in conjunction as in this case study or in isolation by setting the parameters of the reward function accordingly (e.g., $\beta_1 = 0$ makes the reward function exclusively focused on the accuracy).

Results from this study showed that by using differential weights on speed and accuracy of navigation in calculating feedback scores, DeFINE succeeded in eliciting different responses from participants. The effect of the feedback score was evident in the time group in which participants increased their scores by improving the speed of their responses first and then keeping the same speed while maintaining or slightly worsening the accuracy. In the accuracy group, participants appeared to care less for making a speedy response toward the end of the experiment, even though this was not visibly reflected in the feedback scores. It is likely that the feedback scores were more effective in causing the SAT in favor of speed

because of the specific way in which the current experiment was designed—that is,

participants could be self-aware of the speed of their response, but the accuracy was never

explicitly revealed to them, making it harder for participants to improve the accuracy by

slowing down the response. Importantly, this pattern is a result of one particular installation of

DeFINE, and its architecture flexibly enables researchers to set up a suitable balance between

the effects of speed and accuracy according to the objectives of their studies. For example, by

giving heavier weights to accuracy in the feedback function, researchers can make feedback

scores more directly informative about how well participants are reaching the goal location.

Similarly, by demarcating the goal location more specifically by using different goal markers

(e.g., a static marker or a dynamic marker with less variability) and environmental features

(e.g., walls that provide spatial cues), researchers can run experiments in which focus is

entirely on speed (i.e., accuracy is a given) or subtle changes in accuracy are scrutinized.

This case study also examined participants' experience in using DeFINE. Results from

the SSQ indicated that DeFINE caused no major symptoms of cybersickness. The NASA

TLX showed that the participants found doing the navigation task in DeFINE moderately

challenging but not unreasonably taxing. In the feedback survey, the participants gave a

positive evaluation to DeFINE itself and the design of the experiment. Generally, these results

did not differ between the two groups of participants, suggesting that DeFINE provided a

versatile platform that accommodates different types of experiments.

In sum, this case study demonstrated that DeFINE was able to yield different responses

from the same navigation task by using feedback scores, without explicitly revealing the

differential importance of speed and accuracy of the responses to participants. In addition,

both objective and subjective measures of participants' experience indicated that they found

DeFINE easy to use and the navigation task it implemented well tolerable in terms of

cybersickness and task workload, irrespective of the ways in which they carried out the

navigation task. Together, these results validated DeFINE's capability as a tool for

investigating goal-directed navigation in humans under a variety of conditions.

Conclusions

This paper presented the open-sourced Delayed Feedback based Immersive Navigation Environment (DeFINE) for studying goal-directed navigation behaviors in humans using VR. Although similar frameworks have already been developed (Brookes et al., 2019; Commins et al., 2019; Machado et al., 2019; Vasser et al., 2017; Wiener et al., 2019), they are based on an open-loop stimulus–response architecture that omits performance feedback to participants. DeFINE distinguishes itself from the previous frameworks by implementing the closed-loop stimulus–response–feedback architecture as its core element (Figure 2). The feedback is delayed by default in order to suit the needs of typical navigation experiments, but it is also possible to make it real-time so that the stimulus–response–feedback loop is even more tightly closed.

A key strength of DeFINE as compared to the previous frameworks is the reduced load of the experimentalist. This was achieved by focusing primarily on goal-directed navigation tasks, and also by making it possible to interact with DeFINE mostly through intuitive GUIs and simple settings files (demonstrated in the video clips available online). The ease of use of DeFINE was further demonstrated by designing a case study that examined the SAT in goal-directed navigation. As summarized above, this study showed the effectiveness of performance feedback provided through DeFINE's built-in feedback function, and also the general user-friendliness of the entire system. Additionally, this study demonstrated DeFINE's potential as a platform for testing hypotheses about the speed of navigation behavior. The optional feature of seamlessly administering surveys within an HMD enhances the immersion of participants in VR, thereby improving the quality of data collected via DeFINE. Similarly, the optional leader-board enables further investigation of the effect of gamification on spatial navigation. Previous studies have shown its impact in other domains of learning (Barata et al., 2013; Looyestyn et al., 2017), but it is yet to be thoroughly explored for navigation-related applications (Coughlan et al., 2019; Coutrot et al., 2018). These out-of-the-box features of DeFINE, together with its great customizability via the Unity software, open up many new possibilities for human navigation research.

Open Practices Statements

The software used in the experiment reported in this article—i.e., the Delayed Feedback based Immersive Navigation Environment (DeFINE)—is available at https://gitlab.com/aalto-qut/environment. The data and other materials for the experiment are available upon request. The experiment was not preregistered.

