Big data in spatial navigation – towards personalised cognitive diagnostics of 'at-genetic-risk' Alzheimer's disease

Coughlan Gillian.[a], Coutrot Antoine [b], Khondoker Mizanur.[a], Minihane Anne-Marie.[a], Spiers Hugo [c], Hornberger Michael.[a,d]

[a] *Norwich Medical School, University of East Anglia, Norwich, UK*

[b] *CNRS - Laboratoire des Sciences du Numérique de Nantes, University of Nantes, France*

[c] *Institute of Behavioural Neuroscience, Department of Experimental Psychology, University College London, London, UK*

[d] *Dementia and Complexity in Later Life, NHS Norfolk and Suffolk Foundation Trust, UK*

Correspondence: Prof. Michael Hornberger, Norwich Medical School, Bob Champion Research and Education Building, University of East Anglia, James Watson Road, Norwich, NR4 7TJ, UK m.hornberger@uea.ac.uk

Tel: +441603597139, Fax: +441603593752

Running Title: Cognitive diagnostics of 'at-risk' AD

Abstract
Spatial navigation is emerging as a critical factor in identifying preclinical Alzheimer's disease (AD). However, the impact of inter-individual navigation ability and demographic risk factors (eg APOE, age, sex) on spatial navigation make it difficult to identify 'at-high-risk' of AD people in the preclinical stages. In the current study we use spatial navigation Big Data (n=27,108) from the Sea Hero Quest (SHQ) game to overcome these challenges by investigating whether Big Data can be used to benchmark a highly phenotyped healthy ageing lab cohort into high vs. low risk people based on their genetic (APOE) and demographic (sex, age, educational attainment) risk factors. Our results replicate previous findings in APOE ε4 carriers, indicative of grid-cell coding errors in the entorhinal cortex, the initial brain region affected by AD pathophysiology. We also show that although baseline navigation ability differs between men and women, sex does not interact with the APOE genotype to influence the manifestation of AD related spatial disturbance. Most importantly, we demonstrate that such high-risk preclinical cases can be reliably distinguished from low-risk participants using Big Data spatial navigation benchmarks. By contrast, participants were undistinguishable on neuropsychological episodic memory tests. Taken together, we present the first evidence to suggest that in the future, SHQ normative benchmark data can be used to more accurately classify spatial impairments in 'at-high-risk' of AD healthy participants at a more individual level, therefore providing the stepping stone for individualised diagnostics and outcome measures of cognitive symptoms in preclinical AD.

Significance Statement
We report that assessment of navigational behaviour using the Sea Hero Quest App provides a means of discriminating healthy ageing from genetically at-risk individuals of Alzheimer's disease. It further highlights that the global Sea Hero Quest database can be employed as a normative benchmark data set to efficiently determine the significance of spatial abnormality suspected to be indicative of incipient AD on an individual level.

\body

INTRODUCTION

Spatial navigation is a promising cognitive fingerprint for underlying Alzheimer's disease pathophysiology (1–8) and has been adopted by many high profile clinical trials (such as the the European Prevention of Alzheimer's Dementia Consortium) to improve the sensitivity of neurocognitive testing and assess the efficacy of potentially disease-modifying treatments. In fact, brain areas affected by AD pathophysiology in the preclinical stage (including the entorhinal cortex, posterior cingulate cortex, precuneus) form the key nodes in the spatial navigation network (6, 9–13). Recent evidence suggests that abnormal spatial navigation patterns may be present before episodic memory deficits, which are the current gold standard for AD diagnosis (6, 14, 15).

