BRILL *Multisensory Research 0 (2020) 1–25* brill.com/msr

How Much of What We Learn in Virtual Reality Transfers to Real-World Navigation?

Lukas Hejtmanek [1,2], **Michael Starrett** [2,3,4], **Emilio Ferrer** [3] **and Arne D. Ekstrom** [2,3,4,*]

[1] Third Faculty of Medicine, Charles University, Ruská 87, Prague 10, 100 00, Czech Republic
[2] Center for Neuroscience, University of California, Davis, 1 Shields Ave, Davis, CA 95618, USA
[3] Department of Psychology, University of California, Davis, 1 Shields Ave, Davis, CA 95618, USA
[4] Department of Psychology, University of Arizona, 1503 E. University Blvd., Tucson, AZ 85719, USA

Received 15 February 2019; accepted 18 December 2019

Abstract

Past studies suggest that learning a spatial environment by navigating on a desktop computer can lead to significant acquisition of spatial knowledge, although typically less than navigating in the real world. Exactly how this might differ when learning in immersive virtual interfaces that offer a rich set of multisensory cues remains to be fully explored. In this study, participants learned a campus building environment by navigating (1) the real-world version, (2) an immersive version involving an omnidirectional treadmill and head-mounted display, or (3) a version navigated on a desktop computer with a mouse and a keyboard. Participants first navigated the building in one of the three different interfaces and, afterward, navigated the real-world building to assess information transfer. To determine how well they learned the spatial layout, we measured path length, visitation errors, and pointing errors. Both virtual conditions resulted in significant learning and transfer to the real world, suggesting their efficacy in mimicking some aspects of real-world navigation. Overall, real-world navigation outperformed both immersive and desktop navigation, effects particularly pronounced early in learning. This was also suggested in a second experiment involving transfer from the real world to immersive virtual reality (VR). Analysis of effect sizes of going from virtual conditions to the real world suggested a slight advantage for immersive VR compared to desktop in terms of transfer, although at the cost of increased likelihood of dropout. Our findings suggest that virtual navigation results in significant learning, regardless of the interface, with immersive VR providing some advantage when transferring to the real world.

* To whom correspondence should be addressed. E-mail: adekstrom@email.arizona.edu

 DOI:10.1163/22134808-20201445

Keywords
Navigation, virtual reality, spatial cognition, proprioception, transfer

1. Introduction

Many studies on the cognitive and neural basis of human spatial navigation involve virtual environments (VEs) rendered on desktop computers. A limitation of these studies is that they provide a limited approximation of the wealth of multisensory cues available during real-world navigation. Specifically, free ambulation involves critical body-based cues not represented during virtual reality (VR) navigation with a joystick on a desktop computer: cues derived from body turns and movements that displace fluid in our vestibular system (termed here 'vestibular cues') and also affect sensory receptors and proprioceptors in our body, particularly our legs (termed here 'proprioceptive cues') (Loomis and Beall, 1998; Starrett and Ekstrom, 2018). Vestibular input, though, is critical to our representation of bearing and acceleration, with lesions to the vestibular nuclei in both rats and humans significantly impairing navigation (Brandt *et al.*, 2005; Russell *et al.*, 2003; Taube *et al.*, 2013; Valerio and Taube, 2012). Similarly, although less well researched, proprioceptive cues associated with moving our legs and feet are important for estimating our velocity and turning angle (Loomis and Beall, 1998) and successful navigation more generally (Chrastil and Warren, 2013; Gallistel, 1990; Matthis *et al.*, 2018). In particular, the somatosensory responses of the feet are critical for estimating gait and other important aspects of movement in space, such as orientation (Lackner and DiZio, 2005; Matthis *et al.*, 2018; Visell *et al.*, 2011). Others have argued, however, that body-based cues may not all be necessary for normal expression of some forms of spatial representations, such as topological graph knowledge (Chrastil and Warren, 2015). At present, the degree to which body-based cues are necessary for spatial learning remains unresolved.

Immersive VR tools, such as head-mounted displays (HMDs) and omnidirectional treadmills, allow a novel opportunity to study immersive navigation in the lab and more completely model body-based input in a controlled setting. Specifically, omnidirectional treadmills (see Fig. 1) allow participants to freely ambulate while wearing an HMD with the full spectrum of rotations that can render any number of different VEs while presenting them with a wide field of view (about 100 degrees in the HTC Vive HMD (HTC Corp., Taoyuan, Taiwan) in comparison to 60–70 degrees on a desktop monitor). With these two important components, we can study both enriched visual input (*via* the HMD) and the effect of enriched body-based cues on navigation (*via* the omnidirectional treadmill). Specifically, the omnidirectional treadmill

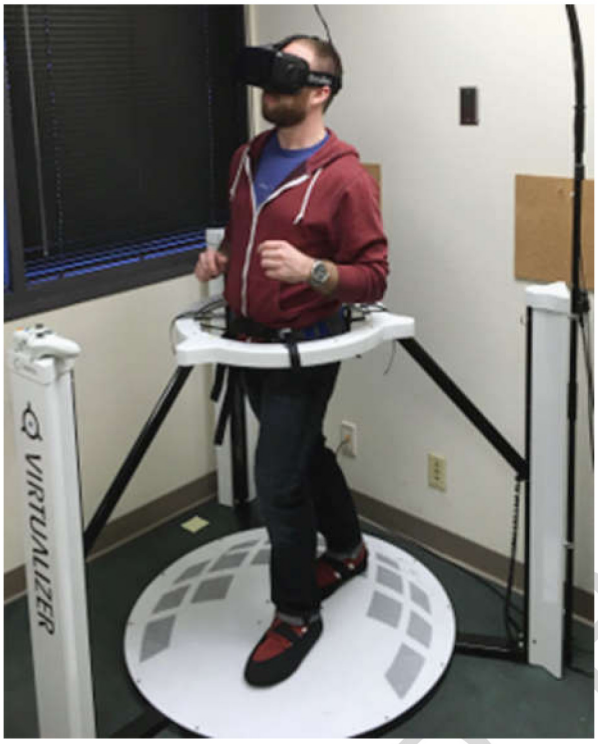


Figure 1. Immersive VR treadmill setup. Participant wearing an HTC Vive headset while being strapped into the omnidirectional treadmill. Participants can rotate 360 degrees and are allowed independent body and head rotation. Movement is achieved by sliding one's feet across the treadmill base.

provides both body and turn input, proprioceptive input from the legs, and somatosensory feedback from the feet as participants walk on the surface of the treadmill. Previous studies investigated navigation by providing participants with a full range of proprioceptive cues using an HMD and ambulation in a room (Ruddle and Lessels, 2006). Although omnidirectional treadmills only approximate the real-world experience (for example, walking on the omnidirectional treadmill does not engage the full range of somatosensory receptors and muscles compared to stepping in the real world), they have the advantage of allowing participants to explore environments larger than a VR/AR-enabled room would offer. Therefore, an important question regards to what extent information learned in such immersive VR contexts compares to real-world navigation.

One way to assess how well information acquired during virtual navigation applies to the real world involves 'transfer'. This typically involves subjects first learning environments in a virtual setting and then determining how effectively participants can apply acquired spatial knowledge to the real world (Montello *et al.*, 2004; Richardson *et al.*, 1999; Waller *et al.*, 1998; Witmer *et al.*, 1996). Studies investigating transfer of spatial knowledge acquired in VR to the real world also suggest advantages for conditions involving some amount of body-based input. One study by Waller et al. compared continuous

exposure to an environment on a desktop VE, navigating the same environment with an HMD with full head control and a joystick, and navigating the same real-world environment. HMD exploration with full head turning led to greater transfer than desktop exploration, although real-world navigation led to the highest transfer (best spatial learning). These findings are similar to some reported in other past studies (Waller *et al.*, 1998; Witmer *et al.*, 1996; see also Richardson *et al.*, 1999), suggesting the importance of body-based cues to navigation and transfer of information to real-world environments. One possibility, though, is that because these studies did not involve walking in the VR conditions, the limited proprioceptive cues led to overall poor encoding of head direction information due to a mismatch with the joystick movements. Thus, we might expect that the presence of more enriched body-based cues could lead to better transfer to the real world, an issue we will address in this study.

In support of the importance of richer body-based cues to navigation, Grant et al. compared navigation with an HMD, head turns, and walking in place by shuffling the feet compared to movement with a joystick (Grant and Magee, 1998). While the HMD/walking condition did result in improvements in taking shortcuts compared to desktop VE, both conditions resulted in worse direction estimates compared to real-world navigation. One limitation of the Grant and Magee study, however, is that the seated shuffling movement of the feet they employed should read out to the proprioceptive system as a net-zero forward translation. In contrast, on an omnidirectional treadmill, standing and moving the feet in a walking motion should provide a sense of net-positive forward translation. Thus, while studies of transfer from VR to real-world environments suggest advantages to some vestibular input rendered by an HMD, the absence and mismatch, in some cases, with other body-based cues (i.e., walking-based input from sensory and muscle receptors) may be a possible reason for incomplete transfer. Alternatively, it is also possible that visual rendering in VR always limits transfer (Thompson *et al.*, 2004), and thus it is possible that, even with richer body-based cues, visual rendering always limits transfer, to some extent.

In this study, we contrast two different forms of learning in VR and how they transfer to real-world navigation. As a control comparison and means of determining the theoretical ‘maximum’ transfer, we included continuously navigating in the real world for the entire experiment. To avoid issues with navigating in outdoor environments (trip hazards and environmental complications, like variable levels of lighting and rain), all testing occurred in a campus building, the UC Davis Center for Neuroscience (Figs 2, 3). One-third of our participants navigated through the building on desktop, the second third in immersive VR (i.e., HMD + omnidirectional treadmill), and the final third of

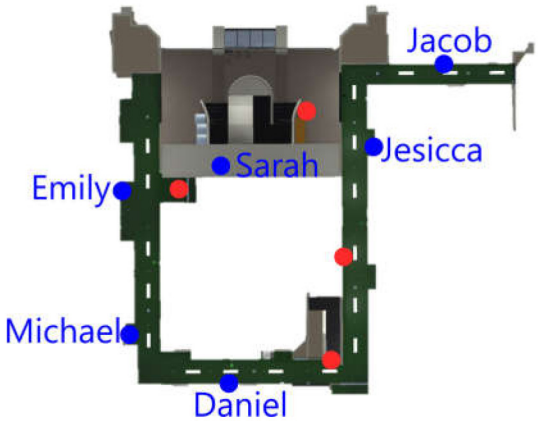


Figure 2. Illustration of the building layout with six offices (dots with names in blue) and four pointing locations.

Figure 3. Comparison of the visual and geometrical fidelity for the 3D visualisation. Center for Neuroscience entrance hall is depicted on the left and the modelled environment on the right.

participants navigated the real-world Center for Neuroscience. After completing three blocks of navigation (by finding specific targets in the building), all participants navigated the Center for Neuroscience in the real world. By tracking their position in the real-world building using Bluetooth trackers beacons (Estimote Inc., San Francisco, CA, USA), we were able to determine the deviation of their path from the ideal path they should walk, as well as any errors in visiting the wrong target.

Our main hypothesis, which we term the body-based enrichment hypothesis, suggests that providing additional body-based cues with an omnidirectional treadmill and HMD should at least partially mimic real-world navigation. In this case, we would predict that transfer should be higher from immersive VR to the real world than desktop VR. Immersive VR, according to the body-based enrichment hypothesis, should result in comparable, or maybe less transfer compared to the real world, depending on the degree to which participants experience the immersive experience as fully mimicking the real world. To briefly preface our results, we found that learning in the real-world

setting led to the highest transfer overall, with some evidence for immersive VR leading to greater transfer than desktop navigation.

2. Methods

2.1. Design Overview

To investigate the effects of learning modality on spatial knowledge acquisition, we devised an experiment with three different conditions involving active exploration and ambulation. Participants learned positions of six office doors (out of a total 32) within the Center for Neuroscience at UC Davis. The reason we had participants learn the labels on doorways was to avoid them having to remember objects they could easily see. Participants experienced the same building either in the real world (i.e., actually navigating the Center for Neuroscience), in immersive body-based VR with an omnidirectional treadmill (i.e., navigating a virtual version of the Center for Neuroscience with an HMD on an omnidirectional treadmill), or on a desktop computer (i.e., navigating the same virtual version but on a desktop computer with a mouse and a keyboard). After three blocks of learning, all participants continued the task in the real world. Our dependent measures involved (1) normalized walked path distance, which we assayed in the (a) real world using Estimote Bluetooth trackers and (b) in VR by recording position (termed 'path accuracy'); (2) visitation errors, which were quantified based on visiting the wrong doors; and (3) unsigned pointing error.

2.2. Walking Task

At the beginning of each walking trial, participants were given the name of a person whose office they had to find. Instructions were provided by the experimenter in the real-world condition or displayed in the user interface in the treadmill VR and desktop conditions. Participants selected the door by walking directly into it in virtual conditions and standing close and pointing to it in the real-world condition. Participants were not told which office belonged to whom, and therefore they had to walk around the building and keep checking doors until they arrived at the correct one. Once they arrived or pointed at the correct door participants were provided with a new name and a new trial started. Each office was visited three times in each phase, six times over the entire experiment. The order in which offices occurred was randomized in a way such that the order was balanced across all conditions.

2.3. Pointing Task

After finishing three blocks of finding doors (the learning phase), the participant was either moved (VR and desktop conditions) or walked to two different viewpoints within the building and asked to point to each office one by one.

This was repeated once more after the transfer phase, for a total of four pointing blocks (24 pointing in total).

2.4. Building a Realistic 3D Model of the Center for Neuroscience

We built the virtual Center for Neuroscience to scale for rendering in both immersive VR and on the desktop VE. An example of a viewpoint in the virtual rendition and the real world can be seen in Fig. 3. We set the walking speed in the desktop so that walking the length of a corridor in the real world would take the same time as walking the corridor in the desktop version. In the immersive VR condition, we set the speed so that walking down the corridor would take approximately the same number of steps in real world as it did on the treadmill. The experiment was built in Unity3D (Unity Technologies, San Francisco, CA, USA). The building can be experienced online at https://hejtmy.github.io/CFNS-task/. For the virtual conditions in the immersive VR condition, we used the HTC Vive HMD to render the building and participants moved using the Cyberith Virtualizer omnidirectional treadmill (Cyberith GmbH, Herzogenburg, Austria) to provide body-based cues (Fig. 1). Participants had a full range of rotational motion available to them. The desktop condition was presented on a 21″ monitor (resolution 1920 × 1080) and participants controlled it with a keyboard and a mouse. All participants were informed about possible side effects of the VR and in case of uncomfortable or prolonged dizziness, such participants were dismissed from the study and their data removed. Participants in the immersive VR condition in the learning phase were significantly more likely to drop out due to the cybersickness, fatigue, or other factors related to discomfort. We observed this both during the first [$\chi^2(1, 82) = 9.91$, $p = 0.002$] and the second experiment [$\chi^2(1, 70) = 14.72$, $p < 0.001$]. Detailed dropout rates for all conditions can be found in Tables 1 and 2.

Table 1.
Cyber sickness dropout table for Experiment 1 (all retests in the real world): Finished/total (percent finished). One participant in each group was removed due to recording failure during analyses, but they finished the experiment without issues. Participants in the immersive VR condition in the first phase were significantly more likely to drop out [$\chi^2(1, 82) = 9.91$, $p = 0.002$]

Learning condition	Male	Female
Real world	4/5 (80%)	19/19 (100%)
Desktop	11/11 (100%)	9/11 (82%)
Immersive VR	9/13 (69%)	12/23 (52%)

Table 2.
Cyber sickness dropout table for Experiment 2 (all transfer phase in the VR): Finished/total (percent finished). One participant from the immersive VR group and two from the real-world group were removed later due to recording failure during analyses, but they finished the experiment without issues and one participant in the immersive VR condition who did not finish did not disclose their gender. Participants in the immersive VR condition in the first phase were significantly more likely to drop out [$\chi^2(1, 70) = 14.721$, $p < 0.001$]

Learning condition	Male	Female
Real world	8/9 (89%)	14/16 (88%)
Immersive VR	8/17 (47%)	13/27 (48%)

2.5. *Real-World Tracking*

We tracked participants' position within the building using Bluetooth and a custom-made app running on the iPhone 6. Participants were given the phone to hold directly in front of them to track both position and rotations of their body. As participants walked around the building, the app tracked the strength of the Bluetooth signal towards the closest beacons and triangulated position within given building constraints. Participants were accompanied by an experimenter who provided instructions and feedback during door selection and noted down errors and special events. For the ambulatory paths travelled in the real world as well as on the VR treadmill, trajectories were smoothed using a median value in a 5-s moving window to remove slight jitter introduced by fluctuating strength of the Bluetooth signal in the real world or the shuffling of the feet on the treadmill.

2.6. *Dependent Measures*

We calculated participants' walked distance in the building, time spent in each task, and the number of incorrectly visited doors during walking trials. For the pointing tasks, we calculated unsigned pointing error. The three conditions were not directly comparable in their measured distance and time, with the VR trials taking a little bit longer in both time and distance than the real-world condition. The increased time was due to the novelty of the movement control that persevered even after the training session. In contrast, the larger absolute distance was due to differences in tracking precision of the VR and desktop compared to the real-world setup. We also observed feet shuffling and occasional small stumbles on the treadmill that added to the absolute walked distance for immersive VR in comparison to the desktop condition.

We therefore min-normalized the distance travelled and the time. We used the shortest path for each task (pair of offices, e.g., Jacob to Sarah) that any participant demonstrated in a single environment (e.g., real world) and then divided other participants performance in the same task (Jacob to Sarah) and

in the same environment (real world) by this number. Therefore, we had three baselines (desktop, immersive VR and real world) for each task (pair of offices). For example, a min-normalized distance value of 2.00 would mean the participant travelled double the distance than what was the best performance in each task (pair of offices) and in a certain modality (desktop, immersive VR or real world). Unless stated otherwise, we calculated all statistics on these normalized measures. We removed trials where distance or time measurements were more than three standard deviations away from the mean (a total of 1.80 percent of trials).

2.7. Experimental Procedure

Participants walked to six different office doors in the Center for Neuroscience building at UC Davis, in real or virtual versions. Positions of all the office doors can be seen in Fig. 2. The task itself consisted of a set of 36 walking trials and 24 pointing trials, separated into two equal phases, the learning phase and the transfer phase. Each phase had 18 walking trials consisting of three different blocks (each office visited once per block, three times per phase). After each phase, participants did a short pointing session of 12 trials (pointing to each of the six offices from two viewpoints). In the transfer phase, the immersive VR and desktop group switched their environment modality to the real world, whereas the real-world group remained in the same environment. Each participant, regardless of experimental condition, received a short training session on the VR treadmill before the learning phase.

Because our initial analyses observations and analyses suggested that participants might learn more slowly in VR, we devised a second experiment to better understand this issue. One group learned the UC Center for Neuroscience building in immersive VR while the second group learned in the real world. They both then transferred to immersive VR. Although this experiment does not directly address our main hypothesis (transfer from immersive/desktop VR to reality), this allowed us to assess if the immersive VR can achieve the same performance given additional time. It also allowed us to look at the conceptual 'inverse' process of transfer.

Schema of the entire procedure can be seen in Fig. 4.

2.8. Participants (Experiment 1)

A total of 82 undergraduate students at the University of California, Davis participated in the study in exchange for a course credit. Each participant was randomly assigned a condition and a randomized set of goals before arrival. Eighteen participants did not finish due to motion sickness and three were removed due to technical problems with the real-world tracking system. This resulted in a final sample size of 61 participants (37 female) (age $M = 20.42$, $\mathrm{SD} = 2.16$) used for all reported analyses. The participant completion rates for

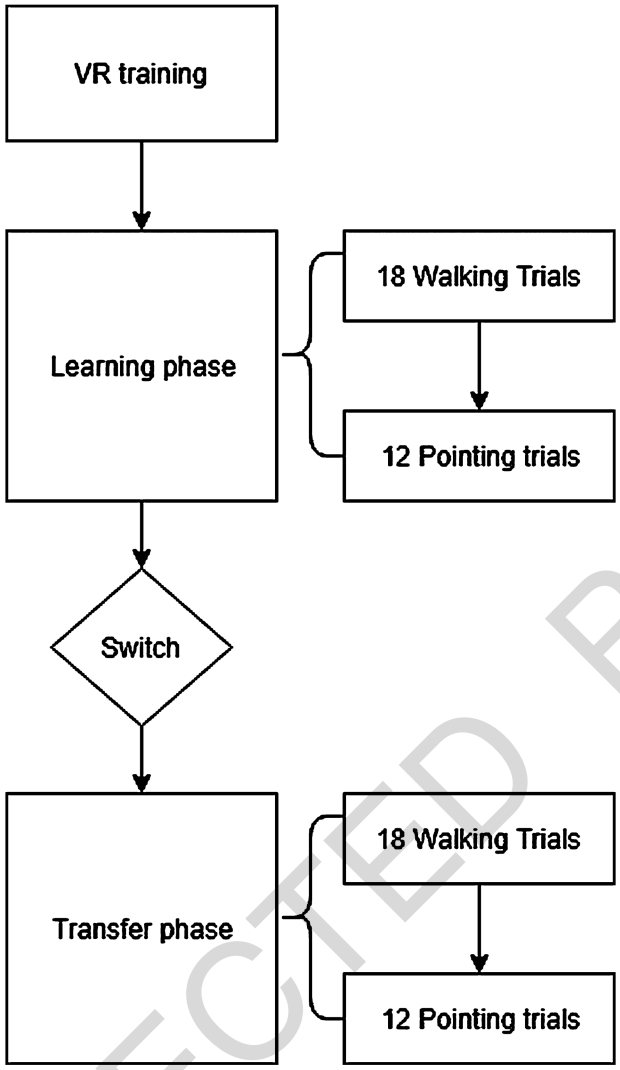


Figure 4. Visualisation of the experimental procedure. Session consisted of training, learning and transfer phases. Both learning and transfer phases included 18 walking trial and 12 pointing trials.

Experiment 1 are in Table 1. All procedures were approved by the UC Davis Institutional Review Board (IRB).

2.9. Participants (Experiment 2)

A total of 70 undergraduate students at UC Davis participated in the second experiment. Twenty-nine participants did not finish due to motion sickness or other issues with the VR interfaces. Analyses were then conducted on a final set of 41 (14 female) participants (age $M = 19.96$, SD $= 1.55$).

2.10. Analyses

We analysed the data in R (R Core Team, R Foundation for Statistical Computing, Vienna, Austria), with the help of the *ez* package for ANOVAs and the ggplot package for plots. Based on our *a-priori* hypothesis that real world $\geqq$

immersive > desktop VR, we performed t-tests uncorrected for multiple comparisons. This allowed us to explore whether effects were consistent with predictions. To determine learning rates (slopes) over the experiment, we performed mixed-model analyses of the entire experiment in MPlus (https://www.statmodel.com/company.shtml).

3. Results

3.1. Experiment 1: Quantifying Learning Rates Prior to Transfer (Blocks 1 to 3)

We first considered blocks 1–3, when all participants navigated in separate modalities (i.e., desktop, immersive, or real world). Considering the very first block of exposure to the environment, we found that participants in the different learning modalities did not differ in their performance either in normalized distance [$F(2, 323) = 1.99$, mean squared error (MSE) = 16.13, $p = 0.139$, $\hat{\eta}^2_G = 0.012$] nor the number of visitation errors [$F(2, 363) = 0.10$, MSE = 64.52, $p = 0.903$, $\hat{\eta}^2_G = 0.001$]. Some caution is needed, however, with these comparisons because participant variability was likely high on the first exposure to the environment. The conditions also did not differ in the second block for normalized distance [$F(2, 317) = 0.47$, MSE = 4.61, $p = 0.627$, $\hat{\eta}^2_G = 0.003$] although we saw a significantly lower number of visitation errors for real-world group ($M = 4.24$) compared to immersive-VR ($M = 7.54$) [$t(249.98) = 4.01$, $p < 0.001$, $d = 0.504$] and desktop ($M = 8.83$) [$t(241.41) = 5.39$, $p < 0.001$, $d = 0.687$].

We then compared the last block before transfer (block 3) with independent sample t-tests. We found a significant difference in visitation errors between real-world and desktop learning [$t(198.7) = 6.84$, $p < 0.001$, $d = 0.896$] and real-world and immersive VR learning [$t(179.9) = 7.40$, $p < 0.001$, $d = 0.958$], but no difference between desktop and immersive learning [$t(223.1) = 1.44$, $p = 0.150$, $d = 0.188$]. The comparison was the same for normalized distance: we found a significant difference in distances travelled in block 3 between real-world and desktop conditions [$t(161.80) = 1.99$, $p = 0.048$, $d = 0.259$] and real-world and immersive learning [$t(140.58) = 3.09$, $p = 0.002$, $d = 0.389$], but no difference between desktop and immersive learning [$t(203.90) = 1.42$, $p = 0.156$, $d = 0.187$]. These findings indicate that prior to transfer (blocks 1–3), both immersive VR and desktop resulted in less spatial knowledge than real-world navigation, particularly for visitation errors. These findings suggested that participants in both VR conditions ended the learning phase performing slightly but significantly worse than real-world participants, particularly for visitation errors.

Our findings for pointing errors (which were only collected on blocks 3 and 6) echoed the same basic pattern described above for normalized path and visitation errors. We again observed significant differences between conditions in pointing performance at the end of the first phase [$F(2, 729) = 17.78$, MSE $= 1412.53$, $p < 0.001$, $\hat{\eta}^2_G = 0.047$] (Fig. 5c). Specifically, participants in the real-world learning group ($M = 21.93$, SD $= 26.34$) performed significantly better compared to both the immersive VR group ($M = 35.23$, SD $= 37.48$) [$t(424.16) = 4.57$, $p < 0.001$, $d = -0.414$] and the desktop group ($M = 41.62$, SD $= 47.47$) [$t(342.85) = 5.57$, $p < 0.001$, $d = 0.523$]. Numerically, this trend followed real world > immersive > desktop in terms of errors, although the differences between desktop and immersive VR were not statistically significant ($p = 0.108$). Overall, these findings suggested that participants acquired information most readily in the real world, although significant degrees of learning occurred in all three conditions, with the immersive condition showing at least numerically better pointing error than desktop.

3.2. Transfer Effects (Block 3 to Block 4)

The critical area of interest in our study related to how participants transferred information from VR modalities to the real world. We addressed this issue by directly comparing normalized distance and visitation errors on the last block before the transfer phase (block 3) with the first block of the transfer phase (block 4). Visitation errors and normalized distance were each entered into a 3 (Learning Condition: real world, desktop, immersive VR) × 2 (Block: block 3, block 4) mixed-measures ANOVA. For visitation errors, there was a main effect for block [$F(1, 58) = 42.67$, MSE $= 2.19$, $p < 0.001$, $\hat{\eta}^2_G = 0.112$], a main effect of condition [$F(2, 58) = 13.66$, MSE $= 10.61$, $p < 0.001$, $\hat{\eta}^2_G = 0.281$] and a condition by block interaction effect [$F(2, 58) = 9.22$, MSE $= 2.19$, $p < 0.001$, $\hat{\eta}^2_G = 0.052$]. For normalized distance, there was a main effect of condition [$F(2, 58) = 6.65$, MSE $= 0.90$, $p = 0.003$, $\hat{\eta}^2_G = 0.110$], but no effect of block [$F(1, 58) = 2.61$, MSE $= 0.78$, $p = 0.111$, $\hat{\eta}^2_G = 0.020$] and the interaction also failed to reach significance [$= F(2, 58) = 1.89$, MSE $= 0.78$, $p = 0.161$, $\hat{\eta}^2_G = 0.029$]. These findings suggest that only visitation errors showed a significant improvement during transfer, unhindered by the change of modality, although the transfer modality (desktop *vs* immersive *vs* real world) impacted the degree of transfer.

We then ran pairwise t-tests to compare the improvement in normalized distance and visitation errors change from block 3 to block 4. For visitation errors, there were significant changes for all conditions, with effect sizes higher (in other words, more transfer) for going from the immersive VR to the real world ($d = 1.33$) compared to desktop to the real world ($d = 0.56$) (see Fig. 6a and Table 3). Here, a larger effect size means a greater reduction from block 3 to

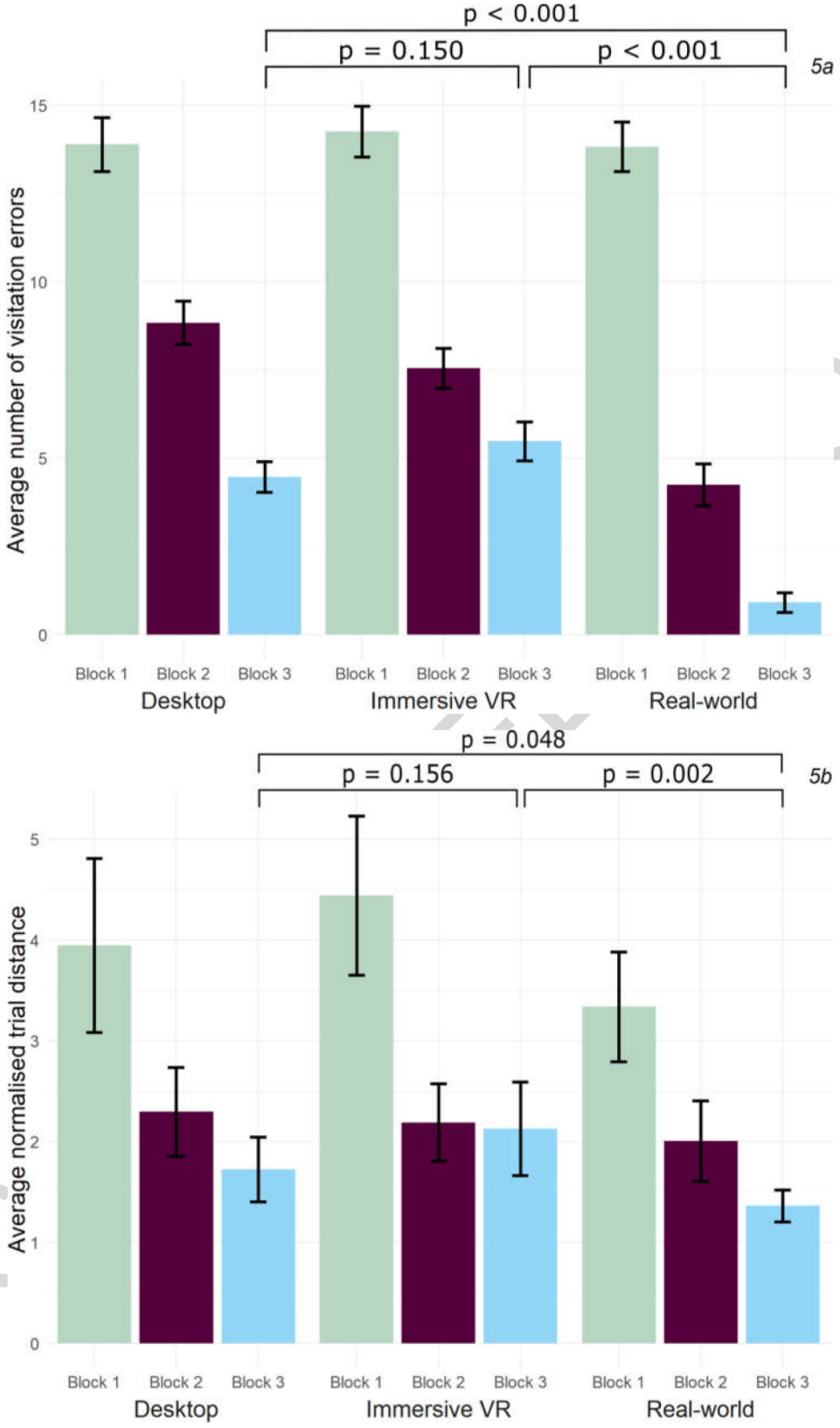


Figure 5. (a) Average number of visitation errors per block in the learning phase (blocks 1 to 3) for Experiment 1, split by learning condition (SEM error bars). (b) Average normalised distance per block in the learning phase (blocks 1 to 3) for Experiment 1, split by learning condition (SEM error bars). (c) Average absolute pointing error at the end of the learning phase for Experiment 1, split by learning condition (SEM error bars).

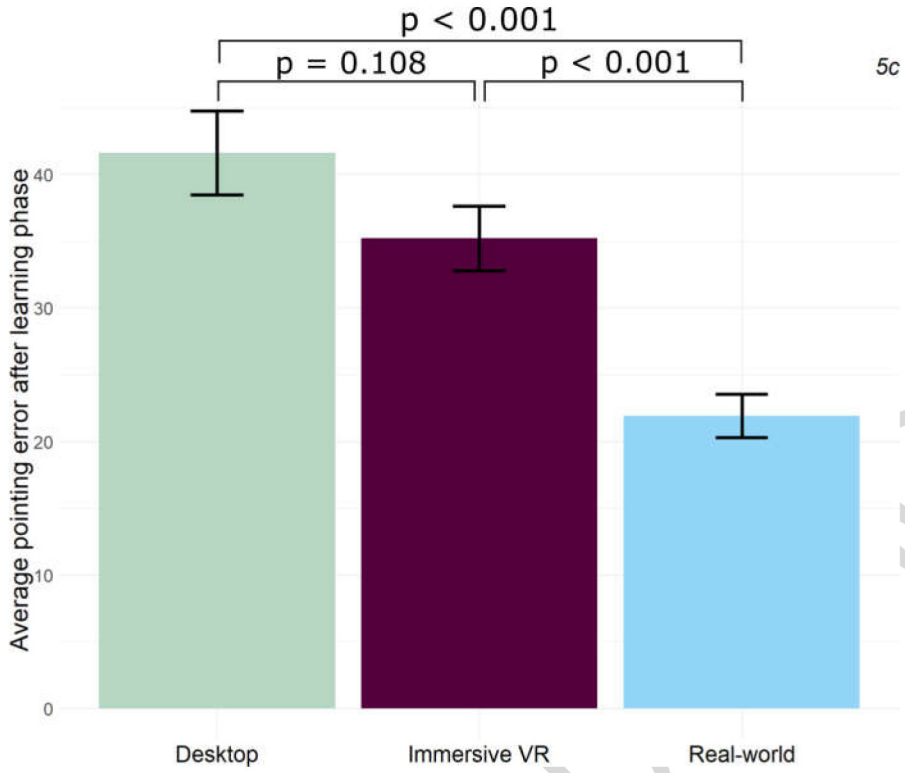


Figure 5. (Continued.)

4 and thus more transfer, suggesting that the immersive conditions resulted in greater transfer. Note that while the effect sizes were lower for real world to real world ($d = 0.55$); this was likely because participants in this condition were already significantly better at real-world navigation by block 3 (0.91 errors on average compared to 5.47 in VR and 4.46 on the desktop), and thus had less to learn. In contrast, for normalized distance, there were no significant changes from block 3 to block 4 (see Fig. 6b and Table 4). These findings suggest a slight advantage for transfer from immersive VR compared to the desktop.

Because an important goal for similar projects is to study participant's performance after being trained under different conditions of immersion, we compared performance in the first block of transfer (block 4). Because all participants navigated the real building during block 4, this also allowed us to control for potential differences between learning conditions and remove potential confounds for block 3 to block 4 comparisons. During block 4, we observed significant differences in visitation errors when comparing real world to the desktop [$t(130.84) = 5.21$, $p < 0.001$, $d = 0.709$] and immersive VR groups [$t(141.81) = 3.74$, $p < 0.001$, $d = 0.491$]. In this case, greater effect sizes between conditions mean a greater difference, i.e., a larger discrepancy between real world and desktop than real world and immersive VR. This suggests again that the immersive condition was closer in terms of transfer,

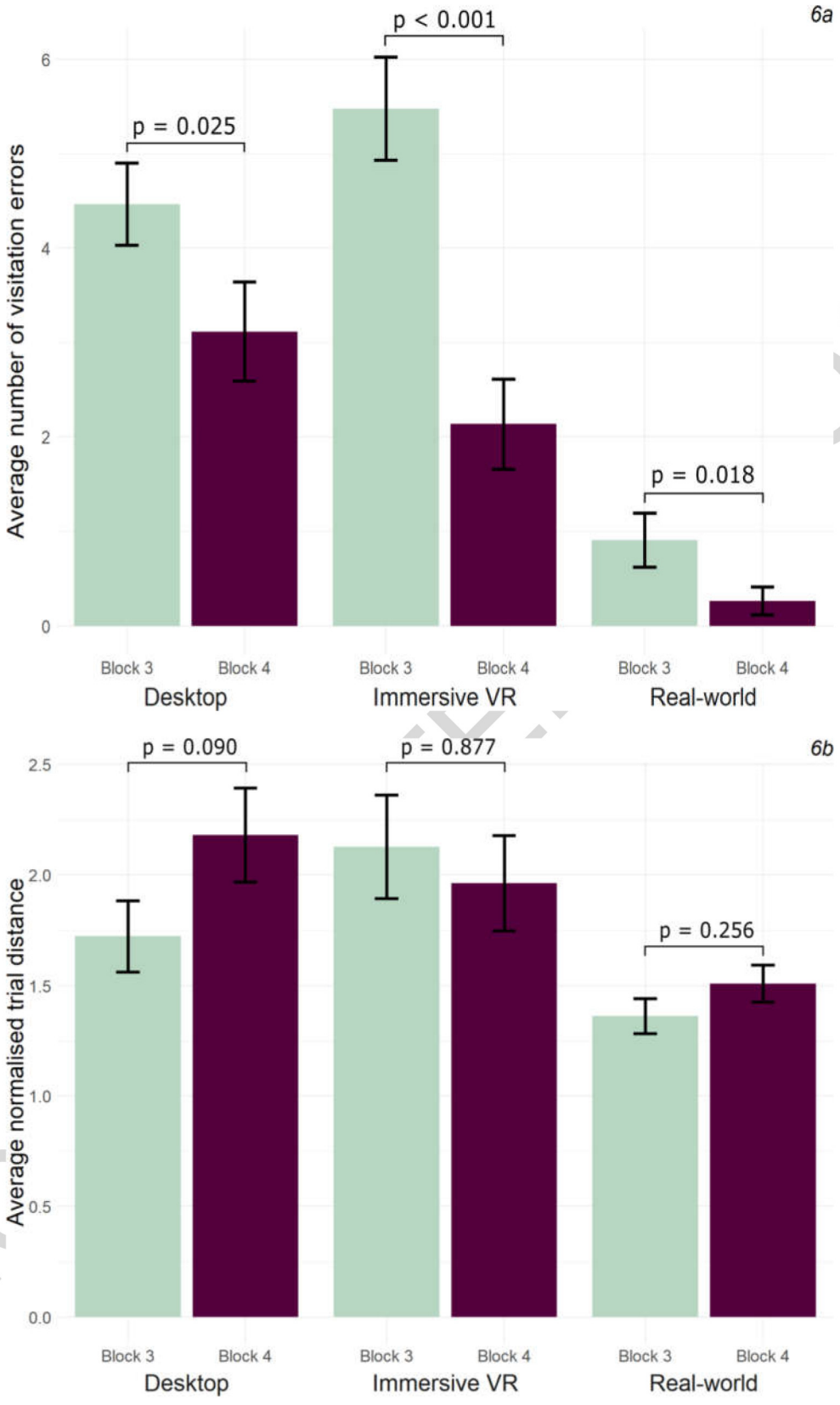


Figure 6. (a) Average number of visitation errors in blocks 3 and 4 (last learning block to first transfer block), split by learning condition (SEM error bars). (b) Average normalised distance in blocks 3 and 4 (last learning block to first transfer block), split by learning condition (SEM error bars).

Table 3.
Visitation error improvements across the three learning conditions from block 3 to block 4 — all conditions transfer to the real world in block 4

Learning condition	M	SD	M_{diff}	df	t	p	d	95% CI
Real world								
Block 3	0.91	3.27	0.64	21	2.56	0.018	0.55	0.12, 1.17
Block 4	0.27	1.70						
Desktop								
Block 3	4.46	4.65	1.35	18	2.44	0.025	0.56	0.19, 2.52
Block 4	3.11	5.62						
Immersive VR								
Block 3	5.47	6.00	3.34	19	5.97	<0.001	1.33	2.17, 4.51
Block 4	2.13	5.22						

Table 4.
Normalized distance improvements across the three learning conditions from block 3 to block 4 — all conditions transfer to the real world in block 4

Learning condition	M	SD	M_{diff}	df	t	p	d	95% CI
Real world								
Block 3	1.36	0.75	−0.15	21	−1.17	0.256	−0.25	−0.41, 0.12
Block 4	1.51	0.87						
Desktop								
Block 3	1.72	1.73	−0.70	18	−1.79	0.090	−0.41	−1.53, 0.12
Block 4	2.18	2.02						
Immersive VR								
Block 3	2.13	2.52	0.04	19	0.16	0.877	0.04	−0.55, 0.63
Block 4	1.96	2.09						

based on effect sizes, to the real world than desktop. A similar relationship was observed in normalized distance, with participants' learning in the real world performing better than on desktop [$t(118.02) = 2.95$, $p = 0.004$, $d = 0.446$] and immersive VR, although this comparison failed to reach significance [$t(119.40) = 1.96$, $p = 0.052$, $d = -0.292$]. Note, however, that immersive and desktop VR did not differ for either dependent measure [visitation errors: $t(228.45) = 1.38$, $p = 0.168$, $d = 0.181$] or normalized distance [$t(181.98) = 0.72$, $p = 0.472$, $d = 0.106$]. Given that the effect sizes suggested larger differences for desktop VR *vs* real world compared to immersive VR *vs* real world, our findings again point to a slight advantage for immersive VR in terms of transfer.

3.3. Final Performance (Block 6)

To understand how the learning modality affected the last block performance, we compared dependent measures on the 6th block, after each group had finished three blocks of navigation in the real world. Because we also collected pointing accuracy on the final block, we also included this measure. In the final testing block (block 6), we found no differences among the groups in either visitation errors [$F(2, 363) = 1.67$, MSE $= 3.33$, $p = 0.191$, $\hat{\eta}^2_G = 0.009$] nor pointing performance [$F(2, 726) = 1.56$, MSE $= 982.05$, $p = 0.210$, $\hat{\eta}^2_G = 0.004$]. These findings suggested that by the 6th repetition, all groups had reached comparable levels on visitation errors and pointing accuracy. We did find, however, a small but significant difference in normalized distance when comparing all three conditions in an ANOVA [$F(2, 255) = 3.21$, MSE $= 0.10$, $p = 0.042$, $\hat{\eta}^2_G = 0.025$]. Specifically, by the last block, we found that the real-world group ($M = 1.25$, SD $= 0.19$) performed better than immersive VR group ($M = 1.37$, SD $= 0.38$) [$t(102.13) = 2.55$, $p = 0.012$, $d = -0.426$], but not better than desktop group ($M = 1.28$, SD $= 0.37$) [$t(131.62) = 0.90$, $p = 0.369$, $d = 0.135$]. For visitation and pointing errors, these findings support the idea that all participants reached the same level of knowledge by the last block. For normalized distance, our findings suggest that real-world learning continued to confer an advantage.

3.4. Modelling Changes Over All Six Learning Blocks

Although we performed t-tests earlier based on our *a-priori* hypotheses, as explained in the introduction, we thought it also important to look at whether subjects learned over the entire experiment. We employed a mixed-effects linear model over all six learning blocks, which involved looking at changes in normalized distance and visitation errors as separate dependent measures for Experiment 1. For visitation errors and normalized path, all three conditions showed significant negative slopes over six blocks of learning, indicating that all three conditions resulted in improvements in learning (desktop $\rightarrow$ real world, immersive VR $\rightarrow$ real world, and real world $\rightarrow$ real world; Supplementary Table S1). For visitation errors, we observed an interaction effect [difference in slopes: $t(3565) = 5.38$, $p < 0.0001$] while for normalized path, there was no significant difference in slopes. The interaction effect suggested that the conditions were learned at different rates across the three conditions, which were explored earlier with t-tests.

3.5. Experiment 2: Comparing Transfer of Real World to Immersive VR

Findings from Experiment 1 suggested that the VR interfaces, particularly the treadmill, might be difficult for participants to learn as quickly as the real world. Experiment 2 assessed the 'opposite' transfer (both conditions

transferred to the immersive VR) and we therefore investigated differences between the two learning conditions (real world, immersive VR) in the last (6th) block using t-tests. The conditions did not differ in normalized distance [$t(152.12) = 0.88$, $p = 0.378$], visitation errors [$t(228.89) = 0.29$, $p = 0.771$, $d = 0.037$], nor pointing error [$t(485.14) = 0.17$, $p = 0.863$, $d = -0.016$]. These findings suggested that both immersive VR and real-world learning lead to the same eventual performance when transferring to immersive VR. Comparing both conditions immediately after transfer to the immersive VR (block 4), we found a trending relationship in average visitation errors [$t(240.09) = 1.90$, $p = 0.058$, $d = 0.242$], with the real-world learning group committing fewer errors ($M = 3.28$) than the immersive VR learning group ($M = 4.31$), but no difference for normalized distance [$t(241.53) = 0.46$, $p = 0.647$, $d = 0.059$].

We then compared all conditions from Experiment 1 (real world, immersive VR and desktop transferring to the real world) and Experiment 2 (real world and immersive VR transferring to the immersive VR). We found a significant effect of condition on final normalized distance [$F(4, 498) = 3.68$, MSE $= 0.19$, $p = 0.006$, $\hat{\eta}^2_G = 0.029$], with Tukey post-hoc tests revealing a significant difference between the real world to immersive VR group *vs* immersive VR to real world ($p < 0.001$) (see Fig. 7a and Table 5). No other comparisons reached significance for normalized distance. For visitation errors, we found significant differences between conditions in the last block across experiments [$F(4, 607)$, MSE $= 4.62$, $p < 0.001$, $\hat{\eta}^2_G = 0.181$]. Tukey's post-hoc tests suggested that the difference was due to the transfer phase modality, e.g., participants in Experiment 1 who learned the real-world condition in the transfer phase always performed better than participants in Experiment 2, who had immersive VR in the transfer phase (see Fig. 7b and Table 5).

Together, these results are consistent with our earlier findings suggesting that real-world navigation, particularly real world to real world navigation, conferred a 'transfer' advantage compared to all VR conditions. These findings also reinforce the idea that, despite capturing some aspects of real-world navigation, the treadmill interface failed to fully substitute for it. The findings, however, argue against the idea that our results are an artefact of difficulty with the treadmill interface, as participant clearly learned and transferred information from the real world to immersive VR and *vice versa*.

4. Discussion

In our experiment, we were able to directly compare acquisition of spatial knowledge in three different modalities: navigating the real world, navigating an immersive virtual interface involving walking on a treadmill and viewing

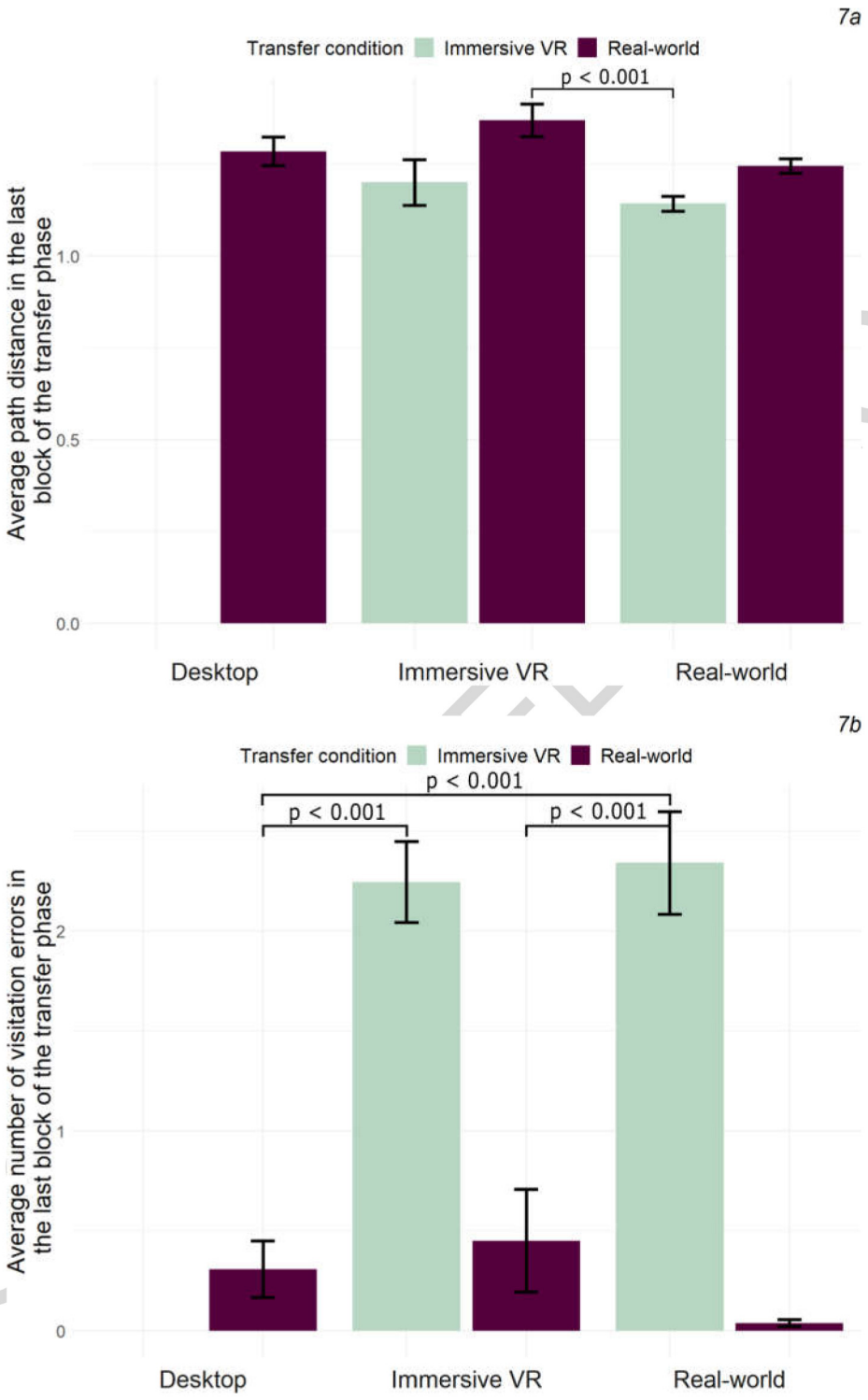


Figure 7. (a) Mean normalised distance in the last block (block 6) for each combination of learning phase (blocks 1–3, bottom labels) and transfer phase (blocks 4–6) conditions. (b) Average number of visitation errors made in the last block (block 6) for each combination of learning phase (blocks 1–3, bottom labels) and transfer phase (blocks 4–6) conditions.