References

Acharya, L., Aghajan, Z. M., Vuong, C., Moore, J. J., & Mehta, M. R. (2016). Causal influence of visual cues on hippocampal directional selectivity. *Cell*, *164*, 197–207. doi: 10.1016/j.cell.2015.12.015

Ainge, J. A., & Langston, R. F. (2012). Ontogeny of neural circuits underlying spatial memory in the rat. *Frontiers in Neural Circuits*, *6*, 8. doi: 10.3389/fncir.2012.00008

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving participation and learning with gamification. In L. E. Nacke, K. Harrigan, & N. Randall (Eds.), *Gamification 2013: Proceedings of the first International Conference on Gameful Design, Research, and Applications* (pp. 10–17). New York, NY: Association for Computing Machinery. doi: 10.1145/2583008.2583010

Bird, C. M., Capponi, C., King, J. A., Doeller, C. F., & Burgess, N. (2010). Establishing the boundaries: The hippocampal contribution to imagining scenes. *Journal of Neuroscience*, *30*, 11688–11695. doi: 10.1523/JNEUROSCI.0723-10.2010

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed–accuracy trade-off that maximizes reward rate? *The Quarterly Journal of Experimental Psychology*, *63*, 863–891. doi: 10.1080/17470210903091643

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence: Teleoperators & Virtual Environments*, *11*, 404–424. doi: 10.1162/105474602760204309

Brand, M. (2008). Does the feedback from previous trials influence current decisions? A study on the role of feedback processing in making decisions under explicit risk conditions. *Journal of Neuropsychology*, *2*, 431–443. doi: 10.1348/174866407x220607

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2019). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, Advance online publication. doi: 10.3758/s13428-019-01242-0

Carton, D., Nitsch, V., Meinzer, D., & Wollherr, D. (2016). Towards assessing the human

trajectory planning horizon. *PLoS ONE*, *11*, e0167021. doi:

    10.1371/journal.pone.0167021

Chen, G., King, J. A., Lu, Y., Cacucci, F., & Burgess, N. (2018). Spatial cell firing during

    virtual navigation of open arenas by head-restrained mice. *eLife*, *7*, e34789. doi:

    10.7554/eLife.34789

Chen, X., He, Q., Kelly, J., Fiete, I., & McNamara, T. (2015). Bias in human path integration

    is predicted by properties of grid cells. *Current Biology*, *25*, 1771–1776. doi:

    10.1016/j.cub.2015.05.031

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human

    spatial navigation. *Cognitive Psychology*, *95*, 105–144. doi:

    10.1016/j.cogpsych.2017.04.003

Cherep, L. A., Lim, A. F., Kelly, J. W., Acharya, D., Velasco, A., Bustamante, E., . . . Gilbert,

    S. B. (2020). Spatial cognitive implications of teleporting through virtual environments.

    Journal of Experimental Psychology: Applied, Advance online publication. doi:

    10.1037/xap0000263

Cheung, A. (2014). Estimating location without external cues. *PLOS Computational Biology*,

    10, e1003927. doi: 10.1371/journal.pcbi.1003927

Chrastil, E. R., & Warren, W. H. (2014). Does the human odometer use an extrinsic or

    intrinsic metric? *Attention, Perception, & Psychophysics*, *76*, 230–246. doi:

    10.3758/s13414-013-0549-3

Commins, S., Duffin, J., Chaves, K., Leahy, D., Corcoran, K., Caffrey, M., . . . Thornberry, C.

    (2019). Navwell: A simplified virtual-reality platform for spatial navigation and

    memory experiments. *Behavior Research Methods*, Advance online publication. doi:

    10.3758/s13428-019-01310-5

Cornwell, B. R., Johnson, L. L., Holroyd, T., Carver, F. W., & Grillon, C. (2008). Human

    hippocampal and parahippocampal theta during goal-directed spatial navigation predicts

    performance on a virtual morris water maze. *Journal of Neuroscience*, *28*, 5983–5990.

    doi: 10.1523/JNEUROSCI.5001-07.2008

Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.-M., Spiers, H., & Hornberger, M.

(2019). Toward personalized cognitive diagnostics of at-genetic-risk AlzheimerâĂŹs

disease. *Proceedings of the National Academy of Sciences of the United States of

America*, *116*, 9285–9292. doi: 10.1073/pnas.1901600116

Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., . . . Spiers, H. J.