A major challenge at this stage, however, is to understand how inter-individual and demographic factors that affect spatial navigation in order to identify earliest pathological spatial navigation changes in AD (16–19). Understanding diversifying factors that influence variability in spatial ability in the *healthy* population and individuals at risk to develop AD will advance the diagnostic power of the spatial tests and support more personalised diagnostics and treatment approaches (17, 20–23). Among factors underlying navigation, age is a well-documented predictor of declining spatial abilities, as older adults show a strong bias toward egocentric rather than allocentric strategies (24, 25) leading to suboptimal navigation performance (26). Age-related decline in allocentric process are due to changes in coding patterns of place, grid, border and head direction cells that underpin our ability to form cognitive maps of the environment and integrating environmental and self-motion cues to optimise navigational performance (27–29). However, decline in other cognitive domains such as general planning and cognitive control abilities(30) also contribute to spatial deficits in old age, suggesting that like most diagnostic tests, age-range normative cut-off scores are required (30, 31). Similarly, sex differences in navigation behaviour and underlying neuroanatomy have generated arguments for sex-specific clinicopathological AD phenotypes (17, 21, 32–35). Rodent models of the Morris Water Maze have shown that male rats consistently outperform females (36) and human studies display similar sex differences favouring males (37–40) across 57 countries in both map-dependent allocentric and map-independent egocentric navigational strategies (27). Therefore, although spatial navigation tools must retain sensitivity and specificity to preclinical AD pathophysiology, it will be critical to develop diagnostic tools that can adjust for underlying sex differences.

Finally, one of the biggest challenges in preclinical AD studies is to identify those who are at-high-risk to develop symptomatic AD in the future. Genetic variation in the apolipoprotein E 4 allele carriers is currently the strongest known genetic risk factor for sporadic AD (7, 41–43). Compared to the ε3/ε3 carriers, those with the ε3ε4 show a three-four fold increased risk for AD (43, 44). Phenotypic characteristics of apoE e4 allele show that the cognitive profile of e4 carriers changes over the lifespan, with some cognitive advantage seen in young adulthood (39) and cognitive disturbances in mnemonic and spatial process in mid adulthood (45–47). Recent findings also show that temporal grid-cell like representation in the entorhinal cortex of apoE4 carriers are functionally unstable leading to a boundary-driven error correction during wayfinding (48).

Taken together, there is increasing evidence that spatial deficits, in particular related to wayfinding, are present in preclinical AD long before episodic memory symptom emerge. However, at this stage it is very difficult to employ such knowledge on a clinical level, due to unknown inter-individual variability in navigation behaviour across people, which is vital for sensitive and specific diagnostics on an individual level. In the current study we address this issue by using Big Data (n=27,308) for navigation behaviour from the Sea Hero Quest App (49) to: i) determine whether we can replicate previous wayfinding affects in APOE ε3ε4 carriers compared to the Big Data; ii) to further disentangle inter-individual the effects of genetic risk for AD from the effects of sex, age and baseline cognition on spatial discrepancies; and iii) to explore whether AD specific spatial navigation changes can be detected on an individual level, when using the Big Data as benchmark comparison. We predicted that i) we would replicate previous APOE spatial navigation findings (7); ii) sex differences would make a significant impact on navigation behaviour; and iii) AD specific navigation changes

can be detected in an individual level when using the normative benchmark Big Data of Sea Hero Quest.

RESULTS

Background Characteristics and Neuropsychology

In the lab-based cohort, the ε3ε3 and ε3ε4 groups did not differ in terms of their demographic characteristics (see supplementary Table 1) or their neuropsychological examination (Table 2). We examined the relationship between the three SHQ outcome variables (Fig. 1): Wayfinding distance travelled and wayfinding duration correlate (Pearson $r = 0.61$, $p < 0.001$); duration and flare accuracy correlate ($r = -0.309$, $p < 0.001$); but wayfinding distance travelled and flare accuracy are not correlated $r = 0.04$, $p = .795$); suggesting dissociable neural correlates that underlie performance, corroborating current notions that wayfinding distance relies more on grid-cell based navigational processes (51), and flare accuracy relies more on retrosplenial mediated processes (15). We consider wayfinding distance as the primary outcome measure (and the other outcomes are secondary) as early AD is characterised by abnormal changes in the grid cell code of the entorhinal cortex.