Table 5.
Mean normalised distances and visitation errors in the last transfer block for both experiments. Standard deviation in parentheses

Learning condition	Transfer condition	Last block mean normalised distance	Last block mean visitation errors
Real world	real world	1.25 (0.19)	0.04 (0.19)
Desktop	real world	1.28 (0.37)	0.31 (1.51)
Immersive VR	real world	1.37 (0.38)	0.45 (2.82)
Immersive VR	immersive VR	1.20 (0.7)	2.25 (2.28)
Real world	immersive VR	1.14 (0.22)	2.34 (2.82)

the environment on an HMD, and navigating the same environment on a desktop computer with a joystick. We investigated how navigating in these three different conditions 'transferred' spatial knowledge to the same real-world environment. To fully capture the transfer process, we employed two different dependent measures interspersed with each of the six different blocks, visitation errors (visiting the wrong door) and normalized path; we also employed one dependent measure every three blocks, pointing error. For the first three blocks of learning prior to transfer, our findings show an advantage for real-world navigation compared to both immersive VR and desktop navigation on all three dependent measures. These findings suggest that participants learned the most information during real-world navigation, although, in all cases, they also showed improvements on visitation errors and normalized distance over the first three blocks in the two virtual navigation conditions, suggesting they acquired spatial information effectively from these two modalities as well.

After completing the first three blocks of navigation in a specific modality, participants transferred to the real world. Particularly for visitation errors, we found differences as a function of modality. While both virtual conditions showed less transfer to the real world than prior real-world navigation, we observed a greater decrease in visitation errors (higher effect size) during transfer for the immersive VR compared to the desktop condition. We also found that immersive VR resulted in numerically lower pointing errors after three blocks of navigation than desktop VR (Fig. 5c), although the different was not statistically significant. These findings are consistent with the idea that immersive technologies, particularly those capturing rotational and somatosensory cues associated with walking, lead to some improvements in learning compared to desktop VE (Chrastil and Warren, 2013; Grant and Magee, 1998; Klatzky *et al.*, 1998; Ruddle and Lessels, 2006; Waller *et al.*, 2004). At the same time, we did not find differences in transfer for normalized distance. We measured normalized distance using positional information in virtual conditions and Bluetooth tracking using an iPhone in the real world. Comparing normalized

distance from the second experiment involving navigating in the real world and transferring to immersive VR (or continuously navigating in immersive VR) similarly showed no differences in normalized distance, although continuous navigation in the real world did show an advantage over any navigation involving immersive VR.

Comparing the results from both experiments, we established that all learning conditions reached comparable levels of performance in the end. But we also observed that visitation errors in the last block were lower in the real world than in immersive VR. Also, participants learning in immersive VR and transferring to the real world demonstrated worse walking performance than participants transferring to immersive VR from the real world. Our findings therefore corroborate and elaborate on previous studies (Farrell *et al.*, 2003) which suggest that path and directional knowledge can be learned and transferred from VR, in many cases, quite effectively for building-sized environments, although real-world navigation still confers some advantages.

The effects we observed in terms of learning rates and transfer were more pronounced for visitation errors than normalized path. Why would we see differences in memory for specific doors for the different virtual conditions but less so, and not at all in some cases (transfer metric) for path error? Notably, the environment we tested was a medium sized campus building, the Center for Neuroscience. It is possible that the paths were relatively easier for subjects to master, and thus if we had a used a larger-scale environment, we might have seen differences in the distances of paths walked. Similarly, directional knowledge could be acquired with sufficient exposure to the building, regardless of the interface, consistent with past findings comparing desktop and real-world navigation (Richardson *et al.*, 1999). Visitation errors, in contrast, which involved remembering specific locations and names within the building, were arguably more sensitive measures of learning. This is because subjects not only needed to remember a specific location (a door) but also what name was associated with that door. Indeed, for visitation errors, when participants transferred from immersive VR to the real world we found some evidence for the expected differences as a function of modality: real world > immersive VR > desktop VE. We note, however, that it is possible that if we had used a larger-sized environment, differences would have been evident in normalized path error and directional knowledge as well.

There are two important implications. For building-sized environments, our findings, similar to past studies that compared desktop and real-world navigation (but not immersive VR), suggest that significant spatial learning can indeed take place in VR. Particularly for the accuracy of paths to targets, often taken as a measure of environment-specific knowledge (Newman *et al.*, 2007), our findings demonstrate that subjects can readily acquire such knowledge regardless of the modality. Because we only collected pointing performance on

the third and sixth blocks (to avoid disrupting subjects from the VR interfaces to have them point) we cannot say whether transfer might have differed for pointing performance throughout the learning process. Rather, whatever knowledge was needed for the real task could be learned either in the virtual conditions or rapidly acquired in the real world. This also suggests, however, that both immersive VR and desktop VE generally mimic sufficient cues, particularly visual ones, to allow subjects to acquire spatial knowledge regardless of testing modality.

The second implication is that, for arguably more sensitive measures, such as visitation errors, the 'immersiveness' of the modality can impact transfer and knowledge. As discussed previously, we suspect that visitation errors overall required more intimate memory for the spatial layout, and thus the immersive VR interface did confer a slight advantage over the desktop interface for this dependent measure. Similarly, for pointing error, rotational cues in immersive VR would confer an advantage. Notably, however, the real-world condition continued to perform better than the two virtual interfaces. This suggests that real-world navigation continues to capture cues that we could not fully emulate in VR. One reason for this might relate to encoding specificity (Tulving and Thomson, 1973), i.e., that learning will be better for conditions that more fully recapitulate visual and other multisensory cues in the 'encoded' condition compared to those that do not. Although our design did not allow us to specifically investigate this issue, we speculate that visual (i.e., the rendering of depth and other visual cues in VR), somatosensory, and proprioceptive cues, particularly related to movements of the legs and feet, could have potential implications for this difference (Matthis *et al.*, 2018). Another factor might be 'presence' (Kiryu and So, 2007), i.e., the sense that the other VR interfaces do not involve the same rules and contingencies as the real world (e.g., collisions). As VR technology improves and better approximates real-world experiences, future studies will be better able to address this issue as well.

Because of the difficulty of directly comparing virtual and real-world navigation, our setup necessarily involved some limitations. We employed a between-subjects design to try to limit any effect of re-exposure to an environment, although one limitation with between-subjects designs is increased variance due to employing different subjects in each condition. It is possible that a within-subject design could have picked up more subtle differences in the efficiency of walked paths. For immersive VR, we did find increased incidence of cybersickness and dizziness. Possibly, even low levels of discomfort in participants who completed the study could have limited the efficacy of immersive VR, to some extent. This is also a potential drawback of using the immersive VR as a replacement for desktop navigation. Although immersive VR can bring certain benefits to spatial acquisition, some participants

might be less likely to finish it. The specific reasons for cybersickness and potential ways to diminish it are therefore important to address in future studies, although we expect such issues to be less of a factor as the technology continues to improve. Finally, our use of a campus building could have potentially obscured differences, particularly in path efficiency, that would have been present during navigation of larger-scale space. Past studies have shown differences in how participants learn different scales of spaces (Meilinger *et al.*, 2016; Montello, 1998; Starrett and Ekstrom, 2018; Starrett *et al.*, 2019). It is quite possible that if participants had to navigate longer distances, such as would be required for a park or city, we would have seen an advantage for immersive over desktop VE.

5. Conclusion

Our results support the view that both immersive VR and desktop navigation are effective spatial learning tools, and, given time, participants readily acquire useable spatial representations in both modalities. Our findings also suggest, however, that immersive technologies confer a slight advantage for when tasking more sensitive measures, like remembering the location of a specific office, although this might be at the cost of an increased dropout.

Acknowledgements

We would like to thank Charlotte Mehaffey for helping to design the virtual Center for Neuroscience. This research was supported by grants from the National Science Foundation (NSF BCS-1630296, PI: ADE) and a Fulbright scholarship (Komise J. W. Fulbrighta, Prague, CZ) to LH.

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.11611632

References

Brandt, T., Schautzer, F., Hamilton, D. A., Brüning, R., Markowitsch, H. J., Kalla, R., Darlington, C., Smith, P. and Strupp, M. (2005). Vestibular loss causes hippocampal atrophy and impaired spatial memory in humans, *Brain* **128**, 2732–2741.

Chrastil, E. R. and Warren, W. H. (2013). Active and passive spatial learning in human navigation: acquisition of survey knowledge, *J. Exp. Psychol. Learn. Mem. Cogn.* **39**, 1520–1537.

J. Exp. Psychol. Learn. Mem. Cogn. **41**, 1162–1178.

Farrell, M. J., Arnold, P., Pettifer, S., Adams, J., Graham, T. and MacManamon, M. (2003). Transfer of route learning from virtual to real environments, *J. Exp. Psychol. Appl.* **9**, 219–227.

Gallistel, C. R. (1990). *The Organization of Learning. Learning, development, and conceptual change, Vol. viii*. MIT Press, Cambridge, MA, USA.

Grant, S. C. and Magee, L. E. (1998). Contributions of proprioception to navigation in virtual environments, *Hum. Factors* **40**, 489–497.

Kiryu, T. and So, R. H. Y. (2007). Sensation of presence and cybersickness in applications of virtual reality for advanced rehabilitation, *J. Neuroeng. Rehab.* **4**, 34. DOI:10.1186/1743-0003-4-34.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S. and Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion, *Psychol. Sci.* **9**, 293–298.

Lackner, J. R. and DiZio, P. (2005). Vestibular, proprioceptive, and haptic contributions to spatial orientation, *Annu. Rev. Psychol.* **56**, 115–147.

Loomis, J. M. and Beall, A. C. (1998). Visually controlled locomotion: its dependence on optic flow, three-dimensional space perception, and cognition, *Ecol. Psychol.* **10**, 271–285.

Matthis, J. S., Yates, J. L. and Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain, *Curr. Biol.* **28**, 1224–1233.

Meilinger, T., Strickrodt, M. and Bulthoff, H. H. (2016). Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation, *Cognition* **155**, 77–95.

Montello, D. R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments, in: *Spatial and Temporal Reasoning in Geographic Information Systems*, M. J. Engenhofer and R. G. Golledge (Eds), pp. 143–154. Oxford University Press, New York, NY, USA.

Montello, D. R., Waller, D., Hegarty, M. and Richardson, A. E. (2004). Spatial memory of real environment, virtual environments, and maps, in: *Human Spatial Memory: Remembering Where*, G. L. Allen (Ed.), pp. 251–285. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Newman, E. L., Caplan, J. B., Kirschen, M. P., Korolev, I. O., Sekuler, R. and Kahana, M. J. (2007). Learning your way around town: how virtual taxicab drivers learn to use both layout and landmark information, *Cognition* **104**, 231–253.

Richardson, A. E., Montello, D. R. and Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments, *Mem. Cogn.* **27**, 741–750.

Ruddle, R. A. and Lessels, S. (2006). For efficient navigational search, humans require full physical movement, but not a rich visual scene, *Psychol. Sci.* **17**, 460–465.

Russell, N. A., Horii, A., Smith, P. F., Darlington, C. L. and Bilkey, D. K. (2003). Long-term effects of permanent vestibular lesions on hippocampal spatial firing, *J. Neurosci.* **23**, 6490–6498.

Starrett, M. J. and Ekstrom, A. D. (2018). Perspective: assessing the flexible acquisition, integration, and deployment of human spatial representations and information, *Front. Hum. Neurosci.* **12**, 281. DOI:10.3389/fnhum.2018.00281.

Starrett, M. J., Stokes, J. D., Huffman, D. J., Ferrer, E. and Ekstrom, A. D. (2019). Learning-dependent evolution of spatial representations in large-scale virtual environments, *J. Exp. Psychol. Learn. Mem. Cogn.* **45**, 497–514.

Taube, J. S., Valerio, S. and Yoder, R. M. (2013). Is navigation in virtual reality with fMRI really navigation?, *J. Cogn. Neurosci.* **25**, 1008–1019.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M. and Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments?, *Presence (Camb.)* **13**, 560–571.

Tulving, E. and Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory, *Psychol. Rev.* **80**, 352–373.

Valerio, S. and Taube, J. S. (2012). Path integration: how the head direction signal maintains and corrects spatial orientation, *Nat. Neurosci.* **15**, 1445–1453.

Visell, Y., Giordano, B. L., Millet, G. and Cooperstock, J. R. (2011). Vibration influences haptic perception of surface compliance during walking, *Plos One* **6**, e17697. DOI:10.1371/journal.pone.0017697.

Waller, D., Hunt, E. and Knapp, D. (1998). The tranfer of spatial knowledge in virtual environment training, *Presence (Camb.)* **7**, 129–143.

Waller, D., Loomis, J. M. and Haun, D. B. M. (2004). Body-based senses enhance knowledge of directions in large-scale environments, *Psychon. Bull. Rev.* **11**, 157–163.

Witmer, B. G., Bailey, J. H., Knerr, B. W. and Parsons, K. C. (1996). Virtual spaces and real world places: transfer of route knowledge, *Int. J. Hum. Comput. Stud.* **45**, 413–428.

Mapping the scene and object processing networks by intracranial EEG

Vlcek Kamil[1]*, Fajnerova Iveta[1,2], Nekovarova Tereza[1,2], Hejtmanek Lukas[1], Janca Radek[3], Jezdik Petr[3], Kalina Adam[4], Tomasek Martin[5], Krsek Pavel[6], Hammer Jiri[4]§, Marusic Petr[4]§

[1]Department of Neurophysiology of Memory, Institute of Physiology, Czech Academy of Sciences, Prague, Czech Republic

[2]National Institute of Mental Health, Klecany, Czech Republic

[3]Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

[4]Department of Neurology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

[5]Department of Neurosurgery, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

[6]Department of Paediatric Neurology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

'§' Both authors contributed equally

**** Correspondence:***
Kamil Vlček
kamil.vlcek@fgu.cas.cz

Abstract length: 349 words

Figures and tables: 15

Manuscript length: 10192 words

Keywords: stereoencephalography; high-frequency gamma activity; parahippocampal place area; lateral occipital complex; human brain; visual processing; scenes; objects

1 Abstract

Human perception and cognition are based predominantly on visual information processing. Much of the information regarding neuronal correlates of visual processing has been derived from functional imaging studies, which have identified a variety of brain areas contributing to visual analysis, recognition and processing of objects and scenes. However, only two of these areas, namely the parahippocampal place area and the lateral occipital complex, were verified and further characterized by intracranial electroencephalogram (iEEG). iEEG is a unique measurement technique that samples a local neuronal population with high temporal and anatomical resolution. In the present study, we aimed to expand on previous reports and examine brain activity for selectivity of scenes and objects in the broadband high-gamma frequency range (50 - 150 Hz). We collected iEEG data from 27 epileptic patients while they watched a series of images, containing objects and scenes, and we identified 375 bipolar channels responding to at least one of these two categories. Using K-means clustering, we

delineated their brain localization. In addition to the two areas described previously, we detected significant responses in two other scene-selective areas, not yet reported by any electrophysiological studies; namely the occipital place area and the retrosplenial complex. Moreover, using iEEG we revealed a much broader networks underlying visual processing than that described to date, using specialized functional imaging experimental designs. Here we report the selective brain areas for scene processing include the posterior collateral sulcus and the anterior temporal region, which were already shown to be related to scene novelty and landmark naming. The object-selective responses appeared in the parietal, frontal, and temporal regions connected with tool use and object recognition. The temporal analyses specified the time course of the category selectivity through the dorsal and ventral visual streams. The receiver operating characteristic analyses identified the parahippocampal place area and the fusiform portion of the lateral occipital complex as being the most selective for scenes and objects, respectively. Our findings represent a valuable overview of visual processing selectivity for scenes and objects based on iEEG analyses and thus, contribute to a better understanding of visual processing in the human brain.

2 Introduction

Scene and object visual perception form the fundamentals of our understanding of the world around us. Scenes can be understood as a view of space within which we can move and act, while objects are individual parts of these scenes that we can manipulate. Early functional imaging studies revealed preferential responses to scenes in brain areas along the collateral sulcus, designated the parahippocampal place area (PPA) (Epstein and Kanwisher 1998; Aguirre et al. 1998; Ishai et al. 1999). Another scene-responsive region was described in the retrosplenial-medial parietal region (O'Craven and Kanwisher 2000), named the retrosplenial complex, or medial place area (MPA) to avoid confusion with the retrosplenial cortex (Epstein and Baker 2019). Preferential responses to scenes have also been described in the occipital cortex, in the proximity of the transverse occipital sulcus (Nakamura et al. 2000; Hasson et al. 2003). Originally, this region was labeled anatomically as the TOS by the sulcus name, but it was later renamed the occipital place area (OPA) to stress its functional localization (Dilks et al. 2013). In contrast, visual perception of everyday objects evokes a larger hemodynamic response than the perception of scrambled objects in the lateral occipital cortex extending to the posterior lateral and the basal temporal regions. This area was originally described as the lateral occipital complex (LOC) (Malach et al. 1995), and later subdivided into two functional portions (Grill-Spector et al. 1999): the posterior (labeled LO), and the anterior, localized in the posterior fusiform gyrus (labeled pFs). Nevertheless, scene and object perception are highly interconnected; object perception is dependent on scene context, and the incorporated objects influence scene recognition (Brandman and Peelen 2017).

While some of the regions responding selectively to scenes and objects are well documented in functional imaging studies, they are only partially supported by direct intracranial EEG (iEEG) recordings with high (milliseconds) temporal resolution and, in the implanted areas with a high anatomical resolution. The selectivity for scenes, around 250-300 ms after stimulus presentation, has been confirmed in the parahippocampal gyrus for both local field potentials and single-unit activity (Mormann et al. 2017), and also along the collateral sulcus near the parahippocampal/lingual boundary in the broadband gamma range (Bastin et al. 2013a; Bastin et al. 2013b). However, confirmation of the scene selectivity of the MPA and OPA, by iEEG analysis is lacking. Nonetheless, selective activity, associated with scene presentation, has been described in the hippocampus for both the firing rate and local field potential (Kraskov et al. 2007). Responses to objects within the fusiform portion of the LOC

area (pFs) were described in an early electrocorticography study with a larger N200 component in the inferior lingual, fusiform, and inferior occipital gyri (Allison et al. 1999) and later in an iEEG study for broadband gamma activity (Vidal et al. 2010). Single unit object-selective activity from the LO, with a delay of about 225 ms after the stimulus, was reported in a recent study using microelectrode grids (Decramer et al. 2019).

Most functional imaging studies focusing on scene and object perception reported the properties of the PPA, MPA, OPA, and LOC areas. However, other brain regions involved in scene and object processing have been identified using specific experimental fMRI designs. Structures of the anterior part of the medial temporal lobe, hippocampus, and parahippocampal gyrus, seem to be more active for novel, rather than familiar scenes (Rombouts et al. 2001; Kohler et al. 2002). Also, similarly to the PPA area, the anterior hippocampal region showed higher activation for scenes than for objects (Kohler et al. 2002). On the other hand, the naming of unique landmarks seems to be associated with the left temporal pole (Tranel 2006). Other cortical areas are involved in the visual processing of objects, depending on their type. Passive viewing of familiar tools is connected with higher activity in the premotor cortex and the inferior frontal gyrus (Grafton et al. 1997). The activity of the premotor cortex, together with the middle temporal gyrus and intraparietal sulcus, was increased during presentation of novel manipulatable objects after training (Weisberg et al. 2006). In contrast, recognition of familiar objects has been associated with higher activity in the inferior frontal gyrus, along the occipitotemporal sulcus and anterior parts of the fusiform and parahippocampal gyri (Bar et al. 2001) and perirhinal cortex (Clarke and Tyler 2014).

In our study, we aimed to identify the brain networks and anatomical areas facilitating scene and object processing using iEEG. To this end, we examined recordings from 27 epilepsy patients implanted with intracerebral electrodes while they were engaged in a simple visual detection task with stimuli including pictures of scenes and objects. In the analysis, we focused on the broadband gamma activity (BGA, 50-150Hz) responses, correlating with both the fMRI BOLD signal (Mukamel et al. 2005; Ojemann et al. 2009) and local neuronal firing rate (Manning et al. 2009; Hammer et al. 2016). We analyzed the iEEG data to identify the category-selective processing within a few hundred milliseconds after stimulus onset and employed the K-means clustering algorithm to group the localization of category-selective responses without any prior neuroanatomical assumptions. Using ROC analysis we evaluated the degree of discrimination between both categories. Our results confirm the scene responding PPA and object responding LOC areas, similarly to previous iEEG studies (Bastin et al. 2013a; Decramer et al. 2019). Furthermore, we describe electrical activity in two scene-selective areas, the OPA and MPA, not yet reported by electrophysiological studies. In addition, our results reveal a much broader network for scene-selective processing in the anterior temporal lobe, as well as for object-selective processing in the parietal, frontal, and temporal cortices.

3 Methods

3.1 Patients and recordings

Twenty-seven patients (15 women, median age 30 years, range 17-48 years, education level: 3 primary school, 20 secondary school, 3 college) with drug-resistant epilepsy investigated before epilepsy surgery, were recruited from the Motol Epilepsy Center in Prague. For precise localization of the seizure onset zone, the patients underwent intracranial EEG recordings (iEEG), and stereoencephalography, employing stereotactically implanted multi-contact electrodes. Recording sites were selected solely according to clinical indication with no

reference to the presented experiment. This study was approved by the Ethics Committee of Motol University Hospital and all patients gave their informed consent to participate. All patients had normal or corrected to normal vision.

3.2 Electrode implantation

11 to 15 semi-rigid electrodes per patient were implanted intracerebrally, and positioned dependent on the suspected origin of their seizures. Each electrode had a diameter of 0.8 mm and consisted of 8 to 18 contacts of 2 mm length, 1.5 mm apart (DIXI Medical Instruments). Electrode contacts were identified on patient postimplantation CT and coregistered to preimplantation MRI. The contact anatomical positions were visually verified by an experienced neurologist. The brain was normalized to Montreal Neurological Institute (MNI) space using standard Statistical Parametric Mapping algorithms (SPM 12) and all contacts were localized in the standard MNI space. The iEEG signal was recorded using a video-EEG monitoring system (Natus NicoleteOne in 22 or Natus Quantum in five patients). The data were sampled at 512, 2048 or 8000 Hz, using reference electrodes located in the white matter.

3.3 Stimuli and Task

All the patients voluntarily participated in a series of experiments focused on visual recognition and spatial orientation. The results we present here were obtained from a task exploring visual recognition of four categories of objects, designed according to the previously published PPA localizer (Vidal et al. 2010; Bastin et al. 2013b). The task lasted approximately 25 minutes and consisted of 650 pictures in total. We used pictures of three categories: scenes (referred to as 'Scenes'), small objects of daily life (referred to as 'Objects'), and faces (see Figure 1). This study focuses on responses to Scenes and Objects only. The pictures were selected from the Bank of Standardized Stimuli (BOSS) (Brodeur et al. 2010) and the SUN Database (Xiao et al. 2010). To control for a potential decrease in attention, patients were instructed to press a button each time a picture of a fruit or vegetable appeared on the screen (fourth category, visual oddball paradigm). Each category consisted of 100 different pictures (except fruits/vegetables with 25 different pictures), each repeated twice, with a pseudorandom number of other pictures in-between. All stimuli were grayscale squares, 11 cm wide, with normalized average luminance and contrast by ImageMagick® software. Stimuli were presented for a duration of 300 ms every 1100 ms in blocks of 5 pictures interleaved by 3-s pause periods to rest the eyes. Patients reported the detection of a target (fruit/vegetable) by pressing the space-bar on a keyboard and were given feedback on their performance (number of correct responses and their average reaction time) after each block. The analysis was only performed on trials in which participants did not press a key.

Visual stimuli were delivered on a 15.6 inch TFT notebook monitor with a refresh rate of 60 Hz, using the PsychoPy 1.84 environment (Peirce et al. 2019). The monitor was positioned about 60 cm from the subject's eyes, making the stimuli cover 10 deg of the visual field. We synchronized stimulus presentation and the EEG recording, using TTL pulses sent to the EEG acquisition PC with each stimulus.

3.4 Data Analysis

Time-frequency analyses of the EEG data were performed using a custom package (freely available at https://github.com/kamilvlcek/iEEG_scripts/releases/tag/v1.1.0) in MATLAB 9.4 (Mathworks, Inc.). The data were resampled to 512 Hz unless recorded at this frequency, and channels with obvious artifacts were excluded. From the EEG recording of the whole experiment, bipolar derivations were computed between adjacent electrode contacts to

suppress contributions from distant neuronal assemblies and further assumed that the bipolar EEG signals can be considered as originating from a cortical volume centered between the two contacts. We refer to the bipolar contact pairs further as 'channels'. The time-frequency analysis was focused on a broadband gamma activity (BGA, 50-150 Hz). Instantaneous amplitude was estimated using the following procedure (similar to Bastin et al. 2013a): the entire recording dataset was band-pass filtered (3rd order Butterworth filter, zero phase shift) in consecutive non-overlapping 5 Hz frequency bands in the broad gamma range (e.g., 50–55, 55–60, …, 145–150 Hz). For each band, we extracted the amplitude envelope using a Hilbert transform. The obtained envelope was down-sampled to 64 Hz. For each frequency band, the envelope was then divided by its mean value over the entire recording session, channel-wise, to whiten the EEG power spectrum and compensate for the frequency 1/f-power decay of EEG signals (Miller et al. 2009). This yielded 20 amplitude time-series between 50 and 150 Hz (one for each frequency band), which were subsequently averaged together and multiplied by 100 to obtain a single time-series of BGA power for each channel expressed in percent of the mean value. This signal was then epoched into data segments between -200 and 800 ms relative to the stimulus onset. The mean of the prestimulus interval (-50 to 0 ms) was subtracted from each epoch to remove signal changes independent of the respective stimulus. For each channel independently, epochs containing interictal epileptiform discharges identified by a spike detector implemented in MATLAB (Janca et al. 2015) were excluded from further analysis.

The BGA responses were used to identify channels selective for each stimulus category for further analysis, as follows. For all recorded EEG channels, we calculated the average BGA during the prestimulus interval (-200 – 0 ms) for all trials of the respective category and compared it with all time points between 0 and 800 ms post-stimulus using the two-sided Wilcoxon rank-sum test corrected for multiple comparisons across the time dimension and across all channels with a false discovery rate (FDR) procedure (Genovese et al. 2002). As a conservative estimate, we used a sliding window of six samples (93.75 ms) with the highest p-value. If there was a significant difference at any time point relative to the baseline for a selected stimulus category, the channel was considered as responding to that category. Channels that showed a significant response to any of the two categories (Scenes, Objects) were considered to be 'active channels'. After exclusion of channels localized in the white matter or heterotopic cortex or with a response containing obvious artifacts or appearing too late (still increasing at 800 ms, therefore with an impossible to determine magnitude for our epoch length), these channels were subject to further analysis.

To evaluate the differences in response between the two categories, we compared each channel response in both categories for all time points using the same procedure as above. The two-sided Wilcoxon rank-sum tests comparing the response to both categories were computed for all recorded EEG channels and all post-stimulus time points, and again FDR corrected for multiple comparisons across all channels and across the time dimension. A channel with a significant difference in its response to both categories was considered category-selective, either Scene- or Object-selective. The latencies of these effects were compared using two complementary methods. First, we compared the time course of each channel response to both stimulus categories by averaging the response over 100-ms time bins (similar to Bastin et al. 2013a). These means were then analyzed using a three-way repeated measure ANOVA (stimulus category vs. time bins vs. brain region/cluster) with post hoc Tukey HSD test and are reported with the effect size (η^2). Second, we used three measures of the temporal dynamics of the channel selectivity (all in ms): (1) The 'time of discrimination' (*tsig*) is the the first time point when the difference in response to both categories reached the significance level. (2) The ‘length of discrimination’ (*lensig*) is the total length of significant

difference in response to both categories. Finally, (3) the 'time of maximal discrimination' (*t90*) is the time when the difference in power change in response to both stimulus categories reached 90% of its maximum for the first time. As this last measure (*t90*) is computed from the difference magnitude , and not time course of significance as *tsig* and *lensig*, it can occasionally give distinct results.

To compare the magnitude of the individual channel responses, we calculated the maximum positive power change for each channel for both stimulus categories. This value is referred to as 'response magnitude' in the following text. ANOVA with post hoc Tukey HSD test was used to compare this value between groups of channels and is reported with the effect size (η^2). χ^2 was used to test the unequal distribution of channel selectivity between the brain regions. In all statistical tests, we used the significance level of $p<0.05$.

We used K-means clustering with cityblock distance metrics to segment the MNI locations of the category-selective channels, as implemented with the 'kmeans' function in Matlab, according to a procedure published previously for iEEG data (Engell and McCarthy 2014). Using silhouette analysis, we estimated the optimal number of clusters, with all channels being closest to the assigned cluster centroid and most far from others. If these clusters were unstable (i.e., with different centroid positions or different assignment of channels to clusters) over several runs of 'kmeans', we lowered their number until a stable solution was reached. To increase the cluster stability, we implemented a recent seed initialization method (von Luxburg 2010). Because of the rather low number of category-selective channels, the right and left hemisphere channels were pooled together by using absolute values of MNI 'x' coordinates. Therefore, each cluster can contain both left and right hemisphere channels.

To assess the response selectivity for individual stimulus categories, we used a receiver operating characteristic (ROC) binary classifier from signal detection theory (Green and Swets 1966). The area under the curve (AUC) index was estimated from the response size to Scenes and Objects for each time point for each channel. For channels responding more to Scenes than Objects, we evaluated the power to discriminate Scenes from Objects and vice versa.

4 Results

4.1 Behavioral results

The patients mainly responded correctly to fruits or vegetables (error rate 5.3 ± 1.9%) and did not respond to other categories (error rate 0.78 ± 0.2%). The average response time for fruits or vegetables was 542 ± 13 ms.

4.2 Significantly activated channels

Overall, 2707 bipolar channels (Figure 2) were obtained from the 27 patients, with more recording sites being in the right hemisphere (64%) than the left. A significant response to at least one category, Scenes or Objects, relative to the baseline (-200 to 0 ms, relative to stimulus onset), was identified in 448 (16.5%) channels. Of these, 73 were excluded due to; white matter or heterotopic cortex localization, the response being an artifact or appearing too late (see Methods). The remaining 375 channels constitute the basic set for the analysis. Out of these, 71 were labeled 'epileptic', i.e. either located in the seizure onset zone or manifesting high interictal epileptiform activity. To compare epileptic and non-epileptic channels we used two-way ANOVA for the channels, which responded to both Scenes and Objects, with the Scene vs. Object response as repeated measures factor. To compare the response time of the

individual channel responses, we calculated the time in ms when the positive power change for both stimulus categories reached 90% of its maximum for the first time. We found no difference in the response magnitude ($F(1,175)<=0.001$, $p=0.98$, $\eta^2<0.01$) or the response time ($F(1,175)=1.060$, $p=0.31$, $\eta^2<0.01$). Similarly, two-way ANOVA for the channel responding to either Scenes or Objects, with the Scene vs. Object as a factor, did not reveal a significant difference in response magnitude ($F(1,194)=0.368$, $p=0.54$, $\eta^2<0.01$) nor response time ($F(1,194)=1.66$, $p=0.20$, $\eta^2<0.01$). Despite the epileptic activity, these channels seemed to be functional and the epileptic activity did not correlate with our visual oddball paradigm. The epileptic channels were therefore included in the analysis. Note, however, that all epochs showing epileptic activity were excluded (see Methods).

Of the 375 channels, relative to the baseline, the highest number of channels (177, 47%) responded to both categories, 123 channels (33%) responded to Objects exclusively and 75 (20%) to Scenes only. The mean responses to each stimulus category are shown in Figure 3. The channels responding to both Scenes and Objects (see Figure 3C) showed larger response magnitude and faster time of discrimination than channels responding only to Scenes (see Figure 3A, magnitude, t-test: $t(250)=4.58$,$p<0.001$; *tsig*, t-test: $t(250)=6.57$,$p<0.001$), or only to Objects (see Figure 3B, magnitude: t-test: $t(298)=6.68$,$p<0.001$; *tsig*, t-test: $t(298)=8.57$,$p<0.001$). On the contrary, the response magnitude and time of response was similar for channels responding only to Objects (magnitude 43%; time 146 ms) and only to Scenes (magnitude 42%, t-test: $t(176)=0.47$, $p=0.63$; *tsig* 152 ms, t-test: $t(176)=1.03$, $p=0.30$). Also, channels responding to both Scenes and Objects responded similarly to both categories (magnitude: Scenes 21%, Objects 20%, t-test: $t(196)=0.58$, $p=0.55$; *tsig* : Scenes 245 ms Objects 244 ms, t-test: $t(196)=0.04$, $p=0.97$).

Subsequently, we mapped the distribution of all these channels to anatomical regions of interest (ROIs) in the cortex. We grouped the anatomical location of the active 375 channels into the following 11 brain regions (see also Figure 4 and 9): (1) OC - occipital cortex (but without primary visual cortex) including the OPA (36 channels), (2) PHLG -parahippocampal and inferior lingual gyri, including the collateral sulcus and the PPA (57 channels), (3) FUG - fusiform cortex without the lateral bank of the collateral sulcus (17 channels), (4) RSC - retrosplenial cortex, superior lingual gyrus and precuneus including the MPA (25 channels), (5) PC - parietal cortex, other parts of the superior parietal lobule and inferior parietal lobule (46 channels), (6) HIP - hippocampus (22 channels), (7) LTC - lateral temporal cortex - superior, middle and inferior temporal gyrus (69 channels), (8 ATC - anterior temporal cortex - amygdala, entorhinal gyrus, temporal pole (28 channels), (9) FC - frontal cortex (61 channels), (10) INS - insula (6 channels) and (11) CC - cingulate and paracingulate cortex (8 channels).

These regions differed in the average time course of their response (see Figure 5). A three-way repeated-measures ANOVA (stimulus category vs. time bin vs. brain region) for all channels showed a significant effect of all factors and interactions (the three-way interactions: $F(80, 2912)=6.71$, $p<0.001$, $\eta^2=0.16$), except the main factor of stimulus category. Figure 5 shows the differences in response to both categories for all time bins brain labels, with marked significance. Channels in two regions responded more to Scenes than Objects; in PHLG from 100 to 400 ms and in RSC from 200 to 600 ms (post hoc test on the three-way interaction). Channels in the other three regions responded significantly more to Objects than Scenes (FUG, 100-600 ms; LTC, 200-500 ms; FC, 200-500 ms)

4.3 Selectivity of channels to Scenes and Objects and its cortical distribution

To evaluate the channel response selectivity, we directly compared responses to Scenes and Objects, at all time points after the stimulus presentation and within the epoch. Most channels (217, 58%) did not show significant differences between the two categories. However, 92 (25%) channels responded to Objects significantly more than to Scenes and 66 (18%) channels responded significantly more to Scenes than to Objects.

Scene and Object selectivity were not evenly distributed in the brain regions ($\chi^2(9, N=158)=55.40$, $p<0.001$). The Scene-selective channels were localized predominantly in the PHLG (30%), RSC (24%), OC (11%), ATC (11%) and HIP (11%) regions, while the Object-selective channels were mainly in the LTC (27%), PC (2015%) and FC (12%) regions. From another point of view, the HIP (7 of 9 channels), RSC (16/17) and PHLG (18/28) region predominantly contained the Scene-selective channels, while more Object-selective channels were observed in the FC (11/12), LTC (25/28), PC (18/22) and FUG (10/13) regions. As the INS region contained only one Object-selective region and the CC region did not contain any category-selective regions, both were excluded from further analyses. Visual inspection of the distribution of Scene- and Object-selective channels in the brain suggested differences in their mediolateral and anteroposterior position (see Figure 8). Analyzing the MNI coordinates, we found that the Object-selective channels were located more laterally (with a larger absolute MNI 'x' coordinate, $t(156)=8.35$, $p<0.001$) and more anteriorly (with a larger MNI y coordinate, $t(156)=2.01$, $p<0.05$) than the Scene-selective channels.

Temporal dynamics of selective channels in anatomical regions

One of the advantages of iEEG analysis is the possibility to analyze the precise temporal dynamics of Scene and Object selectivity. Initial information about the response time course we revealed using analysis of response differences in 100-ms time bins. Two, three-way repeated measures ANOVAs (stimulus category vs. time bin vs. brain region) for Scene- and Object-selective channels showed a significant effect of all factors and interactions (both three-way interactions: $F(56,456/656)>3.3$, $p<0.001$, $\eta^2>0.2$). Figure 6 shows the differences in response to both categories for all time bins and brain labels, with marked significance. For the Scene-selective channels (post hoc test on the three-way interaction), the first difference in response to Scenes and Objects was in PHLG (100 to 200 ms), followed by OC and RSC (200 to 300 ms). In the HIP region the selectivity emerged later (300 to 400 ms). As for the duration of the difference in the significance, the longest difference was in the PHLG and RSC region (400 ms) and shortest in the HIP (100 ms). For the Object-selective channels, the post-hoc test revealed the first significant differences in response to Objects and Scenes in OC, PHLG and LTC (100 to 200 ms), followed by PC, FUG and FC (200-300 ms) regions. The longest difference was in the LTC region (500 ms), followed by PHLG, OC, FUG and LTC regions (400 ms) and shortest in PC (300 ms) region.

To specify the time course of category selectivity with a higher time resolution, we used three measures based on our BGA sampling frequency (64Hz, see Figure 7). First, we compared the time of discrimination (*tsig*) for regions with at least two channels in both channel groups (i.e., excluding RSC and FC). A two-way ANOVA (brain region vs. category) on the time of discrimination did not reveal a significant effect of the category ($F(1,114)=0.12$, $p=0.73$), and the interaction was close to significance ($F(6,114)=2.11$, $p=0.06$). However, we found differences between the brain regions by two separate one-way ANOVAs for both categories, including the RSC and FC regions (see Table 2 for individual values). The time of discrimination of Scenes from Objects in the PHLG region was earlier than in the MTL and HIP regions ($F(7,57)=4.20$, $p<0.001$, $\eta^2=0.34$, post hoc both $p<0.05$). In contrast, the time of discrimination of Objects from Scenes was the earliest in the OC region and latest in the FC

region, later than in the FUG region ($F(7,83)=4.21$, $p<0.001$, $\eta^2=0.26$, post hoc all $p<0.05$). Concerning the length of discrimination (*lensig*), similar analysis revealed longer period of significant difference in PHLG than ATC regions ($F(7,57)=3.28$, $p<0.01$, $\eta^2=0.29$, post hoc $p<0.01$), but no differences between regions in Object-selective channels ($F(7,83)=2.16$, $p=0.05$, $\eta^2=0.15$, post hoc all $p>0.05$). There were also no differences between the regions in the time of maximal discrimination (*t90*). In general, these results parallel and confirm those using 100-ms time bins, with some exceptions in Object-selective channels. These channels in the FC regions showed a significant difference between the response to Objects and Scenes from the 200-300 ms time bin, but its time of discrimination for Objects was around 350 ms.

4.4 MNI based clustering of channel selectivity for Scenes and Objects

The 11 anatomical brain regions did not adequately portray the distribution of response selectivity seen (see Figure 8 and Figure 9). To further summarize the category-selective channel locations and avoid any prior assumptions of anatomical localization, we used the K-means clustering algorithm (Engell and McCarthy 2014, see Methods for more details). The K-means algorithm, explaining 70.9% of the total spatial variance, segmented the 66 Scene-selective channels by their MNI coordinates to seven clusters (marked as S1-S7, see Table 1, Figures 8 and 10). The centroids of these clusters were localized to the following structures: the posterior angular and medial occipital gyrus (S1), the posterior collateral sulcus at the junction with the lingual sulcus (S2), the lingual and fusiform gyrus along the middle collateral sulcus (S3), the parahippocampal and fusiform gyrus along the anterior collateral sulcus (S4), the precuneus (S5), the superior lingual gyrus and precuneus next to the retrosplenial region (S6), the anterior hippocampus (S7). Based on the anatomical position and MNI coordinates, the S1 cluster partially overlapped with the OPA, the S3 cluster with the PPA, and the S6 cluster with the MPA.

Similarly, we used the K-means algorithm to further specify the locations of the 92 Object-selective channels. The algorithm segmented these channels to seven clusters (marked as O1-O7, see Table 1, Figures 8 and 10), explaining 70.7% of the total spatial variance. The centroids of these clusters were localized in: around the posterior inferior temporal sulcus (O1), the orbitofrontal gyrus (O2), area around the anterior end of the collateral sulcus (O3), the anterior part of the fusiform gyrus (O4), the posterior part of the angular gyrus (O5), near the anterior intraparietal sulcus (O6) and near the inferior frontal sulcus(O7). The clusters O1 and O3 partially overlapped with the LOC area, its posterior (LO) and anterior (pFs) portions, respectively (but see Discussion).

Temporal dynamics of selective channels in MNI based clusters

We aimed to compare the temporal characteristics of the Scene and Object selectivity in the clusters with the anatomically defined regions. Similarly to brain regions above, we started with the analysis of response differences in 100-ms time bins. Two three-way repeated measures ANOVAs (stimulus category vs. time bin vs. cluster) for Scene- and Object-selective channels showed a significant effect of all factors and interactions (both three-way interactions: $F(48,472/680)>3.7$, $p<0.001$, $\eta^2>0.2$). The differences in response to both categories for all time bins and brain labels, with marked significance, can be seen in Figure 11. For the Scene-selective channels, the post hoc test on the three-way interaction revealed the first difference in response to Scenes and Objects in S2 and S3 clusters (100 to 200 ms), followed by S1, S4, S5 and S6 clusters (200 to 300 ms), with the last cluster S7 (300 to 400 ms). The cluster with the longest difference between both categories was S3 (500 ms), while the shortest one was cluster S4 (100 ms). For the Object-selective channels, the post hoc test

revealed the first significant differences in response to Objects and Scenes in clusters O1 and O3 (100 to 200 ms, followed by O4, O5 and O7 (200-300 ms), with the last cluster O6 (300-400 ms). The cluster with the longest difference between categories was O6 (600 ms), followed by O1 and O7 (400 ms).

To specify, with a higher time resolution, how the category selectivity develops in clusters, we again used three measures based on our BGA sampling frequency (64Hz, see Figure 12). We compared them by separate one-way ANOVAs for both groups of clusters (see Table 2 for individual values). The time of discrimination (*tsig*) of Scenes from Objects in the S3 cluster was earlier than in the S6 and S7 clusters ($F(6,59)=5.28$, $p<0.001$, $\eta^2=0.35$, all post hoc $p<0.01$). The length of discrimination (*lensig*) was longer in the S3 cluster than in the S4 and S7 clusters ($F(6,59)=4.61$, $p<0.001$, $\eta^2=0.32$, all post hoc $p<0.05$). The time of maximal discrimination (*t90*) was shorter in S4 cluster than S6 and S7 clusters and in S3 th in S6 cluster ($F(6,59)=3.58$, $p<0.005$, $\eta^2=0.27$, all post hoc $p<0.05$). In Object-selective channels, the time of discrimination (*tsig*) was the earlier in the O1 than in O7 cluster ($F(6,85)=3.72$, $p<0.005$, $\eta^2=0.21$, both post hoc $p<0.05$). The length of discrimination (*lensig*) was similar in all clusters ($F(6,85)=2.22$, $p=0.05$, $\eta^2=0.14$, no post hoc $p<0.05$), while the time of maximal discrimination (*t90*) was shorter in O1 cluster than O4 and O7 clusters ($F(6,85)=4.18$, $p<0.001$, $\eta^2=0.23$, both post hoc $p<0.05$).

These results again closely parallel and confirm the results from the analysis using 100-ms time bins, with some exceptions in object clusters. The O7 cluster showed a significant difference between the response to Objects and Scenes from the 200-300 ms time bin, earlier than cluster O6, but its time of discrimination was around 350 ms, while in cluster O6 the *tsig* was below 300 ms. Besides, cluster O3 showed the longest difference in response to both stimulus categories, 400 ms longer than cluster O6, but there were no differences in the length of discrimination (*lensig*) between the clusters.

Interestingly, we found more diverse measures of temporal dynamics in the MNI based clusters than in the anatomical brain regions. The time course in the S3 cluster, overlapping the PPA area, was similar to the PHLG region, with an early start and long discrimination between Scenes and Objects. This discrimination started late with a late maximal difference in the cluster S6, with a centroid near retrosplenial region, but these differences we did not find in the RSC region, including the retrosplenial cortex and precuneus. The time of discrimination in cluster S7, with a centroid near the anterior hippocampus, was late, similarly to the HIP region, but with also the late time of maximal discrimination, which was not paralleled in the HIP region. Also, the cluster S4, with a centroid near anterior collateral sulcus, showed the fastest time of maximal discrimination and short time of discrimination of Scenes from Objects, with no similar characteristics in any of the anatomical regions. Concerning the Object clusters, we found a fast time of discrimination and time of maximal discrimination in cluster O1 near the posterior inferior temporal sulcus, partially overlapping with the LO area, similarly to OC region (but with no *t90* difference). Late discrimination was also found in the O7 cluster with a centroid near inferior frontal sulcus, paralleled in the FC region, which, however, included more channels. The cluster O4, with a centroid near anterior fusiform gyrus, showed late category discrimination, not different from the O7 cluster, in contrast to the FUG region with faster category discrimination than the FC region. Finally, the cluster O3 near middle fusiform gyrus, partially overlapping with the pFs area, shower a very long time of difference in response to both stimulus categories in the time bins analyses, with no such long time of difference in any of the anatomical regions.

4.5 ROC analysis of the stimulus categories discrimination

Finally, we were interested in how reliably we could distinguish if the stimulus was Scene or Object from the single-trial individual channel responses. To this end, we used a ROC analysis to illustrate how well the responses of the two categories were separated, for a series of BGA magnitude thresholds (for a similar procedure see Bastin et al. 2013a). The ROC area under the curve (AUC) is a summary measure of the separation across all thresholds levels. We computed the AUC values for all post-stimulus time samples, for the separation of Scenes from Objects, as well as Objects from Scenes, and compared the maximal AUC values between the seven brain regions with more than two Scene- as well as Object-selective channels . The two-way ANOVA (brain region vs. category) revealed significant differences between the brain regions ($F(76,114)=3.83$, $p<0.005$, $\eta^2=0.17$), and significant interaction ($F(76,114)=2.61$, $p<0.05$, $\eta2=0.12$) with no differences between categories (Figure 13A). The post hoc test of the brain region factor showed that category discrimination was better for channels in the PHLG region than for the channels in the ATC and HIP regions (all $p<0.05$) in Scene-selective channels. In contrast, the FUG regions shower better category discrimination than PC ($p<0.05$) in Object-selective channels. Then, in a similar way, we also compared the maximal AUC values for the Scene and Object clusters, using two independent one-way ANOVAs (Figure 13B). The discrimination of Scenes from Objects was better for channels in the S3 cluster than in clusters S1 and S7 ($F(6,59)=3.81$, $p<0.005$,$\eta^2=0.30$, both post hoc $p<0.05$). Similarly, the discrimination of Objects from Scenes was better for channels in the O3 cluster than channels in any other Object cluster ($F(6,85)=4.28$, $p<0.001$,$\eta^2=0.23$, post hoc all $p<0.05$), except O2 and O6.

This good discrimination corresponds well with the position of the S3 and O3 clusters near the PPA and the anterior portion of the LOC (pFs), respectively. In parallel, the PHLG and FUG regions showed the best discrimination for Scenes and Objects, respectively.

5 Discussion

Our study provides a broad survey of the human cortex, searching for regions that respond in a category-selective fashion to scenes or objects with BGA power increase. We did not restrict our analysis to previously identified category-selective regions of interest, but instead tested all implanted areas for any scene or object-selective regions, whether previously identified or not. Our results reveal much broader brain networks involved in scene and object processing, than previously reported from functional imaging studies with similar experimental designs. Besides the visual perception areas in the ventral stream, we found significant activity in areas previously reported to be associated with scene novelty, scene construction, object recognition and object tool use. Taking advantage of the fast temporal resolution of iEEG, we used two complementary methods to analyze the time course of discrimination of object from scenes of vice versa. Employing ROC analysis, we also showed how reliably the analyzed areas discriminate between scenes and objects.