(2018). Global determinants of navigation ability. *Current Biology*, *28*, 2861–2866.

doi: 10.1016/j.cub.2018.06.009

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory

network. *Nature*, *463*, 657–661. doi: 10.1038/nature08704

Du, Y., Mou, W., & Zhang, L. (2018). Unidirectional influence of vision on locomotion in

multimodal spatial representations acquired from navigation. *Psychological Research*,

Advance online publication. doi: 10.1007/s00426-018-1131-3

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L., &

Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*,

184–188. doi: 10.1038/nature01964

Ekstrom, A. D., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human spatial

navigation*. Princeton, NJ: Princeton University Press.

Geva-Sagiv, M., Las, L., Yovel, Y., & Ulanovsky, N. (2015). Spatial cognition in bats and rats:

From sensory acquisition to multiscale maps and navigation. *Nature Reviews

Neuroscience*, *16*, 94–108. doi: 10.1038/nrn3888

Harris, M. A., & Wolbers, T. (2012). Ageing effects on path integration and landmark

navigation. *Hippocampus*, *22*, 1770–1780. doi: 10.1002/hipo.22011

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the

Human Factors and Ergonomics Society Annual Meeting*, *50*, 904–908. doi:

10.1177/154193120605000909

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and

behavior. *Frontiers in Neuroscience*, *8*, 150. doi: 10.3389/fnins.2014.00150

Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*,

76, 616–628. doi: 10.1016/j.neuron.2012.08.030

Hollup, S. A., Kjelstrup, K. G., Hoff, J., Moser, M.-B., & Moser, E. I. (2001). Impaired

recognition of the goal location during spatial navigation in rats with hippocampal

lesions. *Journal of Neuroscience*, *21*, 4505–4513. doi:

10.1523/JNEUROSCI.21-12-04505.2001

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: Increasing productivity

using simple framing manipulations. *Management Science*, *58*, 2151–2167. doi:

10.1287/mnsc.1120.1544

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X.-X., . . . Kahana,

M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial

navigation. *Nature Neuroscience*, *16*, 1188–1190. doi: 10.1038/nn.3466

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical

Association*, *90*, 773–795. doi: 10.1080/01621459.1995.10476572

Kelly, J. W., Ostrander, A. G., Lim, A. F., Cherep, L. A., & Gilbert, S. B. (2020). Teleporting

through virtual environments: Effects of path scale and environment scale on spatial

updating. *IEEE Transactions on Visualization and Computer Graphics*, Advance online

publication. doi: 10.1109/TVCG.2020.2973051

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness

questionnaire: An enhanced method for quantifying simulator sickness. *The

International Journal of Aviation Psychology*, *3*, 203–220. doi:

10.1207/s15327108ijap0303_3

Koene, R. A., Gorchetchnikov, A., Cannon, R. C., & Hasselmo, M. E. (2003). Modeling

goal-directed spatial navigation in the rat based on physiological data from the

hippocampal formation. *Neural Networks*, *16*, 577–584. doi:

10.1016/S0893-6080(03)00106-0

Kunz, B. R., Creem-Regehr, S. H., & Thompson, W. B. (2009). Evidence for motor

simulation in imagined locomotion. *Journal of Experimental Psychology: Human

Perception and Performance*, *35*, 1458–1471. doi: 10.1037/a0015786

Loomis, J., & Knapp, J. (2003). Visual perception of egocentric distance in real and virtual

environments. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and adaptive

environments: Applications, implications, and human performance issues* (pp. 21–46).

Mahwah, NJ: Lawrence Erlbaum Associates.

Loomis, J. M., & Philbeck, J. W. (2008). Measuring spatial perception with spatial updating

and action. In R. L. Klatzky, B. MacWhinney, & M. Behrmann (Eds.), *Embodiment,*

ego-space, and action (pp. 1–43). New York, NY: Psychology Press.

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does

gamification increase engagement with online programs? A systematic review. *PLoS*

ONE, *12*, e0173403. doi: 10.1371/journal.pone.0173403

Machado, M., Lefèvre, N., Philoxene, B., Le Gall, A., Madeleine, S., Fleury, P., . . . Besnard,

S. (2019). New software dedicated to virtual mazes for human neurocognitive

investigations. *Journal of Neuroscience Methods*, *327*, 108388. doi:

10.1016/j.jneumeth.2019.108388

Mokrisova, I., Laczo, J., Andel, R., Gazova, I., Vyhnalek, M., Nedelska, Z., . . . Hort, J.