Genotype effects on wayfinding

There was a main effect of genotype (b=0.22; p=0.004; Fig. 2 A) on wayfinding distance, with ε3ε3 carriers (M=3.79, SD=0.63) travelling a shorter distance during wayfinding relative to ε3ε4 carriers (M=4.45, SD=0.94) after controlling for age and sex. The mixed model for wayfinding duration (i.e. time taken to complete wayfinding levels) showed no main effect of genotype between ε3ε3 (M=4.66, SD=2.65) and ε3ε4 carriers (M=4.97, SD=1.36; Fig. 2 B). See Table 2 for group mean values and Table 3 for the effects of genotype on wayfinding distance and duration. Please refer to SI Appendix for results including a small high-risk ε4/ε4 carrier group, which showed an even larger effect for distance travelled (SI Appendix, Fig. S1).

To further examine the different routes taken by the two genetic groups, we plotted the exact trajectory of each participant on wayfinding level 6, 8 and 11 using (x,y) coordinates generated during gameplay and found that ε3/ε4 carriers show a lower average distance to border than their ε3/ε3 counterparts (Fig. 2 D-F). On level 6 and 8, *ε3/ε4* carriers deviate from the shortest distance between the checkpoints and travel toward the border of the environment compared to the *ε3ε3* carriers, who tend to navigate along the centre of the virtual environment. To check if the increase in wayfinding distance in ε3/ε4 carriers compared to the ε3/ε3 group was driven by any specific level, fixed effects linear models were fitted for level 6, 8 and 11 to test if the properties in one specific level captured this effect, or if this effect was an accumulative error over the three wayfinding levels. Using the same explanatory variables as in the final base model, the e4 allele was found to increase wayfinding distance on level 6 (F_{60}=5.48, p=0.023) and level 8 (F_{60}=4.08, p=0.04) but not on level 11 (see SI Appendix, Fig. S2; also see Fig. S5 for diagnostic plots underlying key assumptions of the linear mixed models).

Genotype and sex effect on wayfinding

No effects of sex were found on wayfinding distance as men (M=4.06, SD=0.87) and women (M=4.22, SD=0.91; b=0.02, p=0.12) took similarly efficient paths, but sex did affect duration taken to complete wayfinding levels, with men (M=4.33, SD=1.09) requiring less time to complete levels than women (M=5.26, SD=2.17; b=0.39, p=0.02; SI Appendix, Fig. S3(A)). Importantly, no interactive effects of genotype and sex on wayfinding distance or wayfinding duration were uncovered.

Genotype and sex effects on path integration

We then tested the effects of genotype and sex levels on flare accuracy, a measure of path integration. No main effect of genotype (b=0.04, p=0.14; Fig. 2 C) and no genotype*sex interactions were found. However, sex had a significant main effect on flare accuracy, with men (M=5.11, SD=1.3) scoring higher than women (M=4.31, SD=1.4; b=-0.36, p=0.04; SI Appendix, Fig. 3(B)).

Memory and spatial navigation as predictors of APOE genotype

The sensitivity and specificity of a traditional memory task to predict APOE genotype compared to spatial navigation on SHQ was done using logistic regression and ROC curves. This was motivated by the prediction that memory deficits would not be detectable on current gold standard episodic memory tasks. Covarying for gender, non-verbal episodic memory (three-minute total recall score for the ROCF) and wayfinding distance in SHQ were used separate predictors in two logistic regression analyses. The regression model for wayfinding distance x^2 (2) = 9.1, p=0.03, was statistically significant and correctly classified 71.3% of the APOE genotyped cohort (75%:*ε3ε3* 63.3%:*ε3ε4*). As predicted, the model for ROCF delayed recall was not significant x^2 (2) = 9.1, p=0.393. An ROC curve was then computed showing both navigation and delayed recall as predictors of APOE genotype (Fig. 3). Consistent with the above, area under the curve values indicated that wayfinding distance (AUC .714, SE .068, 95% CI .555 - .822; pink curve), but not delayed recall (AUC .541, SE .074, 95% CI .286 - .578; gold curve) has a significant level of diagnostic accuracy.