While almost half of the active channels responded to both categories with increased BGA power, a significant proportion of them were selective for either scenes or objects. Channels responding to scenes more than objects comprised 18% of all active channels. Most functional imaging studies have defined the PPA, MPA, and OPA as regions selective for the scenes and landscapes, when contrasting their responses to object stimuli (Epstein and Kanwisher 1998; O'Craven and Kanwisher 2000; Nakamura et al. 2000). In our study, using intracranial EEG data, we confirmed these three regions to be scene-specific. Most of our scene-selective channels were localized in the PHLG region, in the RSC and in the OC regions. However, we also found numerous scene-selective channels in other, previously unreported, brain areas, especially in other parts of the temporal lobe (the HIP and ATC regions), with most channels

in the hippocampus being selective for scenes. In addition, scene-selective channels were also localized in the parietal, frontal, and lateral temporal cortices.

Channels selective for objects constituted 25% of all active channels. Their position was generally more lateral and anterior compared to the scene-selective channels; most were found in the LTC, PC and FC regions , but object-selective channels were also apparent in the FUG, PHLG and OC regions. This distribution corresponded with the results of another human intracranial study (Vidal et al. 2010). The object selectivity in the LTC and FUG regions overlapped with the LOC area, defined by functional imaging studies.

5.1 Areas selective for scenes

To further summarize the channel locations, we segmented them into spatially defined clusters, seven with both scene-selective and object-selective channels, and identified the locations of the cluster centroids.

The first scene-selective region to be described was the PPA (Epstein and Kanwisher 1998; Aguirre et al. 1998), which typically includes portions of the posterior parahippocampal, anterior lingual, and medial fusiform gyri (Epstein and Baker 2019), along the collateral sulcus. Our S3 cluster was localized to an area with similar MNI coordinates to the PPA recently published location (Spiridon et al. 2006). It almost completely included channels only in the PHGL region. According to our ROC analysis, the degree of discrimination of scenes from objects was largest in this cluster, approaching 0.8. Another functionally defined scene-selective area, the MPA, was described near the cingulate gyrus (O'Craven and Kanwisher 2000), mostly comprising the retrosplenial cortex and the anterior precuneus. These data agree with the localization of our S6 cluster in the RSC region, specifically in the precuneus and the superior part of the lingual gyrus near retrosplenial cortex, along the banks of the parietal-occipital sulcus. The third most commonly reported scene-selective region is the OPA in the occipital lobe (Nakamura et al. 2000; Hasson et al. 2003), typically near the transverse occipital sulcus. Originally labeled the TOS, it was later renamed the OPA (Dilks et al. 2013) to emphasize its functional localization. Our S1 cluster was localized to a similar area, in the middle occipital gyrus, also encompassing channels in the posterior angular gyrus. It included channels from OC and also PC regions. Surprisingly, the degree of discrimination (i.e., the average maximal AUC value) of scenes from objects in this cluster was below 0.7, significantly lower than in the cluster S3.

Many scene-selective channels in our study were localized to the HIP region, forming about half of the S7 cluster, together with the ATC region. The hippocampus has not been routinely described as a scene-selective region; however, its association with scene processing is well known. An early PET study showed anterior hippocampal activation in response to novel scenes and also a larger response to scenes than to objects in a scene-learning task (Kohler et al. 2002). Another element of the hippocampus, the presubiculum/parasubiculum, was also found to be active during scene recall and imagination (Zeidman et al. 2015). Selectivity for spatial layouts has been described for about 30% of hippocampal neurons (Kreiman et al. 2000). The scene construction theory even proposes the main hippocampal function to be the facilitation of scene construction (Hassabis and Maguire 2009). As our task included a series of one hundred unique scenes, each repeated twice, it may have induced hippocampal activity due to estimating the novelty of the scene, although this was not the subjects' task. The individual hippocampal units seem to be highly selective in their responses, even within a category (Mormann et al. 2008), possibly explaining the lack of hippocampal activation revealed by many visual perception fMRI studies. The degree of discrimination of scenes

from objects in the cluster S7 was the lowest one, significantly lower than of the cluster S3, between a little higher discrimination in the HIP region and lower in the ATC region.

Another scene-selective area in our experiment was the region along the anterior collateral sulcus, mostly comprising the anterior parahippocampal, fusiform, and entorhinal cortex. The channels in this area formed the S4 cluster and were dispersed over PHLG, FUG, HIP and ATC regions. This area, together with the anterior hippocampus, was described to be more active in a scene recall task during correct judgments about scene novelty (Rombouts et al. 2001). Its activation in our experiment could, therefore, be connected to the novelty of half of the presented scenes and a weak familiarity with the other half.

The largest BGA responses l, were found in the S2 cluster containing five channels from three patients in the posterior lingual gyrus, at the junction of the collateral and lingual sulcus. In spite of this large response, the discrimination of scenes from objects was rather low. The channels in this cluster were located more posteriorly than the most recent probabilistic localization of the PPA area (Weiner et al. 2018). T

The last scene-selective area was localized around in the posterior precuneus (cluster S5). The precuneus activity in scene object discrimination could be associated with its role in spatial attention and its shifts (Cavanna and Trimble 2006), which are more probably in spatial scenes than single objects without background. Besides, the precuneus is involved in spatial judgments using egocentric reference frame and translation between egocentric and allocentric coordinates (Moraresku and Vlcek 2020; Byrne et al. 2007), which are also the processes more likely to occur when viewing spatial scenes than centered single objects.

5.2 Areas selective for objects

Functional imaging studies defined the lateral occipital complex (LOC) as an area responding more strongly to photographs of everyday objects than shapeless textures (Malach et al. 1995). It covers a large area from the lateral occipital cortex to the posterior temporal regions, both ventral and lateral. Subsequently, it was subdivided into two areas discriminated by their functional properties (Grill-Spector et al. 1999). Using the fMRI adaptation paradigm, the authors showed that while the more posterior part (named LO) distinguishes between the same object being translated or transformed in size, the anterior portion in the fusiform gyrus (named pFs) preferentially displays position and size invariant responses.

The O1 and O3 clusters in our analysis were localized to a similar area. Cluster O1 was comprised of channels around the anterior occipital sulcus, in the middle and superior temporal, middle occipital, inferior temporal gyri, and also the temporo-occipital transition zone and posterior angular gyrus. It included mostly channels in the LTC region. It was, therefore positioned slightly more anteriorly than the fMRI localized LO area near the lateral occipital sulcus. Cluster O3 was localized more anteriorly, covering channels mostly in the posterior part of the fusiform gyrus, but also in the inferior temporal gyrus, corresponding to the pFs area. Channels in this cluster were dispersed mostly over the FUG, PHLG and LTC regions. They showed superior object-scene discrimination, agreeing with its previously reported strong shape selectivity (Grill-Spector et al. 1999). An earlier human iEEG study demonstrated a BGA response selective for tools localized to a similar area (Vidal et al. 2010). A large number of object-selective channels in the O4 cluster was also positioned in a more anterior temporal area, comprising anterior parts of the parahippocampal gyrus, entorhinal and perirhinal cortex, temporal pole and also the inferior temporal gyrus (PHLG, LTC and ATC regions). A similar area in the temporal pole responded to the familiarity of

faces and scenes in an early PET study (Nakamura et al. 2000), suggesting its connection to recognition memory. Moreover, the perirhinal cortex seems to represent object-specific semantic information, as documented by an fMRI study (Clarke and Tyler 2014). This study also showed the gradient of semantic specificity along the ventral stream, increasing anteriorly. Object recognition was also associated with brain activity in the anterior regions of the temporal lobe in another fMRI study (Bar et al. 2001). The contrast of successful to almost successful object recognition revealed activity in the anterior parahippocampal gyrus (besides activation of the LOC area), close to our O4 cluster. The same contrast also showed activity in the inferior frontal gyrus, which, according to an earlier publication, reflects the general effort, semantic analysis, and/or general feedback processes (Bar et al. 2001). We found a number of similarly localized object-selective channels in the O7 cluster. The activity of channels in these two clusters could reflect object recognition described in the above-mentioned studies. An important distinction between the objects and scenes in our test was that the objects were all familiar from everyday life, in contrast to the scenes, which were selected to be generally unfamiliar. Therefore, the results could reveal sites responding to familiarity instead of the object specifically.

Another important characteristic of all object stimuli in our test was that they could be grasped and manipulated by hand, as we excluded any pictures containing furniture or animals. This difference relative to the scene stimuli seems to manifest in the activity of brain areas related to tool use. Several such brain regions were revealed by an fMRI study, where subjects learned how to manipulate novel objects and were scanned during their visual presentation both before and after the training (Weisberg et al. 2006). This training increased activity in four areas: mainly the fusiform gyrus (LOC area), but also the middle temporal gyrus, the left intraparietal sulcus, and the left premotor cortex. These areas correspond to the location of object-selective channels in our data: the O3 cluster in the fusiform gyrus, mentioned above and O5 in the posterior part of the angular gyrus. . The areas around the intraparietal sulcus, mainly posterior, have been associated with object graspability. In one study, the activity in the posterior intraparietal sulcus was induced by the presentation of both tools and graspable objects, relative to animals (Mruczek et al. 2013). An additional area, devoted to the execution and observation of tool action is anterior supramarginal gyrus (Orban and Caruana 2014), overlapping with next cluster O6 in our data. In a meta-analysis of seven PET studies, tools activated the left posterior middle temporal region and to a lesser degree, the supramarginal gyrus (Devlin et al. 2002). In an earlier PET study (Grafton et al. 1997), passive viewing of familiar tools was connected with activation of the premotor cortex, and also the left inferior frontal gyrus, which formed the majority of channels in cluster O7 in our data.

A small cluster O2 of three channels also appeared in the orbitofrontal gyrus. Orbitofrontal cortex is known to be involved in reward learning and decision making (Rolls 2004), but it was also shown to be activated by confidently identified visual objects bearing meaningful associations in humans (Chaumon et al. 2013). The orbitofrontal cortex also appeared in the contrast of recognized and unrecognized objects in an fMRI study (Bar et al. 2001).

5.3 Temporal scheme of processing

Using the results of two complementary analyses of the temporal dynamics of scene and object discrimination, we can discuss the overall scheme of these two categories processing. We recorded the first discrimination of Scenes from Objects in the PPA (cluster S3 and the PHLG region) at 164 ms after the stimulus. It was also the longest one in our data, with the length of discrimination of 338 ms (or spanning for 400-500 ms according to the 100-ms time

bin analysis). The onset was close to the latency of discrimination between buildings and non-building objects (170 ± 34 ms) seen in broadband gamma of a previous iEEG study focused on the PPA (Bastin et al. 2013b). The length of this effect was also similar to the length of discrimination in our data, lasting until about 550 ms and was also consistent with another intracranial EEG study documenting multiple processing stages in the PPA (Bastin et al. 2013a). The latency of response in our data was also similar to the onset of scene-selective responses previously observed in the parahippocampal LFP (Mormann et al. 2017). Higher stages of visual scene processing in the ventral cortex were estimated to occur at a similar time (141 ms) by a classification analysis on MEG data, although early visual areas discriminated individual scene images before 100 ms (Cichy et al. 2017). A more posteriorly located cluster S2 in the posterior lingual gyrus showed similarly early but shorter scene-object discrimination.

Next, the discrimination of scenes from objects appeared in several areas of both the ventral and dorsal visual streams. Ventrally, the cluster S4 near the anterior collateral sulcus showed only short duration difference between scenes and objects, at 211 ms after the stimulus. Dorsally, scene-object discrimination appeared in two areas, first close to transverse occipital sulcus near the OPA area (cluster S1) at 242 ms and second in the posterior part of precuneus (cluster S5) at 215 ms. The onset latency in cluster S1 was markedly longer than the onset of discriminated scene layout appearance in the OPA in an fMRI-MEG study (60 ms) (Henriksson et al. 2019). According to this and another study (Kamps et al. 2016), the OPA is specialized in discrimination of spatial boundaries (see also Julian et al. 2018). All our scene stimuli were mainly outdoor views of landscapes and buildings with indistinct spatial boundaries, possibly explaining the long latency of OPA scene-object discrimination. But, similarly late responses, with a latency around 300 ms, were observed for scene presentation using MEG (Sato et al. 1999), with one of the sources estimated to be a parieto-occipital junction, close to our S1 cluster.

This time range over 200 ms is in agreement with the scalp EEG experiment focused on temporal dynamics of scene processing. The P2 component, peaking at 220 ms, was described as an ERP marker for scene processing (Harel et al. 2016), showing the earliest discrimination between scenes and both objects and faces. In a follow-up parallel ERP and fMRI study, this component was localized to the scene-selective areas, OPA and PPA (Kaiser et al. 2020). But surprisingly, cluster S3 in our data, localized to the PPA, showed earlier discrimination at 164ms. Kaiser et al. (2020) also described earlier discrimination of spatially intact scenes starting at 55ms and localized to V1, close to the time of discrimination of global scene properties at 84ms in Oz channel in scalp EEG study (Lowe et al. 2018). Results in this time range below 100 ms support conclusions from an earlier iEEG study showing decoding of five visual categories at around 100 ms after the stimulus (Liu et al. 2009). In the data set of the current study, none of the active channels was in the primary visual cortex.

At a later stage of scene processing, the cluster S6, encompassing the scene-selective MPA area, showed scene object discrimination at 307 ms, which lasted for the next 244 ms. The retrosplenial cortex showed higher activation for allocentric to egocentric processing at a close time interval from 350 to 650 ms in an intracranial EEG study (Bastin et al. 2013a). In the cluster S7 near the anterior hippocampus scene-object discrimination appeared with a similar onset of 326 ms. It was about 100 ms slower than in the more posterior clusters S2 to 4, which agrees with the differences in response latency reported for single-unit hippocampal activity (Mormann et al. 2008; Quiroga 2012). The cluster with fastest Objects from Scenes discrimination was O1 partially overlapping with the object-selective area LO, having an onset latency of 215 ms. This was noticeably longer than LFP latency reported in a similar

region near the posterior portion of the inferior temporal sulcus in patients with subdural electrodes (73 ms) (Yoshor et al. 2006). This study, however, reported the latency of a simple response onset, not an analysis of the difference between other types of stimuli. Besides, our cluster O1 included channels from two regions, with OC having a significantly shorter time of discrimination (153 ms) than the LTC region (263 ms). Therefore, the O1 cluster is not probably a functionally homogeneous unit.

The object scene discrimination appeared next in one ventral and two dorsal clusters. The ventral one was cluster O3 in the posterior fusiform gyrus, corresponding to the pFs area. The discrimination between object and scene activation in this cluster occurred later than in the O1 cluster (at 255 ms) and about 100 ms later than in the OC region. A similar timing was reported for a surface-negative potential with a peak latency of around 200 ms selective to objects found in the lingual, fusiform and inferior occipital gyri in patients with subdural electrodes (Allison et al. 1999). The discrimination in the cluster O3 was also the longest one, according to the 100-ms time bins analysis, it lasted for 500-600 ms. The two dorsal clusters with a similar time of discrimination were O5 in the posterior part of the angular gyrus (257 ms) and O6 in the anterior supramarginal gyrus (291 ms). EEG and MEG studies recording from these areas have focused on the sources of the visual and auditory oddball task and localized here some of the generators of the P3 ERP component with a similar latency (Halgren et al. 1998; Brazdil et al. 2005).

At a later stage, with a significantly slower time of discrimination than the O1 cluster, three other clusters showed object scene selectivity. Cluster O4 (317 ms) was located around the anterior end of the collateral sulcus, while two clusters were in the frontal cortex, O7 (345 ms) near inferior frontal sulcus and O2 (345 ms) in the orbitofrontal gyrus. These times are considerably longer than in a MEG study focused on top-down control of object recognition in the ventral cortex (Bar et al. 2006). In this study, the orbitofrontal LFP activity associated with object recognition peaked at 130ms. This short latency may be explained by specific experimental design, with a very short masked object presentation for 26 ms and the repeated presentation of the same object.

In general, these processing schemes for scene and object recognition are in agreement with the concept of dorsal and ventral visual streams. Specifically, the high overlap in activations in individual areas, especially the long processing in the S3 (PPA) and O3 (pFs) clusters, overlapping in time with all other clusters, are consistent with the view of the visual pathway as highly interactive and recurrent network (Kravitz et al. 2013). The late and prolonged discrimination of scenes in the S6 (MPA) cluster is in agreement with its position the parieto-medial temporal pathway (Kravitz et al. 2011).

5.4 Specificities of our experimental design

Our study raises questions about the comparability of fMRI and electrophysiological experiments focused on visual image processing. The fMRI BOLD signal correlates with the local field potential signal optimally when in the BGA range (Mukamel et al. 2005), even though the correlation seems to depend on the cortex region (Conner et al. 2011). Both broadband gamma, in the range of 80-150Hz, and BOLD signal, seem to reflect inputs from neighboring neural circuits (Ojemann et al. 2009). However, our experiment, which used a simple oddball design, without any memory requirements, revealed more brain areas associated with scenes and object perception than any previously reported study using similar experimental designs. Scene responding channels appeared not only in the expected PPA, OPA and MPA regions (Spiridon et al. 2006), but also in areas associated with scene novelty,

such as the hippocampus and the anterior parahippocampal gyrus. Object responding channels were found in areas documented to be object-selective (both portions of LOC) (Grill-Spector et al. 2000), as well as in areas associated with tool use (intraparietal sulcus, supramarginal gyrus, middle temporal cortex) and object recognition (inferior frontal and orbitofrontal gyri, perirhinal cortex). This disparity may be explained by the unique approaches of our study. Firstly, we used a specific experimental setup with only outdoor scene stimuli, only graspable objects and 100 unique stimuli from each category, each repeated twice. Future studies will be necessary to clarify this view. We also defined the object selectivity relative to scene stimuli, which is different from many functional imaging studies. Secondly, we evaluated the difference between the categories at all time points within the 800 ms long epochs of the 64Hz BGA signal, and a significance noted at any time point (FDR corrected) resulted in the channel being identified as selective. This time precision is not possible in functional imaging and was not used in most iEEG studies Bastin, 2013 3397 /id;Bastin, 2013 4124 /id; Mormann, 2017 5408 /id /pt "but see "}. Thirdly, we used an event-related design with each block containing stimuli from all target categories, which is different from many functional imaging studies using blocks of single category stimuli.

Possibly, our results could be influenced by low-level spatial properties of visual stimuli. We used grayscale images with normalized average luminance and contrast, to prevent the potential effect of low-level properties of individual stimuli categories. However, both image categories still differed in some characteristics. In contrast to scenes, the object images contained a uniform background and presumably had a lower power of high spatial frequencies (SF). Responses of the scene- and object-selective areas are reportedly influenced by image SF, but to a different degree and with some controversy. According to some studies, the PPA seems to be more strongly activated by high SF (Rajimehr et al. 2011; Zeidman et al. 2012), but others report stronger response to lower SF (Peyrin et al. 2004). This contradiction may be explained by another finding of the PPA being sensitive to the interaction between SF and image contrast (Kauffmann et al. 2015). In this study, PPA responded more to low SF, but it responded more to high SF under normalized contrast. According to Kauffmann et al. (2015), the effect of SF on the MPA region seems to be also dependent on the image contrast, while OPA was activated more by high SF independently on the image contrast. High SF was preferential for also the two object-selective areas (LO and pFs) (Canario et al. 2016). Therefore, we argue that the effects observed in our study were connected to the semantic content rather than the low-level properties of the images.

5.5 Study limitations

Despite its merits, this study has a number of limitations. A disadvantage of intracranial recordings, in humans, is that coverage of the brain is inevitably limited. Although we used data from 2707 bipolar channels from 27 patients, the posterior occipital cortex was not implanted and while the coverage of the right temporal cortex was dense, it was much lower in areas such as the left frontal or parietal cortex. These regions may contain other category-selective channels that we could not investigate at this time.

The validity of the clustering method has a number of caveats. Firstly, the clusters were computed from a limited number of category-selective channels with a highly varying density over the brain. Using data from a different group of implanted patients may result in a different set of clusters, with probably higher overlap with our clusters in densely implanted regions and low overlap in more sparsely implanted areas of the brain. With a higher channel density, some clusters could split while others could merge. Secondly, similarly to an earlier publication (Engell and McCarthy 2014), we computed the clusters using channel MNI

coordinates, which however do not take into account anatomical boundaries (sulci and fissures) or cortex gyrification. The use of cortical surface distances could change some channel-cluster assignments. Thirdly, our approach using the absolute MNI 'x' coordinates would not reveal any laterality in cluster localization. This approach enabled us to cluster the low number of category-selective channels and helped avoid false cluster laterality due to uneven channel distribution in both hemispheres. However, it could erroneously merge clusters with slightly distinct localization in the left and right hemispheres. Despite these limitations, we believe the clustering method we applied revealed valuable information about the distribution of scene and object selectivity in the brain.

The data were collected from epilepsy patients and there is a possibility they could reflect inherent pathological conditions. However, trials showing any type of epileptiform activity were discarded for every channel and 'epileptic' channels did not display a different response magnitude or time from non-epileptic channels, thus, we believe that our data reflect primarily physiological mechanisms.

6 Conclusions

Here, we describe and characterize the electrophysiological activity in areas selective for scenes or objects over many cortical regions by direct iEEG recording. We confirm the scene and object response selective PPA and LOC areas, consistent with reports from functional imaging, and extend the previous iEEG reports of two scene-selective areas: the MPA and OPA. Moreover, our results extend this network into other brain areas, which have not yet been described by iEEG. Selective processing for scenes was apparent in parts of the anterior temporal lobe, associated with scene novelty, such as the hippocampus and the anterior parahippocampal gyrus. In addition, selectivity for objects appeared in areas associated with tool use, the intraparietal sulcus, supramarginal gyrus, and middle temporal cortex, as well as areas connected with object recognition, such as the inferior frontal gyrus and the perirhinal cortex. By the detailed analyses of the time course of category selectivity, we document its progress through the dorsal and ventral visual streams. The high overlap in the time of processing is consistent with the view of the visual pathway as a highly interactive network. Consequently, the main contribution of this study is our description of direct electrophysiological activity selective for scenes and objects in numerous areas of the human brain. Future studies could address the functional connectivity between these areas to shed further light on the novel network dynamics of the brain during visual perception.

7 Data Availability Statement

The datasets generated for this study are available on request from the corresponding authors.

8 Ethics Statement

The study was reviewed and approved by the Ethics Committee of Motol University Hospital (02-APR-2014). All patients gave their informed consent to participate. The protocol was conducted according to the principles and guidelines of the Declaration of Helsinki.

9 Author Contributions

KV, PM, and JH conceptualized the study. KV and IF developed the test. KV, RJ, PJ, LH, and JH implemented data analysis techniques. KV, IF, and TN conducted the experiments with the help of colleagues listed in the Acknowledgments. MT, AK, PK and PM verified the

neuroanatomy localization. KV analyzed the data and created figures with the help of IF, LH, and PJ. KV wrote the manuscript with the help of PM, JH, IF, LH and TN. All authors have critically reviewed the work and provided their final approval for publication.

10 Funding

This work was supported by GACR grants 19-11753S and 16-07690S. Institutional support for IPHYS was provided by RVO: 67985823.

11 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

12 Acknowledgments

We thank all patients for their participation, and Naďa Bednárová, Helena Buchtová, Lucie Paterová, and Jana Kalinová for their help with collecting the iEEG data.

13 Abbreviations

ATC, anterior temporal cortex brain region; AUC, area under curve; BGA, broadband gamma activity; CC, cingulate and paracingulate cortex brain region; FC, frontal cortex brain region; FDR, false discovery rate; FLPG, fusiform, lingual and parahippocampal gyri brain region; HIP, hippocampus brain region; iEEG, intracranial electroencephalography; INS, insula brain region; LOC, lateral occipital complex; LO, lateral occipital portion of LOC; LTC, lateral temporal cortex brain region; MPA, medial place area; OC, occipital cortex brain region; OPA, occipital place area; PC, parietal cortex brain region; pFs, posterior fusiform sulcus portion of LOC; RSC, retrosplenial cortex and precuneus brain region; PPA, parahippocampal place area; SF, spatial frequency

14 Footnotes

http://www.ucl.ac.uk/dot-hub

15 References

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* 21, 373-383. doi: 10.1016/S0896-6273(00)80546-2.

Allison, T., Puce, A., Spencer, D.D., and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex* 9, 415-430. doi: 10.1093/cercor/9.5.415.

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmid, A.M., Dale, A.M., Hamalainen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., and Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America* 103, 449. doi: 10.1073/pnas.0507062103.

Bar, M., Tootell, R.B.H., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., and Dale, A.M. (2001). Cortical Mechanisms Specific to Explicit Visual Object Recognition. *Neuron* 29, 529-535. doi: 10.1016/S0896-6273(01)00224-0.

Bastin, J., Committeri, G., Kahane, P., Galati, G., Minotti, L., Lachaux, J.P., and Berthoz, A. (2013a). Timing of posterior parahippocampal gyrus activity reveals multiple scene processing stages. *Hum.Brain Mapp.* 34, 1357-1370. doi: 10.1002/hbm.21515.

Bastin, J., Vidal, J.R., Bouvier, S., Perrone-Bertolotti, M., Benis, D., Kahane, P., David, O., Lachaux, J.P., and Epstein, R.A. (2013b). Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *J Neurosci* 33, 10123-10131. doi: 10.1523/JNEUROSCI.4646-12.2013.

Brandman, T., and Peelen, M.V. (2017). Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J Neurosci* 2017/07/07, 7700-7710. doi: 10.1523/Jneurosci.0582-17.2017.

Brazdil, M., Dobsik, M., Mikl, M., Hlustik, P., Daniel, P., Pazourkova, M., Krupa, P., and Rektor, I. (2005). Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. *Neuroimage* 26, 285-293. doi:10.1016/j.neuroimage.2005.01.051.

Brodeur, M.B., onne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PloS one* 5, e10773. doi:10.1371/journal.pone.0010773.

Byrne, P., Becker, S., and Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol.Rev.* 114, 340-375. doi: 10.1037/0033-295X.114.2.340.

Canario, N., Jorge, L., Loureiro Silva, M.F., berto Soares, M., and Castelo-Branco, M. (2016). Distinct preference for spatial frequency content in ventral stream regions underlying the recognition of scenes, faces, bodies and other objects. *Neuropsychologia* 87, 110-119. doi: 10.1016/j.neuropsychologia.2016.05.010 .

Cavanna, A.E., and Trimble, M.R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564-583. doi:10.1093/brain/awl004.

Chaumon, M., Kveraga, K., Barrett, L.F., and Bar, M. (2013). Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. *Cerebral Cortex* 24, 2899-2907. doi: 10.1093/cercor/bht146.

Cichy, R.M., Khosla, A., Pantazis, D., and Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage* 153, 346-358. doi: 10.1016/j.neuroimage.2016.03.063.

Clarke, A., and Tyler, L.K. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. *The Journal of Neuroscience* 34, 4766. doi: 10.1523/JNEUROSCI.2828-13.2014.

Conner, C.R., Ellmore, T.M., Pieters, T.A., DiSano, M.A., and Tandon, N. (2011). Variability of the Relationship between Electrophysiology and BOLD-fMRI across Cortical Regions in Humans. *The Journal of Neuroscience* 31, 12855. doi: 10.1523/Jneurosci.1457-11.2011.

Decramer, T., Premereur, E., Uytterhoeven, M., Van Paesschen, W., van Loon, J., Janssen, P., and Theys, T. (2019). Single-cell selectivity and functional architecture of human lateral occipital complex. *PLoS Biol* 17, e3000280. doi: 10.1371/journal.pbio.3000280.

Dempsey, L.A., Cooper, R.J., Roque, T., Correia, T., Magee, E., Powell, S., Gibson, A.P., and Hebden, J. (2015). Data-driven approach to optimum wavelength selection for diffuse optical imaging. *Journal of Biomedical Optics* 20, 1-11. 10.1117/1.JBO.20.1.016003.

Devlin, J.T., Moore, C.J., Mummery, C.J., Gorno-Tempini, M.L., Phillips, J.A., Noppeney, U., Frackowiak, R.S.J., Friston, K.J., and Price, C.J. (2002). Anatomic Constraints on Cognitive Theories of Category Specificity. *Neuroimage* 15, 675-685. doi: 10.1006/nimg.2001.1002.

Dilks, D.D., Julian, J.B., Paunov, A.M., and Kanwisher, N. (2013). The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. *The Journal of Neuroscience* 33, 1331. doi: 10.1523/Jneurosci.4081-12.2013.

Engell, A.D., and McCarthy, G. (2014). Face, eye, and body selective responses in fusiform gyrus and adjacent cortex: an intracranial EEG study. *Front Hum Neurosci* 8, 642. doi: 10.3389/fnhum.2014.00642.

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598-601. doi: 10.1038/33402.

Epstein, R.A., and Baker, C.I. (2019). Scene Perception in the Human Brain. *Annu.Rev.Vis Sci* 5, 373-397. doi: 10.1146/annurev-vision-091718-014809.

Genovese, C.R., Lazar, N.A., and Nichols, T. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *Neuroimage* 15, 870-878. doi: 10.1006/nimg.2001.1037.

Grafton, S.T., Fadiga, L., Arbib, M.A., and Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage* 6, 231-236. doi: 10.1006/nimg.1997.0293.

Green, D.M., and Swets, J.A. (1966). *Signal detection theory and psychophysics.* Wiley New York.

Grill-Spector, K., Kushnir, T., Hendler, T., and Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3, 837-843. doi: 10.1038/77754.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. *Neuron* 24, 187-203. doi: 10.1016/S0896-6273(00)80832-6.

Halgren, E., Marinkovic, K., and Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and clinical neurophysiology* 106, 156-164. doi: 10.1016/S0013-4694(97)00119-3.

Hammer, J., Pistohl, T., Fischer, J., Krsek, P., Tomasek, M., Marusc, P., Schulze-Bonhage, A., Aertsen, A., and Ball, T. (2016). Predominance of Movement Speed Over Direction in

Neuronal Population Signals of Motor Cortex: Intracranial EEG Data and A Simple Explanatory Model. *Cerebral Cortex* 26, 2863-2881. doi: 10.1093/cercor/bhw033.

Harel, A., Groen, I.I., Kravitz, D.J., Deouell, L.Y., and Baker, C.I. (2016). The Temporal Dynamics of Scene Processing: A Multifaceted EEG Investigation. *eNeuro.* 3. doi: 10.1523/ENEURO.0139-16.2016.

Hassabis, D., and Maguire, E.A. (2009). The construction system of the brain. *Philos Trans R Soc Lond B Biol Sci* 364, 1263-1271. doi: 10.1098/rstb.2008.0296.

Hasson, U., Harel, M., Levy, I., and Malach, R. (2003). Large-Scale Mirror-Symmetry Organization of Human Occipito-Temporal Object Areas. *Neuron* 37, 1027-1041. doi: 10.1016/s0896-6273(03)00144-2.

Henriksson, L., Mur, M., and Kriegeskorte, N. (2019). Rapid Invariant Encoding of Scene Layout in Human OPA. *Neuron* 103, 161-171. doi: 10.1016/j.neuron.2019.04.014.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., and Haxby, J.V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci USA* 96, 9379. doi: 10.1073/pnas.96.16.9379.

Janca, R., Jezdik, P., Cmejla, R., Tomasek, M., Worrell, G.A., Stead, M., Wagenaar, J., Jefferys, J.G., Krsek, P., Komarek, V., Jiruska, P., and Marusic, P. (2015). Detection of interictal epileptiform discharges using signal envelope distribution modelling: application to epileptic and non-epileptic intracranial recordings. *Brain Topogr* 28, 172-183. doi: 10.1007/s10548-014-0379-1.

Julian, J.B., Keinath, A.T., Marchette, S.A., and Epstein, R.A. (2018). The Neurocognitive Basis of Spatial Reorientation. *Current Biology* 28, R1059-R1073. doi: 10.1016/j.cub.2018.04.057.

Kaiser, D., Haberle, G., and Cichy, R.M. (2020). Cortical sensitivity to natural scene structure. *Hum.Brain Mapp.* 41, 1286-1295. doi: 10.1002/hbm.24875.

Kamps, F.S., Julian, J.B., Kubilius, J., Kanwisher, N., and Dilks, D.D. (2016). The occipital place area represents the local elements of scenes. *Neuroimage* 132, 417-424. doi: 10.1016/j.neuroimage.2016.02.062.

Kauffmann, L., Ramanoel, S., Guyader, N., Chauvin, A., and Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *Neuroimage* 112, 86-95. doi: 10.1016/j.neuroimage.2015.02.058.

Kohler, S., Crane, J., and Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus* 12, 718-723. doi: 10.1002/hipo.10077.

Kraskov, A., Quiroga, R.Q., Reddy, L., Fried, I., and Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. *J Cogn Neurosci* 19, 479-492. doi: 10.1162/jocn.2007.19.3.479.

Kravitz, D.J., Saleem, K.S., Baker, C.I., and Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nat Rev Neurosci* 12, 217-230. doi:10.1038/nrn3008.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences* 17, 26-49. doi: 10.1016/j.tics.2012.10.011.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience* 3, 946-953. doi: 10.1038/78868.

Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 62, 281-290. doi: 10.1016/j.neuron.2009.02.025.

Lowe, M.X., Rajsic, J., Ferber, S., and Walther, D.B. (2018). Discriminating scene categories from brain activity within 100 milliseconds. *Cortex* 106, 275-287.

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92, 8135-8139. doi: 10.1073/pnas.92.18.8135.

Manning, J.R., Jacobs, J., Fried, I., and Kahana, M.J. (2009). Broadband Shifts in Local Field Potential Power Spectra Are Correlated with Single-Neuron Spiking in Humans. *The Journal of Neuroscience* 29, 13613-13620. doi: 10.1523/JNEUROSCI.2041-09.2009.

Miller, K.J., Sorensen, L.B., Ojemann, J.G., and den Nijs, M. (2009). Power-Law Scaling in the Brain Surface Electric Potential. *PLOS Computational Biology* 5, e1000609. doi:10.1371/journal.pcbi.1000609.

Minear, M., and Park, D.C. (2004). A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers 36, 630-633. doi: 10.3758/BF03206543.

Moraresku, S., and Vlcek, K. (2020). The use of egocentric and allocentric reference frames in static and dynamic conditions in humans. *Physiological Research* in press.

Mormann, F., Kornblith, S., Cerf, M., Ison, M.J., Kraskov, A., Tran, M., Knieling, S., Quian Quiroga, R., Koch, C., and Fried, I. (2017). Scene-selective coding by single neurons in the human parahippocampal cortex. *Proc Natl Acad Sci USA* 114, 1153. doi: 10.1073/pnas.1608159113.

Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., and Koch, C. (2008). Latency and Selectivity of Single Neurons Indicate Hierarchical Processing in the Human Medial Temporal Lobe. *The Journal of Neuroscience* 28, 8865. doi: 10.1523/JNEUROSCI.1640-08.2008.

Mruczek, R.E.B., von Loga, I.S., and Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *Journal of Neurophysiology* 109, 2883-2896. doi: 10.1152/jn.00658.2012.

Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., and Malach, R. (2005). Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309, 951-954. DOI: 10.1126/science.1110913.

1014 Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K.,
1015 Ito, K., Fukuda, H., Schormann, T., and Zilles, K. (2000). Functional delineation of the
1016 human occipito-temporal areas related to face and scene processing: A PET study. *Brain* 123,
1017 1903-1912. doi: 10.1093/brain/123.9.1903.

1018 O'Craven, K.M., and Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates
1019 Corresponding Stimulus-Specific Brain Regions. *J.Cogn Neurosci.* 12, 1013-1023. doi:
1020 10.1162/08989290051137549.

1021 Ojemann, G.A., Corina, D.P., Corrigan, N., Schoenfield-McNeill, J., Poliakov, A., Zamora,
1022 L., and Zanos, S. (2009). Neuronal correlates of functional magnetic resonance imaging in
1023 human temporal cortex. *Brain* 133, 46-59. doi: 10.1093/brain/awp227.

1024 Orban, G.A., and Caruana, F. (2014). The neural basis of human tool use. *Frontiers in*
1025 *Psychology* 5, 310. 10.3389/fpsyg.2014.00310.

1026 Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Hochenberger, R., Sogo, H., Kastman, E.,
1027 and Lindel+Şv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior*
1028 *Research Methods* 51, 195-203. doi: 10.3758/s13428-018-01193-y.

1029 Peyrin, C., Baciu, M., Segebarth, C., and Marendaz, C. (2004). Cerebral regions and
1030 hemispheric specialization for processing spatial frequencies during natural scene recognition.
1031 An event-related fMRI study. *Neuroimage* 23, 698-707. doi:
1032 10.1016/j.neuroimage.2004.06.020.

1033 Quiroga, R.Q. (2012). Concept cells: the building blocks of declarative memory functions.
1034 *Nat Rev.Neurosci* 13, 587-597. doi: 10.1038/nrn3251.

1035 Rajimehr, R., Devaney, K.J., Bilenko, N.Y., Young, J.C., and Tootell, R.B.H. (2011). The
1036 "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in
1037 Humans and Monkeys. *PLOS Biology* 9, e1000608. doi:10.1371/journal.pbio.1000608.

1038 Rolls, E.T. (2004). The functions of the orbitofrontal cortex. *Brain and cognition* 55, 11-29.
1039 doi: 10.1016/S0278-2626(03)00277-X.

1040 Rombouts, S.A., Barkhof, F., Witter, M.P., Machielsen, W.C., and Scheltens, P. (2001).
1041 Anterior medial temporal lobe activation during attempted retrieval of encoded visuospatial
1042 scenes: an event-related fMRI study. *Neuroimage* 14, 67-76. doi: 10.1006/nimg.2001.0799.

1043 Sato, N., Nakamura, K., Nakamura, A., Sugiura, M., Ito, K., Fukuda, H., and Kawashima, R.
1044 (1999). Different time course between scene processing and face processing: a MEG study.
1045 *Neuroreport* 10. doi: 10.1097/00001756-199911260-00031.

1046 Spiridon, M., Fischl, B., and Kanwisher, N. (2006). Location and spatial profile of category-
1047 specific regions in human extrastriate cortex. *Human Brain Mapping* 27, 77-89. doi:
1048 10.1002/hbm.20169.

1049 Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., and Leahy, R.M. (2011). Brainstorm: a user-
1050 friendly application for MEG/EEG analysis. *Comput.Intell.Neurosci* 2011, 879716. doi:
1051 10.1155/2011/879716.

Tranel, D. (2006). Impaired naming of unique landmarks is associated with left temporal polar damage. *Neuropsychology.* 20, 1-10. doi: 10.1037/0894-4105.20.1.1.

Vidal, J.R., Ossandon, T., Jerbi, K., Dalal, S.S., Minotti, L., Ryvlin, P., Kahane, P., and Lachaux, J.P. (2010). Category-Specific Visual Responses: An Intracranial Study Comparing Gamma, Beta, Alpha, and ERP Response Selectivity. *Front Hum Neurosci* 4, 195. doi: 10.3389/fnhum.2010.00195.

von Luxburg, U. (2010). Clustering Stability: An Overview. *Foundations and Trends in Machine Learning* 2, 235-274. doi:10.1561/2200000008.

Weiner, K.S., Barnett, M.A., Witthoft, N., Golarai, G., Stigliani, A., Kay, K.N., Gomez, J., Natu, V.S., Amunts, K., Zilles, K., and Grill-Spector, K. (2018). Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *Neuroimage* 170, 373-384. doi: 10.1016/j.neuroimage.2017.04.040.

Weisberg, J., van Turennout, M., and Martin, A. (2006). A Neural System for Learning about Object Function. *Cerebral Cortex* 17, 513-521. doi: 10.1093/cercor/bhj176.

Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., and Torralba, A. (2010}. SUN database: Large-scale scene recognition from abbey to zoo. in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 3485-3492 doi: 10.1109/CVPR.2010.5539970.

Yoshor, D., Bosking, W.H., Ghose, G.M., and Maunsell, J.H.R. (2006). Receptive Fields in Human Visual Cortex Mapped with Surface Electrodes. *Cerebral Cortex* 17, 2293-2302. doi: 10.1093/cercor/bhl138.

Zeidman, P., Mullally, S.L., and Maguire, E.A. (2015). Constructing, Perceiving, and Maintaining Scenes: Hippocampal Activity and Connectivity. *Cerebral Cortex* 25, 3836-3855. doi: 10.1093/cercor/bhu266.

Zeidman, P., Mullally, S.L., Schwarzkopf, D.S., and Maguire, E.A. (2012). Exploring the parahippocampal cortex response to high and low spatial frequency spaces. *Neuroreport* 23, 503. doi: 10.1097/WNR.0b013e328353766a.

16 Figure Captions

Figure 1. Schematic representation of the timing in our task containing a set of images displaying objects, faces, and spatial scenes presented in a pseudorandom order. The stimuli were organized into groups of five, with a 3-s pauses in-between. The infrequent images of fruits or vegetables (8.3%) requiring active responses were used to keep subjects focused on the presented stimuli. All images were selected from the publicly available databases: objects, fruits and vegetables from the SUN Database (Xiao et al. 2010), spatial scenes from the Bank of Standardized Stimuli (BOSS) (Brodeur et al. 2010) and faces from the lifespan database of adult facial stimuli (Minear and Park 2004).

Figure 2. Plot of all 2707 recorded channels across 27 patients on a standard MNI brain in A) axial, B) sagittal and C) coronal plane. The channels responding to Scenes or Objects are plotted in shades of red (higher response magnitude is darker), non-responding in black. The channels were distributed over most of the cortex, but with variable density and excluding the posterior occipital cortex. L, left; R, right; A, anterior; P, posterior

Figure 3. Mean BGA power responses to both categories for all channels responding to A) Scenes, B) Objects and C) both. Left two columns: mean over channels. Right column: mean ± SEM over both channels and frequency bands 50-150 Hz, responses to Scenes are in green, responses to Objects are in blue. The red line marks the region of significant difference by FDR corrected Wilcoxon signed rank at $p < 0.05$. Note the different scale for A and B than for C.

Figure 4. Positions of all 375 active channels responding to Scenes and/or Objects plotted in standard MNI brain in sagittal (A) and axial (B) views. The size of each point corresponds to the maximum magnitude of each channel's response (to either Scenes or Objects), with the scale at the bottom left in percent signal change. Channels are marked by different colors according to the 11 brain regions . As a background, we used the adult MNI-ICBM152 head model (Dempsey et al. 2015)[1].

Figure 5. Time course of the group averaged BGA response (mean ± SEM) for all active channels as a function of the anatomically defined brain regions and stimulus type. Significance markers (*) reflect the difference between the response to Scene and Object in each 100-ms time bin. Grey dotted vertical lines mark the stimulus onset. The x-axis ticks mark the centers of the time bins, while the x-axis labels mark the boundaries of the time bins.

Figure 6. Time course of the group averaged BGA response (mean ± SEM) for Scene-selective (squares) and Object-selective (diamonds) channels, as a function of the anatomically defined brain regions and stimulus type. Solid lines mark responses to the preferred stimulus category, while the dotted line responses to the non-preferred category. Significance markers (*) reflect the difference between response to Scenes (green) and Objects (blue) in each 100-ms time bin. Same convention as in Figure 5. See also Figure 9 for the numbers of Scenes- and Object-selective channels in each brain region.

Figure 7. Measures of the timing of the BGA response of Scene- and Object-selective channels sorted by brain region. The time of discrimination (*tsig*) is marked by red squares, while the blue diamonds mark the time of maximal discrimination (*t90*). The length of discrimination (*lensig*) is marked by the dotted orange vertical line stacked on the mean time of discrimination (*tsig*). Therefore, the red triangle marks the time of discrimination plus the length of discrimination (*tsig+lensig*), with the variability of the length of discrimination. All values are mean ± SEM. The same letter (a, b, c) between measures of the same color denotes a lack of significant difference (one-way ANOVA). For example, the time of discrimination of the HIP region (a) was larger than that of the PHLG region (b) but not that of the RSC region (ab). A) Scene-selective channels. B) Object-selective channels.

Figure 8. Positions of the channels responding to Scenes more than Objects (S1-S7, shades of green, A, C and E) and channels responding to Objects more than Scenes (O1-O2, shades of blue, B, D and F) plotted in standard MNI brain in axial (A, B), sagittal (C, D) and coronal (E, F) planes. The size of each point corresponds to the maximum magnitude of each channel's response, with the scale at the bottom left in percent signal change. Channels in each cluster are marked by a different shade of green or blue. Note that centroids of clusters (crosses) are

[1] http://www.ucl.ac.uk/dot-hub

bilaterally symmetrical (see main text). As background, we used the adult MNI-ICBM152 head model (Dempsey et al. 2015)[2].

Figure 9. Distribution of channels responding to Scenes more than Objects (green), to Objects more than Scenes (blue) or to both similarly (grey) across the ten brain regions. Overlaid in white are the numbers of channels for each of these three groups of channels with the number of patients in parentheses. Green and blue letters (right) represent the clusters of Scene- and Object-selective channels. Overlaid in green and blue areas are the numbers of channels in each cluster and their distribution across the brain regions. *Legend: OC - occipital cortex, PHLG - parahippocampal and lingual gyri, FUG - fusiform gyrus, RSC - retrosplenial cortex and precuneus, PC - parietal cortex, HIP - hippocampus, LTC - lateral temporal areas, ATC - anterior temporal cortex, FC - frontal cortex, INS - insula, CC - cingulate and paracingulate cortex.*

Figure 10. Positions of all channels responding to Scenes more than Objects (S1-S7, shades of green) or to Objects more than Scenes (O1-O7, shades of blue) plotted on inflated FSAverage subject brain (produced using Brainstorm (Tadel et al. 2011)). Channels in each cluster are marked by a different shade of green or blue. Lateral (A and B) and medial (C and D) view of the left (A and C) and right (B and D) hemisphere.

Figure 11. Time course of the group averaged BGA response (mean ± SEM) for Scene-selective and Object-selective channels, as a function of the MNI based clusters and stimulus type. Significance markers (*) reflect the difference between the response to Scenes (green) and Objects (blue) in each 100-ms time bin. Same convention as in Figure 5. (A) Clusters of Scene-selective channels. (B) Clusters of Object-selective channels.

Figure 12. Measures of the timing of the BGA response of Scene- and Object-selective channels, sorted by clusters. Same convention as in Figure 7.

Figure 13. The average maximal AUC values of category-selective channels for Scenes (green circles) and Objects (blue squares). Error bars represent SEM across channels. A) Across the seven analyzed brain regions (see Figure 9 for their list). Green and blue asterisks represent significant differences in the brain region factor (two-way ANOVA at $p < 0.05$) in Scene- and Object-selective channels, respectively. Category discrimination was better for channels in the FLPG region than for the channels in the ATC and FC regions. B) Across the channels with the S1-7 and O1-7 clusters. Similarly to Figure 7, the same letter (a, b, c) between clusters of the same color mark a lack of significant difference.