(2016). Real-space path integration is impaired in AlzheimerâĂŹs disease and mild

cognitive impairment. *Behavioural Brain Research*, *307*, 150–158. doi:

10.1016/j.bbr.2016.03.052

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation

impaired in rats with hippocampal lesions. *Nature*, *297*, 681–683. doi:

10.1038/297681a0

Mou, W., & Zhou, R. (2013). Defining a boundary in goal localization: Infinite number of

points or extended surfaces. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, *39*, 1115–1127. doi: 10.1037/a0030535

Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., . . . Rosenberg,

E. S. (2018). 15 years of research on redirected walking in immersive virtual

environments. *IEEE Computer Graphics and Applications*, *38*(2), 44–56. doi:

10.1109/MCG.2018.111125628

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus

with lidocaine differentially affects expression of place and response learning.

Neurobiology of Learning and Memory, *65*, 65–72. doi: 10.1006/nlme.1996.0007

Paludan, A., Elbaek, J., Mortensen, M., Zobbe, M., Nilsson, N. C., Nordahl, R., . . . Serafin, S.

(2016). Disguising rotational gain for redirected walking in virtual reality: Effect of visual density. In T. Höllerer, V. Interrante, A. Lécuyer, & E. Suma (Eds.), *2016 IEEE virtual reality (VR)* (pp. 259–260). Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi: 10.1109/VR.2016.7504752

Pezzulo, G., van der Meer, M. A., Lansink, C. S., & Pennartz, C. M. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends in Cognitive Sciences*, *18*, 647–657. doi: 10.1016/j.tics.2014.06.011

Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 72–85. doi: 10.1037/0096-1523.23.1.72

Ravassard, P., Kees, A., Willers, B., Ho, D., Aharoni, D., Cushman, J., . . . Mehta, M. R. (2013). Multisensory control of hippocampal spatiotemporal selectivity. *Science*, *340*, 1342–1346. doi: 10.1126/science.1232655

Regal, G., Schatz, R., Schrammel, J., & Suette, S. (2018). VRate: A Unity3D asset for integrating subjective assessment questionnaires in virtual environments. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–3. doi: 10.1109/QoMEX.2018.8463296

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. doi: 10.1016/j.jmp.2012.08.001

Schatz, R., Sackl, A., Timmerer, C., & Gardlo, B. (2017). Towards subjective quality of experience assessment for omnidirectional video streaming. *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. doi: 10.1109/QoMEX.2017.7965657

Spiers, H. J., & Maguire, E. A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. *NeuroImage*, *31*, 1826–1840. doi: 10.1016/j.neuroimage.2006.01.037

Tcheang, L., Bülthoff, H. H., & Burgess, N. (2011). Visual influence on path integration in

darkness indicates a multimodal representation of large-scale space. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 1152–1157. doi: 10.1073/pnas.1011843108

Vasser, M., Kängsepp, M., Magomedkerimov, M., Kilvits, K., Stafinjak, V., Kivisik, T., . . . Aru, J. (2017). VREX: An open-source toolbox for creating 3D virtual reality experiments. *BMC Psychology*, *5*, 4. doi: 10.1186/s40359-017-0173-4

Wiener, J. M., Carroll, D., Moeller, S., Bibi, I., Ivanova, D., Allen, P., & Wolbers, T. (2019). A novel virtual-reality-based route-learning test suite: Assessing the effects of cognitive aging on navigation. *Behavior Research Methods*, Advance online publication. doi: 10.3758/s13428-019-01264-8

Williams, B., Narasimham, G., Rump, B., McNamara, T. P., Carr, T. H., Rieser, J., & Bodenheimer, B. (2007). Exploring large virtual environments with an HMD when physical space is limited. *APGV07: Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, 41–48. doi: 10.1145/1272582.1272590

Yamamoto, N., Mach, D. E., Philbeck, J. W., & Van Pelt, J. (2018). Why is the duration of imagined walking underproduced? A dual-representation view on the mental imagery of locomotion. *PsyArXiv*. doi: 10.31234/osf.io/rm298

Yamamoto, N., Meléndez, J. A., & Menzies, D. T. (2014). Homing by path integration when a locomotion trajectory crosses itself. *Perception*, *43*, 1049–1060. doi: 10.1068/p7624

Yamamoto, N., Philbeck, J. W., Woods, A. J., Gajewski, D. A., Arthur, J. C., Potolicchio, S. J., Jr., . . . Caputy, A. J. (2014). Medial temporal lobe roles in human path integration. *PLoS ONE*, *9*, e96583. doi: 10.1371/journal.pone.0096583

Yartsev, M. M., & Ulanovsky, N. (2013). Representation of three-dimensional space in the hippocampus of flying bats. *Science*, *340*, 367–372. doi: 10.1126/science.1235338

Yechiam, E., & Hochman, G. (2014). Loss attention in a dual-task setting. *Psychological Science*, *25*, 494–502. doi: 10.1177/0956797613510725