Benchmark data validates an effect of APOE4 on wayfinding. Having determined the diagnostic utility of SHQ for APOE genotype compared to standard memory test, we wanted to examine the utility of the population-level benchmark dataset as a normative control sample which could be used by clinicians in diagnostic settings. We took advantage of the fact that the benchmark SHQ dataset-as a representative of the population-predominantly includes ε3/ε3 carriers (75%) and performed a ROC curve with the *ε3ε4* and the benchmark data as a representative of non-risk controls. Area under the curve values indicated a very similar significant level of diagnostic accuracy as was demonstrated with the lab only cohort (AUC .701 SE .031 95% CI .639 - .759; see Fig. 3 [dark pink curve]). Finally, to further representation the diagnostic utility of the benchmark population, we plotted each *ε3ε4* carrier's score over their age sex, education matched sub-population from the normal distribution of the UK population (see Fig. 4).

DISCUSSION

Our results show that i) we can replicate previous wayfinding changes in APOE gene carriers; ii) sex differences significantly impact on wayfinding behaviour but the effect of sex is negligible compared to APOE genetic risk; iii) healthy 'at-genetic-risk' of AD with no memory deficits can be distinguished on wayfinding measures on an individual level.

In more detail, using navigation benchmark Big Data and smaller APOE genotyped cohorts, we show that adults 'at-genetic-risk' of AD with no clinically detectable cognitive deficits, not only navigate further during wayfinding, but show a bias in navigating towards the border of the virtual SHQ environment in large open areas. This supports the hypothesis that suboptimal navigation performance is present in preclinical AD and that this is detectable on levels of the SHQ game, even when a closely matched demographic sample is provided by the global SHQ data set. We also show that while sex accounts for variation in navigation performance, sex does not reduce the sensitivity of SHQ to discriminate healthy ageing from genetically at-risk individuals of Alzheimer's disease.

Although adults at-genetic risk of AD deviate from the shortest route (often the Euclidean between the checkpoints) towards the environmental border of the SHQ environment, they successfully completed the wayfinding levels albeit sub-optimally. Thus, we hypothesis that the navigational deficits detected here reflect an error corrective strategy (48) for which environmental boundaries hold valuable navigational cues that aid the navigators' ability to self-localise and find their way through the environment when navigational uncertainty ensues. The neural substrates that give rise to the navigational uncertainty in the genetically at-risk group is most likely induced by errors in the grid cell system within the entorhinal cortex (see SI for further discussion). The entorhinal cortex is not only one of the first cites of AD pathology in the brain (13) but is also crucial for facilitating shortcut wayfinding behaviours and optimal navigation behaviour (56). Given that grid cells compute large-scale information (30, 31) and encode representations of self-location by measuring distance travelled by the navigator (32, 33), it is not surprising grid cell dysfunction results in navigational discrepancies in at-risk individuals of AD.

Given that phenotypic heterogeneity currently reduces the diagnostic and prognostic power of neurocognitive evaluations for early AD, we also sought to investigate if demographic and neuropsychology diversity impact navigation. The effect of the genotype that was most prominent when the environmental space was large and open (level 6 and 8). In terms of sex, we did find strong evidence of better performance in males on baseline navigation ability but no evidence to suggest that males at-genetic-risk were less vulnerable (in the preclinical stage at least) to the effect of the APOE e4 genotype than women at-genetic-risk. In our opinion, this is a critical finding as it suggests that sex difference may not act on the phenotypic presentation of navigation deficits in the early asymptomatic stage of the disease. A recent meta-analysis (54) reports that women are particularly vulnerable to early underlying pathology between the ages of 55 and 70. Thus whether sex and genotype interact to predict navigational ability on SHQ in later preclinical or prodromal stages of AD remains to be investigated. In the interest of diagnostic sensitivity, the time at which an increased female susceptibility to underlying pathology manifests behaviourally is a high priority. Although we found a sex-independent navigational deficit in adults at genetic risk of AD, evidence for strong spatial disparities on navigation performance across the sexes globally (55) suggest that it is indeed appropriate to consider the need to stratify risk assessment by sex. For example, when genotype status is unknown, considering sex difference may hold prognostic value as many high profile previous studies already suggest (17, 21, 56).