17 Table captions

Table 1. List of clusters of Scene- and Object-selective channels. N, number of channels, L/S, number of channels in the left/right hemisphere, P, number of patients, X, Y, Z, MNI coordinates of the cluster centroids ± SEM. The brain structures column lists anatomical labels for all channels in the cluster, with number of channels in parentheses, the structure in bold being closest to the cluster centroid*: amg, amygdala; ang, angular gyrus; cun, cuneus; ent, entorhinal cortex; fg, fusiform gyrus; fop, frontal operculum; hi, hippocampus; ifg, inferior frontal gyrus; ins, insula; itg, inferior temporal gyrus; lg, lingual gyrus; mfg, middle frontal gyrus; mog, middle occipital gyrus; mtg, middle temporal gyrus; ofg, orbitofrontal*

[2] http://www.ucl.ac.uk/dot-hub

gyri; pcun, precuneus; phg, parahippocampal gyrus; pog, postcentral gyrus; prg, precentral gyrus; rsc, retrosplenial cortex; sfg, superior frontal gyrus; smg, supramarginal gyrus; spl, superior parietal lobule; stg, superior temporal gyrus; sub, subiculum; totz, temporo-occipital transition zone; tp, temporal pole;

					MNI coordinates		
Cluster	**N**	**L/R**	**P**	**Brain structures**	**Abs(X)**	**Y**	**Z**
S1	8	8/0	2	**ang**(4),mog(4)	38 ± 3	-76 ± 2	24 ± 2
S2	5	3/2	3	**lg**(4),cun(1)	25 ± 11	-72 ± 1	-8 ± 4
S3	13	3/10	5	**fg**(6),lg(3),phg(3),mtg(1)	30 ± 7	-45 ± 2	-7 ± 1
S4	6	1/5	5	fg(2),**phg**(2),ent(1),hi(1)	32 ± 10	-26 ± 1	-22 ± 2
S5	9	7/2	3	**pcun**(7),cun(2)	15 ± 4	-61 ± 2	27 ± 2
S6	10	0/10	4	**pcun**(4),lg(5),rsc(1)	16 ± 1	-53 ± 1	12 ± 2
S7	15	4/11	9	**hi**(6),ent(3),mtg(2),amg(1),ofg(1),sub(1),tp(1)	25 ± 7	-12 ± 3	-20 ± 1
O1	25	8/17	5	**mtg**(9),mog(5),itg(4),stg(3),ang(2),totz(2)	46 ± 9	-63 ± 2	4 ± 2
O2	3	2/1	3	**ofg**(3)	39 ± 18	29 ± 1	-17 ± 1
O3	24	13/11	9	**fg**(16),itg(5),hi(2),mtg(1)	37 ± 8	-34 ± 1	-20 ± 1
O4	12	7/5	7	itg(3),ent(2),fg(2),**phg**(2),amg(1),tp(2)	31 ± 9	-5 ± 2	-34 ± 1
O5	14	11/3	3	**ang**(10),cun(2),pcun(1),spl(1)	27 ± 6	-63 ± 3	34 ± 2
O6	5	5/0	3	smg(2),**spl**(2),pog(1)	53 ± 6	-30 ± 2	43 ± 3
O7	9	4/5	5	**ifg**(4),prg(3),fop(1),ins(1)	38 ± 13	17 ± 2	23 ± 3

Table 2: Time of discrimination (*tsig*), time of maximal discrimination (*t90*) and length of discrimination (*lensig*) (all in ms, mean ± SEM) for Scene-selective (Scene>Object) and Object-selective (Object>Scenes) channels by brain region (left) and clusters (right). The rows are sorted by the mean *tsig* value. The same letter (a, b, c) between regions and clusters in the same group, or no letter, marks a lack of significant difference. The labels for clusters present approximate locations of cluster centroids; but see Figure 9 and Table 1 for more specific cluster localization. *aCoS, anterior collateral sulcus; aHip, anterior hippocampus; aIPS, anterior intraparietal sulcus; IFS, inferior frontal sulcus; LO, posterior portion of lateral occipital complex; MPA, medial place area; OFG, orbitofrontal gyri; OPA, occipital place area; pAnG, posterior angular gyrus; PCun, precuneus; pFs, anterior portion of lateral occipital complex; pLG, posterior lingual gyrus; PPA, parahippocampal place area;*

	Scene > Object					**Scene > Object**		
	Tsig	***t90***	***lensig***	**Label**		***tsig***	***t90***	***lensig***
PHLG	166 ± 29 [b]	260 ± 29	305 ± 31 [a]	S3	PPA	164 ± 23 [ab]	256 ± 14 [a]	338 ± 41 [b]
PC	228 ± 18	296 ± 4	242 ± 55	S2	pLG	189 ± 35	244 ± 22	191 ± 36
FUG	235 ± 5	258 ± 5	208 ± 66	S4	aCoS	211 ± 33 [a]	195 ± 43	128 ± 38 [a]
LTC	235 ± 59	274 ± 45	89 ± 28	S5	PCun	215 ± 17	284 ± 14	253 ± 31
RSC	249 ± 19	317 ± 18	236 ± 36	S1	OPA	242 ± 13	294 ± 7	232 ± 28
OC	251 ± 18	298 ± 9	243 ± 39	S6	MPA	307 ± 41 [c]	380 ± 43 [b]	244 ± 56
HIP	302 ± 10 [a]	250 ± 66	152 ± 42	S7	aHip	326 ± 27 [bc]	334 ± 34 [b]	125 ± 22 [a]
ATC	359 ± 38 [a]	374 ± 36	92 ± 19 [b]					

	Object > Scene					**Object > Scene**		
	Tsig	***t90***	***Lensig***	**Label**		***Tsig***	***t90***	***lensig***
OC	153 ± 21 [c]	454 ± 112	243 ± 40 [a]	O1	LO	215 ± 19 [b]	262 ± 14 [b]	165 ± 22 [a]
FUG	233 ± 29 [bc]	389 ± 86	267 ± 48 [a]	O3	pFs	255 ± 20	320 ± 14	232 ± 31 [a]
PHLG	262 ± 21	377 ± 91	159 ± 36 [a]	O5	pAnG	257 ± 31	302 ± 29	152 ± 30 [a]
LTC	263 ± 21 [ab]	420 ± 62	155 ± 25 [a]	O6	aIPS	291 ± 33	344 ± 30	197 ± 51 [a]
PC	268 ± 23 [ab]	326 ± 77	168 ± 23 [a]	O4	aCoS	317 ± 22 [a]	348 ± 19 [a]	117 ± 25 [a]
ATC	310 ± 48 [ab]	440 ± 137	113 ± 41 [a]	O7	IFS	345 ± 19 [a]	378 ± 17 [a]	137 ± 32 [a]
HIP	336 ± 16	488 ± 313	109 ± 31 [a]	O2	OFG	363 ± 28	378 ± 24	52 ± 21 [a]
FC	353 ± 16 [a]	409 ± 82	121 ± 28 [a]					

Fig1.tif

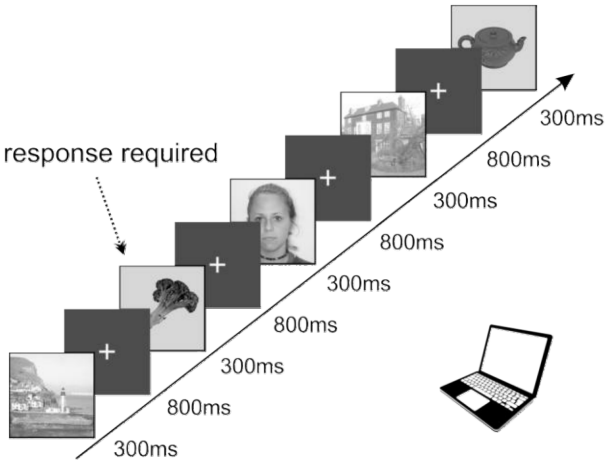

response required
300ms
800ms
300ms
800ms
300ms
800ms
300ms
800ms
300ms

Fig2.tif

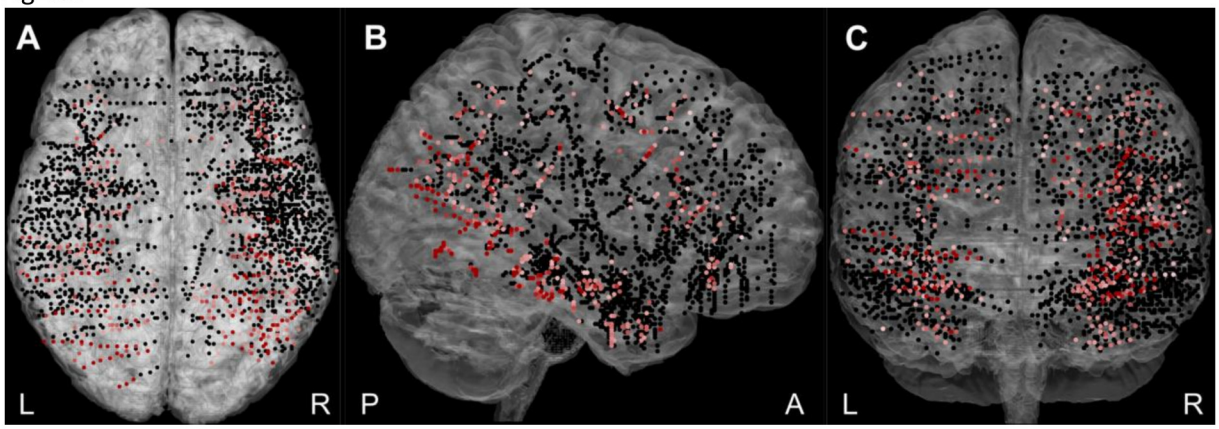


Fig3.tif

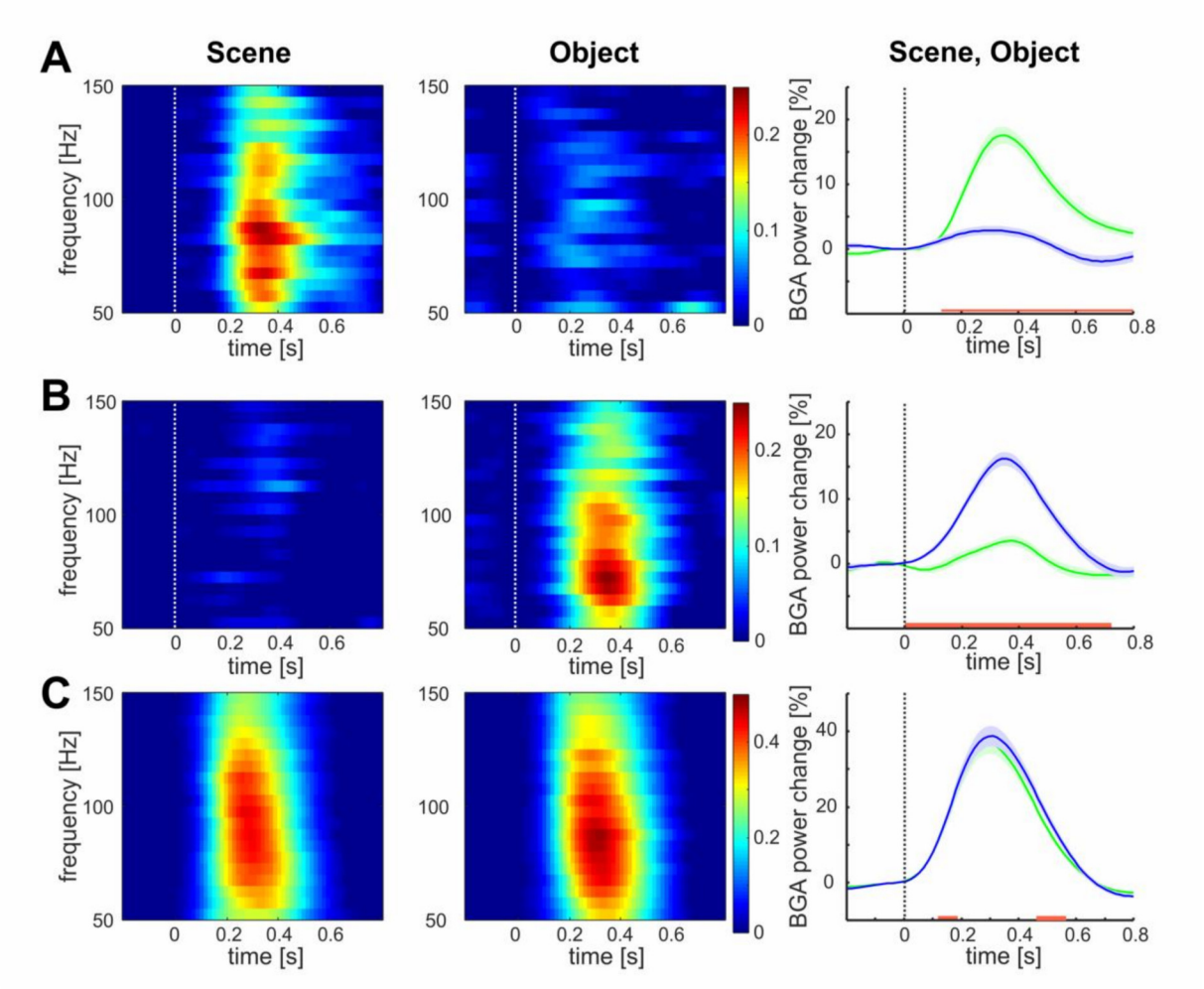

A
Scene
Object
Scene, Object
B
C
frequency [Hz]
time [s]
BGA power change [%]

Fig4.tif

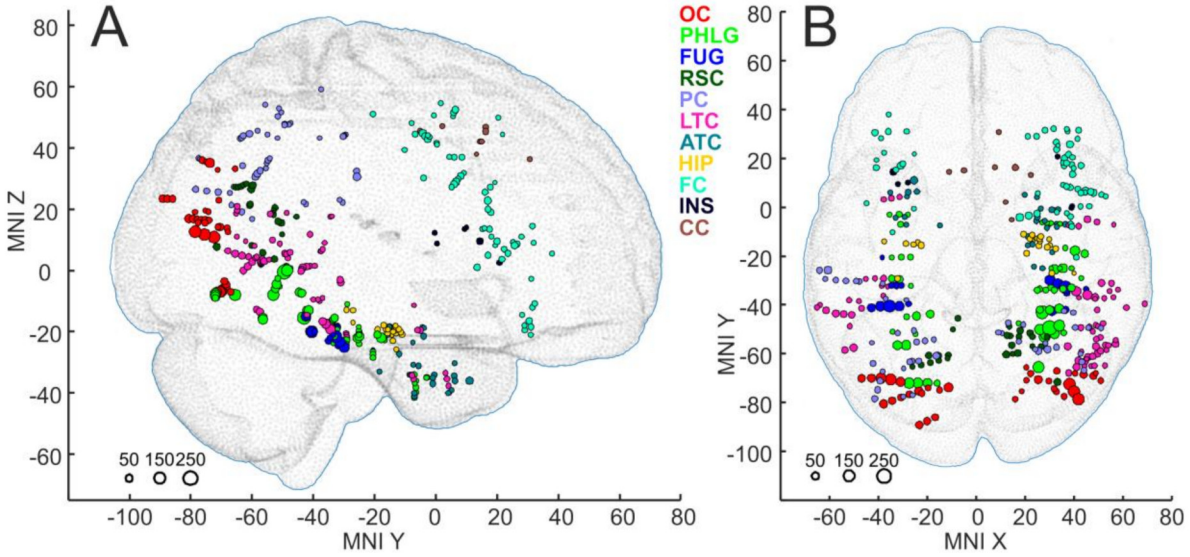

A
B
OC
PHLG
FUG
RSC
PC
LTC
ATC
HIP
FC
INS
CC
MNI Z
MNI Y
MNI X
50 150 250
50 150 250

Fig5.tif

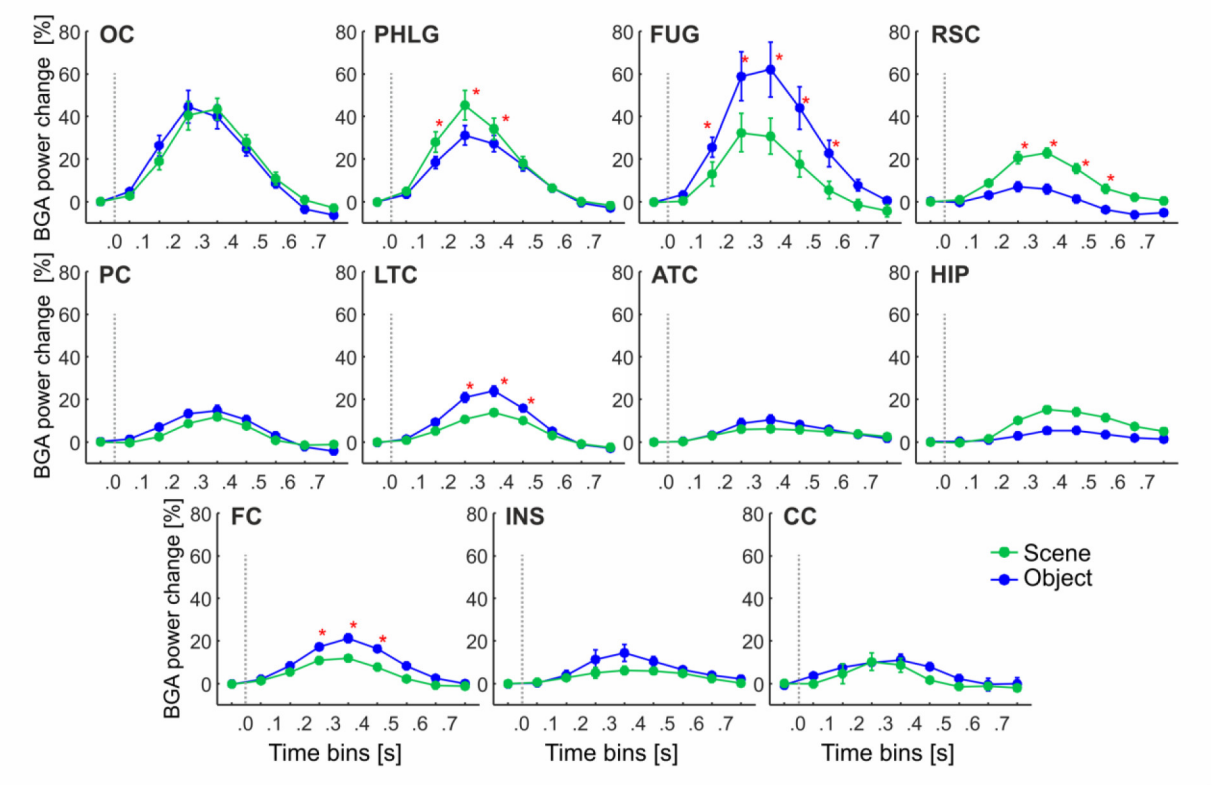

OC
PHLG
FUG
RSC
PC
LTC
ATC
HIP
FC
INS
CC
BGA power change [%]
Time bins [s]
Scene
Object

Fig6.tif

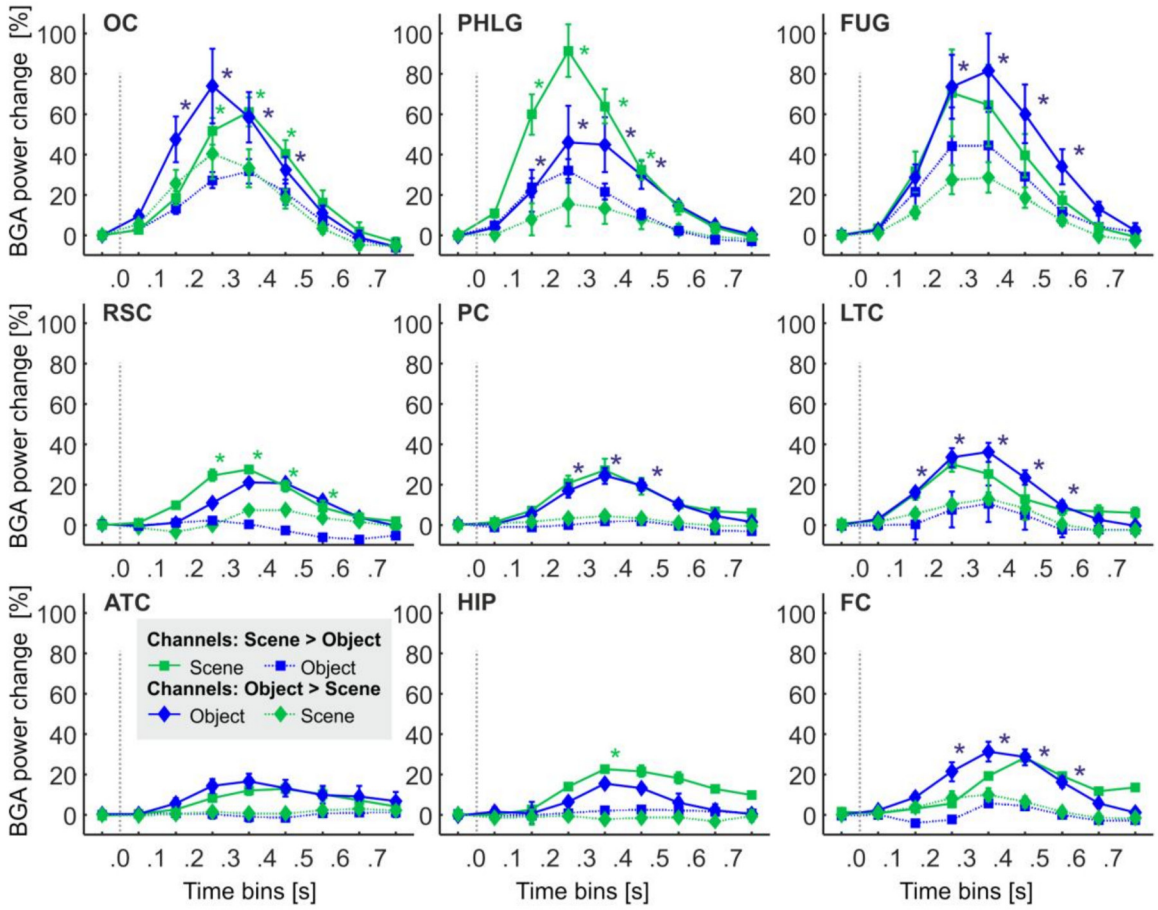

OC
PHLG
FUG
RSC
PC
LTC
ATC
HIP
FC
BGA power change [%]
Time bins [s]
Channels: Scene > Object
Scene
Object
Channels: Object > Scene
Object
Scene

Fig7.tif

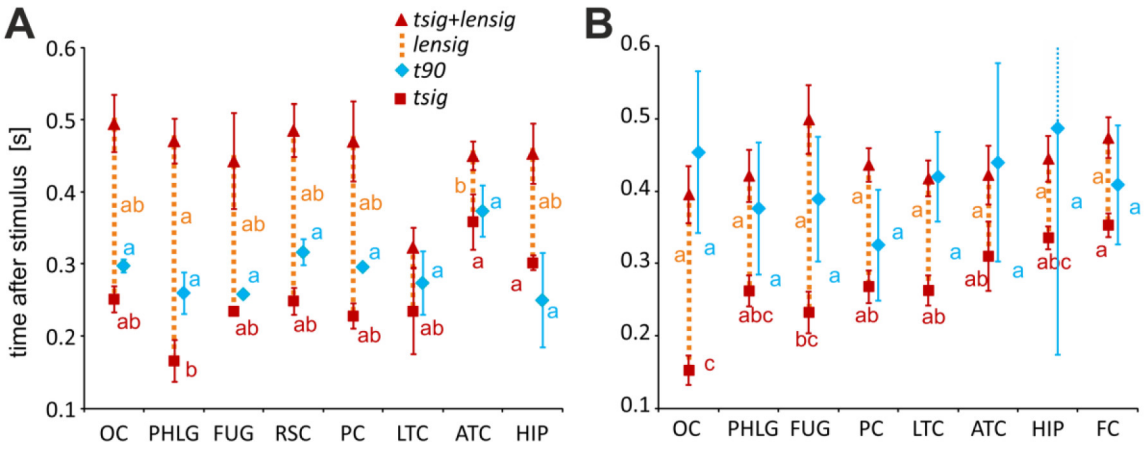

A
B
tsig+lensig
lensig
t90
tsig
time after stimulus [s]
0.6
0.5
0.4
0.3
0.2
0.1
OC
PHLG
FUG
RSC
PC
LTC
ATC
HIP
FC

Fig8.tif

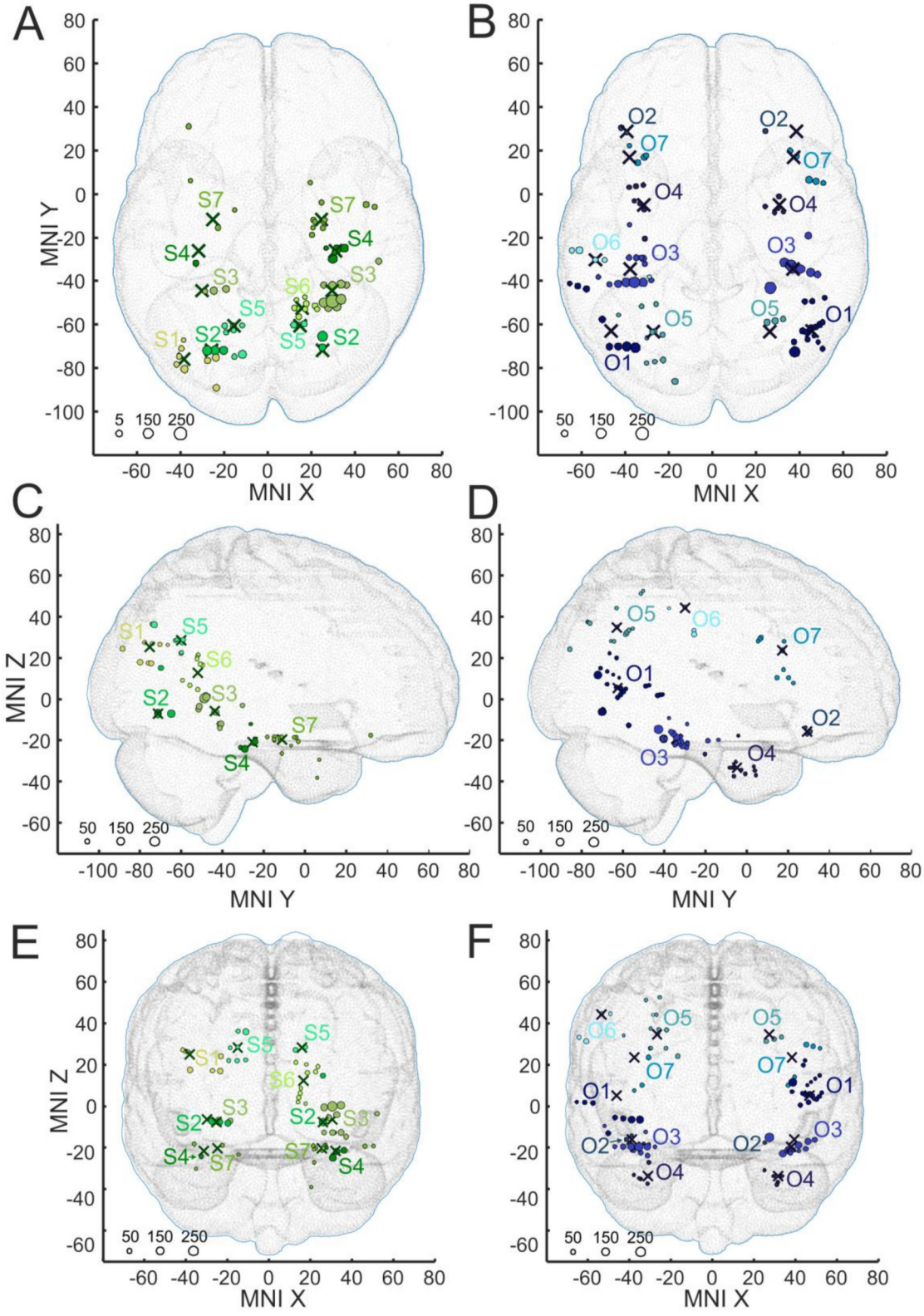

A
B
C
D
E
F
MNI X
MNI Y
MNI Z
S1
S2
S3
S4
S5
S6
S7
O1
O2
O3
O4
O5
O6
O7
5 150 250
50 150 250

Fig9.tif

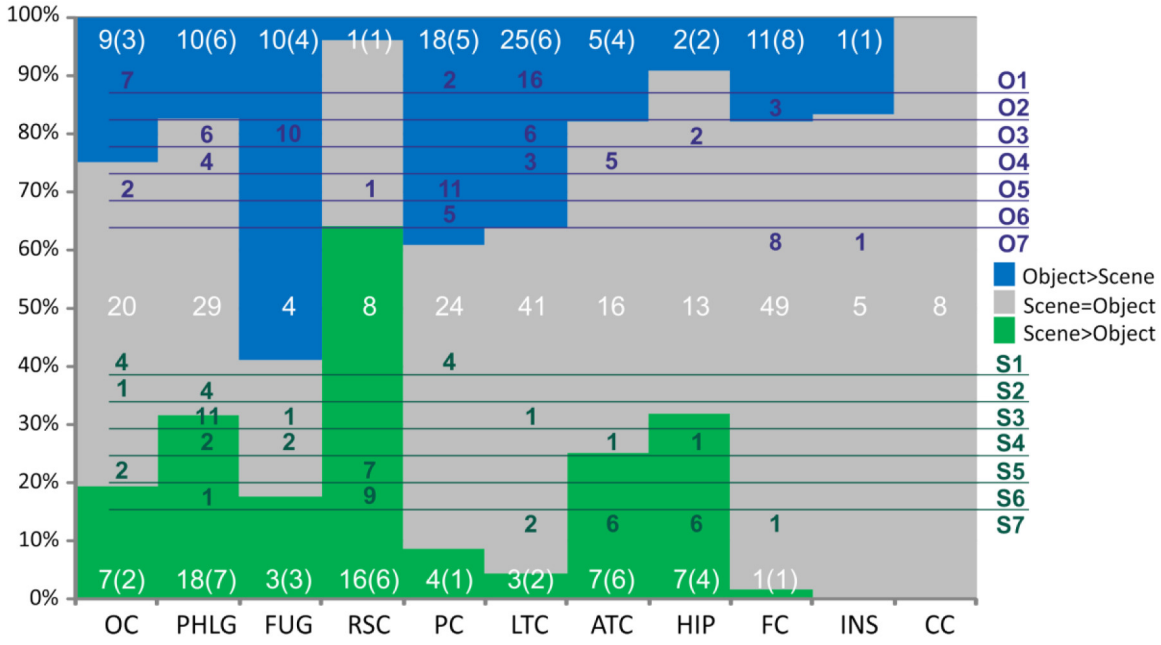

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
9(3) 10(6) 10(4) 1(1) 18(5) 25(6) 5(4) 2(2) 11(8) 1(1)
20 29 4 8 24 41 16 13 49 5 8
7(2) 18(7) 3(3) 16(6) 4(1) 3(2) 7(6) 7(4) 1(1)
OC PHLG FUG RSC PC LTC ATC HIP FC INS CC
O1 O2 O3 O4 O5 O6 O7
S1 S2 S3 S4 S5 S6 S7
Object>Scene
Scene=Object
Scene>Object

Fig10.tif

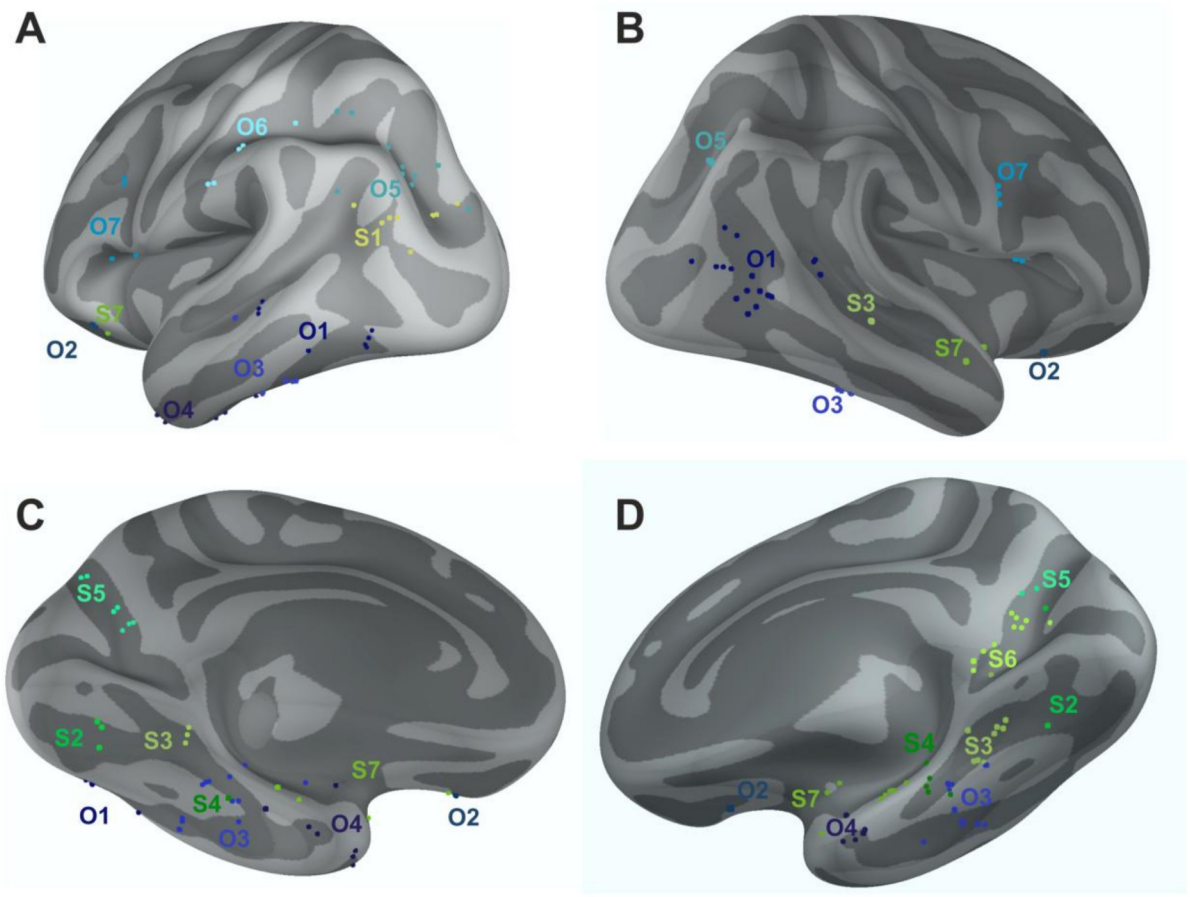

A
O6
O5
O7
S1
S7
O1
O2
O3
O4
B
O5
O7
O1
S3
S7
O2
O3
C
S5
S2
S3
S7
S4
O1
O3
O4
O2
D
S5
S6
S2
S4
S3
O2
S7
O3
O4

Fig11.tif

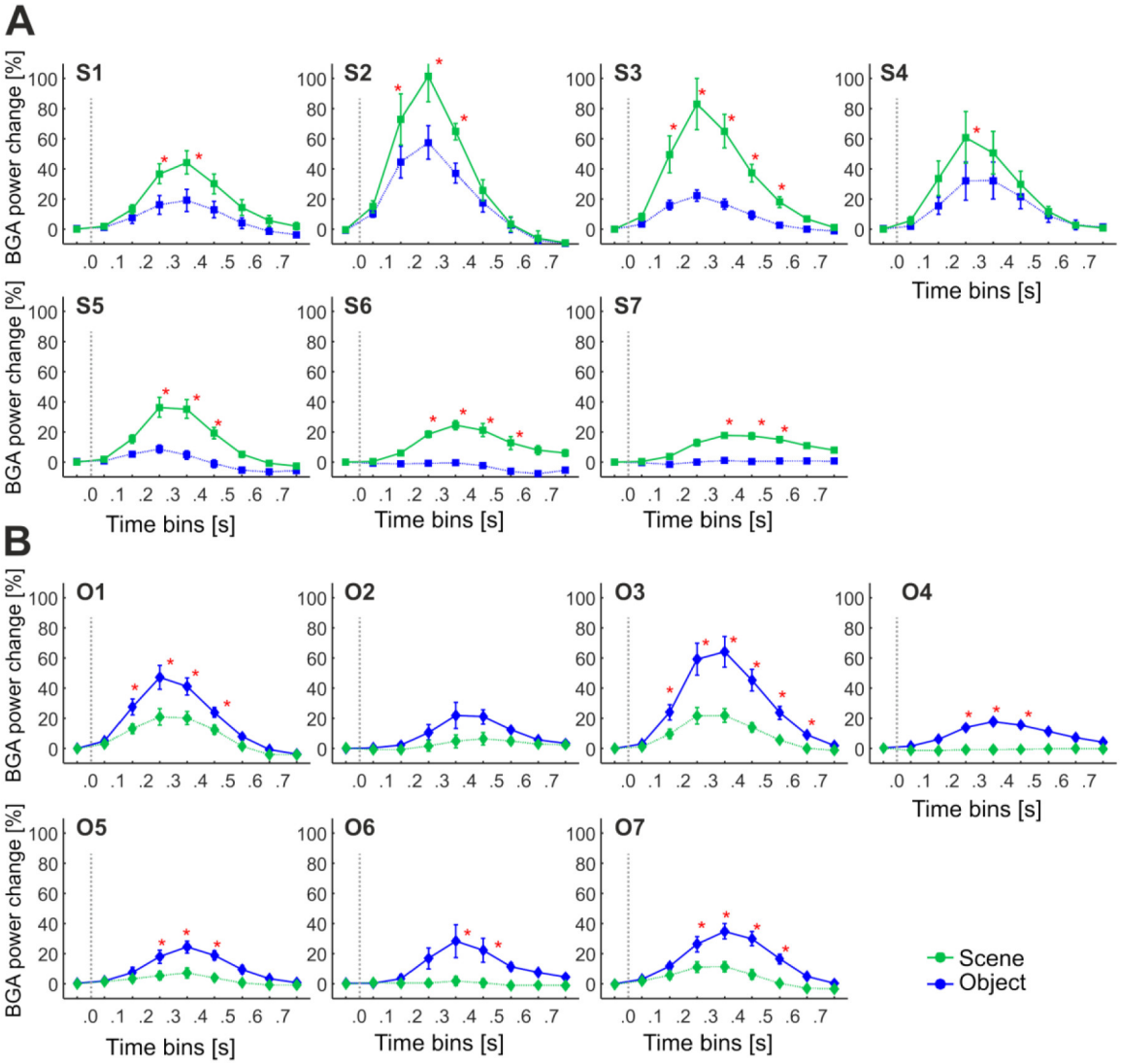

A
S1
S2
S3
S4
S5
S6
S7
BGA power change [%]
Time bins [s]
B
O1
O2
O3
O4
O5
O6
O7
Scene
Object

Fig12.tif

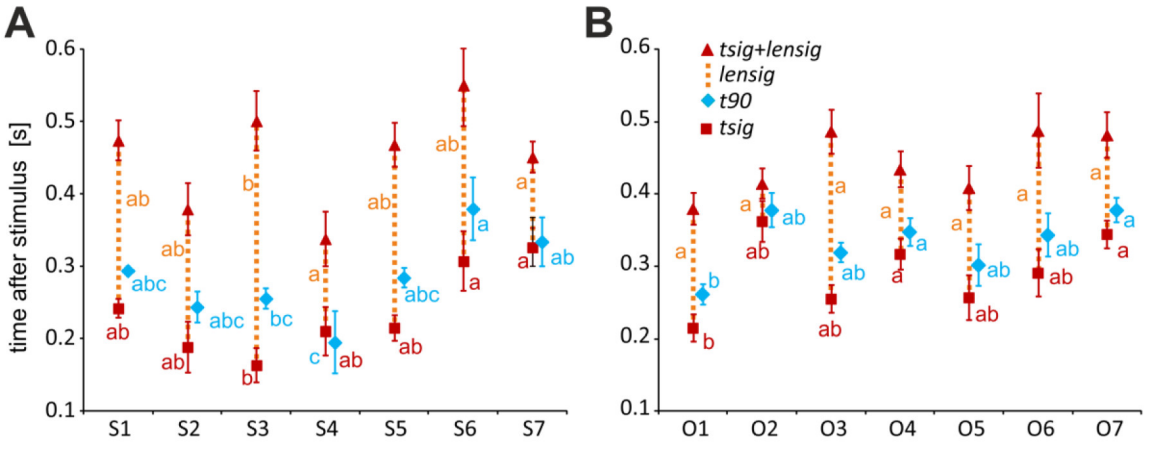
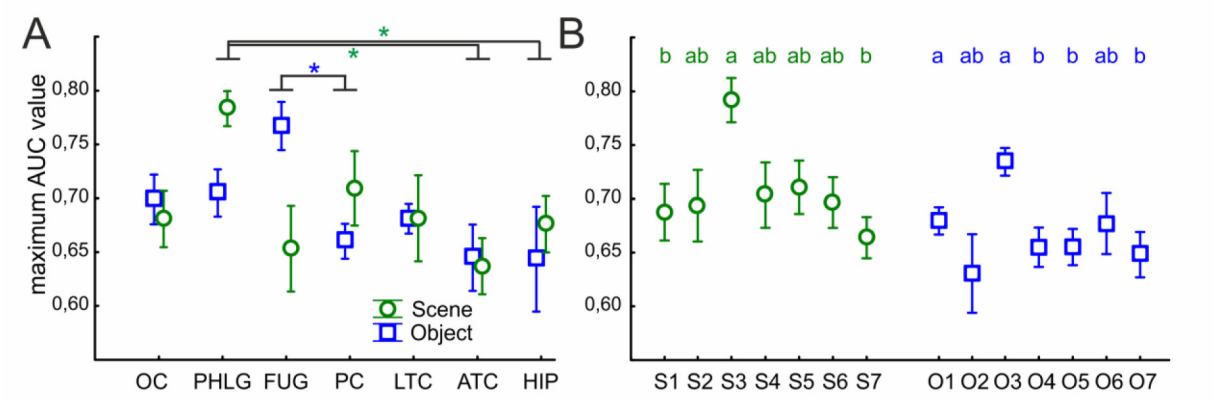

A
B
time after stimulus [s]
0.6
0.5
0.4
0.3
0.2
0.1
S1
S2
S3
S4
S5
S6
S7
O1
O2
O3
O4
O5
O6
O7
tsig+lensig
lensig
t90
tsig

Fig13.tif


A
B
maximum AUC value
0,80
0,75
0,70
0,65
0,60
Scene
Object
OC PHLG FUG PC LTC ATC HIP
b ab a ab ab ab b
a ab a b b ab b
S1 S2 S3 S4 S5 S6 S7
O1 O2 O3 O4 O5 O6 O7

Int. J. Human-Computer Studies 116 (2018) 15–24

Contents lists available at ScienceDirect

International Journal of Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town

Lukáš Hejtmánek[a,b,*], Ivana Oravcová[a], Jiří Motýl[a], Jiří Horáček[a,b], Iveta Fajnerová[a]

[a] National Institute of Mental Health, Topolová 748, Klecany, 250 67, Czech Republic
[b] Third Faculty of Medicine, Charles University in Prague, Ruská 87, Prague 10, 100 00, Czech Republic

ARTICLE INFO

Keywords:
Eye tracking
Navigation
GPS
Virtual environment
Spatial memory

ABSTRACT

There is a vibrant debate about consequences of mobile devices on our cognitive capabilities. Use of technology guided navigation has been linked with poor spatial knowledge and wayfinding in both virtual and real world experiments. Our goal was to investigate how the attention people pay to the GPS aid influences their navigation performance. We developed navigation tasks in a virtual city environment and during the experiment, we measured participants' eye movements. We also tested their cognitive traits and interviewed them about their navigation confidence and experience. Our results show that the more time participants spend with the GPS-like map, the less accurate spatial knowledge they manifest and the longer paths they travel without GPS guidance. This poor performance cannot be explained by individual differences in cognitive skills. We also show that the amount of time spent with the GPS is related to participant's subjective evaluation of their own navigation skills, with less confident navigators using GPS more intensively. We therefore suggest that despite an extensive use of navigation aids may have a detrimental effect on person's spatial learning, its general use is modulated by a perception of one's own navigation abilities.

1. Introduction

We have all been there. Feeling that after driving all the way across a city with a help of a GPS device, we remember nothing of the road and we feel that we won't be able to return without it. Recently, one of the authors heard few ambulance drivers from Prague discussing the difference between older and younger drivers. They agreed that many of the new drivers, who use GPS on a daily basis, do not know their way around the city even after five years of service, whereas their older superiors scarcely even look at the device. This anecdotal evidence points to the detrimental effect of GPS use over the long run (McKinlay, 2016), but the truth seems to be more complicated.

Efficient navigation in both small and large-scale environments requires formation of a mental representation (Shapiro, 2015; Taylor et al., 2008) that may be combined from several components, as was suggested by the route, landmark and survey theory (RLS) by Siegel and White (1975). According to this theory, landmarks serve as external reference points, routes connect these landmarks and survey knowledge binds everything together in a relational space representation. It was hypothesised that the survey knowledge grows with time, as routes are discovered between different landmarks and different relations are put into perspective (O'Keefe and Nadel, 1978; Siegel and White, 1975). This theory was later challenged by other studies demonstrating that route and survey knowledge can be actually acquired simultaneously (Aginsky et al., 1997; Hirtle and Heidorn, 1993). Some authors argue (Münzer et al., 2012; Shelton and McNamara, 2004; Willis et al., 2009; Zhang et al., 2014) that the quality of the representation may be affected by the modality or cues used during the encoding process (e.g. acquisition from a map versus directly from the environment), while other studies found no major differences (Huang et al., 2012). In particular, there are some indications that more abstract 2D representations (Münzer et al., 2006) can actually promote learning better than 3D simulations (Dillemuth, 2005). Moreover, while map-supported navigation positively stimulates forming of both route and survey knowledge, guided navigation forms only route knowledge (Münzer et al., 2006). While the use of maps, in general, could be beneficial in supporting survey knowledge and therefore the accuracy of the formed mental representation, extensive use of GPS based navigation may have a detrimental impact. Possible effects of GPS-based navigation systems should be therefore studied in more detail.

Mobile devices are currently taking over our everyday tasks as their adoption rate increases. Navigation systems surely bring a lot of relief to many users (Kim and Dey, 2009), but voices suggesting their negative impact on spatial abilities of everyday GPS users are growing (McKinlay, 2016). Such concerns are not without a cause. Indeed, spatial representations learned through a mobile interface have differ-

* Corresponding author.
E-mail address: lukas.hejtmanek@nudz.cz (L. Hejtmánek).

https://doi.org/10.1016/j.ijhcs.2018.04.006
Received 1 May 2017; Received in revised form 12 April 2018; Accepted 14 April 2018
Available online 22 April 2018

ent and arguably worse properties than layouts acquired by actively walking around or from maps (Münzer et al., 2006; Willis et al., 2009). Moreover, recent fMRI study (Javadi et al., 2017) demonstrated different hippocampal activity in situations where participants were required to make navigation decisions in contrast to those where instruction was provided. It was argued that the GPS supported navigation requires less occupation with the task, which allows us to focus on other stimuli and become less mindful of the surrounding environment (Parush et al., 2007). Such disengagement of a navigator from an environment has been shown as a relevant source of spatial knowledge impairment (Leshed et al., 2008).

Some researchers have tried various techniques to increase the engagement with the environment in computer simulations but to a limited success (Parush et al., 2007). It was demonstrated that changing navigation instructions to personally relevant ("Turn left in front of the cinema, where they screen your favourite movie Star wars") or salient descriptions of landmarks can yield better results than plain directional statements ("Turn left in front of that building" Gramann et al. (2017). Studies also demonstrated that navigation with a map, requiring shared map-environment attention, leads to better survey knowledge than when participants are completely guided by technology (Aslan et al., 2006).

Overall, there is sufficient evidence pointing towards the fact that people using GPS devices or simply following directions tend to form worse representations of environments than those who engage with the world or who learn actively using GPS or other navigation aids (Ishikawa et al., 2008; Parush et al., 2007). However, there is a lack of information about how much time people actually spend using these aids. Also, not all studies control for cognitive correlates such as memory, which was shown to significantly influence navigation aptitude (Hegarty et al., 2006). It might easily happen that some people use GPS only sparsely, while others rely on the aid heavily during their navigation.

Using a virtual city environment (VE) presented in first-person view and an eye-tracking device, we aimed to address these gaps in knowledge in usage of GPS-based devices during acquisition of spatial information. We assessed spatial knowledge through return path length and time needed to finish each trial (Ishikawa et al., 2008), pointing tasks and blank maps (Münzer et al., 2012; Wang and Spelke, 2000). In order to simulate GPS based navigation system used e.g. during driving, the GPS-like map of the virtual city was displayed in the left corner of the monitor. Such combination of methods allowed us to precisely determine how much time people spend looking at the GPS cues, how much they relied on it during recollection of spatial relationships and possibly differentiate between various use strategies. In addition, we assessed participants' spatial and memory skills using various neurocognitive methods.

1.1. Aims

This study aims to answer several principal questions related to the impact of using GPS devices on speed and accuracy of human spatial navigation.

Firstly, we aim to determine individual differences in GPS use by measuring how long participants spend looking at the aid during their navigation in case it is directly required by the task itself or when its use is optional.

Secondly, we aim to identify the effect of GPS use on the accuracy of acquired spatial knowledge using various measures of navigation performance - path and time redundancy, pointing accuracy, a number of deliberation stops, and a blank map. We hypothesise that the more time people spend looking at the navigation aid during learning, the worse wayfinding skills they will demonstrate during the recall. We predict that their pointing accuracy will be worse and they will show slower navigation times and longer paths. We also hypothesise that their memory for location positions will be affected, but a memory for location names (non-spatial semantic information) will be unaffected.

And **thirdly**, we focus on the relationship between one's own subjective evaluation of navigation abilities, as well as objective assessment of navigation abilities with cognitive tests, and how much they use the navigation aids. Our assumption is that those who either feel less skilled or those who are objectively worse navigators would use the GPS more than those who trust in their navigation skills. Combined with the first aim, this may help us to understand where individual differences in navigation aid use originate from and how they relate to GPS' use in general.

As a necessary prerequisite for answering these questions, we also tested intraindividual stability in navigation performance. Participants therefore attended two sessions over the course of 3 months, each with a different virtual city simulation consisting of 21 there-and-back tasks (described in Section 2.6). Navigation in the virtual city was accompanied by eye-tracking measurements to identify the proportion of the time participants focused on the provided GPS-like map. Participant navigation skills were self-reported and assessed by a battery of psychometric tests.

2. Methods

2.1. Participants

Healthy participants were recruited through a local advertisement which included questionnaire addressed to persons interested in human cognitive enhancement technologies such as virtual reality and GPS navigation. This per protocol recruitment process affected the unbalanced gender distribution in the final sample of 42 participants, 36 males ($M_{age} = 27.3, SD = 6.5$) and 6 females ($M_{age} = 29.75, SD = 7.9$). Although we consider this sample suitable for our study, any gender related differences should be interpreted cautiously.

All participants had completed at least high school education, 17 subjects finished and 12 were attending university. Participants with ophthalmologist reported sight defects that could affect the eye tracking measurement were excluded from the study. All participants signed an informed consent approved by the Ethical Committee of NIMH-CZ. Participants were financially rewarded at the end of the second session.

2.2. Procedure

All participants attended two sessions over the course of 10–12 weeks. During each session, participants completed the virtual city task (described in Section 2.6) accompanied by eye-tracking. After a delay of approximately 90 min, each participant received a blank map and was asked to fill in locations and names of all remembered landmarks. To control for individual differences in cognitive performance, specifically in spatial abilities and delayed memory, participants were assessed using a short psychometric battery of tests and they were also interviewed about their gaming habits and subjective navigation experience. Both sessions had the same structure, but the virtual city environment was different to remove possible bias in location placement and to create more tasks overall.

2.3. Psychometric battery

Spatial abilities were assessed using the Money Road-Map Test (RMT, number of errors from total 32 left-right decisions and completion time) Money et al. (1965) and the Perspective Taking Spatial Orientation Test (PTSOT, number of correct trials completed in 5 min) Hegarty and Waller (2004). Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) Randolph et al. (1998) was used to evaluate delayed memory performance (RBANS memory score calculated from List Recall, List Recognition, Story Memory, and Figure Recall).