Based on data presented here on a population level and elsewhere, we now know that demographic diversity based on age, sex and nationality act on navigation proficiency, and men perform better at digital and real-life spatial navigation tasks (57). This finding, coupled with a plethora of pre-existing evidence for natural age-related decline in spatial navigation (26), means that we must establish personalised normative measures to accurately assess spatial disturbances that have not been well-established as a underlying feature in preclinical AD pathology. From a clinical standpoint, clinicians and researchers should be advised to consider not only age but also the sex of their putative patient before inferring pathological related spatial impairment. From a research perspective, researchers should work towards providing demographically corrected benchmarked scores for standardised neuropsychological test. To date, obtaining normative data of this nature has been challenged by heterogeneity in methodological approaches used to measure spatial navigation and uncertainty about population level differences in cognitive performance. Consistency across our non-risk control group and the benchmark scores is compelling evidence that SHQ may provide unique benchmarking data, on a global scale, by controlling for the demographical factors such as sex, advanced age and cultural background; factors which will alter how individuals perform on SHQ. Although level of education was included to refine the population data, education did not have a compelling effect on navigation performance in the global SHQ database. Further research is required to determine what demographic factors beyond age, sex and nationality will increase the sensitivity and specificity of navigation test for underlying preclinical AD.

Despite illustrating for the first time the clinical utility of new epidemiological data gathered on a global scale using the SHQ game, our study has several limitations. Firstly, we focus on preclinical rather than symptomatic Alzheimer's disease, seeking to evaluate the prognostic value of SHQ rather than validate SHQ data as a potential diagnostic tool. However, given that many excellent cognitive diagnostics measures exist for symptomatic AD, we question whether navigation measures have true utility in this aspect. Instead, identification of subtle cognitive preclinical changes will be of greater future importance to complement other biomarkers as diagnostic and treatment outcome measures. Secondly, only 47% of all ε3/ε4 carriers develop symptomatic AD. This is consistent with about 50% of the ε3/ε4 individuals in this study being impaired relative to the demographically corrected benchmark L longitudinal studies are needed to truly determine how predictive spatial navigation combined with genotypic information is in the preclinical stages of the disease however. Further replication of our findings with preclinical cohorts defined by multiple cognitive, genetic and neurological markers is desirable, although it is promising that we replicate previous boundary findings (Kunz 2015). Moreover, although education was considered in the individualised approach to diagnosis of 'at-risk' AD, approx. 40% of the genotyped cohort has 15 years+ of education and 50%

of the cohort working in "professional" fields vs. skilled or low-skilled / manual, potentially leading to an over-representation at the educated individuals in this genotyped sample. Lastly, although best efforts were made to control for gaming proficiency, we cannot completely rule out a potential influence of previous gaming experience contributing to the observed male advantage in the data. Still, considering that we are investigating a 50-75 year old cohort, gaming proficiency should not play such a large role. More importantly, the difference of male and females in the SHQ data across ages does not change, suggesting that gaming proficiency plays only overall a minor role in assessing spatial navigation via an online App.

In conclusion, our work supports the hypothesis that navigational discrepancies are present in preclinical AD and can be captured by Sea Hero Quest available on iOS and Android platforms. We show for the first time promising evidence that normative data generated from the 3.7 million people who played SHQ worldwide, may in the future help us to create a prognostic test based on navigational proficiency – to help us to understand how the very earliest symptoms of AD is isolation of potentially confounding demographic factors such as sex, advancing age, educational attainment or cultural background. This should reduce the problematic nature of phenotype variation obscuring the assessment of spatial disorientation as a first symptom of AD and offer the promise of individually tailored solutions in healthcare settings. Thus, spatial navigation emerges as a promising cognitive fingerprint, which can complement existing biomarker for future AD diagnostics and disease intervention outcome measures.

MATERIALS AND METHODS

Participants

APOE genotyped cohort.

Between Feb, 2017 and June, 2017, 150 people between 50 to 75 years of age were recruited to participate in a research study at the University of East Anglia. All 150 participants were pre-screened for a history of psychiatric or neurological disease, history of substance dependence disorder or any significant relevant comorbidity. All participants had normal or corrected-to-normal vision. Family history of AD and history of antidepressant treatment with serotonin reuptake inhibitor (SSRI) drugs was retrospectivity obtained. Saliva samples were collected from those who passed this screening and apoE genotype status was determined.

In total, 69 participants underwent cognitive testing. As just 23% of the population carry APOE ε3/ε4, all participants in our sample who tested positive for the ε3/ε4 allele completed cognitive testing. We selected a subset of the ε3/ε3 carriers that form the majority of the population (75%) to match the ε3/ε4 risk group for age and sex (see SI Appendix, Table S1 for group background characteristics). We did not include a third genetic subgroup of homozygous APOE-ε4carriers from the tested cohort, because they were too rare (n=5) although their scores are reported in the SI Appendix. E2 carriers were also excluded.

During testing, three participants showed signs of distress and their data was excluded from subsequent analyses. One participant scored lower than 86 on the Addenbrooke's Cognitive Examination and was classified as mildly cognitively impaired and excluded from the study. The final group sizes (post-exclusion) were: apoE ε3/ε3, n=29 and apoE ε3/ε4, n=31). Written consent was obtained from all participants and ethical approval was obtained from Faculty of Medicine and Health Sciences Ethics Committee at the University of East Anglia Reference FMH/2016/2017–11.

The Benchmark Population.

A unique population level benchmark dataset was generated by extracting a subset of the global Sea Hero Database (50) that matched the demographic profile of our lab-based genotype cohort, namely players from the UK aged 50-75 years old. Following extraction, 14,470 British men and 12,710 British women (N = 27,108) remained as a representative normative sample of heathy navigation performance on the basis that epidemiological studies have shown that the majority of the general population (~75%) are non-apoE4 carriers (36). Participants from the benchmark sample were given the option to opt in or opt out of the data collection when they played the game on their personal

mobile phone, iPad or tablet. If a participants' response was to opt in, their SHQ data was anonymised and stored securely by the T-Systems' datacentre under the regulation of German data security law. Ethical approval was previously granted by Ethics Research Committee CPB/2013/015. *For more information on the global SHQ database see www.seaheroquest.com

Outcome Measure

Sea Hero Quest (SHQ)

The SHQ app was developed in 2015 by our team and funded by Deutsche Telekom and Alzheimer's Research UK. The app was created to be a reliable and valid measure of spatial navigation performance both in monitored research settings and unmonitored at-home settings (49). It was made available for free on the App Store and Play Store from May 2016 and since then over 4 million people have downloaded the App worldwide. The game performance is divided into two main domains: goal-oriented wayfinding and path integration.

Goal-orientated wayfinding. In wayfinding levels, players initially see a map featuring a start location and several checkpoints to find in a set order, as illustrated in **Fig 1**. Checkpoints are buoys with flags marking the checkpoint number. Participants study a map of the level for a recorded number of seconds. When participants exit the map view, they are asked to immediately find the checkpoints (or goals) in the order indicated on the map under timed conditions. As participants navigate the boat through the level, they must keep track of their location using self-motion and environmental landscape cues such as water-land separation. The initiation time is zero as the boat accelerates immediately after the map disappears. If the participant takes more than a set time, an arrow appears pointing in the direction along the Euclidean line to the goal to aid navigation. To familiarize themselves with the virtual environment and game controls, participants started with two easy learning levels 1 and 2. Wayfinding levels generate two measures of interest:

- *Wayfinding distance* travelled to visit all required checkpoints is defined as the Wayfinding distance between all points recorded and is a proxy for navigation efficiency. To navigate efficiently, individuals need to form and retain a cognitive map of the environment (after viewing the map at the start of the level) and then consistently update self-location in that cognitive map based on the visual cues from the SHQ game.
- *Wayfinding duration*, is defined as the time in seconds to complete a wayfinding level. While inefficient navigation also results in longer time to visit all checkpoints, increased duration is primarily due to the amount of acceleration that the player used. By "swiping up", one can increase the speed of the boat temporarily, therefore reducing travel time but not changing the distance travelled at all. Since speeding up requires confidence in one's sense of direction, the resulting wayfinding duration score we take duration as less representative of participants' ability to navigate along the shortest path and more representative of non-navigational factors such as confidence or the tendency to sample more cues before speeding up.