2.4. Interview

We briefly interviewed each participant to determine their subjective navigation expertise on a scale from 1 to 5 (5 being the best). We also

Fig. 1. Illustration of our experimental setup.

asked about their experience with first-person video games and assigned each participant into one of three groups: 0 - none at all; 1 - have some experience; 2 - playing often.

2.5. Apparatus

Virtual city task was displayed on 21 in 1920 × 1080 monitor positioned 70 cm away from the participant (See Fig. 1). The experimental PC was equipped with i5-4440 CPU, 8GB DDR3 RAM and a GTX 750Ti graphics card. Task was running at an average of 60 frames per second with no trials dropping below 50. Participants moved around the environment using a keyboard and a mouse with standard key bindings common in first-person video games.

Eye movements were recorded using EyeLink®1000 Plus eye-tracking device at 1000 Hz. We recorded from a single eye (the one with better accuracy during calibration) as the area of interest was quite large. Calibration was based on 9 points across the monitor and we accepted any angular precision better than 5°.

2.6. Virtual city task

Virtual city task consisted of 21 there-and-back navigation tasks in a city environment. The environment was built using Unity 3D (Unity Technologies, 2017) and its street layout was inspired by a small town in Slovakia, which none of our participants knew. The city was 2500 by 3000 virtual units large with participants' speed set to 20 units per second. We populated the city with forty-two locations (e.g. hospital, school, train station, main square, church) and filled the rest of the environment with buildings, parks, roads and other features to create a natural representation of a city. We changed the location placement and the city visuals between the first and the second session to eliminate possible bias in the first design and to corroborate results across different allocentric representations. However, the same road layout was used in both sessions in order to acquire comparable complexity of the environment. For the names and positions of individual locations see the illustration (Fig. 7 in Supplementary materials).

Individual locations were placed in specific city parts with distinguishable landmarks and/or thematic visuals (e.g. university campus, hospital complex), but some locations were more exactly defined than others (e.g. church or statue vs apartment block or satellite town). Locations were always tied to a single spot of 20 units in diameter (e.g. church location was positioned in front of the church's main gate, satellite town location at a central crossing within the town etc.). We chose both salient and less salient locations to better simulate real city environments, where parts of cities are not necessarily defined by major landmarks, but by smaller features within those parts.

User interface included instructions in the right upper corner and a GPS-like map in the bottom left corner (see Fig. 2). As the GPS-like map did not always show directions, we often refer to it as a *map* and its use as *map use*, to indicate that it shows the city layout without a plotted path. For the eye-tracking purposes, we modeled the user interface with position of the verbal instruction distinct from the position of the GPS-like map, as having both UI elements close to each other would make it impossible to clearly distinguish which one were participants paying attention to. The map covered 250 by 250 units large city section (approximately 1% of area coverage). Participant's position and orientation were displayed on the map throughout the experiment (participant was depicted with a red arrow, see Fig. 3), but participants were given directions (guiding trail) only during the learning phase. The GPS-like map displayed buildings in grey, roads in yellow, path to be followed in red and terrain in the same colour as in the environment (green, brown etc.).

A common task in navigation aid research usually consists of participants following the same route with and without the provided aid (Ishikawa et al., 2008). Our tasks, on the other hand, were designed so that participants learned the path in one direction and then returned to the point of origin, i.e. there-and-back task. We constructed 21 there-and-back tasks in each city, each with two consecutive phases: learning and recall.

During the learning phase, the participant was asked to follow the guiding trail path to the destination. We created these paths so that they had comparable lengths and complexity and they were always longer than the shortest connection between the two landmarks. This allowed participants to improve on the learning path during the recall. During the recall, participant had to return back to the starting position. We consider these tasks well suited for guided navigation research, as they remove view-dependent navigation (Burgess, 2008; Gillner and Mallot, 1998). Each task had the following structure:

1. Learning phase
   A. Participant was teleported to the start location and shown a target location name and a map with plotted navigation path (see Fig. 2).
   B. Participant was asked to follow the guiding trail from the start location to the target location and to learn the environment as they went.

Fig. 2. (A) Learning phase with the guiding trail. GPS-like map was positioned in the left bottom corner, instructions in the right upper corner and the crosshair in the middle of the screen.

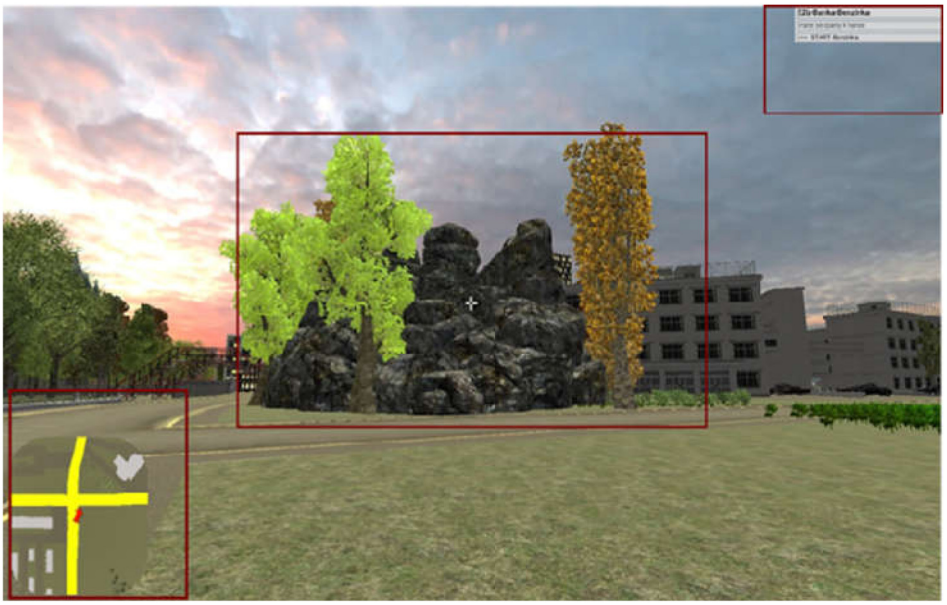


Fig. 3. (B) Task state during recall, with illustrative red boxes designating areas of interest captured by eye tracking device: GPS-like map, center, instruction panel (from left to right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

C. Upon reaching their destination, learning phase ended and a recall phase started

2. Recall phase
   A. 1st pointing - The map was hidden and the participant was asked to point towards the start location (using crosshairs in the center of the screen).
   B. After confirming the direction, the map was revealed and the participant was asked to navigate to the start location in the shortest way possible. If they were unable to finish the task within five minutes, researcher instructed them how to reach the goal.
   C. 2nd pointing - Upon reaching their destination, the map was hidden again and the participant was asked to point towards the start of the recall path (i.e. initial target location).
   D. After confirming the direction, new trial started.

In order to prevent participant's fatigue and to recalibrate the eye-tracking camera, trials were presented in three blocks, each consisting of 7 there-and-back tasks with short breaks in between. Block order was randomized, while trial order within each block was presented in the same succession. Example of a single trial performance can be seen in Fig. 4.

Each location was exposed twice throughout the task, both times within the same trial; once as a starting location and once as a target. Additionally, locations were positioned so that participants would pass them during following trials in order to complete allocentric representation of the entire city.

2.7. Blank map

Approximately ninety minutes after concluding the virtual city task, participants were asked to fill in a blank map. Participants were informed they could provide either location name, location position or both. Each answer was scored both on location placement and correct

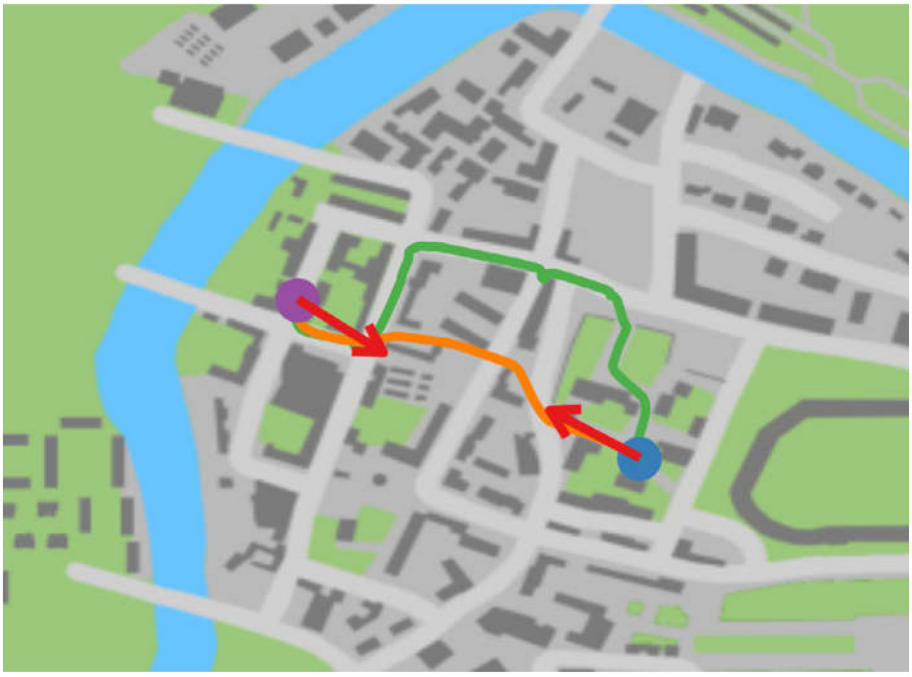


Fig. 4. Example of a single trial. Blue dot represents starting position and violet target position. The green line represents the learning path and the orange line represents the recall path. Arrows are depictions of participant's pointing direction measured at the start and at the end of the recall phase. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Blank map scoring.

Points	Location	Name
0	No location provided	No name provided
1	Different district	Valid synonym
2	Correct district	Correct name
3	Precise location	–

naming. For detailed scoring, see Table 1. This was participants' first encounter with the entire city layout.

2.8. Data acquisition and preprocessing

2.8.1. Gaze

Data from the eye-tracker were parsed and synchronised with participant's behavioural data from the virtual city task using unique timestamps. We calculated the amount of time spent (fixations) in three areas of interest (AOI) - GPS-like map, instruction panel and centre of the screen. Because the tasks were not equally long, we calculated ratios of the time spent at each AOI in each task, rather than absolute times.

2.9. Virtual city task

We extracted several relevant characteristics used in navigation research - path and time redundancy, deliberation stops, pointing accuracy and pointing speed (Burgess, 2006; Huang et al., 2012; Ishikawa et al., 2008; Parush et al., 2007; Wang and Spelke, 2000).

2.9.1. Path and time redundancy

We calculated path redundancy as a difference between the distance travelled by a participant during a task and the optimal path for that particular task. Optimal path is not a clear concept in a complex city layout, as multiple different trajectories might lead to the target destination with a similar distance covered. We chose the optimal path to be the shortest distance travelled by any participant in that particular task. We considered this approach better than using a route that was hypothetically the shortest, but which none of the participants was able to discover.

Path redundancy could, therefore, be a positive number ranging from zero upward, which designated how much farther each participant travelled than was necessary. It also provided a simple baseline for comparing trials of different lengths. Time redundancy was calculated in the same manner. Linear regression models and correlations were calculated with normalised z-scores of path and time redundancy.

2.9.2. Pointing accuracy and speed

At the start and at the end of each recall phase, participants were asked to point towards their point of origin. Difference between given direction and the correct direction was recorded as an absolute angle error, therefore any positive number between 0° and 180°. Pointing speed was defined as the time it took the participant to confirm his or her choice.

2.9.3. Deliberation stops

We calculated deliberation stops as a number of stops during each task where participant did not move (they could rotate) for a period of at least 3 s.

2.10. Outliers

As was mentioned earlier, some participants failed to finish the recall phase under 5 min and were instructed to the target location and those particular trials were excluded. This happened in 2% of the tasks with a total of 21 participants needing help in at least one trial, but no participant needed help in more than three trials. Trials which were more than four standard deviations away from the mean in either time or path redundancy were also excluded from the analysis. Some trials had to be discarded because of temporary loss of the eye-tracking signal.

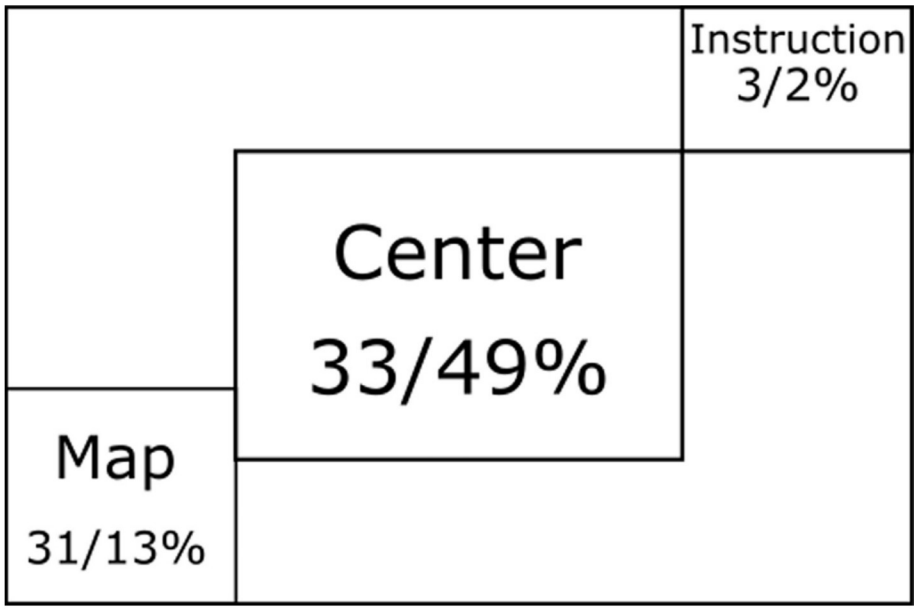


Fig. 5. Areas of interest with the ratio of gaze time during learning/recall tasks.

2.11. Session aggregation

Participants fulfilled a total of 42 two-phased virtual city tasks across the two sessions. Tasks in both sessions were comparable, yielding Cronbach's alpha of 0.75 for path efficiency, 0.87 for time efficiency and 0.93 for pointing accuracy. Therefor, we used all tasks and averaged participant's performance across both sessions.

2.12. Statistical analysis

All statistical analyses were conducted in R version 3.3 (R Core Team, 2017) using core components of the *stats* package. We fitted linear models using R's *lm* function, and reported Pearson's correlation coefficients were calculated with *cor* function. Graphs were constructed using *ggplot2* package (Wickham, 2009).

3. Results

3.1. Map use

Because of the design, participants were required to observe the map during the learning phase but its use during the recall phase was optional. During the learning phase, participants spent approximately third of their time looking at the map ($M = 0.31, SD = 0.11$), during recall this ratio fell to 0.13 ($SD = 0.10$) and this difference was significant ($t(2, 837.08) = 48.04, p < .001$). We also calculated Cronbach's alpha to determine internal consistency of individual map use throughout the experiment. This yielded Cronbach's alpha of 0.94 for the map use during learning and 0.97 overall, suggesting that participants didn't change their map use strategy as the experiment progressed. Because the time spent with instructions was very small ($M = 0.03, SD = 0.02$), we didn't use this measure in our analyses. Comparison of average times spent at each AOI can be found in Fig. 5.

3.2. Effect of map use on navigation performance

We observed a significant effect of map use during the learning phase on recall phase path redundancy ($R^2 = .15, F(1, 42) = 7.40, p = .009$) (can be seen in Fig. 6). We did not observe a similar effect either with time redundancy ($R^2 = .02, F(1, 42) = 0.80, p = .376$), or with deliberation stops ($R^2 =< .01, F(1, 42) = 0.06, p = .812$). Importantly, even though map use during learning and recall were correlated ($r = .57, t(1, 419) = 25.85, p < .001$), suggesting people used the map in a consistent manner, we observed no significant effect of how much participants used the map during recall on their path redundancy ($R^2 = .05, F(1, 42) = 2.33, p = .134$).

We found a significant effect of map use during learning on pointing error, both before ($R^2 = .31, F(1, 42) = 19.30, p < .001$) and after recall ($R^2 = .14, F(1, 42) = 6.86, p = .012$), but we saw no effect of map use on pointing times either before recall ($R^2 = .04, F(1, 42) = 1.85, p = .181$) or after recall ($R^2 = .02, F(1, 42) = 1.05, p = .311$).

In the blank map test we observed an effect of map use during learning on participant's score for location placement ($R^2 = .15, F(1, 41) = 7.01, p = .011$) and also their score for location naming ($R^2 = .12, F(1, 41) = 5.43, p = .025$). Correlation coefficients between map use and individual parameters of navigational performance can be seen in Table 2.

Table 2
Correlation coefficients calculated between map use during learning and specific behavioural measures and their descriptive statistics.

Measurement	Cor. coef.	*p*-value	Mean	SD
Path redundancy	0.39	0.01	376.53	185.24
Time redundancy	0.14	0.38	36.39	15.20
Deliberation stops	−0.04	0.81	1.57	0.71
Pointing error overall	0.41	0.00	24.72	16.1
Pointing speed overall	−0.03	0.78	5.71	2.89
Pointing error before recall	0.56	0.00	22.00	9.31
Pointing error after recall	0.37	0.01	26.48	19.78
Self assessed navigation skills	−0.36	0.02	3.9	0.87
Blank map location score	−0.38	0.01	55.52	26.42
Blank map name score	−0.34	0.02	44.72	12.69

3.3. Subjective and objective navigation skills

Lastly, we focused on participants' navigation skills. We found an effect of self assessed navigation ability on map use during learning ($R^2 = .25, F(1, 37) = 12.22, p = .001$), as well as path redundancy during recall ($R^2 = .13, F(1, 37) = 5.56, p = .024$). We also observed an effect of RMT completion time ($M = 54.09, SD = 20.47$) on path redundancy

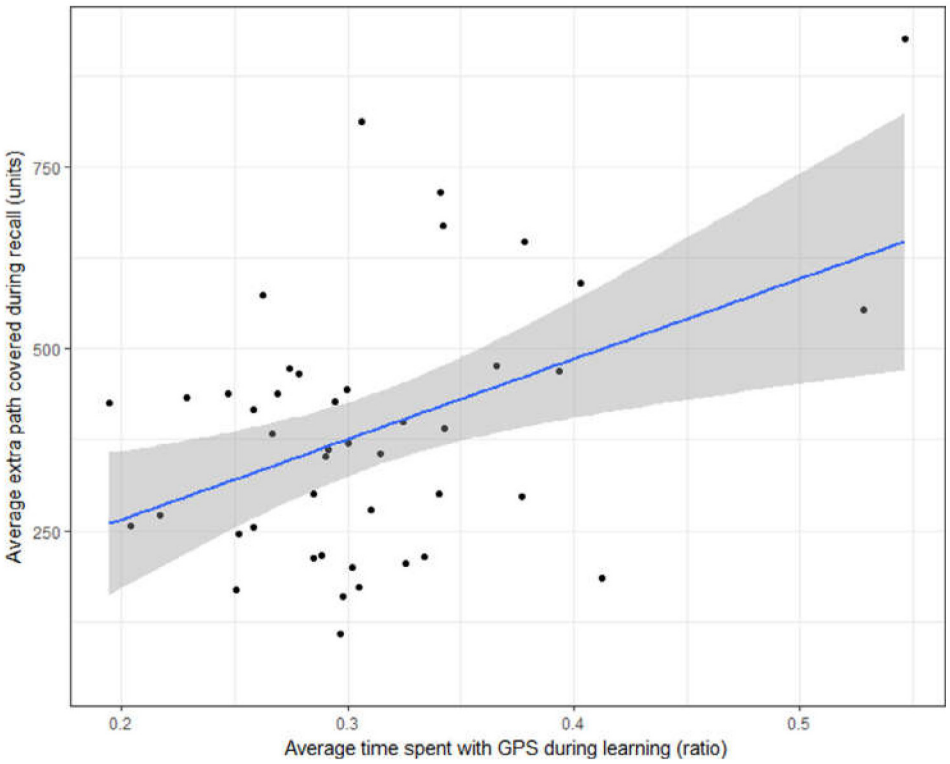


Fig. 6. Average time participants spent looking at the GPS during learning phase and average extra path (path redundancy) they covered during recall phase.

($R^2 = .14, F(1, 39) = 6.54, p = .015$) as well as PTSOT score (M = 10.61, SD = 1.91) on path redundancy ($R^2 = .28, F(1, 39) = 14.98, p < .001$). Moreover, our data showed a moderate negative correlation between number of correct answers on PTSOT and how much participants used the map during learning tasks ($r = -.46, t(38) = -3.16, p = .003$).

3.4. Controlling for the effects of gender, gaming experiences and delayed memory performance

We found no gender differences in average path redundancy ($t(10.04) = -1.02, p = .329$) and time spent with the map during learning or recall ($t(13.13) = 0.15, p = .885$). However, there was a significant difference in speed, with men finishing tasks faster ($M_{male} = 35.884, M_{female} = 46.571, t(12.75) = -2.28, p = .040$). This can be attributed to video games experience, with men having significantly more experience with video games ($chi^2(2) = 14.99, p < .001$). Because an effect of video games on navigation performance could be expected, we controlled for this by including map use and gaming experience in a linear regression model. This revealed only map use during learning was a predictor of path redundancy ($t(37) = 2.66, p = .012$) and pointing accuracy ($t(37) = 2.40, p = .021$).

Since some participants might just have better memory for names, we fitted linear model with both map use during learning and RBANS score for delayed recall. Indeed, only map use was a significant predictor of location placement score ($t(37) = -2.27, p = .029$), but both map use ($t(37) = -2.13, p = .040$) and RBANS delayed memory score ($t(37) = 2.09, p = .044$) were significant predictors of locations' name recollection.

Additionally, in order to control for repeated measurement of navigation performance, we looked at performance changes throughout the virtual city task, but observed no effect of task order on path redundancy ($R^2 =< .01, F(1, 2876) = 0.80, p = .371$) nor pointing accuracy ($R^2 =< .01, F(1, 2876) = 0.80, p = .371$). Neither we found any correlation between map use and task order ($r = .03, t(1, 426) = 1.27, p = .205$).

4. Discussion

Our main finding was that the amount of time dedicated to the GPS-like map during learning was negatively influencing navigation and pointing performance measured during recall as well as delayed spatial knowledge (both location naming and placement in the blank map). This result is in line with previous reports (Ishikawa et al., 2008; Leshed et al., 2008; Münzer et al., 2006; Willis et al., 2009), but at the same time provides new insights into how specifically this impairment happens.

4.1. The effect of GPS use on spatial knowledge

To address our first aim, we used eye-tracking to determine how long participants spent using the GPS-like map when they were required to do so or when its use was voluntary. Our data show that the GPS-like map use was consistent within subjects across trials. We also observed it being correlated with learning and recall phases, suggesting stable preference towards navigation aid usage in both forced and voluntary conditions.

As was proposed previously (Münzer et al., 2012), people could encode environment differently and adopt various strategies; either to orient themselves either using a map or relative to the cues in an external environment. In our experiment a GPS-like view of the city section was presented to the participants, enabling a combination of both mentioned strategies. Münzer and colleagues propose Münzer et al. (2012) that those who use a map extensively during learning might benefit from its presence more as they could be encoding the environment as map-like representation and navigating "through the map". Our results, however, challenge this. Despite our participants used the GPS-like map fairly consistently during both learning and recall phases, we found that its increased use during recall had no positive effect on their performance.

The two pointing episodes (before and after recall) proved to be useful in providing an important insight into the just formed spatial knowledge and our second aim. We found that the pointing error measured immediately following guided navigation was more affected by

the map use ($r = 0.56$) than after self guided return ($r = 0.37$). This can be explained by the fact that at the time of the second pointing participants had travelled between the locations twice, possibly forming better spatial representation (independent of the GPS aid). Yet, we observed their performance after the recall was slightly worse than after learning ($M(SD) = 22.0(9.31) vs. 26.5(19.80)$), though this difference was not significant.

Our findings suggest that the route knowledge obtained during recall did not help to improve participant's survey knowledge needed for correct pointing. It is possible that route knowledge and survey knowledge might be completely independent as other researchers propose Meilinger et al. (2013). Moreover, besides providing a route, standard GPS navigation devices also remove the necessity of angular orientation, as the map can rotate in respect to the egocentric view. The stable north-oriented map view used in our study could facilitate orientation in the environment, which could lead to better pointing performance when tested immediately after guided navigation.

A surprising discovery in the light of previous studies Ishikawa et al. (2008) and our second hypothesis was that the use of the GPS-like map influenced both the pointing accuracy and the path length, but did not slow participant's decision making. We found that neither time redundancy, pointing speed nor the number of deliberation stops were affected. This suggests that although participants were incorrect in their judgements, they considered these judgements acceptable and did not reflect on their choices.

In respect to the RLS theory by Siegel (Siegel and White, 1975), we argue that our virtual city design enabled acquisition of both survey and route knowledge and participants could thus choose the preferred strategy. Moreover, participants were aware from the beginning that they will be tasked with navigating in the environment without the GPS aid (during the recall and pointing) and asked about names and positions of the visited locations during the blank map test. In line with our second aim hypothesis, the GPS-like map usage clearly affected location placement in the blank map (spatial memory) with participants who used it more either remembered fewer locations or placed them incorrectly. This points to the importance of the engagement of the navigator with the navigated environment (Leshed et al., 2008). It was yet surprising that the map use partially affected also delayed memory for location names. Nevertheless, unlike location placement, location naming was also related to the delayed memory performance in the RBANS test.

Even though it is arguable to what extent are results from virtual environments transferable to the real world environments (Meilinger et al., 2013; Richardson et al., 1999; Witmer et al., 1996), it is necessary to stress that real world navigation takes longer, provides much more visual content (and distractions) and uses more than just visual cues (Witmer et al., 1996). We chose virtual environment out of necessity to closely control how much are people using the provided GPS aid, but advances in technologies might soon allow us to monitor people's gaze even in the real world. When that possibility arrives, it would be interesting to replicate our study in a real city.

4.2. Individual navigation skills

In respect to our final study aim, we analysed both objectively assessed spatial skills and subjectively evaluated navigational abilities and their effect on individual preference to use GPS-like map if available. Participants demonstrated interesting differences in how much they used the navigation aid and what it meant for their performance. Their map use during learning was related to their subjective evaluation of navigation abilities as well as to objective spatial skills. The less participant scored in neurocognitive tests assessing spatial skills or the less confident they were in their abilities, the more they focused on the map, thus effectively disengaging from the environment. This possibly demonstrates that people might choose to rely on the navigation aid because they simply do not believe in their own skills. As a result, their navigation capabilities might be impaired by shifted attention (Leshed et al., 2008). Future studies should address the relationship between confidence in one's own spatial skills and GPS use strategies in more detail.

Importantly, the amount of map use during the learning phase was along with objective spatial abilities one of strongest predictor of individual's performance, suggesting that participants could have navigated better, had they focused on the map less. Indeed, previous study demonstrated that active participation in navigation rather than just driving the vehicle according to guidance allows for better spatial knowledge (von Stülpnagel and Steffens, 2012), with the person giving directions achieving better route and survey knowledge than the driver. In concordance, our study showed that the participant's active engagement with the environment (increased fixations to central AOI) during learning phases resulted in better navigation performance as well as more precise spatial representation of the city.

Previous studies (Parush et al., 2007; Zhang et al., 2014) have documented that the environmental knowledge grows with repetition, regardless of acquisition methods. Other researchers also pointed out that strategies might shift to more efficient ones as the experiment progresses (Andersen et al., 2012). This sounds intuitive, as a failure, getting lost or inefficiency in navigation usually motivates to change approach. Nevertheless, we did not find any signs of improvement in performance in early and later trials and we noticed that participants relied on the map fairly consistently throughout the study and did not significantly change their strategy. This is more in concordance with results of Ishikawa and Montello (2006), who observed that most participants either formed good metric representation at the start, or they did not at all.

Needless to note, we did not give a clear feedback about how well participants navigated. The only feedback was potential feelings of frustration and experimenter intervening after 5 's of being lost. In following studies, we will consider displaying a measure of path redundancy directly in the user interface to inform participants how well they are doing. We chose not to display it because it could change the spontaneous strategy of GPS use, but it will be interesting to see how this change might happen.

4.3. Study limitations

Some limitations of the present study should be pointed out. First, it is important to mention that the act of looking at a GPS does not stipulate how specifically it is used. Some participants can be constantly looking at it only to confirm their heading in the environment, whereas others might rely on its information more. Nevertheless, we argue that the eye glances on our GPS-like map could be considered analogous to the real world GPS usage. The large interindividual differences in how people use the information could have also affected observed effects, which are modest. Another possible reason for moderate correlations could be that the participants were travelling each path only twice, whereas multiple location exposures should be arguably more helpful to those participants who use GPS less.

Second, our GPS-like map was proposed as an analogy of the "real-world" GPS navigation system; however, it differed in some aspects that should be pointed out. In contrast to some navigation systems that allow for directions to be delivered through audio, the verbal task-related information (location names) was presented visually in the right upper corner of the screen. We argue that both visually Kim and Dey (2009) and auditory Jensen et al. (2010) verbal instructions can provide sufficient information without significantly affecting spatial performance (e.g. while driving). Nevertheless, given participants were not paying much attention to the instruction AOI, we will consider auditory instructions in next iterations. Navigation aids also use a variety of different visualisation methods, either allocentric (survey or map like) view or more typical, egocentric view rotating the entire environment depending on user's orientation. Some different presentation modes were already studied Münzer et al. (2012), but it might be interesting to look at them with eye-tracking as well. Our GPS-like map layout was

presented in a bird's-eye view perspective, contrary to arguably more typical ego-perspective view. The map rotation also remained fixed to the north, with participant's orientation being represented by a directional arrow. Despite limited comparability of this layout to a typical GPS, it was chosen to facilitate allocentric (survey-based) knowledge and we would argue that the typical ego-centred visualisation would impair navigation even more. Moreover, it was proposed as a potential drawback of any GPS study Ishikawa et al. (2008) that it is unclear how large the map's coverage area should be. Our design used constant map size covering approx. 1% of the city. Future studies should investigate possible effects of the above listed factors.

Third, our participants were not able to fully immerse into the three-dimensional virtual environment as they experienced it only through a two-dimensional screen. However, similar designs were successfully applied in previous spatial studies (Ekstrom et al., 2005; Maguire et al., 1998). Moreover, our virtual city environment presented together with a navigation aid on a single screen was necessary to achieve our aims. This limitation could be addressed in the future using new head-mounted displays with incorporated eye-tracking.

Finally, it is possible that gender imbalance in our study complicates interpretation of the results, as previous research pointed to some gender differences in navigation depending on a GPS visualisation mode (Lin and Chen, 2013). The unbalanced gender selection was generated by the recruitment procedure; an online questionnaire aimed at acceptance of new technologies. Moreover, multiple females in our sample were concerned about their navigation abilities and/or gaming experiences and we are aware the same reason may have discouraged other women from participating. This limitation should be therefore addressed in the future.

4.4. Implications

In the introduction we mentioned that there are concerns about what navigation aids might do to our ability to navigate, if adopted globally (McKinlay, 2016). However, our results point to an important individual element in the navigation aids use strategies. We suggest that it might be especially those who are unsure or unable to navigate well who rely on these aids extensively. This could mean that GPS devices do not necessarily diminish our navigation abilities, they only offer alternative to those who could not effectively navigate without them.

Also, we were not only able to connect navigation aid use to navigation impairment, but the amount of time it was used to overall navigation performance as well. Therefore, if GPS use diminishes navigation skills to some degree, we suggest it might be better to incentivise users to use it less rather than not use it at all. Our results suggest that reducing the time spent with the GPS might, at least marginally, improve navigation performance and would not collide with individual preferences.

And lastly, our results provide some recommendations for future research of navigation aids. We suggest that such research should also track how much time participants actually spend with their navigation aid to control for possible individual differences in its use.

4.5. Conclusion

Our study confirmed that spatial knowledge is negatively affected by navigation aids. However, we propose, that there is an inner tendency to engage or disengage with the environment or navigation aid when given the possibility. The impairment of spatial knowledge might be partly driven by personal strategies and preferences, with people doubting their skills relying on GPS more. Thanks to our study design using eye-tracking and virtual environment, we were able to relate the effect to the amount of time participants spent using the GPS-like map.

Acknowledgements

The research leading to these results has received funding from the Norwegian Financial Mechanism 2009–2014 and the Ministry of Education, Youth and Sports of the Czech Republic under Project Contract no. MSMT-28477/2014, Project no. 7F14236. The Ministry of Health of the Czech Republic, grant nr. 17-30833A supported additional data analysis and covered the publication fees. The NIMH institution provided for the technical equipment and services by the financial support from the MEYS under the NPU I program project Nr. LO1611. We would like to thank Anna Francová, Adéla Plechatá and Michal Šmotek for help with psychological assessment and Martin Paňko and Jakub Gemrot for their help with the virtual city development. We also would like to acknowledge Eva Žačková, TomáŠ Zítka and Honza Hranička from the Man-Machine Interaction Department of the New Technologies - Research Centre, University of West Bohemia for recruitment of the volunteers through the online questionnaire they created for a related experiment. We thank the reviewers of the original manuscript for their helpful and inspiring comments and suggestions.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijhcs.2018.04.006.

References

Aginsky, V., Harris, C., Rensink, R., Beusmans, J., 1997. Two strategies for learning a route in a driving simulator. J. Environ. Psychol. 17 (4), 317–331.
Andersen, N.E., Dahmani, L., Konishi, K., Bohbot, V.D., 2012. Eye tracking, strategies, and sex differences in virtual navigation. Neurobiol. Learn. Mem. 97 (1), 81–89.
Aslan, I., Schwalm, M., Baus, J., Krüger, A., Schwartz, T., 2006. Acquisition of spatial knowledge in location aware mobile pedestrian navigation systems. In: Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 105–108.
Burgess, N., 2006. Spatial memory: how egocentric and allocentric combine. Trends Cogn. Sci. 10 (12), 551–557.
Burgess, N., 2008. Spatial cognition and the brain. Ann. N. Y. Acad. Sci. 1124, 77–97.
Dillemuth, J., 2005. Map design evaluation for mobile display. Cartogr. Geogr. Inf. Sci. 32 (4), 285–301.
Ekstrom, A.D., Caplan, J.B., Ho, E., Shattuck, K., Fried, I., Kahana, M.J., 2005. Human hippocampal theta activity during virtual navigation. Hippocampus 15 (7), 881–889.
Gillner, S., Mallot, H.A., 1998. Navigation and acquisition of spatial knowledge in a virtual maze. J. Cogn. Neurosci. 10 (4), 445–463.
Gramann, K., Hoepner, P., Karrer-Gauss, K., 2017. Modified navigation instructions for spatial navigation assistance systems lead to incidental spatial learning. Front. Psychol. 8, 193.
Hegarty, M., Montello, D.R., Richardson, A.E., Ishikawa, T., Lovelace, K., 2006. Spatial abilities at different scales: individual differences in aptitude-test performance and spatial-layout learning. Intelligence.
Hegarty, M., Waller, D., 2004. A dissociation between mental rotation and perspective–taking spatial abilities. Intelligence 32 (2), 175–191.
Hirtle, S.C., Heidorn, P.B., 1993. Chapter 7 the structure of cognitive maps: Representations and processes. In: Gärling, T., Golledge, R.G. (Eds.), Advances in Psychology, 96. North-Holland, pp. 170–192.
Huang, H., Schmidt, M., Gartner, G., 2012. Spatial knowledge acquisition with mobile maps, augmented reality and voice in the context of GPS-based pedestrian navigation: results from a field test. Cartogr. Geogr. Inf. Sci. 39 (2), 107–116.
Ishikawa, T., Fujiwara, H., Imai, O., Okabe, A., 2008. Wayfinding with a GPS-based mobile navigation system: a comparison with maps and direct experience. J. Environ. Psychol. 28 (1), 74–82.
Ishikawa, T., Montello, D.R., 2006. Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. Cogn. Psychol. 52 (2), 93–129.
Javadi, A.-H., Emo, B., Howard, L.R., Zisch, F.E., Yu, Y., Knight, R., Pinelo Silva, J., Spiers, H.J., 2017. Hippocampal and prefrontal processing of network topology to simulate the future. Nat. Commun. 8, 14652.
Jensen, B.S., Skov, M.B., Thiruravichandran, N., 2010. Studying driver attention and behaviour for three configurations of GPS navigation in real traffic driving. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1271–1280.
Kim, S., Dey, A.K., 2009. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 133–142.
Leshed, G., Velden, T., Rieger, O., Kot, B., Sengers, P., 2008. In-car gps navigation: Engagement with and disengagement from the environment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1675–1684.

Lin, P.-C., Chen, S.-I., 2013. The effects of gender differences on the usability of automotive on-board navigation systems – a comparison of 2D and 3D display. Transp. Res. Part F Traffic Psychol. Behav. 19 (Supplement C), 40–51.
Maguire, E.A., Frith, C.D., Burgess, N., Donnett, J.G., O'Keefe, J., 1998. Knowing where things are: parahippocampal involvement in encoding object locations in virtual large-Scale space. J. Cogn. Neurosci. 10 (1), 61–76.
McKinlay, R., 2016. Technology: use or lose our navigation skills. Nature 531 (7596), 573–575.
Meilinger, T., Frankenstein, J., Bülthoff, H.H., 2013. Learning to navigate: experience versus maps. Cognition 129 (1), 24–30.
Money, J., Alexander, D., Walker, H.T., 1965. A Standardized Road-Map Test Of Direction Sense: Manual. Johns Hopkins Press.
Münzer, S., Zimmer, H.D., Baus, J., 2012. Navigation assistance: a trade-off between wayfinding support and configural learning support. J. Exp. Psychol. Appl. 18 (1), 18–37.
Münzer, S., Zimmer, H.D., Schwalm, M., Baus, J., Aslan, I., 2006. Computer-assisted navigation and the acquisition of route and survey knowledge. J. Environ. Psychol. 26 (4), 300–308.
O'Keefe, J., Nadel, L., 1978. The hippocampus as a cognitive map.
Parush, A., Ahuvia, S., Erev, I., 2007. Degradation in spatial knowledge acquisition when using automatic navigation systems. In: Spatial Information Theory. Springer, Berlin, Heidelberg, pp. 238–254.
R Core Team, 2017. R: A language and environment for statistical computing.
Randolph, C., Tierney, M.C., Mohr, E., Chase, T.N., 1998. The repeatable battery for the assessment of neuropsychological status (RBANS): preliminary clinical validity. J. Clin. Exp. Neuropsychol. 20 (3), 310–319.
Richardson, A.E., Montello, D.R., Hegarty, M., 1999. Spatial knowledge acquisition from maps and from navigation in real and virtual environments. Mem. Cognit. 27 (4), 741–750.
Shapiro, M., 2015. A limited positioning system for memory. Hippocampus 25 (6), 690–696.
Shelton, A.L., McNamara, T.P., 2004. Orientation and perspective dependence in route and survey learning. J. Exp. Psychol. Learn. Mem. Cogn. 30 (1), 158–170.
Siegel, A.W., White, S.H., 1975. The development of spatial representations of large-scale environments. Adv. Child Dev. Behav.
von Stülpnagel, R., Steffens, M.C., 2012. Can active navigation be as good as driving? a comparison of spatial memory in drivers and backseat drivers. J. Exp. Psychol. Appl. 18 (2), 162–177.
Taylor, H.A., Brunyé, T.T., Taylor, S.T., 2008. Spatial mental representation: implications for navigation system design. Rev. Human Factors Ergon. 4 (1), 1–40.
Unity Technologies, 2017. Unity 3D.
Wang, R.F., Spelke, E.S., 2000. Updating egocentric representations in human navigation. Cognition 77 (3), 215–250.
Wickham, H., 2009. ggplot2: Elegant graphics for data analysis.
Willis, K.S., Hölscher, C., Wilbertz, G., Li, C., 2009. A comparison of spatial knowledge acquisition with maps and mobile maps. Comput. Environ. Urban Syst.
Witmer, B.G., Bailey, J.H., Knerr, B.W., Parsons, K.C., 1996. Virtual spaces and real world places: transfer of route knowledge. Int. J. Hum. Comput. Stud. 45 (4), 413–428.
Zhang, H., Zherdeva, K., Ekstrom, A.D., 2014. Different "routes" to a cognitive map: dissociable forms of spatial knowledge derived from route and cartographic map learning. Mem. Cognit. 42 (7), 1106–1117.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Path integration in large-scale space and with novel geometries: Comparing vector addition and encoding-error models

Sevan K. Harootonian[1,2¤], **Robert C. Wilson**[2,3,4], **Lukáš Hejtmánek**[1,5], **Eli M. Ziskin**[1,2], **Arne D. Ekstrom**[1,2,4]*

1 Center for Neuroscience, University of California Davis, Davis, California, United States of America, **2** Psychology Department, University of Arizona, Tucson, Arizona, United States of America, **3** Cognitive Science Program, University of Arizona, Tucson, Arizona, United States of America, **4** Evelyn McKnight Brain Institute, University of Arizona, Tucson, Arizona, United States of America, **5** Third Faculty of Medicine, Charles University, Ruská, Prague, Czech Republic

¤ Current address: Psychology Department, Princeton University, Princeton, New Jersey, United States of America
* adekstrom@email.arizona.edu

Citation: Harootonian SK, Wilson RC, Hejtmánek L, Ziskin EM, Ekstrom AD (2020) Path integration in large-scale space and with novel geometries: Comparing vector addition and encoding-error models. PLoS Comput Biol 16(5): e1007489. https://doi.org/10.1371/journal.pcbi.1007489

Editor: Francesco P. Battaglia, Radboud Universiteit Nijmegen, NETHERLANDS

Received: October 11, 2019

Accepted: March 24, 2020

Published: May 7, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1007489

Data Availability Statement: All data files are available at: github.com/sharootonian/PA-TCT.

Funding: Research supported by grants from NSF Division of Behavioral and Cognitive Sciences

Abstract

Path integration is thought to rely on vestibular and proprioceptive cues yet most studies in humans involve primarily visual input, providing limited insight into their respective contributions. We developed a paradigm involving walking in an omnidirectional treadmill in which participants were guided on two sides of a triangle and then found their back way to origin. In Experiment 1, we tested a range of different triangle types while keeping the distance of the unguided side constant to determine the influence of spatial geometry. Participants overshot the angle they needed to turn and undershot the distance they needed to walk, with no consistent effect of triangle type. In Experiment 2, we manipulated distance while keeping angle constant to determine how path integration operated over both shorter and longer distances. Participants underestimated the distance they needed to walk to the origin, with error increasing as a function of the walked distance. To attempt to account for our findings, we developed configural-based computational models involving vector addition, the second of which included terms for the influence of past trials on the current one. We compared against a previously developed configural model of human path integration, the Encoding-Error model. We found that the vector addition models captured the tendency of participants to under-encode guided sides of the triangles and an influence of past trials on current trials. Together, our findings expand our understanding of body-based contributions to human path integration, further suggesting the value of vector addition models in understanding these important components of human navigation.

Author summary

How do we remember where we have been? One important mechanism for doing so is called path integration, which refers to the computation of one's position in space with only self-motion cues. By tracking the direction and distance we have walked, we can

[BCS-1630296] awarded to Arne Ekstrom. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

create a mental arrow from the current location to the origin, termed the homing vector. Previous studies have shown that the homing vector is subject to systematic distortions depending on previously experienced paths, yet what influences these patterns of errors, particularly in humans, remains uncertain. In this study, we compare two models of path integration based on participants walking two sides of a triangle without vision and then completing the third side based on their estimate of the homing vector. We found no effect of triangle shape on systematic errors, while the systematic errors scaled with path length logarithmically, similar to Weber-Fechner law. While we show that both models captured participants' behavior, a model based on vector addition best captured the patterns of error in the homing vector. Our study therefore has important implications for how humans track their location, suggesting that vector-based models provide a reasonable and simple explanation for how we do so.

Introduction

"Dead reckoning," first coined by Charles Darwin [1], described a process whereby experienced navigators kept course to a particular spot over long distances and changes in directions, despite being in the featureless arctic tundra. All animal species tested, including spiders [2], bees [3], gerbils [4], hamsters [5], house mice [6], rats [7], birds [8], ants [9], arthropods [10], drosophila [11], dogs [12], cats [13], and humans [14] show the ability to dead reckon, thought to involve a computational process called path integration (please see [15–17] for a review of prior literature). Because humans employ vision as a primary modality to navigate, research on path integration has often focused on experiments in which visual input provides sufficient information to solve most navigational tasks, such as in desktop virtual reality. This is because visually rendered optic flow provides a velocity signal sufficient for some forms of path integration, with desktop virtual reality providing the opportunity to decouple visual from body-based cues in freely navigating humans. A limitation, however, with desktop VR is that it lacks the enriched cues and codes that we obtain when we freely move our body throughout space, thought to contribute critically to path integration [18–20], and does not allow comparison of the effects of visual vs. body-based inputs on path integration.

Past experiments on path integration have often involved a path completion task in which the navigator is guided in physical space and must return using the shortest trajectory back to the origin with no guidance [2, 21, 22]. Behavioral results both in vertebrates and invertebrates show a systematic bias in the return trajectory which is independent of random accumulated noise [23–27]. Two sets of computational models have been proposed as potential strategies participants might use, in some form, for path integration. The Homing Vector Model (also called a continuous strategy) assumes that the navigator does not encode information about the outbound path but rather continuously updates their current position relative to the origin by computing a continuous homing vector. To capture systematic errors in return paths, Homing Vector Models assume different variations of memory decay or leaky integration which are in theory independent of outbound path configurations [28–30].

Configural Models, in contrast, suggest a continuous encoding of the outbound paths (guided path) which are then manipulated using trigonometric or vector calculations to compute a return trajectory (configural homing vector) when required. One of these models, the Encoding-Error Model, suggests that systematic errors in the return path are due to errors in encoding of the outbound paths and their relative configurations [31, 32]. Recent work, however, suggests that execution errors (during the unguided path) make the largest contribution to systematic errors [33]. Furthermore, Wiener et. al. systematically compared configural and

continuous updating strategies by explicitly telling participants to focus on the outbound path or the origin. The results suggest that participants can deploy either configural or continual strategies during a path completion task [34]. Given the widespread application of configural models to both human and non-human path integration [23, 24, 26, 31, 32, 35] and the evidence that humans employ both configural and homing strategies [34], we directly compared three different configural models of path integration to better understand the patterns of errors that accumulate as participants dead-reckon.

In humans, a frequently employed experimental paradigm is the triangle completion task in which the experimenter guides the participant on two sides of a triangle and then the participant must return, without guidance, to the origin [21, 35]. To model how systematic errors accumulate when human participants perform path completion tasks and specifically the triangle completion task, Fujita et al. 1993 [31] proposed the Encoding-Error Model. This configural model of path integration uses the law of cosines to compute the third side of the triangle given the distances and angles encoded during first two sides (which are the guided portions). Therefore, it assumes that all systematic errors accumulate during encoding. Additionally, the model has four assumptions: (1) the internal representation satisfies Euclidean axioms (2) straight-line segments are encoded as a single value that represents their length (3) turns are encoded as a single value that represents the angle (4) there are no systematic errors during computation or execution of the homeward trajectory.

In support of their model, Fujita et al. fit data collected in [35] and [21] involving the triangle completion task in the absence of vision. The model captured the systematic errors seen in both studies to a relatively high degree (see [31] Table 3). As predicted, though, the model performed poorly for paths with more than two sides or paths that crossed each other. The Encoding-Error model was expanded by Klatzky et al. 1999 [32] to test its generalizability, who found that systematic errors were context and experience dependent. They also found that while partial vision increased path accuracy, it did not change the pattern of errors. Another important finding, supported by the Encoding-Error Model and other studies [36], was that systematic errors in path integration, at least in small environments ($\leq$10m), showed a pattern of regression to the mean. Specifically, past paths influenced the current paths and therefore, shorter angles and distances were overestimated and longer angles and distances were underestimated [21, 35]. Petzschner and Glasauer 2011 [36] (using desktop virtual reality) extended these findings by showing that the same angle or distance could be overestimated in some cases and underestimated in others. The degree of under/overestimation depended on the distribution of priors, known as range effects, such that a broader distribution of priors (e.g., distances from 5–100 meters vs. 5–10 meters) increased the effect of the regression to the mean [37].

The issue of how the distribution of priors influences the current trajectories, however, begs the question of how path configurations affect errors in the triangle completion task. Specifically, past work suggests that the geometric properties of shapes can influence navigation [38, 39]. For example, shapes like isosceles or equilateral triangles could serve as "templates" for how we learn paths [12] by providing a means for estimating paths that approximate it. Grid cells, neurons that fire as animals explore spatial environments, show 6-fold symmetry, with equilateral triangles composing part of this structure [40]. Given arguments that neural codes might manifest in spatial representations useful for navigating [41, 42] and the proposed link between path integration and grid cells [43, 44], it could be the case that geometric regularities (such as equilateral triangles) also influence path integration. Indeed, some past studies on the triangle completion task provide support for the idea that geometric regularities can, in some cases, influence path accuracy [35]. Yet, whether and how different types of triangles (equilateral vs. isosceles vs. scalene) influence path accuracy and patterns of errors on the unguided side in the triangle completion task remains unclear.

Another important yet largely unanswered question about human path integration regards the accuracy and patterns of errors over longer distances. The vast majority of studies in human path integration have involved small-scale environments (< = 10 meters) and consistent with this, computational models of path integration largely base their predictions on much smaller scales. For example, Klatzky et al. 1999 [32] suggested that it was unlikely that the same encoding function in their model was used for pathways that were larger than 10 meters [32]. A more recent computational model of path integration that employs grid cells suggests that, in the absence of specific mnemonic aids, path integration codes may rapidly degrade in mammals [45], consistent with the idea that path integration could breakdown dramatically over longer distances. Interestingly, however, other grid cell models assuming continuous rather than configural encoding suggest reliable estimations up to 100 meters [46]. Thus, an important question to test is how well human participants perform at path integration over longer distances (> = 100 meters).