Flare Accuracy. In path integration levels (in the game this is measured by flare accuracy on levels 9 and 14), participants are not provided with an allocentric map. Instead, they immediately navigated along a river to find a flare gun. Once they find the flare gun at the end of the river, the boat rotates by 180°, and participants are asked to choose one of three possible directions (right, front, left) that they believe points to the starting point. This level requires participants to a) form an accurate representation of the starting point relative to their position and b) integrate this representation with a representation of the direction they are facing after the rotation. (see Tu and colleagues for a similar path integration based experimental design(15)). In this case, gaming proficiency was not advantageous because participants simply view navigate a single passage and are then required to choose A,B,C direction as a single response. Depending on their accuracy, players receive either one, two or three stars.

Procedure

Data collection

Spatial navigation data was collected for both the APOE genotyped cohort and benchmark datasets using Sea Hero Quest, a digital game that we pre-designed to measure human navigation ability. Decisions on level selection was made by considering which levels had the most normative data and level type/difficulty (wayfinding or path integration). Level 1 and 2 were included for learning and practice navigating the boat, as well as normalising the data for App interaction with player proficiency. Level 3-5 were excluded as they did not challenge participants' navigation skills and were intended to ease the players into the game. Further, starting with level 14, the sample size of the benchmark population drops substantially. This then left us with three wayfinding levels (6,8,11) and two path integration levels (9 and 14). Participants in the lab-based APOE cohorts provided their demographic information during a screening call and were then invited to the UEA to play SHQ. Participants from the benchmark population provided information regarding their sex, age, location and educational attainment (high-school, college, university) demographics in-app before playing SHQ

APOE Genotyping

DNA was collected using a Darcon tip buccal swab (Fisher Scientific, Leicestershire, United Kingdom, LE11 5RG). Buccal swabs were refrigerated at 2-4°C until DNA was extracted using the QAIGEN QIAamp DNA Mini Kit (QAIGEN, Manchester, United Kingdom, M15 6SH). DNA was quantified by analysing 2 µL aliquots of each extraction on a QUBIT 3.0 Fluorometer (Fisher Scientific, Leicestershire, United Kingdom, LE11 5RG). Successful DNA extractions were confirmed by the presence of a DNA concentration of 1.5µg or higher per 100µg AE buffer as indicated on the QUBIT reading. PCR amplification and plate read analysis was performed using Applied Biosystems 7500 Fast Real-Time PCR System (Thermo Fisher Scientific, Ashford, United Kingdom, TN23 4FD). TaqMan Genotyping Master Mix was mixed with two single nucleotide polymorphisms of APOE (rs429358 at codon 112 and rs7412 at codon 158). These two single nucleotide polymorphisms determine the genotype of APOE2, E3, and E4 (Applied Biosystems, 2007).

Statistical Analysis

The data was analysed using SPSS (Version 23), RStudio (Version 1.0.153) and MATLAB (R2017a). Chi square and simple two tailed t-tests were used to test the significance of any demographic or neuropsychological differences between the genetic groups in our lab cohort. When quantifying the group differences, Cohen's d was used as a measure of effect size. To control for the influence of player proficiency on digital devices, the SHQ data was pre-processed in MATLAB and participant performance on each level within the game was divided by the sum of the two practice levels:

$$level\ N\ normalised = \ln\left(\frac{level\ N}{(level\ 1 + level\ 2)}\right)$$

To assess the fixed effects of genotype and sex, we first compared competing statistical models with the inclusion and exclusion of different demographic factors using the nlme package in R (https://cran.r-project.org/web/packages/nlme/index.html) that allows fitting fixed and random effects to evaluate the most appropriate model for data. In each model, subject-level random effects were included to vary the intercept for each subject and importantly to account for interdependence between repeated measures from playing multiple levels of the game. Three sets of linear models were fitted that included the following outcome variables: a) wayfinding distance and b) wayfinding duration, using scores from SHQ levels 6, 8 and 11 completed by each subject and c) flare accuracy on each of the two path integration levels (9 and 14). Model selection was based on relative goodness of fit and model simplicity (determined using gold standard Akaike and Bayesian information criterion, AIC and BIC, respectively).