In the current study, we employed an omnidirectional treadmill and somatosensory input via handheld controllers (Fig 1A) to determine the extent to which manipulating the angle and distance participants needed to walk affected the accuracy of navigation without vision. The unique advantage of using the omnidirectional treadmill is that it permits manipulation of infinitely large spaces thereby eliminating the need for any boundaries while preserving the input from walking. The issue of boundaries, perceived or imagined, is of potential importance because if a participant were to overshoot a distance they would be stopped before hitting a wall, providing inadvertent feedback on the distances of the room and potentially affecting subsequent performance. In addition, the use of handheld controllers allowed us to carefully manipulate participant trajectories on the guided sides, an issue we return to in greater depth in the discussion.

To attempt to capture the pattern of errors and history effects in the triangle completion task in human participants, we created a simple yet novel configural model based on vector addition often used to understand path integration in other species [23, 24]. The model assumes the outbound path is encoded as discrete sets of vectors, such that individual vectors can be employed to compute new trajectories using a configural homing vector. This model allows for a smaller set of required assumptions and free parameters when compared to Encoding-Error Model as well as a physiologically plausible mechanism by means of head direction accumulation [26]. In Experiment 1, we set out to test the extent to which different shapes of triangles (isosceles, equilateral) influenced how participants learned the configural homing vector by explicitly manipulating triangle type (while keeping configural homing distance constant). In Experiment 2, we explicitly manipulated the distance participants had to walk to reach the origin (while keeping triangle type constant) to determine how participants performed over a range of different distances. Critically, by manipulating these variables, we were able to simultaneously test hypotheses related to 1) triangle type and whether some might perform better than others; 2) configural homing distance and whether path integration would show different properties at ~10m vs. ~100m; 3) which model, one based on vector addition or the Encoding-Error model, would provide a better account of the pattern of findings. We provide a detailed comparison of the assumptions and setup of the different models in the Methods section.

Results

Experiment 1: *Basic behavior*

An example raw trace of a participant's path overlaid on the vector distances is shown in Fig 1D (dashed lines) between the points. We defined angle error as the difference between the

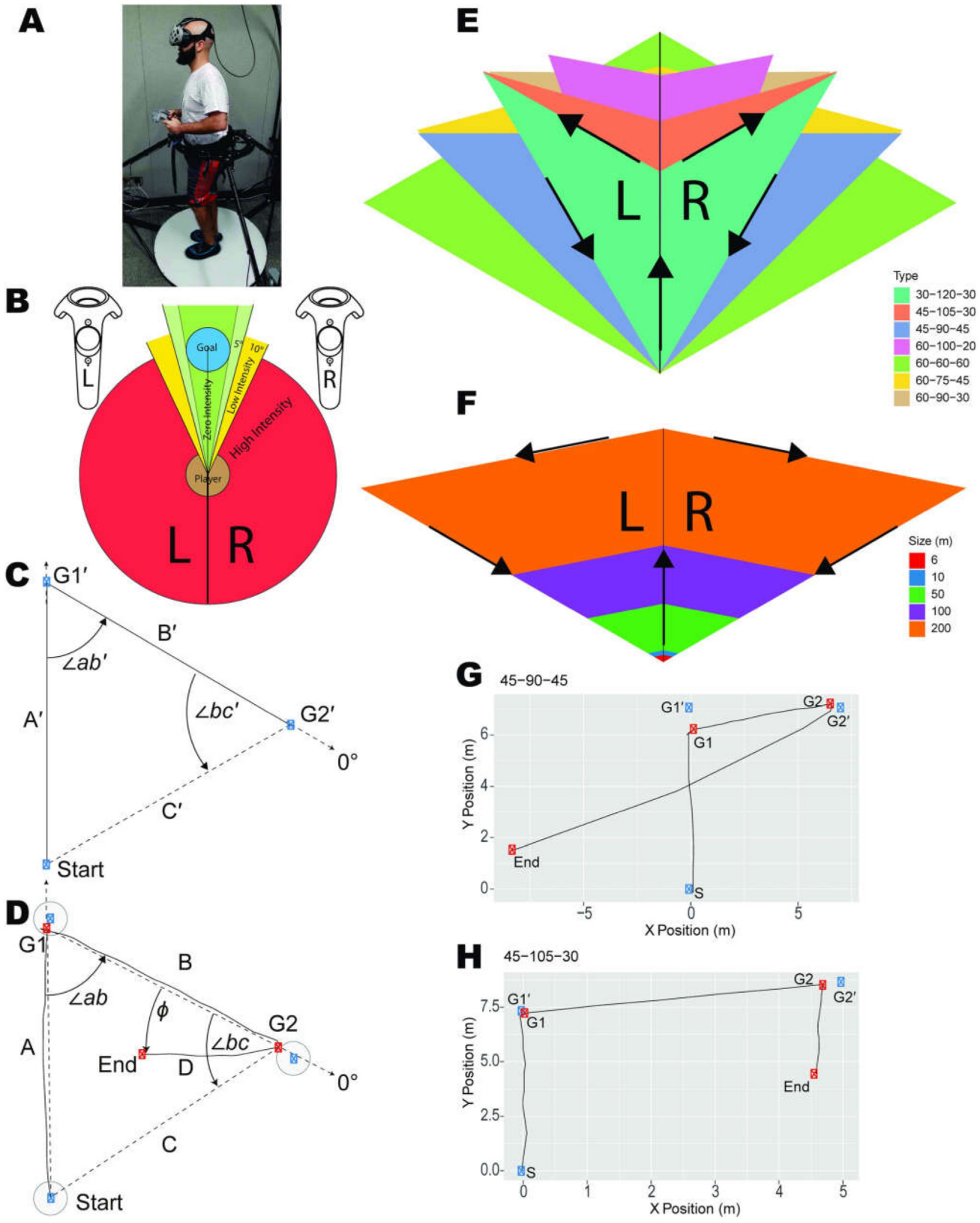

A
B
Goal
5°
10°
Low Intensity
Zero Intensity
High Intensity
Player
L
R
C
G1′
B′
∠ab′
A′
∠bc′
G2′
0°
C′
Start
D
G1
B
∠ab
ϕ
A
End
D
G2
∠bc
0°
C
Start
E
L
R
Type
30–120–30
45–105–30
45–90–45
60–100–20
60–60–60
60–75–45
60–90–30
F
L
R
Size (m)
6
10
50
100
200
G
45–90–45
G1′
G2
G2′
G1
End
S
Y Position (m)
X Position (m)
H
45–105–30
G1′
G2
G2′
G1
End
S
Y Position (m)
X Position (m)

Fig 1. Overview of experiments. (A) HTC VIVE headset along with the handheld controllers used in the experiments. Participants walked on the Cyberith Virtualizer omnidirectional treadmill system allowing them to navigate a much larger space while in stationary ambulation. (B)Visualization of HTC VIVE handheld controllers' feedback intensity based on the deviation of the angle. (C) Depiction of an equilateral triangle used in experiment 1. (D) Raw trace of participant's path overlaid on the vector distances (dashed lines) between the points. The blue point denotes the G1' and G2' locations that the participant is guided to and the red points are the participant's unique G1 and G2 locations for that trial. "D" refers to the actual distance walked by the participant when attempting to return to the origin. (E) Triangle templates used in experiment 1 overlaid on top of each other showing the different triangle types and denoting the internal angles. (F) Triangle templates used in Experiment 2 overlaid on top of each other showing how the length of side C varied over trials. (G) Raw trial where the participant over estimated distance and the angle. (H) Raw trial where the participant underestimated the distance and the angle.

https://doi.org/10.1371/journal.pcbi.1007489.g001

correct angle (∠bc) and participant's heading angle for side D (ϕ). This yields ∠bc–ϕ (Fig 1), in which a positive number denotes an overshoot and negative an undershoot. Distance error is the ratio of side D (the participants unguided walked distance) over the distance of C (configural homing vector from G2); a value greater than 1 is an overshoot and less than 1 an undershoot. As can be seen in the raw example shown (Fig 2A & 2B) and others (S1 Fig), although participants were often quite accurate at completing the triangle, they tended to overestimate the angle and underestimate the distance, regardless of triangle type. We will compare our findings of systematic errors with prior literature, specifically with Klatzky et al. 1999 [32] in the Discussion section.

Participants overestimate angle and underestimate distance. We next addressed the extent to which this overestimation of angle and underestimation of distance was true across the group of participants. As shown in Fig 2A, we found a tendency for participants to overestimate the angle they needed to turn to reach their start point (1-sample t-test against 0: $t(21) = 3.7, p<0.001$, Cohen's $d = 0.79, BF_{10}>10$), with participants, on average, tending to turn about 34.71˚±9.37˚ too far when estimating the angle they would need to turn to reach the origin. In contrast, we found that participants tended to underestimate the distance they needed to walk to get back to the start point, with participants' normalized walked distance significantly less than 1 (see Fig 2B, 1-sample t-test against 1: $t(21) = 16, p<0.001$, Cohen's $d = 3.42$, $BF_{10}>10$). The normalized walked distance was 0.87±0.05 (8.70m±0.50m). To determine the overall accuracy of the walked distance, we regressed the configural homing vector (side C) onto participants' unguided walked vector (side D). The beta values were positive and well above zero (1-sample t-test against 0: $t(21) = 5.4, p < .001$, Cohen's $d = 1.151$, $BF_{10}>10$), demonstrating that participants, despite underestimating distance, were well above chance in their estimates.

Sensory modality of guidance information. To ensure that our results were not due to difficulty with employing the handheld controllers to navigate the guided sides, we compared against a subset of trials in Experiment 1 in which the guided sides involved a visual beacon (note that participants otherwise navigated the unguided sides identically in somatosensory and vision conditions). During the *guided* section of the trials, there was no effect of vision (S2A Fig 2-sample t-test: $t(21) = 1.09$, $p = 0.288$, Cohen's $d = 0.336$ and $BF_{01}>3$), confirming that the handheld controller feedback system provides sufficient guidance. For angle error on the unguided side, as shown in S2B & S2D Fig, we found a slight but significant improvement in the vision-on (SD:43.10˚) compared to vision-off (SD:46.51˚) condition (2-way ANOVA with main effect of vision: $F(1, 21) = 4.9$, $p<0.026$, $\eta 2 = 0.016$ $BF_{10} = 1.16$). For distance error, as shown in S2C & S2E Fig, we also found a decrease in distance error during vision-on (SD:0.256) trials compared to vision-off (SD:0.271; 2-way ANOVA with main effect of vision: $F(1, 21) = 8.2$, $p<0.004$, $\eta 2 = 0.026$ $BF_{10} = 4$). We found no interaction effect between vision and triangle type. These findings suggest that providing vision on the guided sides did improve angle and distance estimates on the unguided side, but that participants still tended to overestimate angle and underestimate distance (see S2D & S2E Fig for additional information). Klatzky et al. 1999 [32] also found partial vision to improve accuracy, though, similarly, it seemed to have little effect on the direction of systematic errors. Thus, the overestimation of

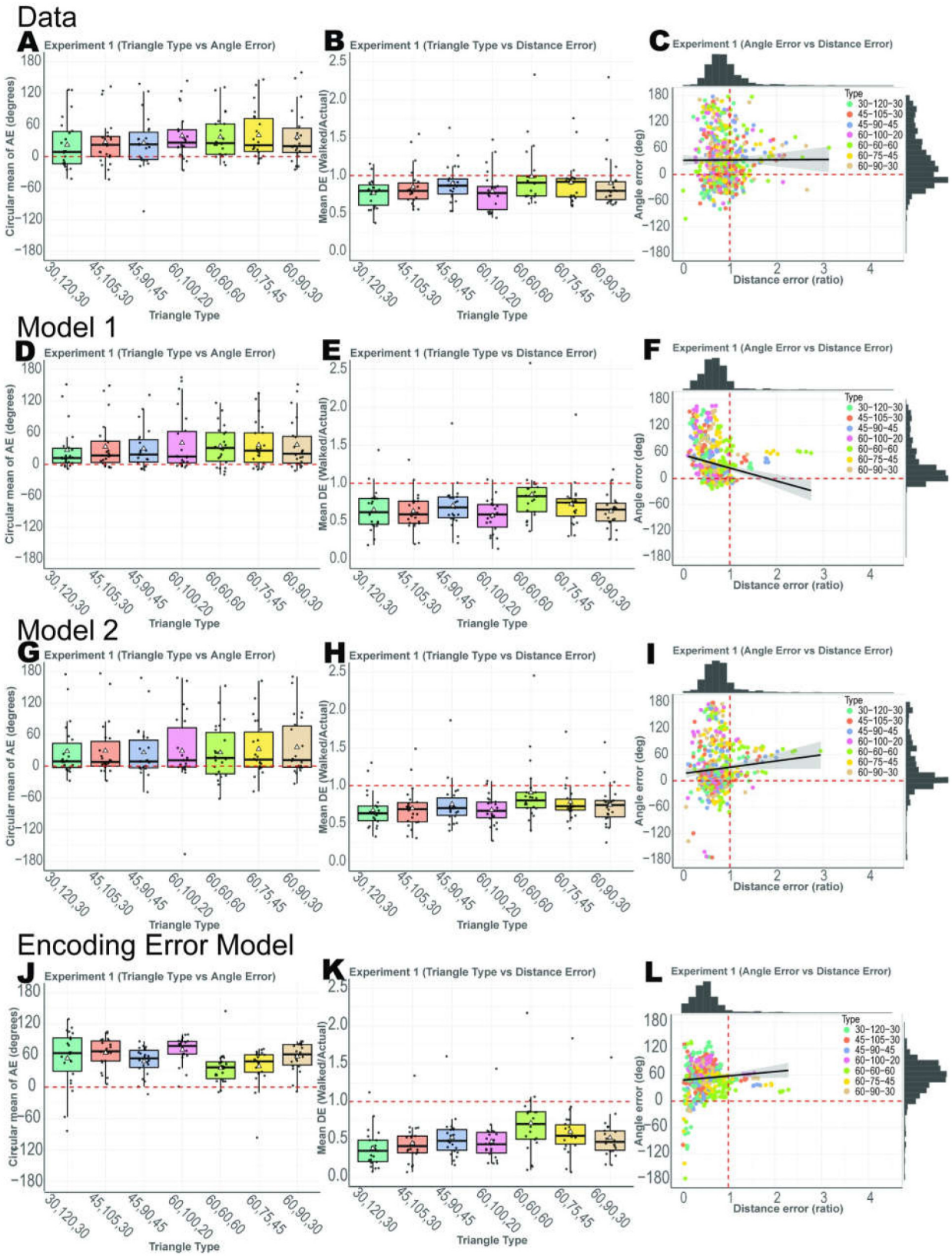

Data
A
Experiment 1 (Triangle Type vs Angle Error)
Circular mean of AE (degrees)
Triangle Type
B
Experiment 1 (Triangle Type vs Distance Error)
Mean DE (Walked/Actual)
Triangle Type
C
Experiment 1 (Angle Error vs Distance Error)
Angle error (deg)
Distance error (ratio)
Type
30-120-30
45-105-30
45-90-45
60-100-20
60-60-60
60-75-45
60-90-30
Model 1
D
Experiment 1 (Triangle Type vs Angle Error)
E
Experiment 1 (Triangle Type vs Distance Error)
F
Experiment 1 (Angle Error vs Distance Error)
Model 2
G
Experiment 1 (Triangle Type vs Angle Error)
H
Experiment 1 (Triangle Type vs Distance Error)
I
Experiment 1 (Angle Error vs Distance Error)
Encoding Error Model
J
Experiment 1 (Triangle Type vs Angle Error)
K
Experiment 1 (Triangle Type vs Distance Error)
L
Experiment 1 (Angle Error vs Distance Error)

Fig 2. Results from Experiment 1 along with model fits. White triangles represent the mean, while the median is shown as a black bar. (A) Participants' circular mean of angle error for the 7 triangle types. (B) Participants' mean distance error for the unguided side. (C) Angle error and distance error for all trials. Model 1 simulated data from fitted values are shown as (D-F). Model 2 simulated data from fitted values are shown as (G-I). Encoding-Error model simulated data from fitted values are shown as (J-L).

https://doi.org/10.1371/journal.pcbi.1007489.g002

angle and underestimation of distance that we observed cannot be accounted for by difficulty in completing the unguided sides using somatosensory input alone.

Effect of triangle shape on patterns of errors in path integration. Next, we wished to address the issue of triangle shape and whether this may have contributed in any way to the patterns of errors for the unguided side, as this might suggest participants used geometric features to anchor their path integration knowledge. For example, it could be that participants were most accurate for distance and angle on one triangle type (for example, right or equilateral triangles). To address this issue, we compared error on the unguided side with triangle type as an independent factor. Overall, we found only a modest effect of triangle type on angle error (2-way ANOVA with main effect of triangle type: $F(6,21) = 2.9$, $p<0.01$, $\eta^2 = 0.058$, $BF_{10} = 1.72$). Distance error, however, showed a fairly robust difference as a function of triangle type (2-way ANOVA with a main effect of triangle type: $F(6, 21) = 5.7$, $p<0.1.33e\text{-}5$, $\eta^2 = 0.109$ $BF_{10}>10$); see Fig 2A and 2B. While we did not find a consistent effect of triangle type across angle and distance errors, it is clear the triangle type contributed significantly to the patterns of participant errors. For example, the isosceles triangle (30,120,30) showed the lowest mean angle error (10.96˚±9.11˚) yet the equilateral triangle demonstrated the lowest mean distance error (0.985±0.062). Whether participants were relying on triangle templates, abstract geometric cues, or instead repeating identical/similar past paths in working memory, is an issue we return to in the Discussion.

As an additional analysis to investigate the use of geometric features of triangles, if participants were using specific shapes over others to perform the task, we might expect that both angle and distance errors would be correlated. This would be consistent with using the shape, rather than individual features, to compute the unguided side. Comparing angle and distance error is also important in determining the extent to which these two estimates were stored in a common vs. independent manner. We found no correlation between angle and distance error across trials and participants (Fig 2C; $r(579) = 0.0035$, $p = 0.933$), suggesting that angle and distance errors were not related to each other. We also observed no clustering of angle and distance error by triangle type (Fig 2C). Finally, we looked at the left and right-handedness of the triangle and found no difference between them (S4A & S4B Fig; angle error 2-sample t-test: $t(21) = 0.7$, $p = 0.485$, Cohen's $d = 0.118$, $BF_{01}>3$ and 2-sample t-test: distance error $t(21) = 1.136$, $p = 0.268$, Cohen's $d = 0.103$ and $BF_{01} = 2.53$). Together, these findings suggest that while geometric features (angle and distance) contribute to participant errors, it is not clear how these features are combined (if at all) to create and deploy triangle templates.

Vector addition model. To better understand the pattern of errors that participants made in Experiment 1, we built a configural computational model to predict the pattern of errors for the unguided sides. We combined angle and distance into a single vector value (see Methods) and employed the vectors for guided sides A and B as predictors for the unguided side. Based on previous findings [31], we would expect the guided sides to strongly predict performance on the unguided side. The modeling approach we employed also allowed us to compare the relative weighting of side A vs. side B and whether past trial history had any impact on unguided side performance.

We employed two versions of a simple vector addition model to fit our results. Model 1 assigns weights to sides A and B and predicts the distance and direction of side C (Eq 5). Model 2, in addition, includes a weighted influence of the participants' past trial history (Eq 6).

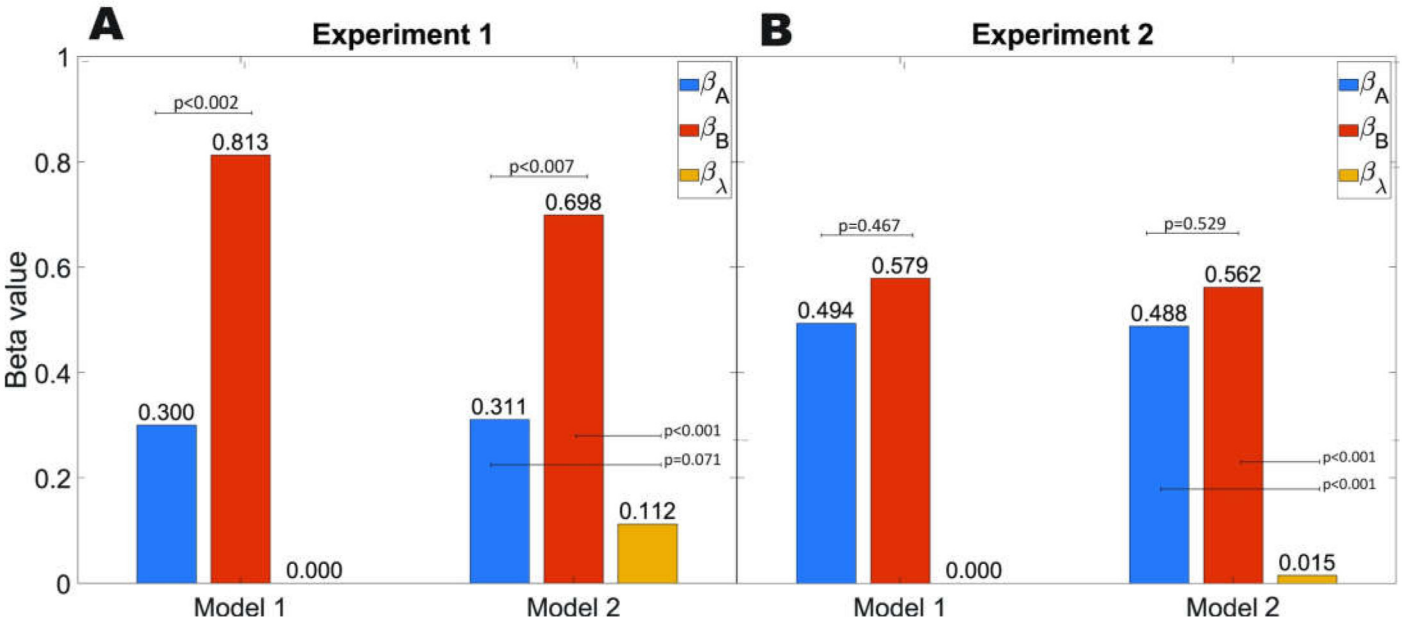


Fig 3. Mean Beta values from the vector models. (A): experiment 1 and (B): experiment 2.

https://doi.org/10.1371/journal.pcbi.1007489.g003

We compared these two vector addition models to the Encoding-Error Model [31]. For Model 1, both guided sides A and B strongly predicted performance on the unguided side (mean β_A = 0.3, 1-sample t-test against 0: t(21) = 2.86 p<0.0001 BF_{10}>3 and mean β_B = 0.813, 1-sample t-test against 0: t(21) = 7.41 p<0.0001, BF_{10}>10; Fig 3A). Notably, only the beta values for side A were significantly less than 1 (1-sample t-test against 1: t(21) = 6.6, p<1.299e-6, BF_{10}>10), suggesting that participants underweighted side A when estimating the return vector. The underweighting of side A could, potentially, account for the angle overestimation. In addition, side B was weighted significantly higher than side A, (2-sample t-test: t(21) = 3.62, p<0.002, Cohen's d = 1.02, BF_{10}>10).

For model 2, which included participants' past trial history, we found mean β_A = 0.311, 1-sample t-test against 0: t(21) = 3.05, p<0.006, BF_{10} = 7.5 and mean β_B = 0.698, 1-sample t-test against 0: t(21) = 7.775, p<0.0001, BF_{10}>10, suggesting similar results in terms of underweighting the guided sides as Model 1. We also found a significant effect of past trials (mean β_λ = 0.112, 1-sample t-test against 0: t(21) = 3.415, p<0.003, BF_{10}>10), suggesting that sequential effects contributed significantly in Experiment 1 (Fig 3A). While the priors (past walked triangles) were similar in size in Experiment 1 (i.e., distance was not explicitly manipulated), the modeling results showed that past trials made a significant contribution to participant's estimation of the unguided side.

Taken together, these findings suggest that the findings from Experiment 1, which involved different triangle types, could be captured by our configural vector-based models, particularly Model 1. Participants underweighted both guided sides A and B, with a tendency to underweight side A to a greater extent, possibly accounting for the tendency of participants to overshoot the unguided angle. We also found evidence for a linear combination of past trials providing explanatory power for the unguided side.

Model validation. Next, we simulated Model 1 to determine whether it could account for the trends observed in the empirical data [47, 48]. We found that Model 1 captured both the angle overestimation (Fig 3A) and distance underestimation (Fig 3B) in Experiment 1. The

simulation results supported the idea that Model 1 provided a better account for the data than Model 2 (Fig 2D–2I) and captured the relevant empirical phenomenon reported here.

Encoding-Error Model. We fitted and simulated our data using the Encoding-Error Model, and, similar to Model 1 and Model 2, were able to capture the systematic errors in angle overestimation (Fig 2J) and distance underestimation (Fig 2K). The Encoding-Error Model, given the limited range of triangle distances in Experiment 1, did not show regression to the mean (effect of past trials). When we directly compared the models (S2J and S2K Fig), however, we found that Model 1 fit the data fairly decisively, at both subject and group level, compared to the other two models. While Model 1 did outperform the other two models in BIC and AIC, the confusion matrix in S7A–S7C Fig showed that simulated data from Encoding-Error model did not fit Encoding-Error model best compared to the two vector addition models. This method of model recovery suggests some limitations with our model comparison (i.e., how well our task can distinguish between models) and was likely due to small number of trials and the fact that the vector addition models involved far fewer free parameters than the Encoding-Error Model [48]. We return to a more detailed comparison between vector addition and Encoding-Error Models in the Discussion.

Experiment 2

Basic behavior. In Experiment 2, we manipulated the distance of the triangles (perimeters = 15.19, 25.32, 126.60, 253.20, and 506.42 meters) while keeping triangle geometry constant. This involved manipulating the distance of the guided sides while maintaining a scalene triangle shape, thus leaving the angles constant. We implemented the same task structure as Experiment 1 but here we kept the shape of the guided path the same and varied the scale across trials.

Participants systematically underestimated distance but accurately estimated angle. For angle error, somewhat in contrast to Experiment 1, we found no significant overestimation or underestimation of angle, with participants showing a mean error of 0.8˚±7.44˚ (1-sample t-test against 0: $t(16) = 0.107$, $p = 0.916$, Cohen's $d = 0.026$, $BF_{01}>3$). We also found no effect of triangle size on angle error (Fig 4A 1-way ANOVA with no main effect of triangle size: $F(4,16) = 0.609$, $p = 0.658$, $\eta^2 = 0.036$,$BF_{01}>10$). We attribute this to the fact that triangle configuration was consistent across Experiment 2 because we primarily manipulated distance.

We found evidence of fairly accurate estimation of distance for smaller triangle perimeters (15-25m perimeter) and considerable underestimation for larger triangle perimeters (126m – 500m perimeter). In fact, we found a trend whereby distance underestimation increased as a function of the unguided distance (Fig 4B, 1-way ANOVA with main effect of triangle size: $F(4,16) = 21.107$, $p<3.913e-11$, $\eta^2 = 0.553$ and $BF_{10}>10$). This is shown in Fig 5A, where the dotted line indicates a slope of 1, with the actual slope well below this value. In other words, the further that participants walked, the more they tended to underestimate the unguided side.

To better understand this phenomenon, we analyzed the spread of the errors as participants walked the unguided side. We found that the distribution of distance error scaled linearly as a function of the walked distance. As shown in Fig 5B, the standard deviation of the walked unguided distances increased linearly, as shown by a regression fit (linear model: $t(4) = 23.6$, $p<0.0001$), suggesting that the greater the walked distance, the proportionately greater the error in distance with variance increasing exponentially. Distance underestimation ($1 - \frac{side\ D}{side\ C}$) increased as a function of distance as well, however, this increase was best fit by a logarithmic function (Fig 5C, linear model: $t(4) = 13.39$, $p<0.001$) rather than linearly, similar to Weber–Fechner and Stevens' power law [49]. Together, these findings suggest that as participants

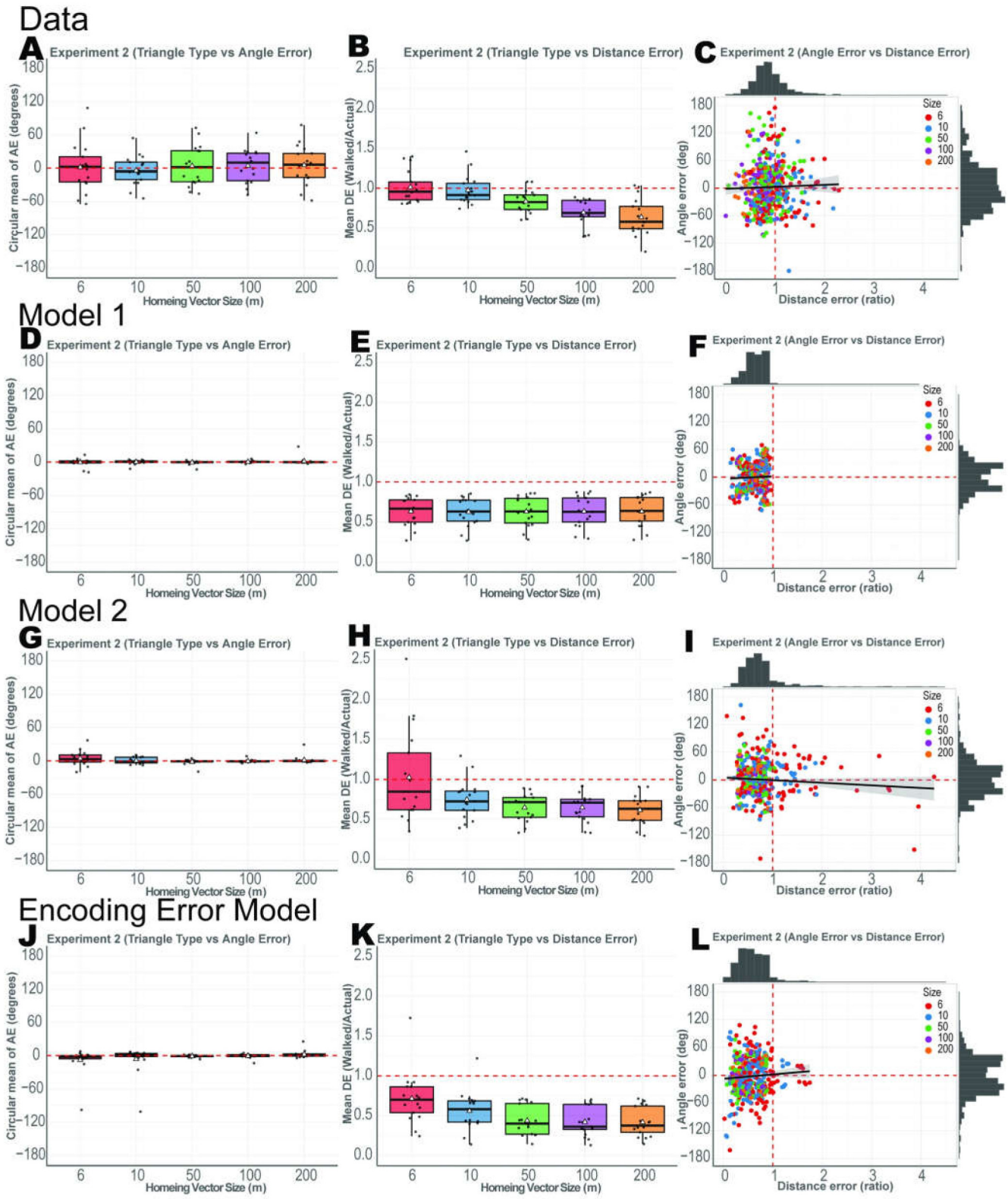

Data
A
Experiment 2 (Triangle Type vs Angle Error)
Circular mean of AE (degrees)
Homeing Vector Size (m)
B
Experiment 2 (Triangle Type vs Distance Error)
Mean DE (Walked/Actual)
Homeing Vector Size (m)
C
Experiment 2 (Angle Error vs Distance Error)
Angle error (deg)
Distance error (ratio)
Size
6
10
50
100
200
Model 1
D
Experiment 2 (Triangle Type vs Angle Error)
E
Experiment 2 (Triangle Type vs Distance Error)
F
Experiment 2 (Angle Error vs Distance Error)
Model 2
G
Experiment 2 (Triangle Type vs Angle Error)
H
Experiment 2 (Triangle Type vs Distance Error)
I
Experiment 2 (Angle Error vs Distance Error)
Encoding Error Model
J
Experiment 2 (Triangle Type vs Angle Error)
K
Experiment 2 (Triangle Type vs Distance Error)
L
Experiment 2 (Angle Error vs Distance Error)

Fig 4. Results from Experiment 2 along with model fits. White triangles represent the mean while the median is shown as a black bar. (A) Participants' circular mean of angle error for the 7 triangle types. (B)Participants' mean distance error for the unguided size. (C)Angle error and Distance error of all trials. Model 1 simulated data from fitted values are shown as (D–F). Model 2 simulated data from fitted values are shown as (G–I). Encoding–Error model simulated data from fitted values are shown as (J–L).

https://doi.org/10.1371/journal.pcbi.1007489.g004

walked longer distances, they tended to increase their underestimation of the distance they would need to walk and scale their errors logarithmically as a function of distance.

Similar to Experiment 1, we also found no correlation between angle and distance error (Fig 4C, 2-sample t-test: $t(487) = 0.623$, $p = 0.533$, $BF_{01} > 7.8$). We also found no effect of right vs. left turns on guided sides (angle error: 2-sample t-test: $t(16) = 1.51$, $p = 0.151$, Cohen's $d = 0.245$, $BF_{01} = 1.55$ and distance error: (2-sample t-test: $t(16) = 0.724$, $p = 0.4797$, Cohen's $d = 0.176$ and $BF_{01} = 3.188$), see S4C & S4D Fig.

Vector addition model. To better understand the effects of the guided sides on the unguided side estimates in Experiment 2, we employed the computational models used in

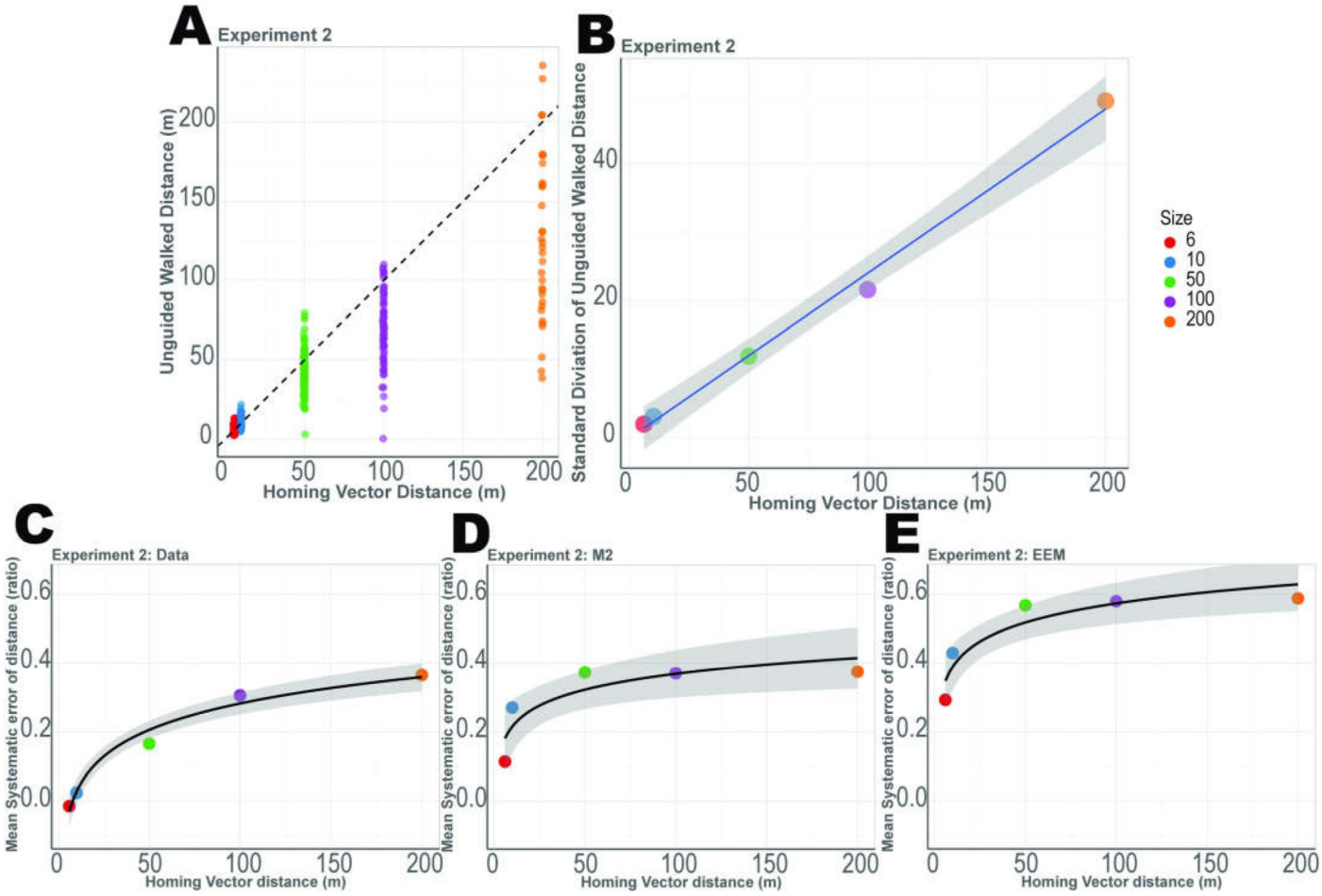


Fig 5. Systematic underestimation of distance from Experiment 2. (A) The distribution of distances for the unguided side (D) for each triangle size with y = x plotted at the dotted line. (B) Standard deviation of distances for the unguided side (D), which show a linear increase. (C) Mean systematic errors of distance (1- distance error), which increase logarithmically. (D) Mean systematic errors of distance of the simulated data from Model 2, which increase logarithmically. (E) Mean systematic errors of distance of the simulated data from Encoding-Error Model, which increase logarithmically.

https://doi.org/10.1371/journal.pcbi.1007489.g005

Experiment 1 to predict the pattern of errors for the unguided sides. The modeling analysis from Model 1 again revealed that both guided sides A and B strongly predicted performance on the unguided side (mean β_A = 0. 494, 1-sample t-test against 0: t(16) = 5.09, p<0.001, BF_{10}>10 and mean β_B = 0. 579, 1-sample t-test against 0: t(16) = 9.33, p<0.001, BF_{10}>10; Fig 3B). Notably, both beta values were less than 1 (1-sample t-test against 1: β_A t(16) = 5.22 p<0.001, BF_{10}>10 and 1-sample t-test against 1: β_B t(16) = 6.80 p<0.001, BF_{10}>10), suggesting that participants underweighted *both* sides when estimating the return vector. In addition, unlike Experiment 1, both sides were weighted evenly (1-sample t-test: t(16) = 0.547,p = 0.591, Cohen's d = 0.218, BF_{01}>3). These findings are perhaps unsurprising because angle was neither under nor overestimated.

Comparing Model 1 (modeling the distance of the guided sides to predict the unguided sides) and 2 (using Model 1 with an additional term for past trial distances), we found significant fits for all three beta terms. In other words, guided sides A & B, as well as past trial history (mean β_A = 0. 488, 1-sample t-test against 0: t(16) = 5.06, p<0.001, BF_{10}>10, mean β_B = 0. 562, 1-sample t-test against 0: t(16) = 8.96, p<0.001, BF_{10}>10 and mean β_λ = 0.015, 1-sample t-test against 0: t(16) = 3.26, p<0.005, BF_{10} = 9.83), all predicted errors in walking the unguided side in Experiment 2. Thus, similar to Experiment 1, trial history provided a significant explanation of error in Experiment 2. Note, β_λ values are smaller in Experiment 2 due to the large values of λ (linear combination of past paths), which in Experiment 2 can go up to 500m.

Model validation. Next, we simulated our data in a manner similar to Experiment 1. Simulated data from Model 1 showed that we were able to capture participant patterns in angle error (Fig 4D). While Model 1 captured the distance underestimation (Fig 4E), it did not capture the trend of increase in underestimation as a function of distance. We hypothesized that this effect could be an influence of past trials, in other words, a form of regression to the mean [35, 36]. Fig 4G shows the simulated angle error from Model 2. We were again able to capture the accurate angle predictions. Importantly, however, simulated distance error for Model 2, as shown in Fig 4H, better captured the pattern of distance underestimation. Model 2, in particular, captured the tendency of participant underestimation of distance to increase as a function of distance while Model 1 (which did not include trial history) was not able to capture this effect. These findings suggest that the increasing underestimation of distance was influenced, in part, by past trials. This effect was likely stronger in Experiment 2 than 1 due to the longer range of distances employed.

Encoding-Error Model. The Encoding-Error Model also captured some of the same patterns in the data as Model 1 and 2. The simulated data from the Encoding-Error Model showed accurate angle error and underestimation of distance errors as a function of distance (Fig 4J & 4K). We also considered how well the Encoding-Error Model compared with Model 2 in terms of capturing the mean systematic error in distance, which was 1- mean distance error (Fig 5C–5E). While the Encoding-Error Model fit the logarithmic function of systematic errors, the values were less accurate than Model 2. Similar to Experiment 1, Model 2 best fit the data, but the BIC and AIC favored Model 1 (S6D–S6F Fig). Notably, though, our analyses (see Fig 2D–2F) suggested that Model 1 did not capture the pattern of systematic errors (distance undershoot) and thus we removed it from the model comparison with the Encoding-Error Model. As shown in S8A–S8D Fig, we can see Model 2 fits 11 subject's data better while the Encoding-Error Model fit the other 5 subjects' data better. Similar to Experiment 1, the confusion matrix (S3A–S3C Fig) showed that Encoding-Error model did not fit its own simulated data well. This was likely due to small number of trials and the fact that the vector addition models involved fewer free parameters than the Encoding-Error Model (S1 Text). We return to a more detailed comparison of the models in the Discussion.

Discussion

In two different experiments, participants were guided on two sides of a triangle and then attempted to return to the origin without any input using a novel interface involving an omnidirectional treadmill. In Experiment 1, we manipulated triangle type (equilateral vs. isosceles vs. right vs. scalene) while holding distance on the unguided side constant to minimize prior effects. Consistent with previous work using the triangle completion task in small-scale room-sized environments [21, 31, 32, 50, 51], we found that participants underestimated distance and overestimated angle. These systematic errors, however, did not show a regression to the mean effect. In Experiment 1, our computational modeling results suggested that this pattern could be explained by a model in which participants underweighted side A compared to side B with an effect of past trial history. In Experiment 2, we found systematic errors in distance as participants accurately estimated the angle they needed to turn while increasingly underestimating the unguided side as a function of distance, consistent with logarithmic scaling described in the Weber-Fechner law. Modeling results for Experiment 2 further suggested equal weighting of both encoded sides. We also found no correlation between angle and distance errors in both experiments, consistent with reports that, at least in part, we derive angular motion from the semicircular canals and linear motion through the otoliths [52]. Our findings thus suggest that participants use independent estimates of direction and distance to estimate a configural homing vector, with the current trial guided sides influencing estimates of the configural homing vector.

In Experiment 1, we found that the triangle type tested had an influence on participants' distance and angle errors on the unguided side. We found that the equilateral triangles showed significantly lower distance error and similarly, we found a weak tendency for the isosceles triangle angles (30,120,30) to show lower angle overestimation. All triangle types observed in Experiment 1, however, showed an angle overestimation and distance underestimation. The question of whether the differences in distance and angle errors observed are due to participants using triangle templates is one we sought to answer in this study.

We believe that If participants were using a triangle template, then we would be able to observe accurate distance *and* angle estimations for one triangle type. For example, it might be possible to predict that equilateral triangles or right triangles would be overall more accurate than scalene triangles. This is because these geometries are far more regular and potentially easier to encode holistically, particularly given their influence on visually-guided navigation [53]. While we did not find one triangle type to have lower distance *and* angle errors at the group level, there may be several explanations for how triangle type had an influence on participants' performance.

First, we found two participants who showed a high degree of accuracy for both distance and angle on equilateral and right isosceles triangles (S3C Fig). This suggests that some participants may have used a strategy that involved deploying templates. Unlike Wiener et al. 2011 [34], we did not explicitly ask participants to focus on the outbound path or the origin, which in turn could possibly have affected some participants' likelihood of spontaneously using a template. Our results, in this way, build on other studies showing individual differences in human navigation studies [54–56].

A second explanation for how triangle type may have influenced participants' performance focuses on the possibility of using abstract geometric features. Studies in young children suggest that rather than using room geometry to navigate, children use certain geometric features (i.e., the distance between walls) to reorient themselves [57, 58]. In our task, participants were not directly exposed to the geometric regularities of the path but rather they had to "trace" it by walking it. This process may conserve some geometric features and not others, which may

in turn influence the pattern of errors. In particular, because participants could not directly perceive the entire shape, they may have attempted to guess it, and if their memory for a side was inaccurate, this would affect their memory for the entire shape. Thus, while we could not determine any correlation between patterns of errors and any obvious geometric features, there may be other features that were not detectable for our experimental task. These could have been influenced by geometry although in a way not detectable in our design due to tracing the path over several minutes rather than directly perceiving it in one glance.

Third, it may be that participants simply do not rely on any geometric features. Instead, it is possible that the small distance errors for the equilateral triangle type can be attributed to a working memory effect based on the equivalence of all three side distances. The overall systematic trends of overestimating angle and underestimating distance may have little to do with encoding geometric properties and instead be a result of under-encoding the guided legs and execution errors [33, 59]. Overall, the lack of any consistent effects in angle and distance for specific triangle types in terms of accuracy and the lack of a correlation between angle and distance suggest, on a group level, that participants may not have been using triangle templates. The fact that triangle type does have separate yet significant influences on individual distance and angle errors leaves room for further investigation.

In Experiment 2, we tested path integration over distances much longer than those typically employed in past human studies. Almost all of our current knowledge base about path integration mechanisms in humans derive from testing in room-sized environments, and therefore, in contrast to what is known about other species, the extent to which path integration mechanisms operate accurately over distances greater than 10 meters remains unclear. We found that participants were fairly accurate in their ability to complete the third side of a triangle, even for triangle perimeters as long as 500 meters. Although we found a systematic increase in error and underestimation as a function of longer distances, these biases increased logarithmically, suggesting that the basic mechanisms underlying path integration were not substantially different at 500 meters compared to 25 meters. In contrast to Experiment 1, we found that both sides A and B contributed equally to errors in unguided side C, although we attribute this effect to the fact that we did not manipulate angle in Experiment 2. We found that past trial history contributed significantly to the pattern of errors at longer distances. These findings suggest that in fact some of the properties of path integration do change somewhat over longer distances, particularly the tendency to erroneously weigh past trials to estimate the current ones. Given that our two models, however, involved the same basic conceptual setup (side A +B = C), these results suggest that the basic mechanism of adding vector values for the guided sides to compute a configural homing vector held constant across experiments.

Our computational modeling results indicated an effect of past trials on participant error patterns in both experiments. In other words, for the longer distance triangles, we found a weak, but significant bias for past trials to influence the extent to which participants underestimated the amount they needed to walk on the current triangle. For large triangles, therefore, shorter past trials would result in a greater tendency to underestimate distance. Notably, including the history term in our model significantly increased our ability to account for the increasing tendency of participants to undershoot the distance they needed to walk on the unguided side. These findings support the idea that for particularly long distances, path integration is also influenced by a form of regression to the mean from past trials, thus explaining why undershoot increased with longer distances. These findings, which, to the best of our knowledge, have not been demonstrated previously at such long distances in humans, suggest that path integration is not merely a function of the current walked triangle, but is also influenced by the memories and experiences of past trajectories.

Because of our strong reliance on visual input, testing humans in the absence of vision is challenging, particularly due to the possibility of trip hazards and collisions. Thus, many researchers have chosen to investigate path integration using desktop VR, which also allows simultaneous brain imaging, for example, using fMRI [54, 60]. One limitation with desktop VR is that although it preserves optic flow, it lacks the rich cues that one obtains from freely moving the body in space [19]. These include vestibular information from head turns, proprioceptive information about body position, efferent copy from motor movements, and somatosensory input from the feet as they move over the surface [16] [61–66]. Our novel interface was able to reproduce many of these cues, particularly those that would be expected from angular vestibular cues, turning, and shuffling the sides of the feet. The treadmill, however, has diminished translational vestibular cues although it is notable that the distance underestimation results we obtained were largely comparable to past studies using the real-world triangle completion task [32]. As such, we were able to capture novel aspects of non-visual navigation otherwise difficult to observe in desktop-VR.

Additionally, participants in our study generated their linear and angular motion, while non-VR versions of the triangle completion task used in the past relied on the experimenter physically guiding the participants' movements. Previous versions of path completion tasks have used an object (rod or rope) in which the experimenter guides the participants by pulling or lowering for turning [21, 32, 35]. In contrast, in our design, participants received feedback from handheld controllers indicating which way to go. We believe that the use of feedback via handheld controllers, rather than external forces to guide participants, better approximates active walking. Specifically, active walking requires one to initiate the movement while outside forces that initiate or guide the movement would typically be referred to as passive. We believe by controlling for active walking during the guided portion, we have better controlled for differences between guided and unguided conditions. While the distinction between active and passive movement is a subtle one, recent work suggests important differences between these two forms of walking in terms of their neural bases [52].

Model comparisons

Vector addition has long been assumed to be the functioning principle for path integration [23, 24, 26]. The vector addition models proposed in this paper (Models 1&2) assume that the configural homing vector is computed by summing vector representations of sides A and B. In contrast, the Encoding-Error Model assumes that the configural homing vector is computed using the distance and angle values experienced during the entire guided portion. While both models are similar in aim, we believe the computational principles for the vector model may be more plausible. To employ the Encoding-Error Model, participants must form a representation of the linear relationship between distance guided and distance walked (distance representation) as well as for turns, for each path configuration. In contrast, vector addition models assume a linear relationship between the guided sides and the configural homing vector, with the possibility of prior trials influencing the current trajectory.

As mentioned in the Introduction, there are other reasons to think that vector addition models confer advantages, particularly in accounting for human path integration findings from the triangle completion task. The Encoding-Error Model has four requirements, with one important assumption being that the internal representations must obey Euclidean axioms. Recent papers, however, suggest that human spatial navigation, in some instances, may be better characterized by representations based on non-Euclidean labeled graphs [67]. Specifically, Warren et al. 2019 [67] described path integration using simple vector manipulations with such manipulations preserved in non-Euclidean spaces. Our model, which can be readily

adapted to non-Euclidean geometries, would therefore also provide greater flexibility than the Encoding-Error Model in terms of fitting violations of Euclidean axioms.

Another requirement of the Encoding-Error Model is the assumption that all systematic errors occur during encoding rather than during spatial reasoning or execution. Vector addition models are more flexible, assuming systematic errors can aggregate at different stages, whether it is during encoding, retrieval or computation of the configural homing vector. The Encoding-Error Model, however, is limited in that side A and B derive from the same linear function, such that side A cannot be underestimated more than side B. There may be instances, however, in which one side is weighted differently in a path with 2 segments compared to 5 segments [59]. This was also true in Experiment 1, where we found that side A was underweighted compared to side B.

In addition, the Encoding-Error Model is limited to 2 segmented triangular paths, based on the law of cosines (S1 Text) and does not perform well with 3 segmented paths [31]. In contrast, vector addition models can readily be extended to n paths with the caveat of adding a free parameter with each segment. Notably, the vector addition models we employed here provided an overall better fit of the actual data (S6A & S6D Fig). We note, though, that the Encoding-Error model cannot be fully differentiated during model recovery (S7 Fig). The likely reason for this is the small number of trials in our task. While both Model 2 and Encoding-Error model can account for some patterns in the data, including systematic errors, importantly, Model 2 has the best log-likelihood fit (S6A & S6D Fig), despite the Encoding-Error model having more free-parameters. Overall, therefore, we think the vector addition models provide a better fit of our data and are more parsimonious although more work is needed to allow a detailed and formal model comparison.

Limitations of the Vector Models

While the vector addition models employed here do a fairly good job of capturing the patterns of our findings in the two experiments in this study, they are not without certain limitations. One issue is that the model in its current form assumes that side A and side B are encoded with similar directions (i.e. $\beta_A x_A^t$ has the same direction as x_A^t) or opposite directions (S10 Fig). Our current versions also assume that only the vector magnitudes affect systematic errors (see S10 Fig). We hope to address the issue of vector directions in more detail in future models.

Methods

Ethics statement

All participants gave written informed consent to participate in the study, which was approved by the Institutional Review Board at the University of California, Davis (Protocol: 816949).

Training and the triangle completion task

We employed a task used previously to investigate human path integration termed the triangle completion task [21]. Briefly, the task involves guiding participants on two sides of a triangle and then completing the third side without guidance or feedback. Based on our goal of studying a variety of different triangle types and sizes, we adapted the task to an omnidirectional treadmill, the Cyberith Virtualizer treadmill. The task involved participants walking on the treadmill, with guidance on two of the sides provided by somatosensory feedback from HTC VIVE handheld controllers. Participants wore the HTC VIVE headmounted headset to allow us to track head and body position, as well as to limit visual input.

To first ensure that participants could walk comfortably in the treadmill, we employed a pre-experimental training session. We employed an HTC VIVE head-mounted display to give visual feedback to ensure balance and comfort on the treadmill. In the first part of the training, we included a 3-stage puzzle game created in Unity 2017.1.1f1 in which participants had to explore an environment to find an object. Once participants completed the 3-stage puzzle game, reported no cybersickness, and the experimenter determined that their walking technique was adequate, they advanced to the next level. At this point, we introduced the HTC VIVE handheld controllers feedback system (Fig 1B) and had subjects walk straight lines with no visual information while receiving feedback from the handheld controllers. This insured that they could accurately perform the guided sides. Following this, they performed a small number of practice triangles. After practicing the triangle completion task on 6 unique triangles, which were not included in the experiment, the experiment started. The training period ranged from 30–60 min. To ensure participant safety, we occasionally questioned them about how they were feeling to guard against issues with cybersickness.

Participants then proceeded to the main experiment. The first experiment involved manipulating triangle geometry (i.e., primarily the angles they turned) and Experiment 2 involved manipulating triangle size (i.e., we manipulated the distance they walked on the third / unguided side). Trial sequences were randomly chosen from 5 pseudorandomized configurations. In both experiments, we guided participants along the first two sides of the triangle using the handheld controllers' feedback system (Fig 1B). The feedback system was designed such that if the participants strayed from their path, the controller vibrated accordingly (for example if the right controller vibrated, they would need to turn left) to help guide them in walking in a straight line. When participants walked in the correct direction, the controller did not send feedback, allowing for active walking (passive guidance). Participants were guided alongside A' and then alongside B' by controller feedback (Fig 1C). At G2', the handheld controller feedback system turned off and participants were instructed to find their way to the start point. Participants pressed the trigger on the handheld controllers once they believed that they reached the start point. We constructed trial specific vectors to capture the performance variability during guided sides (see Fig 1E). We manually inspected these trials, and those which showed a clear deviation from linearity were excluded, which resulted in approximately 16.5% ($\frac{117}{696}$) of removal of trials from Experiment 1 and 6.88% ($\frac{36}{523}$) from Experiment 2 across participants. Participant data that exceeded 25% removed trials were excluded from the analysis. We redid the analysis by including all trials and participants and obtained similar results to what are reported here.

Modeling

Description of models. To further understand how the guided sides contributed to the angle and distance errors of the unguided side, we created a vector model of path integration. In this model, we assume that participants estimate a "configural homing vector", x_C^t, by combining the vectors corresponding to each of the guided sides for that trial (denoted by superscript t), x_A^t and x_B^t. If path integration were optimal, people would combine these vectors in the following way

$$x_C^t = -(x_A^t + x_B^t) \quad (1)$$

and would return perfectly to the point of origin by walking along the vector x_C^t.

We assumed that people could over, or underweight, a given side when computing the sum–perhaps because they integrate evidence unevenly over time [68]. To model participants'

biased responses, we allowed $\tilde{x}_D^t$ to be a *weighted* sum of the vectors from the first two sides:

$$\tilde{x}_D^t = -(\beta_A x_A^t + \beta_B x_B^t) \quad 2$$

Where β_A and β_B denote the weights given to side A and side B respectively (Fig 1D).

Of course, real participants are suboptimal and we modeled these suboptimalities in a number of different ways. First, people may not perfectly encode the vectors from the guided sides and/or may not perfectly implement the desired action, adding noise to the sum in Eq 2. Thus, we assumed that the vector they actually walked x_D^t was sampled from a Gaussian distribution centered on $\tilde{x}_D^t$, i.e.

$$P(x_D^t|\tilde{\sigma}, \beta_A, \beta_B) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x_D^t - \tilde{x}_D^t)^2}{2\sigma^2}) \quad 3$$

Where σ^2 is the variance of the noise. Consistent with Weber's law, we assumed this variance increased with the distance walked to match our finding of increased variance as a factor of distance walked in Experiment 2 (see Results, Fig 5C).

$$\sigma = \tilde{\sigma} * \sqrt{x_A^{t\,2} + x_B^{t\,2}} \quad 4$$

Combining the first two sides gives us Model 1 (Eq 5), which includes noise and the possibility of over and underweighting the sides.

$$x_D^t = -(\beta_A x_A^t + \beta_B x_B^t) + \varepsilon \quad 5$$

Where ε is the noise term sampled from a normal destitution with mean 0 and standard deviation σ.

Finally, we allowed for the possibility that there may be sequential effects in our paradigm, i.e. there was an influence of previous trials on the current response. We modeled these sequential effects by including the vectors walked (x_A^{t-n}, x_B^{t-n} and x_D^{t-n}) from past trials. For simplicity, we assumed that the effect of past trials decayed exponentially into the past [69], thus writing x_D^t as

$$x_D^t = -(\beta_A x_A^t + \beta_B x_B^t + \beta_\lambda[\lambda_A^{t-1} + \lambda_B^{t-1} - \lambda_D^{t-1}])) + \varepsilon \quad 6$$

$$\lambda_n^t = x_n^t + \alpha\lambda_n^{t-1} \quad 7$$

Where λ_n is a linear combination of the previous vectors, fitted with α, which ranges between 0 to 1, to capture the impact of prior trials. Thus, including the possible effect of past trials gave us Model 2.

Fitting the model. We fit the model using a maximum likelihood approach. In particular, we computed the log likelihood of the responses for each subject, as a function of model parameters:

$$LL(\tilde{\sigma}, \beta_A, \beta_B, \beta_\lambda, \alpha) = \sum_{t=1}^{T} \frac{log(2\pi\sigma^2)}{2} \frac{(\tilde{x}_D^t - x_D^t)^2}{4\sigma^2} \quad 8$$

We then found the parameters that maximized the likelihood using Matlab's fmincon function.

Simulating the models. To simulate the model, we used the parameter values fit for each subject to compute the mean x_D^t for each trial. To model the noise in each person's choice, we perturbed the estimate of $\tilde{x}_D^t$ by isotropic Gaussian noise of mean 0 and variance σ^2.

Encoding-Error Model. We recreated the Encoding-Error Model from Fujita et al. 1993. See S1 Text for more details. We used the same fitting and simulation method used for Model 1 and Model 2 with the exception of dividing the data for left and right-handed triangle to better accommodate the parameters of the Encoding-Error Model [32].

Model Comparison Methods. We used two methods of model comparisons: 1) Penalized-Log-likelihood criteria's Bayes Information Criterion (BIC) [70] and Akaike information criterion (AIC) [71]. Both express similar information about the generalizability of the model by penalizing for the number of free parameters. To test how meaningful our model comparisons results are in our task we also tested for model recovery. We did this by simulating each model with randomized parameter values and then fitting the models to the simulated data, allowing comparison of the AIC and BIC (see section 6 and Appendix B in [48]). We performed each simulation at the participant level and then subsequently compared BIC values by calculating exceedance probabilities, which measured how likely it is that the given model fits all of the data [72]. This group level statistic is similar to AIC and BIC. Computed exceedance probabilities on our data as well as each model by simulating 100 times and comparing with the methods mentioned above. These methods are illustrated in S6 Fig where the probability of the model fit for the simulated data ranges from 0 to1. The Exceedance Probability is calculated using SPM 12 spm_BMS function.

Bayes Factor Analyses. We included a Bayes Factor analysis for all statistical analyses [73]. For results below our significance threshold ($p<0.05$), we used a Bayes Factor BF_{10} to indicate the degree of favorability toward the alternative hypothesis. For results that were not below our significance threshold, we employed the Bayes Null factor, BF_{01}. Note that the larger the Bayes Factor, regardless of whether in favor of the alternative or null, the greater the evidence.

Experiment 1

Participants. We tested a total of 26 participants (12m,14f), 4 (1m, 3f) of which were removed due to exceeding 25% of trials removed (see methods), Participants were tested on 7 different triangles described in detail in the methods (i.e., scalene, isosceles, right, equilateral, and isosceles-right). Estimates of sample size were based on the 12 participants used in Loomis et al. 1993 and in subsequent studies by [74]. that employed a similar experimental design: as we were additionally testing a larger range of triangles, we thus approximately doubled the sample size.

Procedure. We outline the basic set up for triangle geometry in Fig 1E, which shows the stacked triangle templates, with a constant 10m side C' (unguided side), while manipulating the angle. The 7 triangle configurations are shown in S1 Table, with 3 scalene, 1 isosceles, 1 right, 1 equilateral, and 1 isosceles-right. To keep side C' at a constant 10m across all 7 triangles, we employed different side A' and side B' (guided portion) sizes to accommodate the different angles. Note side D is the participant's response, and distance errors are calculated as the ratio (side D) / (side C). There were 28 trials, in which 14 of them were left-handed (subjects only made left turn) and 14 right handed (subject only made right turns). We did this to avoid any advantages for right vs. left turns during the task.

In Experiment 1, as part of ensuring the compliance and efficacy of the handheld controllers in following the guided sides, we compared with a condition in which participants walked the guided sides on half the trials using a visual beacon. In this situation, participants saw a large red monolith that they walked to while receiving feedback from the handheld controllers. It is important to emphasize that the vision-guided trials were only present for the *guided* sides and were simply to ensure that participants accurately encoded the guided sides before performing the unguided sides.

Experiment 2

Participants. We tested a total of 21 participants (9m,11f), 3 (1m 2f) of which did not complete the experiment, with additional 1 female participant removed from the analysis for exceeding 25% trials below criterial performance. Given the longer distances in Experiment 2, participants were allowed to take a break, but only at the end of a trial. About 50% of participants took a break at some point during the experiment.

Procedure. Here, we employed scalene triangles with different length perimeters to allow us to manipulate distance while keeping angle constant, testing 5 different triangle sizes. Fig 1F shows the stacked triangle templates we employed with constant internal angles but varying in size. The triangle configurations are shown in S2 Table, with 15m, 25m, 127m, 253m, and 506m perimeters. There were 30 trials, with 15 of them left-handed (participants only made left turn) and 15 right-handed (participants only made right turns). Unlike Experiment 1, there were no vision trials. Due to testing longer distances and wanting to avoid fatigue, we limited the number of trials for the longest distance triangles. The distributions of trials were 10 for the 15m triangle, 10 for the 25m triangle, 8 for the 127m triangle, 4 for the 253m triangle, and 2 for the 506m triangle.

Supporting information

S1 Table. The configuration of each triangle used in experiment 1.
(XLSX)

S2 Table. The configuration of each triangle used in experiment 2.
(XLSX)

S1 Fig. Raw trials from experiment 1 (top 8) and experiment 2 (bottom 8).
(TIF)

S2 Fig. Comparing vision and non-vision trials. (A) Combined distance walked during guided sides during vision on and vision off trial, showing now differences ($t(21) = 1.09$, $p = 0.288$, Cohen's $d = 0.336$ and $BF_{01}>3$) (B) Angle error from experiment 1, showing a small but significant difference between vision on and off condition ($t(21) = 2.46$, $p<0.022$, Cohen's $d = 0.248$ and $BF_{10} = 2.54$) (C) Distance error from experiment 1, showing a significant difference between vision on and off condition ($t(21) = 2.71$, $p<0.013$, Cohen's $d = 0.232$ and $BF_{10} = 3.94$). (D) Angle error from experiment 1, ANOVA significant for triangle type $F(6,21) = 2.9$, $p<0.01$, $\eta^2 = 0.058$ $BF_{10} = 1.72$. and Vision $F(1, 21) = 4.9$, $p<0.026$, $\eta^2 = 0.016$ $BF_{10} = 1.16$, but not for the interaction between Type and Vision $F(6, 21) = 1.454$, $p = 0.194$, $\eta^2 = 0.029$ $BF_{10} = 0.432$. (E) Distance error from experiment 1, ANOVA significant for triangle type $F(6, 21) = 5.7$, $p<0.1.33e\text{-}5$, $\eta^2 = 0.109$ $BF_{10}>10$ and r Vision $F(1, 21) = 8.2$, $p<0.004$, $\eta^2 = 0.026$ $BF_{10}>4$, but not for the interaction between Type and Vision $F(6, 21) = 0.199$, $p = 0.976$, $\eta^2 = 0.004$ $BF_{10}>10$.
(TIF)

S3 Fig. Comparing right and left-handed trials. (A) Angle error from experiment 1 which showed no difference between left and right-handed triangles ($t(21) = 0.7$, $p = 0.485$, Cohen's $d = 0.118$ and $BF_{01}>3$). (B) Distance error from experiment 1, which showed no difference between left and right-handed triangle ($t(21) = 1.136$, $p = 0.268$, Cohen's $d = 0.103$ and $BF_{01} = 2.53$). (C) Angle error from experiment 2, again showing no difference between left and right-handed triangle ($t(16) = 1.51$, $p = 0.151$, Cohen's $d = 0.245$ and $BF_{01} = 1.55$). (D) Distance error from experiment 2, which showed no difference between left and right-handed triangle

($t(16) = 0.724$, $p = 0.4797$, Cohen's $d = 0.176$ and $BF_{01} = 3.188$).
(TIF)

S4 Fig. Angle and distance accuracy. Raster plot (A) showing the percentage of responses with less than 15% angle error (ranging from -27˚ to 27˚) for triangle type (x-axis) and participants (y-axis). Participants were 281.39% more likely to have <15% angle error in their unguided side than <15% total error (angle and distance). (B) percentage of responses with less than 15% distance error (8.5m to 11.5m). Participants are 208.14% more likely to have <15% distance error in their unguided side than <15% total error (angle and distance). (C) percentage of responses with less than 15% angle error (ranging from -27˚ to 27˚) and 10% distance error (8.5m to 11.5m). In (C) we can see that all of participant AT03's responses for triangle 45-90-45 are less than 15% error for both angle and distance error. And 80% for equilateral triangle (60-60-60) for participant AT11.
(TIF)

S5 Fig. Position Error. Mean position error (total distance from the participant's final position and the origin. No main effect of triangle type (left) 1-way ANOVA $F(6,21) = 1.34$, $p<0.24$, $\eta 2 = 0.06$ $BF_{01} = 6.76$.
(TIF)

S6 Fig. Best bit model between Model 1, Model 2, and Encoding-Error Model. Comparing model fitting of the individual participant's data. A) Shows best model fit (highest loglikelihood values) for each subject in experiment 1. B&C) Lowest AIC and BIC values across the 3 models for each subject in experiment 2. D) Shows best model fit (highest loglikelihood values) for each subject in experiment 2. E&F) Lowest AIC and BIC values across the 3 models for each subject in experiment 2
(TIF)

S7 Fig. Model recovery between Model 1, Model 2, and Encoding-Error Model. Model recovery confusion matrices. In column and row, 1 = Model 1, 2 = Model 2, and 3 = the Encoding-Error Model. Probability ranges from 0 to 1. (A & B) Best AIC and BIC for Experiment 1 respectively. The higher value in the diagonal shows better model recovery from this experiment. The Encoding-Error does not fit its own simulated data well. (C)The Exceedance Probability for Experiment 1. (D & E) show best AIC and BIC for Experiment 2 respectively. Again, we see Encoding-Error does not fit its simulated data well. (C)The Exceedance Probability for Experiment 2.
(TIF)

S8 Fig. Best bit model between Model 2 and Encoding-Error Model. Comparing model fitting of the individual participant's data. A) Best model fit (highest loglikelihood values) for each participant in experiment 2. B&C) Lowest AIC and BIC values across the 3 models for each participant in experiment 2. D)The exceedance probability of each model for experiment 2.
(TIF)

S9 Fig. Model recovery between Model 2 and Encoding-Error Model. Model recovery confusion matrices. In rows and columns, 1 = Model 1 and 2 = the Encoding-Error Model. Probability ranges from 0 to 1. (A & B) Show best AIC and BIC for Experiment 2 respectively. We see the Encoding-Error does not fit its own simulated data well. (C)The Exceedance Probability for Experiment 2.
(TIF)

S10 Fig. Vector addition model Encoding error model. Recreation Fig 1 in Fujita et al. 1993 (modified to employ the names of variable used in our study), as shown with the black solid line which represents the participant's walked trajectories. The dashed blue lines are the internal encoded representation as predicted by the Encoding-Error Model (these are represented with subscript r). The dashed red line is a rough overlay for the paths predicted by the vector addition model. As shown below, the Encoding-Error model creates sides A_r and B_r and angle $\angle ab_r$. According to Euclidian properties, this results in side $C_r = C$ and angle $\angle bc_r = \angle bc$. The vector addition model (in red), instead, adjusts A and B accordingly to fit participant's response C. Thus, the angle $\angle ab$ and $\angle bc$ remain relatively the same in the vector addition model. The critical difference here is that the vector addition models assume a different suboptimal encoding of guided distance for sides A and B. In contrast, the Encoding-Error model assumes the same suboptimal encoding of the guided sides A and B and separate suboptimal encoding of $\angle ab$ such that it preserves Euclid's postulates. *Note that this does not include fitted noise which would slightly change the direction of all three sides. This would in fact make it non-Euclidean such that the total sum of all internal angles does not equal 180.
(TIF)

S1 Text. Fitting Encoding-Error Model.
(PDF)

Acknowledgments

The Authors are gratful to E. Erlenbach for helping during data collection.

Author Contributions

Conceptualization: Sevan K. Harootonian, Arne D. Ekstrom.

Data curation: Sevan K. Harootonian, Eli M. Ziskin.

Formal analysis: Sevan K. Harootonian.

Funding acquisition: Arne D. Ekstrom.

Investigation: Sevan K. Harootonian.

Methodology: Sevan K. Harootonian, Robert C. Wilson, Arne D. Ekstrom.

Software: Sevan K. Harootonian, Robert C. Wilson, Lukáš Hejtmánek.

Supervision: Robert C. Wilson, Arne D. Ekstrom.

Writing – original draft: Sevan K. Harootonian, Robert C. Wilson, Arne D. Ekstrom.

Writing – review & editing: Sevan K. Harootonian, Robert C. Wilson, Arne D. Ekstrom.

References

1. Darwin C. Charles Darwin's natural selection: being the second part of his big species book written from 1856 to 1858: Cambridge University Press; 1856/1987.
2. Görner P. Die optische und kinästhetische Orientierung der Trichterspinne Agelena Labyrinthica (Cl.). Zeitschrift für vergleichende Physiologie. 1958; 41(2):111–53. https://doi.org/10.1007/bf00345583
3. Lindauer M. Kompaßorientierung. In: Autrum H, Bünning E, v. Frisch K, Hadorn E, Kühn A, Mayr E, et al., editors. Orientierung der Tiere / Animal Orientation: Symposium in Garmisch-Partenkirchen 17–21 9 1962. Berlin, Heidelberg: Springer Berlin Heidelberg; 1963. p. 158–81.
4. Mittelstaedt M-L, Mittelstaedt H. Homing by path integration in a mammal. Naturwissenschaften. 1980; 67(11):566–7. https://doi.org/10.1007/bf00450672

5. Etienne AS. The Control of Short-Distance Homing in the Golden Hamster. In: Ellen P, Thinus-Blanc C, editors. Cognitive Processes and Spatial Orientation in Animal and Man: Volume I Experimental Animal Psychology and Ethology. Dordrecht: Springer Netherlands; 1987. p. 233–51.
6. Alyan S, Jander R. Short-range homing in the house mouse, Mus musculus: stages in the learning of directions. Animal Behaviour. 1994; 48(2):285–98.
7. Tolman EC. Cognitive maps in rats and men. Psychological Review. 1948; 55(4):189–208. https://doi.org/10.1037/h0061626 PMID: 18870876
8. Mittelstaedt H, Mittelstaedt M-L, editors. Homing by Path Integration1982; Berlin, Heidelberg: Springer Berlin Heidelberg.
9. Wehner R, Srinivasan MV. Searching Behavior of Desert Ants, Genus Cataglyphis (Formicidae, Hymenoptera). Journal of Comparative Physiology. 1981; 142(3):315–38. WOS:A1981LT49400004.
10. Mittelstaedt H. The role of multimodal convergence in homing by path integration. Fortschritte der Zoologie. 1983; 28:197–212.
11. Green J, Adachi A, Shah KK, Hirokawa JD, Magani PS, Maimon G. A neural circuit architecture for angular integration in Drosophila. Nature. 2017; 546:101. https://doi.org/10.1038/nature22343 PMID: 28538731
12. Seguinot V, Cattet J, Benhamou S. Path integration in dogs. Animal Behaviour. 1998; 55(4):787–97. https://doi.org/10.1006/anbe.1997.0662 PMID: 9632467
13. Herrick FH. Homing Powers of the Cat. The Scientific Monthly. 1922; 14(6):525–39.
14. Beritashvili IS. Neural mechanisms of higher vertebrate behavior. 1965.
15. Redish AD. Beyond the Cognitive Map. Boston, MA: MIT Press; 1999.
16. Gallistel CR. The Organization of Learning. Cambridge, MA: MT Press; 1990.
17. Klatzky RL, Loomis JM, Golledge RG. Encoding spatial representations through nonvisually guided locomotion. In: Medin, editor. The psychology of learning and motivation. 37: Academic Press; 1997.
18. Chance SS, Gaunet F, Beall AC, Loomis JM. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. Presence-Teleop Virt. 1998; 7(2):168–78. https://doi.org/10.1162/105474698565659 WOS:000072949600006.
19. Starrett MJ, Ekstrom AD. Perspective: Assessing the Flexible Acquisition, Integration, and Deployment of Human Spatial Representations and Information. Frontiers in human neuroscience. 2018; 12.
20. Taube JS, Valerio S, Yoder RM. Is Navigation in Virtual Reality with fMRI Really Navigation? Journal of Cognitive Neuroscience. 2013. Epub 2013/03/16. https://doi.org/10.1162/jocn_a_00386 PMID: 23489142.
21. Loomis JM, Klatzky RL, Golledge RG, Cicinelli JG, Pellegrino JW, Fry PA. Nonvisual navigation by blind and sighted: Assessment of path integration ability. Journal of Experimental Psychology: General. 1993; 122(1):73–91. https://doi.org/10.1037/0096-3445.122.1.73
22. Müller M, Wehner R. Path integration in desert ants, Cataglyphis fortis. Proceedings of the National Academy of Sciences. 1988; 85(14):5287–90. https://doi.org/10.1073/pnas.85.14.5287 PMID: 16593958
23. Cartwright BA, Collett TS. Landmark maps for honeybees. Biological Cybernetics. 1987; 57(1):85–93. https://doi.org/10.1007/bf00318718
24. Etienne AS, Maurer R, Berlie J, Reverdin B, Rowe T, Georgakopoulos J, et al. Navigation through vector addition. Nature. 1998; 396(6707):161–4. https://doi.org/10.1038/24151 PMID: 9823894
25. Etienne AS, Maurer R, Seguinot V. Path integration in mammals and its interaction with visual landmarks. J Exp Biol. 1996; 199(Pt 1):201–9. PMID: 8576691.
26. Kubie JL, Fenton AA. Heading-vector navigation based on head-direction cells and path integration. Hippocampus. 2009; 19(5):456–79. https://doi.org/10.1002/hipo.20532 PMID: 19072761
27. Wittmann T, Schwegler H. Path integration—a network model. Biological Cybernetics. 1995; 73 (6):569–75. https://doi.org/10.1007/bf00199549
28. Fujita N, Loomis JM, Klatzky RL, Golledge RG. A Minimal Representation for Dead-Reckoning Navigation: Updating the Homing Vector. Geographical Analysis. 1990; 22(4):324–35. https://doi.org/10.1111/j.1538-4632.1990.tb00214.x
29. Benhamou S, Séguinot V. How to find one's way in the labyrinth of path integration models. Journal of Theoretical Biology. 1995; 174(4):463–6.
30. Etienne AS, Jeffery KJ. Path integration in mammals. Hippocampus. 2004; 14(2):180–92. https://doi.org/10.1002/hipo.10173 PMID: 15098724

31. Fujita N, Klatzky RL, Loomis JM, Golledge RG. The Encoding-Error Model of Pathway Completion without Vision. Geographical Analysis. 1993; 25(4):295–314. https://doi.org/10.1111/j.1538-4632.1993.tb00300.x

32. Klatzky RL, Beall AC, Loomis JM, Golledge RG, Philbeck JW. Human navigation ability: Tests of the encoding-error model of path integration. Spatial Cognition and Computation. 1999; 1(1):31–65. https://doi.org/10.1023/a:1010061313300

33. Chrastil ER, Warren WH. Rotational error in path integration: encoding and execution errors in angle reproduction. Experimental Brain Research. 2017; 235(6):1885–97. https://doi.org/10.1007/s00221-017-4910-y PMID: 28303327

34. Wiener JM, Berthoz A, Wolbers T. Dissociable cognitive mechanisms underlying human path integration. Experimental Brain Research. 2011; 208(1):61–71. https://doi.org/10.1007/s00221-010-2460-7 PMID: 20972774

35. Klatzky RL, Loomis JM, Golledge RG, Cicinelli JG, Doherty S, Pellegrino JW. Acquisition of Route and Survey Knowledge in the Absence of Vision. Journal of Motor Behavior. 1990; 22(1):19–43. https://doi.org/10.1080/00222895.1990.10735500 PMID: 15111279

36. Petzschner FH, Glasauer S. Iterative Bayesian Estimation as an Explanation for Range and Regression Effects: A Study on Human Path Integration. The Journal of Neuroscience. 2011; 31(47):17220–9. https://doi.org/10.1523/JNEUROSCI.2028-11.2011 PMID: 22114288

37. Teghtsoonian R, Teghtsoonian M. Range and regression effects in magnitude scaling. Perception & Psychophysics. 1978; 24(4):305–14. https://doi.org/10.3758/bf03204247 PMID: 750977

38. Cheng K. A purely geometric module in the rat's spatial representation. Cognition. 1986; 23(2):149–78. https://doi.org/10.1016/0010-0277(86)90041-7 PMID: 3742991

39. Landau B, Gleitman H, Spelke E. Spatial knowledge and geometric representation in a child blind from birth. Science. 1981; 213(4513):1275–8. https://doi.org/10.1126/science.7268438 PMID: 7268438

40. Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI. Microstructure of a spatial map in the entorhinal cortex. Nature. 2005; 436(7052):801. https://doi.org/10.1038/nature03721 PMID: 15965463

41. Bellmund JLS, Gärdenfors P, Moser EI, Doeller CF. Navigating cognition: Spatial codes for human thinking. Science. 2018; 362(6415):eaat6766. https://doi.org/10.1126/science.aat6766 PMID: 30409861

42. Milivojevic B, Doeller CF. Mnemonic networks in the hippocampal formation: From spatial maps to temporal and conceptual codes. Journal of Experimental Psychology: General. 2013; 142(4):1231.

43. Chen X, He Q, Kelly Jonathan W, Fiete Ila R, McNamara Timothy P. Bias in Human Path Integration Is Predicted by Properties of Grid Cells. Current Biology. 2015; 25(13):1771–6. https://doi.org/10.1016/j.cub.2015.05.031 PMID: 26073138

44. Moser EI, Moser M-B. A metric for space. Hippocampus. 2008; 18(12):1142–56. https://doi.org/10.1002/hipo.20483 PMID: 19021254

45. Cheung A, Ball D, Milford M, Wueth G, Wiles J. Maintaining a Cognitive Map in Darkness: The Need to Fuse Boundary Knowledge with Path Integration. PLoS Comput Biol. 2012; 8(8):1–22.

46. Burak Y, Fiete IR. Accurate Path Integration in Continuous Attractor Network Models of Grid Cells. PLOS Computational Biology. 2009; 5(2):e1000291. https://doi.org/10.1371/journal.pcbi.1000291 PMID: 19229307

47. Palminteri S, Wyart V, Koechlin E. The Importance of Falsification in Computational Cognitive Modeling. Trends in Cognitive Sciences. 2017; 21(6):425–33. https://doi.org/10.1016/j.tics.2017.03.011 PMID: 28476348

48. Wilson RC, Collins AGE. Ten simple rules for the computational modeling of behavioral data. eLife. 2019; 8:e49547. https://doi.org/10.7554/eLife.49547 PMID: 31769410

49. Stevens SS. Psychophysics: introduction to its perceptual, neural, and social prospects. New York,: Wiley; 1975. v, 329 p. p.

50. Philbeck JW, Klatzky RL, Behrmann M, Loomis JM, Goodridge J. Active control of locomotion facilitates nonvisual navigation. J Exp Psychol Hum Percept Perform. 2001; 27(1):141–53. PMID: 11248929.

51. Yamamoto N, Philbeck JW, Woods AJ, Gajewski DA, Arthur JC, Potolicchio SJ, et al. Medial Temporal Lobe Roles in Human Path Integration. Plos One. 2014; 9(5). ARTN e9658310.1371/journal.pone.0096583. WOS:000338029800088.

52. Carriot J, Jamali M, Brooks JX, Cullen KE. Integration of Canal and Otolith Inputs by Central Vestibular Neurons Is Subadditive for Both Active and Passive Self-Motion: Implication for Perception. The Journal of Neuroscience. 2015; 35(8):3555–65. https://doi.org/10.1523/JNEUROSCI.3540-14.2015 PMID: 25716854

53. Moar I, Bower GH. Inconsistency in Spatial Knowledge. Memory & Cognition. 1983; 11(2):107–13. https://doi.org/10.3758/Bf03213464 WOS:A1983QM26800001. PMID: 6865743

54. Chrastil ER, Sherrill KR, Hasselmo ME, Stern CE. There and Back Again: Hippocampus and Retrosplenial Cortex Track Homing Distance during Human Path Integration. Journal of Neuroscience. 2015; 35 (46):15442–52. https://doi.org/10.1523/JNEUROSCI.1209-15.2015 WOS:000366054500020. PMID: 26586830

55. Marchette SA, Bakker A, Shelton AL. Cognitive Mappers to Creatures of Habit: Differential Engagement of Place and Response Learning Mechanisms Predicts Human Navigational Behavior. The Journal of Neuroscience. 2011; 31(43):15264–8. https://doi.org/10.1523/JNEUROSCI.3634-11.2011 PMID: 22031872

56. Weisberg SM, Schinazi VR, Newcombe NS, Shipley TF, Epstein RA. Variations in cognitive maps: understanding individual differences in navigation. J Exp Psychol Learn Mem Cogn. 2014; 40(3):669–82. Epub 2013/12/25. https://doi.org/10.1037/a0035261 PMID: 24364725.

57. Dillon MR, Huang Y, Spelke ES. Core foundations of abstract geometry. Proceedings of the National Academy of Sciences. 2013; 110(35):14191–5. https://doi.org/10.1073/pnas.1312640110 PMID: 23940342

58. Lee SA, Sovrano VA, Spelke ES. Navigation as a source of geometric knowledge: Young children's use of length, angle, distance, and direction in a reorientation task. Cognition. 2012; 123(1):144–61. https://doi.org/10.1016/j.cognition.2011.12.015 PMID: 22257573

59. Wan X, Wang RF, Crowell JA. Effects of Basic Path Properties on Human Path Integration. Spatial Cognition & Computation. 2013; 13(1):79–101. https://doi.org/10.1080/13875868.2012.678521

60. Chadwick MJ, Jolly AEJ, Amos DP, Hassabis D, Spiers HJ. A Goal Direction Signal in the Human Entorhinal/Subicular Region. Current Biology. 2015; 25(1):87–92. https://doi.org/10.1016/j.cub.2014.11.001 WOS:000347409900031. PMID: 25532898

61. Lackner JR, DiZio P. Vestibular, proprioceptive, and haptic contributions to spatial orientation. Annu Rev Psychol. 2005; 56:115–47. https://doi.org/10.1146/annurev.psych.55.090902.142023 PMID: 15709931

62. Loomis JM, Beall AC. Visually controlled locomotion: Its dependence on optic flow, three-dimensional space perception, and cognition. Ecol Psychol. 1998; 10(3–4):271–85. https://doi.org/10.1207/s15326969eco103&4_6 WOS:000077500700006.

63. Matthis JS, Yates JL, Hayhoe MM. Gaze and the Control of Foot Placement When Walking in Natural Terrain. Current Biology. 2018; 28(8):1224–+. https://doi.org/10.1016/j.cub.2018.03.008 WOS:000430694900043. PMID: 29657116

64. Visell Y, Giordano BL, Millet G, Cooperstock JR. Vibration Influences Haptic Perception of Surface Compliance During Walking. Plos One. 2011; 6(3). ARTN e1769710.1371/journal.pone.0017697. WOS:000288813900006.

65. Souman JL, Giordano PR, Schwaiger M, Frissen I, Thümmel T, Ulbrich H, et al. CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments. ACM Transactions on Applied Perception (TAP). 2011; 8(4):25.

66. Waller D, Loomis JM, Haun DBM. Body-based senses enhance knowledge of directions in large-scale environments. Psychon B Rev. 2004; 11(1):157–63. https://doi.org/10.3758/Bf03206476 ISI:000220674200022. PMID: 15117002

67. Warren WH. Non-Euclidean navigation. The Journal of Experimental Biology. 2019; 222(Suppl 1): jeb187971. https://doi.org/10.1242/jeb.187971 PMID: 30728233

68. Keung W, Hagen TA, Wilson RC. Regulation of evidence accumulation by pupil-linked arousal processes. Nature Human Behaviour. 2019; 3(6):636–45. https://doi.org/10.1038/s41562-019-0551-4 PMID: 31190022

69. Lau B, Glimcher PW. Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav. 2005; 84(3):555–79. https://doi.org/10.1901/jeab.2005.110-04 PMID: 16596980.

70. Schwarz G. Estimating the Dimension of a Model. Ann Statist. 1978; 6(2):461–4. https://doi.org/10.1214/aos/1176344136

71. Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974; 19(6):716–23. https://doi.org/10.1109/TAC.1974.1100705

72. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies—revisited. Neuroimage. 2014; 84:971–85. Epub 2013/09/11. https://doi.org/10.1016/j.neuroimage.2013.08.065 PMID: 24018303.

73. Rouder JN, Speckman PL, Sun DC, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon B Rev. 2009; 16(2):225–37. https://doi.org/10.3758/Pbr.16.2.225 WOS:000264915700001. PMID: 19293088

74. Yamamoto N, Philbeck JW. Intrinsic frames of reference in haptic spatial learning. Cognition. 2013; 129 (2):447–56. https://doi.org/10.1016/j.cognition.2013.08.011 WOS:000327368500023. PMID: 24007919

Virtual Supermarket Shopping Task for cognitive rehabilitation and assessment of psychiatric patients: Validation in chronic schizophrenia

Úloha Nákupu ve Virtuálním Supermarketu ke kognitivní rehabilitaci a vyšetření psychiatrických pacientů: Validace u pacientů s chronickou schizofrenií

Adéla Plechatá[1,2*], Lukáš Hejtmánek[1,2], Iveta Fajnerová[1*]

[1] National Institute of Mental Health, Klecany, Czechia

[2] Third Faculty of Medicine, Charles University, Prague, Czechia

Abstract:

Objectives: Schizophrenia has a debilitating impact on patient's cognitive functioning and everyday activities. As a part of the treatment, schizophrenia patients attend sessions of cognitive remediation to restore impaired cognitive abilities. To combine cognitive and real life training, we present a virtual task to use in cognitive rehabilitation and assessment. Virtual Supermarket Shopping Task (VSST) simulates a shopping activity, in which participants have to memorize and collect items from a virtual supermarket. The aim of this study is to establish its validity for use in clinical practice.

Sample and setting: Twenty patients suffering from chronic schizophrenia and twenty healthy controls were tested. Each participant completed the task and a battery of standard neuropsychological tests.

Statistical analyses: Groups' results were compared with Student's t-tests. Validity of VSST was examined using correlations with standard neuropsychological measures. Several VSST metrics, such as trial difficulty, distances and times, and the effect the extraneous variables have on VSST measures were investigated using analyses of variance and mixed effect models.

Results: Our analyses demonstrate that patients perform worse in VSST than healthy controls and their performance corresponds to their mnemonic abilities measured by standard neuropsychological tests. VSST performance relates to the level of executive functioning only in patients. There was no effect of gaming experience on VSST performance. While potential gender effect has to be addressed in future studies, age seems to play a role in the additional VSST measures (trial time and distance).

Study limitations: Subjects were tested only once and therefore long term benefits of using VSST in rehabilitation could not be investigated. Only schizophrenia patients were included in the sample which reduces generalizability of results to other psychiatric and neurologic conditions.

Abstrakt:

Cíle: Schizofrenie je onemocnění výrazně omezující kognitivní schopnosti člověka a jeho každodenní fungování. Pacienti se schizofrenií v rámci léčby podstupují kognitivní remediaci za účelem zlepšení svých kognitivních schopností. Virtuální úloha určená ke kognitivní rehabilitaci a vyšetření kognitivních funkcí byla vytvořena s cílem propojení kognitivního tréninku s tréninkem v reálných životních podmínkách. Úloha Nákupu ve Virtuálním Supermarketu (ÚNVS) simuluje proces nákupu, při kterém si jedinec musí zapamatovat a posbírat produkty ve virtuálním supermarketu. Cílem této studie je stanovit validitu naší úlohy pro její využití v klinické praxi.

Soubor a procedura: V rámci studie bylo otestováno dvacet pacientů trpících chronickou schizofrenií a dvacet zdravých dobrovolníků. Každý respondent absolvoval úlohu společně s baterií standardních neuropsychologických testů.

Statistická analýza: Výsledky skupin byly porovnány pomocí t.testu. Validita ÚNVS byla stanovena na základě Pearsonovy korelace se standardními neuropsychologickými testy. Pomocí lineární regrese a lineárních smíšených modelů byly podrobně prozkoumány jednotlivé proměnné, jako je obtížnost úlohy, ušlá vzdálenost či čas. Ověřen byl i vliv vnějších proměnných na výsledky v testu.

Výsledky: Analýza prokázala, že výkon SZ pacientů v ÚNVS je horší oproti výkonu zdravých dobrovolníků a tento výkon odpovídá úrovni jejich paměťových schopností zachycené

standardními neuropsychologickými testy. Pouze ve skupině SZ pacientů byl zjištěn vztah mezi výkonem v ÚNVS a exekutivními funkcemi. Analýza neodhalila žádný efekt zkušenosti s hraním počítačových her na výkon v ÚNVS. Zatímco efekt pohlaví musí být ověřen v následujících studiích, výsledky studie naznačují, že věk respondenta může mít vliv na vybrané ÚNVS proměnné (ušlá vzdálenost a čas řešení úlohy).

Omezení studie: Participanti absolvovali úlohu pouze jednou, a proto nemohl být ověřen dlouhodobý přínos ÚNVS v rehabilitaci. Výzkumný soubor byl tvořen pouze pacienty se schizofrenií, což omezuje možnost zobecnění výsledků na jiná psychiatrická a neurologická onemocnění.

Key words 3-5: virtual reality, memory deficit, schizophrenia, cognitive remediation, validity

Klíčová slova 3-5: virtuální realita, paměťový deficit, schizofrenie, kognitivní remediace, validita

INTRODUCTION

Schizophrenia (SZ) is a disabling chronic psychiatric illness which affects approximately 1 % of the world population. Besides well-known positive (e.g. hallucinations and delusions) and negative symptoms (such as social withdrawal, abulia or apathy), cognitive deficits represent an important part of schizophrenia psychopathology. The impairment is distributed across all cognitive domains, with the most profound deficit in processing speed, executive functioning and memory functions, especially in episodic memory (Fioravanti, Bianchi, and Cinti 2012; Mesholam-Gately et al. 2009; Dickinson, Ramsey, and Gold 2007). Processing speed decline can further influence memory deficit through its impact on a rehearsal loop (Brébion et al. 2000).

Impairment in neurocognition and social cognition further aggravate SZ patients functioning and quality of life (Green, Kern, and Heaton 2004; Bowie et al. 2008). The cognitive deficit in combination with negative symptomatology leads to difficulties in activities of daily living (ADL), eg. food preparation, handling of finances or shopping (Samuel, Thomas, and Jacob 2018). Pharmacotherapy can significantly reduce positive symptomatology, but has only limited influence on cognitive abilities. So far pharmacotherapy for cognitive enhancement in SZ patients resulted mostly in rather weak effects (Harvey and Bowie 2012; Sinkeviciute et al. 2018).

Currently, cognitive remediation is the method of choice for the cognitive deficit intervention. Computerized cognitive remediation in addition to paper-pencil approaches enables precise repetition of the stimuli, recording participant's performance and automatically adapting the task difficulty. Nevertheless, many computer tasks and paper-pencil methods used in cognitive assessment or rehabilitation predominantly focus on isolated cognitive domains, e.g. verbal working memory, attention shifting or inhibition in executive functions. This approach, although

essential in diagnostics, can prevent patients from transferring the learned abilities into real-life (Karbach and Verhaeghen 2014). Standard neuropsychological methods are therefore criticized for their low ecological validity (Neisser 1978; Parsons 2015) and their separation from real-life functioning.

Recently, virtual environments (VE) and virtual reality (VR) have found their place in cognitive neuroscience and psychiatry (Hejtmánek and Fajnerová 2019; Parsons 2015). VEs enable recreating complex real-life situations while preserving laboratory conditions and control over presented stimuli (Parsons 2015). Moreover, VE can be used as environment enrichment (Kempermann, Gast, and Gage 2002) which can increase cognitive enhancement effects (La Corte et al. 2019; Clemenson and Stark 2015). VEs enable patients to train everyday activities and improve their cognitive functions in a safe and controlled environment. According to previous studies, the opportunity to practice the cognitive skills learned during cognitive remediation enhances SZ patients' everyday functioning (Medalia and Saperstein 2013). VEs enable patients to practice their skills using ADL simulation in a safe environment and represent a suitable tool which can enhance the transfer of learned skills to real life (Rizzo et al. 2004).

Although immersive virtual reality (VR) presented using head-mounted displays (HMD) is currently very popular in rehabilitation, it has a few drawbacks, such as cybersickness, challenging development or a high cost. Using immersive VR during patient's acute psychotic episodes could also be problematic and patients' safety should be always taken into consideration (Valmaggia 2017). Moreover, the results from immersive VR can be influenced by sensory overload or increased cognitive load (Makransky, Terkildsen, and Mayer 2019; Frederiksen et al. 2019). In contrast, VEs presented on a monitor screen with traditional controls (keyboard, mouse, potentially joystick or gamepad) allow us to simulate complex environments

and tasks without the potential drawbacks of HMDs and seem more suitable for cognitive assessment and rehabilitation.

AIMS

The primary aim of this manuscript is to validate a rehabilitation and assessment tool focused on cognitive deficits reported in SZ patients. We developed a task in which participants are asked to remember and later collect a list of items from a virtual supermarket (Virtual Supermarket Shopping Task, VSST). The task was inspired by standard neuropsychological tests assessing declarative memory using words list (e.g. (Rey 1964), and by the concept of the ADL demonstrated to be impaired in SZ patients (Samuel, Thomas, and Jacob 2018).

Based on previous research (Plechatá et al. 2017), we propose that VSST requires a multitude of cognitive skills - declarative and working memory to remember the items, executive functions and semantic memory for item categorization, and navigational and planning skills for route planning and self orientation. However, due to this complexity, the task is not aiming to address individual cognitive processes and is not meant to replace standard diagnostic methods.

To assess VSST's construct validity using convergent and divergent validity approach (Ouellet et al. 2018; Corriveau Lecavalier et al. 2018; Parsons and Rizzo 2008), we administered it to schizophrenia patients and healthy participants along with a battery of standard neuropsychological tests.

We expect significantly lower VSST performance in SZ patients in comparison to healthy participants and the differences between the groups to be more pronounced with the task's increasing difficulty. We hypothesize that participant's VSST performance will correlate with

their memory performance in standard cognitive tests, but due to its multifacetedness we also expect to find relationships with other cognitive measures.

MATERIALS AND METHODS

Participants

We tested a total of 40 participants, 20 patients suffering from chronic schizophrenia (F20.X), and 20 healthy participants paired to the experimental group according to their age, gender and education level. One patient was excluded because of an unfinished protocol. Our final sample had 15 female participants (8 healthy, 7 patients) and 24 male participants (12 healthy, 12 patients). Given the matched pairs design, there was no age difference between the groups ($M = 34.74$ ($SD=10.23$), $t(36.97)=0.19$, $p=.851$).

Sample inclusion criteria

All patients have been diagnosed with schizophrenia according to ICD-10 standard symptom criteria of F20.X (World Health Organization 2004) and chronic schizophrenia was defined as lasting longer than 18 months (Ellison-Wright et al. 2008). We recruited patients from several institutions: NIMH Czech Republic, Psychiatric hospital Kosmonosy and Daily center for psychotic patients in Karvina.

Participants did not suffer from any other psychiatric disease, nor serious somatic or neurological disease which would prevent participation in the study, and they had no history of serious injury or head surgery. Participants had no prior knowledge of the cognitive tests used. Age range was

18-55 years. All participants signed an informed consent. The study was approved by the ethics committee of the NIMH in Klecany.

Methods

Study procedure

Prior to the experiment, we collected participants' basic demographic characteristics (age, education etc.) and inquired about their gaming experience (yes/no). All participants underwent clinical and cognitive evaluation and completed the experimental task (VSST). For more details on study procedure see Supplementary materials.

Virtual Supermarket Shopping Task (VSST)

Task description

Virtual Supermarket Shopping task (VSST) is a simulation of a shopping activity which takes place in a small scale supermarket (29 x 50 meters) (for more details see (Plechata et al. 2019). The supermarket layout is modeled so that products are placed as they would be in a real store, e.g. fruits and vegetables together, cleaning supplies together etc. (see Image 1). The task was developed using Unity3D game engine (Unity Technologies n.d.).

----------INSERT IMAGE 1 HERE----------

VSST consists of several trials of increasing difficulty, with each trial having two phases: ***acquisition phase*** and a ***recall phase***.

During the ***acquisition phase***, the participant is moved to the supermarket lobby and is asked to remember a series of items (shopping list), which is presented as a written list on the monitor. The number of items varies based on trial difficulty and the learning time limit is set to 5s per item (i.e., 15 s for three items; 25 s for five items; etc.). When the time runs out, the administrator can introduce a pause and assign a distractor activity. The idea is to simulate a real life situation of planning the shopping trip prior to the supermarket visit and allows for testing of a delayed, rather than immediate recall.

During the ***recall phase***, the participant walks around the virtual supermarket and collects items. Any items, not just those on the shopping list, can be collected and it is possible for an item to be collected multiple times. Items are visually recognizable and their names show up when directly looked at from a short distance to prevent potential confusion (e.g. cream vs mayonnaise, shampoo vs deodorant). The ***recall phase*** has no time limit and ends when the participant walks to the cashier and confirms their decision to finish.

After completing the ***recall phase***, the participant is shown his or her results (number of errors, trial time, and trial distance) and proceeds to the next trial's ***acquisition phase***. For more details on VSST task see (Plechata et al. 2019; Plechatá et al. 2017).

VSST procedure

VSST was administered on a 17" laptop. Participants controlled the task with mouse and keyboard. The movement velocity was constant. Before the test started, each participant explored the VE until they became familiar with its control system and the supermarket's spatial layout (maximum of 240s). Participants then completed 4 consecutive trials of VSST with increasing difficulty (three, five, seven, and nine items on the shopping list). Participants were instructed to try to solve the trials as fast (short trial time) and as effectively as possible (low trial distance).

Between the ***acquisition*** and the ***recall phases***, participants were administered cognitive tests and questionnaires for approximately three minutes as a distraction task (for details see Supplementary 1).

VSST Measures

VSST has multiple measures of performance:

- number of correctly collected items (or inversely missing items)
- number of *extra items* (items which participant collected but which were not on a list)
- trial time (how long did the participant take to finish the trial)
- trial distance (how long was the distance the participant walked during the trial)

As the number of items to be collected is related to the number of potential mistakes (e.g. forgetting one item out of three is arguably a worse mistake than one out of nine), we decided to evaluate trial's *item performance* as the ratio between the number of items which were correctly collected and the number of items which should have been collected (3, 5, 7 or 9). This measure was used as a primary parameter addressing recall accuracy in the task, while the other less specific variables were analysed as additional measures.

VSST task variants

To allow for repeated assessment in clinical practice, we created two shopping list variants (A and B list) for each difficulty level. Each participant completed only a single randomly assigned VSST variant (healthy subjects 12A, 8B; patients 12A, 7B). We compared the metrics of interest in both variants using two tailed t-tests and found no significant difference in *item performance* ($t(127.38)=-0.78$, $p=.440$), trial time ($t(128.11)=0.00$, $p>.999$), or trial distance ($t(148.82)=1.70$, $p=.091$). We therefore analysed both variants together.

Cognitive evaluation

All participants were evaluated with standard neuropsychological measures to assess their declarative memory, learning abilities, sustained attention, psychomotor speed and executive control.

Rey Auditory Verbal Learning Test (RAVLT) is a standard measure of episodic memory and verbal learning. The participant is asked to remember and recall a list of 15 words which is repeated five times. Delayed recall (RAVLT delayed) is performed after 20-30 minutes (Rey 1964; Preiss 1999). RAVLT is a test which has similar rationale as VSST and was chosen as the gold standard for the convergent validity of VSST.

Logical Memory I, II (LM) is a subtest of Wechsler Memory Scale III for episodic memory assessment (Wechsler 2002). The participant is asked to remember a story, recall it immediately and again after 30 minutes (LM delayed). LM test in contrast to RAVLT measures the ability to remember logically organized material that is repeated once (story A) or twice (story B).

Trail Making Test (TMT) is used to measure psychomotor speed, attention and mental flexibility (Preiss and Preiss 2006; Reitan and Wolfson 1985). The level of psychomotor speed (TMT A) can influence the performance in VSST as the higher psychomotor speed can result in more repetitions during the ***acquisition phase*** list reading. The derived difference score (the difference B-A) indicates the level of executive control (Sánchez-Cubillo et al. 2009)

PEBL Continuous performance task (PCPT) is a vigilance test from PEBL test battery (Mueller and Piper 2014). The PCPT allows us to assess participant's sustained attention. Moreover, the *detectability* measure, indicating the ability to differentiate between signal and noise, was used for divergent validity as it has no direct relationship to performance measured in VSST.

Clinical rating scales

We evaluated the patient's severity of symptoms to address the influence of their mental state on cognitive functioning. While patients with SZ can be in a stabilized state during the remission phase, their cognitive performance can be altered during the relapse episodes (Stratta and Rossi 2013; Brissos et al. 2011) therefore their symptomatology needs to be evaluated. Moreover, the level of negative symptomatology (e.g. abulia) can play a critical role in cognitive performance (Ventura et al. 2009; Bezdicek et al. 2020).

Positive and Negative Syndrome Scale (PANSS) is a standardized interview to assess positive and negative symptoms and general psychopathology in SZ patients (Kay, Fiszbein, and Opler 1987). PANSS was administered only to the SZ patients.

Beck Depression Inventory (BDI-II) is a self-report questionnaire measuring the severity of depression (Beck et al. 1961).

Beck Anxiety Inventory (BAI) is a self-report questionnaire measuring the severity of anxiety (Beck et al. 1988).

Global Assessment of Functioning (GAF) is a 100-point scale measuring illness severity of daily life functioning (Hall 1995).

RESULTS

The statistical analysis was performed using statistical software R version 3.6.0 (R Core Team 2017). We used *ggplot2* package for graphs (Wickham 2009) and *lmerTest (Kuznetsova, Brockhoff, and Christensen 2017)* package for mixed effect modeling. Pearson correlation coefficients with standard neuropsychological measures were used to assess convergent and

divergent validity (Nir-Hadad et al. 2017; Ouellet et al. 2018; Corriveau Lecavalier et al. 2018; Parsons and Rizzo 2008).

Outlier removal

Four trials (in three participants) which did not record properly and one outlier trial which took more than 10 minutes to finish (average time of trials was M=173.79 (SD=106.92) seconds) were removed, leading to 151 trials in total.

Evaluation of VSST measures and group comparisons

Our first aim was to evaluate the VSST measures and their alterations in the SZ group to select task difficulties to focus on. As we expected, the *item performance* differences between groups became more apparent with increasing difficulty. Mixed effect model with group and difficulty as fixed factors and participant as a random factor showed that the task difficulty had an impact on *item performance* (b=−0.07, 95% CI [-0.09, -0.05], t(92.1)=−7.35, p < .001), as well as did the interaction between the experimental group and the trial difficulty (steeper performance decline in patients, b=−0.04, 95% CI [-0.07, -0.02], t(92.77)=−3.21, p=0.002, see Figure 1). This suggests that what best differentiates healthy participants from patients is the rate of *item performance* decline with the increasing difficulty, rather than overall performance. We found no difference in VSST measures between patients and healthy subjects in the lowest difficulty (3 items). At this difficulty level, groups didn't differ in the number of correctly collected items (t(17)=1.46, p=.163), extra items collected (t(34.76)=−0.04, p=.970), nor trial distance (t(28.82)=−0.56, p=.578), although they differed in trial time (t(22.75)=−2.59, p=.016).

----------INSERT FIGURE 1 HERE----------

Running the same mixed model but predicting the *extra items,* we have found an increased number of *extra items* being picked up with increasing difficulty (b=0.19, 95% CI [0.09, 0.29], t(93.22)=3.69, p < .001), but no group by difficulty interaction (b=0.07, 95% CI [-0.08, 0.22], t(93.72)=0.94, p=0.35).

To study the effects of difficulty on item collection more closely, we fitted a linear regression separately for patients and controls to model the total number of collected items (correct items + *extra items*) as a function of difficulty. We observed a significant increase of number of items collected in healthy controls (b=0.55, 95% CI [0.46, 0.65], t(76)=11.70, p<.001), but not in SZ patients (b=0.12, 95% CI [−0.03, 0.27], t(71)=1.64, p=.105). This suggests that patients were collecting approximately 3 items in each trial regardless of difficulty (i.e. number of items on the list) (see Fig 2).

----------INSERT FIGURE 2 HERE----------

VSST distance and time

Using a mixed effect models we modelled the trial distance as a function of a task difficulty and group as fixed effects and participant as a random effect and found a significant effect of task difficulty (b = 17.54, 95% CI [8.76, 26.31], t(92.97) = 3.92, p < .001) and the interaction between task difficulty and group (b = -14.51, 95% CI [-27.45, -1.57], t(93.39) = -2.2, p = 0.03), suggesting increase of travelled distances with increasing number of items. But comparing trial distances between groups at each difficulty using t-tests, we found no differences except at the highest difficulty (t(29.74)=3.20, p=.003).

Modelling the distance as a function of time and group with participant as a random effect, we found a significant effect of time (b=1.06, 95% CI [0.83, 1.29], t(104.23)=9.13, p < .001) and group by time interaction (b=−0.71, 95% CI [-0.99, -0.44], t(109.75)=−5.07, p < .001), but no

group effect (b=60.7, 95% CI [-9.46, 130.86], t(69.83)=1.7, p=0.094). In other terms, for healthy subjects the distance increased more steeply as a function of time than for patients, as can be seen in Figure 3. Note that patients overall walked shorter distances and spent less time shopping, although only trial distance significantly differed from healthy controls (see Table. 1).

----------INSERT FIGURE 3 HERE----------

Neuropsychological evaluation

The results of neuropsychological tests can be found in Table 1. As expected, groups differed in all analysed cognitive measures and clinical scales.

----------INSERT TABLE 1 HERE----------

VSST convergent and divergent validity

As VSST differentiates groups better with increasing difficulty, and groups' performance often did not differ in lower difficulties, we decided to focus solely on the performance from the two most difficult trials (7 and 9 items) to reduce the possibility of false negatives. Although other metrics also demonstrate differences between the groups, we decided to further discuss and analyse the *item performance* as the best and most consistent measure of participant's performance.

We assessed the convergent and divergent validity of VSST *item performance* by obtaining Pearson correlation coefficient against neuropsychological measures. We did so for the average of the two most difficult trials (e.g. 7 and 9 items). The correlation coefficients can be found in Table 2.

----------INSERT TABLE 2 HERE----------

We can pinpoint several important findings. Firstly, the moderate correlations with both standard memory tests for both patients and healthy controls indicate that VSST performance is representative of mnemonic functions in both groups. TMT difference score, evaluating executive control and mental flexibility, correlates moderately with VSST performance only in the patient group, but unfortunately the p-value did not survive correction for multiple comparisons.

The PCPT Detectability d' evaluates the subject's ability to discriminate visual stimuli (target vs. non-target). Although we observed differences between the groups (see Table 1) there was no association with VSST performance.

We didn't observe any significant correlations of VSST performance and measures of mental state. Neither PANSS subscale scores, nor GAF correlate significantly with the *item performance*. However, correlations with PANSS negative scale and GAF score are moderate and we address them more in the Discussion.

Gender, age and gaming experience

We investigated the effect of gender and group and their interaction on average *item performance* in the two most difficult trials using ANOVA and found a significant effect of gender ($F(1,35)=6.91$, $MSE=0.02$, $p=.013$, $\eta^2G=.165$), with female participant performing better ($M = 0.57$, $SD = 0.27$) then male participants ($M = 0.44$, $SD = 0.24$). No gender by group interaction ($F(1, 35) = 0.04$, $MSE = 0.02$, $p = .834$, $\eta^2G = .001$) was identified. We address this later in the Discussion.

We then used ANOVA to compare *item performance* as a function of gaming experience and group. We didn't observe any effect of gaming experience ($F(1,35)=1.07$, $MSE=0.03$, $p=.308$,

η^2G=.030) nor group by gaming experience interaction (F(1,35)=0.27 , MSE=0.03, p=.608, η^2G=.008).

Using a linear regression to model *item performance* as function of age, we didn't find any effect (b=0.00 , 95% CI [−0.01, 0.01], t(37)=−0.39, p=.698), although we did find an effect of age on trial time (b=3.94, 95% CI [1.42, 6.47], t(37)=3.16, p=.003) and marginally on trial distance (b=−2.67, 95% CI [−5.29, −0.05], t(37)=−2.07, p=.046).

DISCUSSION

Our goal was to validate a novel method for cognitive rehabilitation and/or additional assessment of cognitive deficit in psychiatric patients. For this purpose we administered VSST and a battery of standard neuropsychological tests to a group of chronic schizophrenia (SZ) patients and healthy volunteers. We used the *item performance* - calculated as the ratio of number of correctly collected items and number of items on a list for the given trial - as the primary VSST performance measure.

The groups differed significantly in their VSST performance. As hypothesized, SZ patients performed significantly worse than the healthy controls in all, but the easiest trials. Further analyses confirmed that the deterioration in the VSST performance as a function of the task difficulty is more pronounced in SZ patients, although this effect might be solely driven by no significant difference between the groups at the easiest difficulty and a large difference at the highest difficulty.

Groups also differed in the total number of collected items. In each trial, SZ patients collected approximately three items, regardless of their correctness or number of items presented during encoding. Besides presumed inferior ability to encode the items, and therefore impaired ability to

collect the correct ones, we assume that the patients' performance could be influenced by the motivational deficits which are commonly described in SZ (Fervaha et al. 2015).

Our analyses also revealed that what differentiates patients from healthy participants is the relationship between their trial times and distances. In comparison to healthy participants, SZ patients walked shorter distances in longer times. As the movement velocity was constant, this could be the result of longer and more frequent pauses. Trial times and distances could be beneficial variables during repeated clinical examination or intervention, addressing patients' planning and navigation abilities and could offer additional information about their psychomotor speed. However, participants were neither penalized nor rewarded for slow times or optimal trajectories, therefore their performance in these parameters could be influenced by other factors, such as perseverance or impaired attention. These metrics and their association with standard cognitive and clinical measures therefore need to be studied in more detail.

VSST construct validity

Our goal was to determine the construct validity of the VSST using convergent and divergent validity approaches used previously (Nir-Hadad et al. 2017; Ouellet et al. 2018; Corriveau Lecavalier et al. 2018; Parsons and Rizzo 2008). The convergent validity estimates the relationship with a well-established neuropsychological measure of the cognitive domain targeted by the developed method. Conversely, the divergent validity investigates the correlation with a standard method measuring different concepts.

As expected, we have found correlation between VSST *item performance* and the two standard memory measures of delayed recall - Rey Auditory Verbal Learning Task and Logical Memory. In SZ patients *item performance* correlated strongly with RAVLT delayed recall ($r = 0.61$) and

LM delayed recall (r = 0.74). We also found moderate to strong correlations in healthy controls with RAVLT delayed recall (r = 0.63) and LM delayed recall (r = 0.48), although the LM did not survive correction for multiple comparisons. We believe that this result supports the construct validity of the VSST as a memory task.

During the VSST ***acquisition phase***, participants read the shopping list by themselves and the number of repeated readings might be dependent on their processing speed. Due to the processing speed deficit in SZ patients (Brébion et al. 2000), illustrated by longer TMT A time in our sample, we expected that patients might fail to read the list multiple times and their VSST performance can be affected. The missing association between VSST performance and TMT A, however, does not support this assumption.

In the patients' group we have also found a moderate correlation between the *item performance* and the TMT difference score (r = − 0.46), although it did not survive the multiple comparison correction. Given the strength of the correlation, we believe that this might be only due to a relatively small sample. TMT difference score reflects patient's level of executive control (Sánchez-Cubillo et al. 2009). The deficit in executive functioning is common in SZ patients (Green, Kern, and Heaton 2004) and it could inhibit their ability to organize encoded information and prevent them from compensating their memory impairment with a mnemonic strategy. Missing association of the TMT difference score and VSST in healthy subjects suggests that the average level of executive functioning in healthy controls might be sufficient for successful VSST completion and does not relate directly to the VSST performance. But this assumption should be investigated in future studies.

The *item performance* did not correlate with the Continuous performance task Detectability d' measure assessing sustained attention and impulsivity. This missing association of the VSST

performance with specific attentional processes (not addressed by the task) is in line with our hypothesis and supports the divergent construct validity of the task.

Mental status and cognitive performance

It was proposed that the severity of SZ positive symptomatology does not affect cognitive processing (Bezdicek et al. 2020), while some studies demonstrated relationships with negative symptoms (Ventura et al. 2009; Bezdicek et al. 2020). Although we have not found a significant association between negative SZ symptomatology and patient's performance, the observed correlation between PANSS negative scale with *item performance* (r = -0.41) does not seem circumstantial, as a weaker correlation was reported (r = -0.24) in the meta-analysis by Ventura et. al (2009). Similarly, although GAF correlated significantly with cognitive measures in previous studies (Torio et al. 2014), correlation of GAF with *item performance* in our sample (r = 0.42) was not significant. Given that the observed correlations were moderate and in the expected direction (negative for PANSS and positive for GAF), we assume that the correlation coefficients did not reach significance as a result of relatively small sample size (19 patients). These associations should therefore be investigated in future studies.

Gender, age and gaming experience

Interestingly, female participants performed overall better in VSST than male subjects, but we found no interaction between group and gender suggesting the illness affects both genders equally. Previous studies suggest a more profound memory deficit in male patients in comparison to females (Han et al. 2012; Bozikas et al. 2010) This could be however due to high variability and a small sample size and it should be addressed in future studies. Importantly, our

analyses showed no effect of gaming experience on VSST performance. Regarding age related effects, previous studies using VSST in healthy volunteers (comparing young and elderly) showed evidence of age-related performance (Plechatá et al. 2019), and although in our sample age had no significant impact on *item performance*, we found an effect of age on trial time and distance. We therefore suggest that when using these VSST metrics to track a patient's progress or performance, their age should be considered.

Implications for clinical purposes

The main potential of VSST is in its multifacetedness and resemblance to a real life activity. Our data indicate that the task might rely on more cognitive facets than just mnemonic abilities and the performance can be potentially related to executive functions, such as mental flexibility, planning and organisation. This suggests that VSST may be beneficial for multi-domain cognitive rehabilitation. Still, the main targeted domain of the VSST is memory, which was, together with processing speed and executive functions, repeatedly reported as the most impaired cognitive function in SZ (Kraguljac, Srivastava, and Lahti 2013; Fioravanti, Bianchi, and Cinti 2012).

Moreover, we suggest that simulating grocery shopping allows patients to train shopping skills to reduce anxiety or fear associated with this ADL. This might present a benefit for patients' everyday life if resulting in increased internal motivation to continue with rehabilitation sessions. In repeated assessment and combined with standard neurocognitive methods, VSST could provide additional information about patients' everyday performance and functioning.

Importantly, the possible effect of negative and general symptomatology, such as avolition, apathy or disorganized thinking, should be carefully considered in interpretation of the measured performance in neuropsychiatric patients.

Limitations

Previous study with a similar design confirmed that virtual shopping tasks predict real life performance more accurately than standard cognitive measures (Greenwood et al. 2016). At the moment, we cannot confirm the exact relationship between patients' real life functioning and VSST performance as we did not test the performance in real-life situations.

Our sample might be too small for some trending relationships to manifest themselves clearly. We also recruited only chronic schizophrenia patients. It would be interesting to investigate VSST on a larger sample, potentially covering a wider variety of psychiatric or neurological conditions - e.g. mild cognitive impairment, Alzheimer dementia, multiple sclerosis or attention deficit hyperactivity disorder. It is also necessary to confirm some of the observed correlations, particularly the missing association with negative symptomatology.

To address the rehabilitation purpose of the task, it is crucial to conduct repeated assessments and observe its long term impact on cognitive functions. In this study we provide support for our task to be a good indicator of a patient's cognitive state, but long term research is necessary to validate its impact on patients' cognitive abilities and benefits for their wellbeing.

Although the task offers many parameters to be set individually, such as acquisition time, task difficulty, delay time, number of repetitions etc., in this study we have tested only a single setting and a limited number of shopping items. Nevertheless, our results suggest that the difference between groups in their *item performance* and other parameters becomes more pronounced as the task's difficulty increases. While deliberate, slow increase of difficulty is crucial for

rehabilitation purposes, in case of cognitive assessment, it can be beneficial to focus solely on the more challenging trials and to use the lowest trial as training.

Finally, VSST does not allow us to assess isolated cognitive functions, as they can not be entirely separated. For example, the *item performance* can be related to participants spatial orientation and ability to localize the recalled items in the virtual environment. Moreover, the visual recognition of shopping items cannot be addressed and the task thus differs from free recall tasks. The possible solution would be to ask participants to verbally recall encoded items prior to entering the VE, but we believe this would disrupt the task's resemblance to ADL and real-life shopping. As the encoding and recall in VSST are predominantly visual (e.g. items visual recognition during recall), it would be beneficial to investigate the relationship between VSST performance and visual memory tests, eg. Brief Visuospatial Memory Test (Benedict et al. 1996) or written alternative of Rey Auditory Verbal Learning Test (Frydrychová et al. 2018).

CONCLUSION

We developed a method for cognitive remediation and repeated assessment in neuropsychiatric patients, which simulates a real life shopping activity. VSST offers an engaging and stimulating task in which patients can learn and practice valuable skills to use in everyday life and possibly improve their cognitive state. The task offers global evaluation of cognitive abilities through several metrics of patient’s performance. VSST performance is indicative of patients' current cognitive state in mnemonic abilities, and, to a certain degree, executive functions and their level of general functioning. We therefore believe that VSST offers a valuable and an approachable tool for psychiatric practitioners to consider.

Funding

This study was funded with financial support from the European Regional Development Fund project "PharmaBrain" no. CZ.02.1.01/0.0/0.0/16_025/0007444 and by the project Nr. LO1611 with financial support from the MEYS under the NPU I program.

Acknowledgements

We would like to thank MUDr. Martin Matějka and MSc.Martina Bednářová for their assistance in the patients recruitment from Psychiatric hospital Kosmonosy and the Daily Center for Psychotic Patients in Karviná town.

REFERENCES

Beck, A. T., N. Epstein, G. Brown, and R. A. Steer. 1988. "An Inventory for Measuring Clinical Anxiety: Psychometric Properties." *Journal of Consulting and Clinical Psychology* 56 (6): 893–97.

Beck, A. T., C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. 1961. "An Inventory for Measuring Depression." *Archives of General Psychiatry* 4 (June): 561–71.

Benedict, Ralph H. B., David Schretlen, Lowell Groninger, Melissa Dobraski, and Barnett Shpritz. 1996. "Revision of the Brief Visuospatial Memory Test: Studies of Normal Performance, Reliability, and Validity." *Psychological Assessment* 8 (2): 145–53.

Bezdicek, Ondrej, Jiří Michalec, Lucie Kališová, Tomáš Kufa, Filip Děchtěrenko, Miriama Chlebovcová, Filip Havlík, Michael F. Green, and Keith H. Nuechterlein. 2020. "Profile of Cognitive Deficits in Schizophrenia and Factor Structure of the Czech MATRICS Consensus Cognitive Battery." *Schizophrenia Research*, February. https://doi.org/10.1016/j.schres.2020.02.004.

Bowie, Christopher R., Winnie W. Leung, Abraham Reichenberg, Margaret M. McClure, Thomas L. Patterson, Robert K. Heaton, and Philip D. Harvey. 2008. "Predicting Schizophrenia Patients' Real-World Behavior with Specific Neuropsychological and Functional Capacity Measures." *Biological Psychiatry* 63 (5): 505–11.

Bozikas, Vasilis P., Mary H. Kosmidis, Apostolos Peltekis, Maria Giannakou, Ioannis Nimatoudis, Athanasios Karavatos, Kostas Fokas, and George Garyfallos. 2010. "Sex Differences in Neuropsychological Functioning among Schizophrenia Patients." *The Australian and New Zealand Journal of Psychiatry* 44 (4): 333–41.

Brébion, G., M. J. Smith, J. M. Gorman, D. Malaspina, Z. Sharif, and X. Amador. 2000. "Memory and Schizophrenia: Differential Link of Processing Speed and Selective Attention with Two Levels of Encoding." *Journal of Psychiatric Research* 34 (2): 121–27.

Brissos, Sofia, Vasco Videira Dias, Vicent Balanzá-Martinez, Ana Isabel Carita, and Maria Luísa Figueira. 2011. "Symptomatic Remission in Schizophrenia Patients: Relationship with Social

Functioning, Quality of Life, and Neurocognitive Performance." *Schizophrenia Research* 129 (2-3): 133–36.
Clemenson, Gregory D., and Craig E. L. Stark. 2015. "Virtual Environmental Enrichment through Video Games Improves Hippocampal-Associated Memory." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35 (49): 16116–25.
Corriveau Lecavalier, Nick, Émilie Ouellet, Benjamin Boller, and Sylvie Belleville. 2018. "Use of Immersive Virtual Reality to Assess Episodic Memory: A Validation Study in Older Adults." *Neuropsychological Rehabilitation*, May, 1–19.
Dickinson, Dwight, Mary E. Ramsey, and James M. Gold. 2007. "Overlooking the Obvious: A Meta-Analytic Comparison of Digit Symbol Coding Tasks and Other Cognitive Measures in Schizophrenia." *Archives of General Psychiatry* 64 (5): 532–42.
Ellison-Wright, Ian, David C. Glahn, Angela R. Laird, Sarah M. Thelen, and Ed Bullmore. 2008. "The Anatomy of First-Episode and Chronic Schizophrenia: An Anatomical Likelihood Estimation Meta-Analysis." *The American Journal of Psychiatry* 165 (8): 1015–23.
Fervaha, Gagan, Hiroyoshi Takeuchi, Jimmy Lee, George Foussias, Paul J. Fletcher, Ofer Agid, and Gary Remington. 2015. "Antipsychotics and Amotivation." *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology* 40 (6): 1539–48.
Fioravanti, Mario, Valentina Bianchi, and Maria Elena Cinti. 2012. "Cognitive Deficits in Schizophrenia: An Updated Metanalysis of the Scientific Evidence." *BMC Psychiatry* 12 (June): 64.
Frederiksen, Joakim Grant, Stine Maya Dreier Sørensen, Lars Konge, Morten Bo Søndergaard Svendsen, and Steven Arild Wuyts Andersen. 2019. "Cognitive Load and Performance in Immersive Virtual Reality versus Conventional Virtual Reality Simulation Training of Laparoscopic Surgery: A Randomized Trial," June. https://doi.org/10.1007/s00464-019-06887-8.
Frydrychová, Zuzana, Miloslav Kopecek, Ondrej Bezdicek, and Hana Stepankova Georgi. 2018. "Czech Normative Study of the Revised Rey Auditory Verbal Learning Test (RAVLT) in Older Adults (České Normy pro Revidovaný Reyův Auditorně-Verbální Test Učení (RAVLT) pro Populaci Starších Osob)" 62 (4)): 330–49.
Green, Michael F., Robert S. Kern, and Robert K. Heaton. 2004. "Longitudinal Studies of Cognition and Functional Outcome in Schizophrenia: Implications for MATRICS." *Schizophrenia Research* 72 (1): 41–51.
Greenwood, Kathryn E., Robin Morris, Vanessa Smith, Anna-Marie Jones, Douglas Pearman, and Til Wykes. 2016. "Virtual Shopping: A Viable Alternative to Direct Assessment of Real Life Function?" *Schizophrenia Research* 172 (1-3): 206–10.
Hall, R. C. 1995. "Global Assessment of Functioning. A Modified Scale." *Psychosomatics* 36 (3): 267–75.
Han, Mei, Xu-Feng Huang, Da Chun Chen, Mei Hong Xiu, Li Hui, Haibo Liu, Thomas R. Kosten, and Xiang Yang Zhang. 2012. "Gender Differences in Cognitive Function of Patients with Chronic Schizophrenia." *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 39 (2): 358–63.
Harvey, Philip D., and Christopher R. Bowie. 2012. "Cognitive Enhancement in Schizophrenia: Pharmacological and Cognitive Remediation Approaches." *The Psychiatric Clinics of North America* 35 (3): 683–98.
Hejtmánek, Lukáš, and Iveta Fajnerová. 2019. "Využití Virtuální Reality v Psychiatrii." *Psychiatrie* 4. http://www.tigis.cz/images/stories/psychiatrie/2019/Psychiatrie_4_2019/Psychiatrie_4_2019_vzdelavani.pdf.
Karbach, Julia, and Paul Verhaeghen. 2014. "Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults." *Psychological Science* 25 (11): 2027–37.
Kay, S. R., A. Fiszbein, and L. A. Opler. 1987. "The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia." *Schizophrenia Bulletin* 13 (2): 261–76.
Kempermann, Gerd, Daniela Gast, and Fred H. Gage. 2002. "Neuroplasticity in Old Age: Sustained Fivefold Induction of Hippocampal Neurogenesis by Long-Term Environmental Enrichment."

Annals of Neurology 52 (2): 135–43.
Kraguljac, Nina V., Annusha Srivastava, and Adrienne C. Lahti. 2013. "Memory Deficits in Schizophrenia: A Selective Review of Functional Magnetic Resonance Imaging (FMRI) Studies." *Behavioral Sciences* 3 (3): 330–47.
Kuznetsova, Alexandra, Per Brockhoff, and Rune Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software, Articles* 82 (13): 1–26.
La Corte, Valentina, Marco Sperduti, Kouloud Abichou, and Pascale Piolino. 2019. "Episodic Memory Assessment and Remediation in Normal and Pathological Aging Using Virtual Reality: A Mini Review." *Frontiers in Psychology* 10 (February): 173.
Makransky, Guido, Thomas S. Terkildsen, and Richard E. Mayer. 2019. "Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence but Less Learning." *Learning and Instruction* 60 (April): 225–36.
Medalia, Alice, and Alice M. Saperstein. 2013. "Does Cognitive Remediation for Schizophrenia Improve Functional Outcomes?" *Current Opinion in Psychiatry* 26 (2): 151–57.
Mesholam-Gately, Raquelle I., Anthony J. Giuliano, Kirsten P. Goff, Stephen V. Faraone, and Larry J. Seidman. 2009. "Neurocognition in First-Episode Schizophrenia: A Meta-Analytic Review." *Neuropsychology* 23 (3): 315–36.
Mueller, Shane T., and Brian J. Piper. 2014. "The Psychology Experiment Building Language (PEBL) and PEBL Test Battery." *Journal of Neuroscience Methods* 222 (January): 250–59.
Neisser, U. 1978. "Memory: What Are the Important Questions?" In *Practical Aspects of Memory*, edited by M. M. Gruneberg, P. E. Morris, and R. N. Sykes, 3–24. London: Academic Press.
Nir-Hadad, Shira Yama, Patrice L. Weiss, Anna Waizman, Natalia Schwartz, and Rachel Kizony. 2017. "A Virtual Shopping Task for the Assessment of Executive Functions: Validity for People with Stroke." *Neuropsychological Rehabilitation* 27 (5): 808–33.
Ouellet, Émilie, Benjamin Boller, Nick Corriveau-Lecavalier, Simon Cloutier, and Sylvie Belleville. 2018. "The Virtual Shop: A New Immersive Virtual Reality Environment and Scenario for the Assessment of Everyday Memory." *Journal of Neuroscience Methods* 303 (June): 126–35.
Parsons, Thomas D. 2015. "Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences." *Frontiers in Human Neuroscience* 9: 660.
Parsons, Thomas D., and Albert A. Rizzo. 2008. "Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test." *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 11 (1): 17–25.
Plechatá, Adéla, Iveta Fajnerová, Lukáš Hejtmánek, and Václav Sahula. 2017. "Development of a Virtual Supermarket Shopping Task for Cognitive Remediation of Memory and Executive Functions in Schizophrenia." In *2017 International Conference on Virtual Rehabilitation (ICVR)*, 1–2.
Plechata, Adela, Václav Sahula, Dan Fayette, and Iveta Fajnerova. 2019. "Age-Related Differences with Immersive and Non-Immersive Virtual Reality in Memory Assessment." *Frontiers in Psychology* 10: 1330.
Plechatá, Adéla, Václav Sahula, Dan Fayette, and Iveta Fajnerová. 2019. "Age-Related Differences With Immersive and Non-Immersive Virtual Reality in Memory Assessment." *Frontiers in Psychology* 10 (June): 1330.
Preiss, Marek. 1999. "Paměťový Test Učení. [Auditory Verbal Learning Test. Manual.]." *Psychodiagnostika*.
Preiss, Marek, and J. Preiss. 2006. "Test Cesty [Trail Making Test]." In *Psychodiagnostika*. Bratislava: MD.
R Core Team. 2017. *R: A Language and Environment for Statistical Computing* (version 3.1.3). Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Reitan, Ralph M., and Deborah Wolfson. 1985. *The Halstead-Reitan Neuropsychological Test Battery : Theory and Clinical Interpretation*. Tucson Ariz.: Neuropsychology Press.
Rey, André. 1964. *L'examen Clinique En Psychologie*. 2e éd. Paris: Presses universitaires de France.

639 Samuel, Reema, Elizabeth Thomas, and K. S. Jacob. 2018. “Instrumental Activities of Daily Living
640 Dysfunction among People with Schizophrenia.” *Indian Journal of Psychological Medicine* 40 (2):
641 134–38.
642 Sánchez-Cubillo, I., J. A. Periáñez, D. Adrover-Roig, J. M. Rodríguez-Sánchez, M. Ríos-Lago, J. Tirapu,
643 and F. Barceló. 2009. “Construct Validity of the Trail Making Test: Role of Task-Switching,
644 Working Memory, Inhibition/interference Control, and Visuomotor Abilities.” *Journal of the*
645 *International Neuropsychological Society: JINS* 15 (3): 438–50.
646 Sinkeviciute, Igne, Marieke Begemann, Merel Prikken, Bob Oranje, Erik Johnsen, Wan U. Lei, Kenneth
647 Hugdahl, et al. 2018. “Efficacy of Different Types of Cognitive Enhancers for Patients with
648 Schizophrenia: A Meta-Analysis.” *NPJ Schizophrenia* 4 (1): 22.
649 Stratta, Paolo, and Alessandro Rossi. 2013. “Short-Term Remission in Schizophrenia as a Combination of
650 Several Outcome Measures.” *Psychiatry Research* 209 (3): 401–5.
651 Torio, Iosune, Alexandra Bagney, Mónica Dompablo, María José Campillo, Lorena García-Fernández,
652 Javier Rodríguez-Torresano, Miguel Ángel Jiménez-Arriero, Tomas Palomo, and Roberto
653 Rodríguez-Jiménez. 2014. “Neurocognition, Social Cognition and Functional Outcome in
654 Schizophrenia.” *The European Journal of Psychiatry* 28 (4): 201–11.
655 Unity Technologies. n.d. “Unity - Unity.” Unity. Accessed February 17, 2020.
656 https://unity.com/frontpage.
657 Valmaggia, Lucia. 2017. “The Use of Virtual Reality in Psychosis Research and Treatment.” *World*
658 *Psychiatry: Official Journal of the World Psychiatric Association* 16 (3): 246–47.
659 Ventura, Joseph, Gerhard S. Hellemann, April D. Thames, Vanessa Koellner, and Keith H. Nuechterlein.
660 2009. “Symptoms as Mediators of the Relationship between Neurocognition and Functional
661 Outcome in Schizophrenia: A Meta-Analysis.” *Schizophrenia Research* 113 (2-3): 189–99.
662 Wechsler, David. 2002. *Wechsler Memory Scale - Third Edition Abbreviated Manual*. Vol. 3rd
663 Abbreiated edition. The Psychological Corporation.
664 Wickham, Hadley. 2009. “ggplot2: Elegant Graphics for Data Analysis.” Springer-Verlag New York.
665 http://ggplot2.org.
666 World Health Organization. 2004. *International Statistical Classification of Diseases and Related Health*
667 *Problems*. World Health Organization.

668

669

Image 1. Part A (top). Overview of VSST environment layout. Part B (bottom). Participant`s first person view, while collecting an item. Please note that the name of the product is visible after pointing at it. The already collected (shopped) items are visible in the shopping bag in the right bottom corner. The images are in black and white due to the journal requirements.

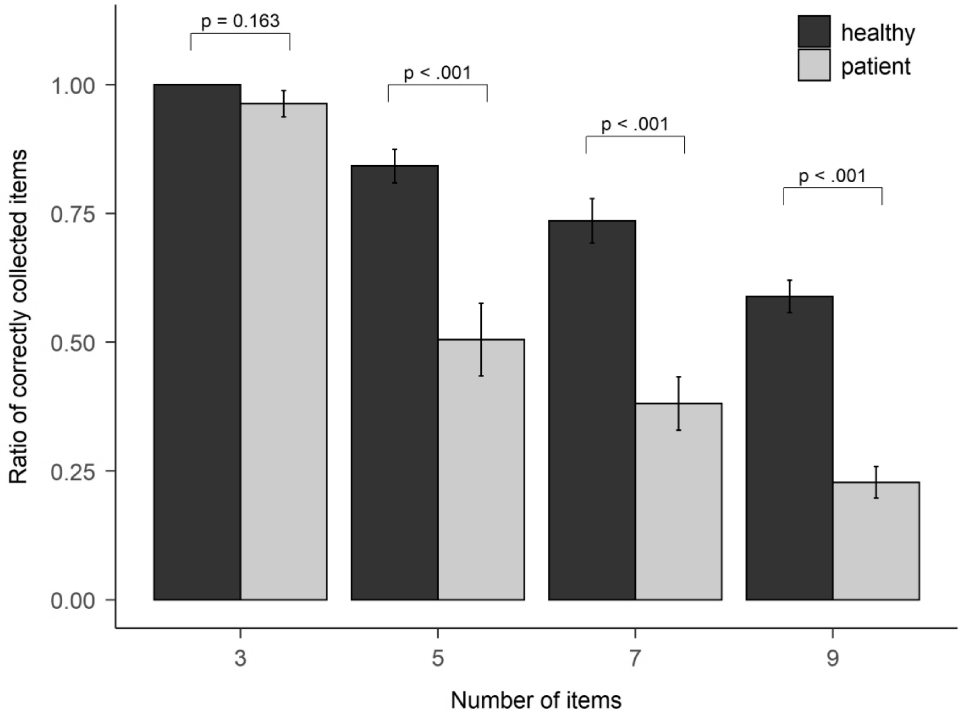


Figure 1. Average item performance (ratio of correctly collected items) with SEM error bars for all VSST trials at increasing difficulties (number of items) split by group. Groups were compared at each difficulty using two-sample t-tests.

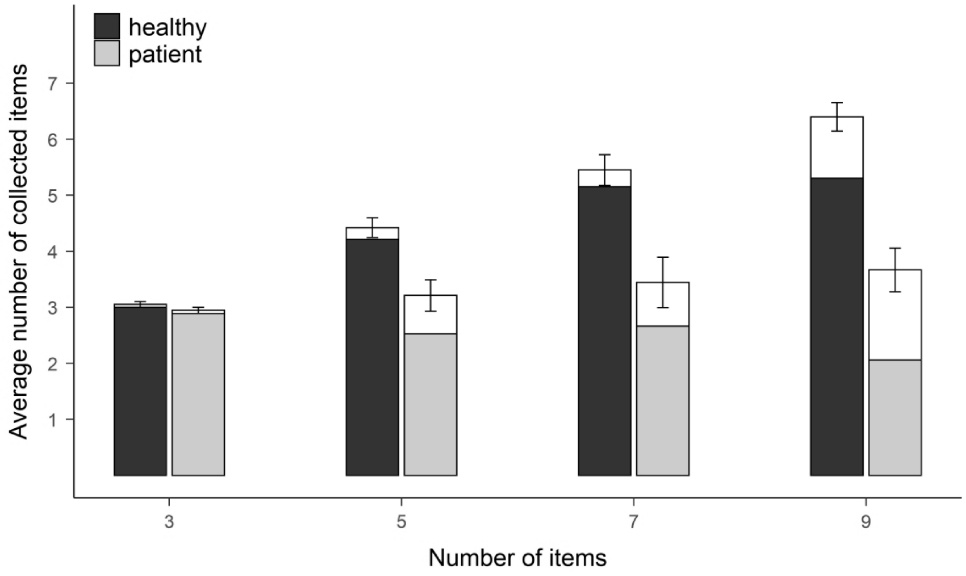


Figure 2. Average number of all collected items (correct + extra) separated by groups and shaded by the item type. The darker section on the bottom represents correct items and the lighter section on top represents extra items. Error bars represent SEM of the total

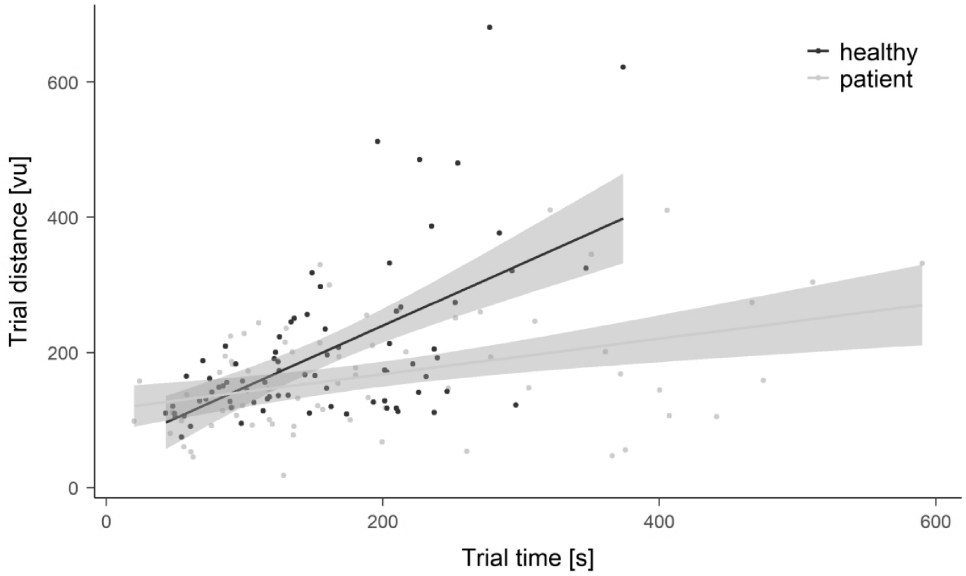


Figure 3. Scatter plot of trial distance (in virtual units) and trial time (in seconds) split by group. Grey areas represent 95% confidence interval of fitted linear regression slope.

Table 1. Descriptive data of neuropsychological assessment, psychiatric scales and VSST assessment (reported for the two most difficult trials) and group comparisons.

	Measure	Healthy group mean (SD)	Patient group mean (SD)	t.value	p.value	cohen.d
VSST measures	Item performance	0.66(0.18)	0.31(0.19)	8.274	< .001	1.907
	Trial time	189.47(73.90)	171.84(118.14)	0.769	0.445	0.181
	Trial distance	224.68(116.76)	160.57(77.02)	2.851	0.006	0.641
Neuropsychological measures	RAVLT delayed	9.95(2.63)	5.42(3.01)	5.001	< .001	1.608
	BAI	5.60(5.27)	12.21(8.30)	-2.949	0.006	-0.955
	BDI	3.00(2.56)	9.58(6.54)	-4.101	< .001	-1.339
	GAF	100.00(0)	66.84(17.97)	8.044	< .001	2.646
	PANSS positive scale	-	11.78(3.52)	-	-	-
	PANSS negative scale	-	16.06(6.67)	-	-	-
	PANSS general psychopathology	-	29.33(6.15)	-	-	-
	LM delayed	27.75(7.21)	15.16(8.18)	5.089	< .001	1.636
	PCPT Detectability d'	2.35(0.72)	0.65(2.34)	2.868	0.01	0.984
	TMT A time	24.20(7.31)	52.26(26.01)	-4.535	< .001	-1.486
	TMT difference score	48.60(28.04)	100.39(64.32)	-3.229	0.004	-1.053

The reported VSST measures were calculated as an average performance in the two most difficult trials (7 and 9 items to remember).

Legend: RAVLT (Rey Auditory Verbal Learning Test) delayed recall, LM (Logical Memory) delayed recall, PCPT - PEBL Continuous Performance Task, TMT (Trail Making Test) difference - TMT B and TMT A time difference, BDI - Beck Depression Inventory, BAI - Beck Anxiety Inventory, GAF - Global Assessment of Functioning.

Table 2. Correlation of VSST *item performance* (average for 7 and 9 items) and neuropsychological measures split by group.

test/correlation	Healthy	patient
Convergent validity		
RAVLT delayed	r = 0.63, p = 0.003*	r = 0.61, p = 0.005*
LM delayed	r = 0.48, p = 0.031	r = 0.74, p < .001*
TMT A time	r = − 0.12, p = 0.617	r = − 0.11, p = 0.667
TMT difference	r = − 0.16, p = 0.489	r = − 0.46, p = 0.046
Divergent validity		
PCPT Detectability	r = 0.08, p = 0.768	r = 0.11, p = 0.18
Mental status		
PANSS negative	-	r = − 0.41, p = 0.088
PANSS positive	-	r = − 0.09, p = 0.732
PANSS general	-	r = - 0.23, p = 0.367
GAF	-	r = 0.42, p = 0.077

Legend: RAVLT (Rey Auditory Verbal Learning Test) delayed recall, LM (Logical Memory) delayed recall, PEBL CPT - PEBL Continuous Performance Task, TMT (Trail Making Test) difference B-A - TMT B and TMT A time difference, PANSS - Positive and Negative Syndrome Scale. GAF - Global Assessment of Functioning. After applying Bonferroni correction on multiple comparisons in each section we set the alpha to 0.0125. Asterisk symbol marks statistically significant correlations at this level.

Supplementary - Experiment protocol

Prior to the experiment, all the participants signed informed consent. The testing using the experimental task of the VSST was preceded by examination by cognitive tests, in the following order:

1. Rey Auditory Verbal Learning Test (RAVLT)
2. Logical Memory I (LP)
3. Continuous performance test (CPT)
4. Rey Auditory Verbal Learning Test (RAVLT del)
5. Logical Memory II (LP-del)

TMT A/B tests were administered during VSST pauses between encoding and shopping list recall, similar to BDI and BAI questionnaires, in both the experimental and comparison groups.

Specifics of the VSST:

The initial exploratory phase lasted until the participant became familiar with the control system and the supermarket spatial layout (maximum of 240s). The task contained a total of four consecutive levels of difficulty (for 3,5,7 and 9 items), word lists included only 50% of food, the remaining 50% consisted of other items (toys, drugstores, office supplies, electronics, etc.). The first trial following exploration with three items to remember served as initial practice. Each trial started with the encoding phase with the presentation of the list. Then there was an approximately 3-minute or slightly longer break (depending on their ability to fulfill the distraction tasks) when the shopping list was not displayed on the screen and the respondents were prevented from repeating the list - they performed the tasks of TMT A / B and filled in the BDI and BAI questionnaires. After the pause, the recall phase followed, followed by the results presentation.

Instructions for the participants prior to the beginning of the experiment:

"Now you will enter a virtual supermarket environment. In the beginning, you will have 4 minutes to see the supermarket and get used to the environment and keyboard and mouse controls (the participant is clearly explained how to interact with the virtual environment)*. After the four-minute limit expires, you will be presented with a shopping list on the screen monitor and your task will be to try to remember it. After a certain amount of time, the list will disappear and you will perform another task with me, so you need to remember the list well. After a 3-minute break, your task will be to find the memorized items in the supermarket and collect them with the mouse click. This part is not limited in time, but you should try to be efficient and walk the shortest possible distance and spend as little time as possible in the store. Failure to collect the correct item and also collecting an item that was not on the list is considered an error."*

The encoding, 3-minutes break and recall phases were repeated 4 times in total (for each difficulty level) as follows:

1.trial - three items on the list encoding
3-minutes break - TMT A
Recall in VSST
2.trial - five items on the list
3 minutes break - TMT B
Recall in VSST
3.trial - seven items on the list
3 minutes break - BAI
Recall in VSST
4.trial - nine items on the list
3 minutes break - BDI
Recall in VSST

After completing the testing using the VSST, the participants were asked about the strategies used to memorize the items. In the case of the experimental group of SZ patients, a structured interview performed by a trained clinician followed, in order to evaluate the patient's condition using a psychiatric PANS scale. Data were obtained from the interview and observation of the patient to obtain a GAF score.

Encoding VSST material

Trial and difficulty	*Item in czech (english translation)*	
	Variant A	*Variant B*
1. Trial - 3 items	Pokladnička (Money-box)	Řepa (Beetroot)
	Třešně (Cherries)	Opice (Monkey)
	Mléko (Milk)	Máslo (Butter)
2. Trial - 5 items	Kečup (Ketchup)	Protlak (Puree)
	Deodorant (Deodorant)	Šampón (Shampoo)
	Smetana (Cream)	Majonéza (Mayonnaise)
	Tužka (Pencil)	Vodovky (Watercolors)
	Jablko (Apple)	Hruška (Pear)
3. Trial - 7 items	Sušenky (Biscuits)	Čokoláda (Chocolate)
	Letadlo (Plane)	Medvídek (Teddybear)
	Těstoviny (Pasta)	Makaróny (Macaroni)
	Talíř (Plate)	Hrnec (Pot)
	Lilek (Eggplant)	Okurka (Cucumber)
	Kniha (Book)	Polštář (Pillow)
	Guma (Rubber)	Štětce (Brushes)
4. Trial - 9 items	Citrón (Lemon)	Švestka (Plum)
	Konvice (Kettle)	Míč (Ball)

	Vajíčka (Eggs)	Losos (Salmon)
	Kartáček (Toothbrush)	Ručník (Towel)
	Šlehačka (Whipped cream)	Mouka (Flour)
	Pánev (Pan)	Mísa (Bowl)
	Tričko (T-shirt)	Kalhoty (Pants)
	Žampion (Champignon)	Kukuřice (Corn)
	Váza (Vase)	Svíčka (Candle)