Age, sex and genotype, were retained as explanatory variables for the final model for each of the outcome variables. ACE defined by total score on the Addenbrooke's Cognitive Examination-III screening tool (52), education, occupation, time spent on viewing the wayfinding maps (see Fig. 1 for maps) and non-verbal episodic memory (defined by 3minute delayed recall on Rey–Osterrieth

Complex Fig. Test; ROCF (53)), were tested in the final model but did not exhibit a significant main effect and were excluded to retain the maximum degrees of freedom (the overall F statistics for explanatory variables in additional models are shown in supplementary Table 2). Once the best fit model was identified, standardised residuals were extracted and plotted against fitted values to examine underlining assumption of normal distribution and heteroscedasticity. We also tested for an interaction between genotype and sex. All statistical tests are two-tailed, $p<0.05$.

To ensure that the benchmark population reflected the demographic profile of our lab-based cohort, we could only use a sub-population of our global SHQ database. We developed a data extraction method using MATLAB (code, data, associated protocols, and materials available from authors on request) that allowed us to generate the population level database. This data was then pre-processed using the same normalisation procedure as detailed above. Linear mixed models examined the effects of sex and age on a population level benchmark. Finally, logistic regression was used to quantify how well SHQ variables such as distance travelled could classify APOE risk status using both the lab-based sample and the benchmark population. ROC curves were used as measures of sensitivity and specificity of SHQ as opposed to standard memory tasks such as the ROCF test to detect preclinical AD.

References

1. Tu S, Spiers HJ, Hodges JR, Piguet O, Hornberger M (2017) Egocentric versus Allocentric Spatial Memory in Behavioral Variant Frontotemporal Dementia and Alzheimer's Disease. *J Alzheimer's Dis* 59(3):883–892.
2. Lithfous S, Dufour A, Després O (2013) Spatial navigation in normal aging and the prodromal stage of Alzheimer's disease: Insights from imaging and behavioral studies. *Ageing Res Rev* 12(1):201–213.
3. Serino S, Morganti F, Di Stefano F, Riva G (2015) Detecting early egocentric and allocentric impairments deficits in Alzheimer's disease: An experimental study with virtual reality. *Front Aging Neurosci* 7(MAY):1–10.
4. Serino S, Riva G (2013) Getting lost in Alzheimer's disease: A break in the mental frame syncing. *Med Hypotheses* 80(4):416–421.
5. DiBattista, Amanda M., Nicolette M. Heinsinger and GWR (2016) Alzheimer's Disease Genetic Risk Factor APOE-ε4 Also Affects Normal Brain Function. *Curr Alzheimer Res* 13(11):1200–1207.
6. Coughlan. G, Laczó J, Hort J, Minihane AM, Hornberger M (2018) Spatial navigation deficits – the overlooked cognitive fingerprint for incipient Alzheimer pathophysiology? *Nat Rev Neurol*. 14(8) 496-506
7. Kunz L, et al. (2015) Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease. *Science (80-)* 350(6259):430–433.
8. Allison SL, Fagan AM, Morris JC, Head D (2016) Spatial Navigation in Preclinical Alzheimer's Disease. *J Alzheimers Dis* 52:77–90.
9. Vlček, K., & Laczó J (2014) Neural correlates of spatial navigation changes in mild cognitive impairment and Alzheimer's disease. *Front Behav Neurosci* 8. doi:10.3389/fnbeh.2014.00089.
10. DeIpolyi AR, Rankin KP, Mucke L, Miller BL, Gorno-Tempini ML (2007) Spatial cognition and the human navigation network in AD and MCI. *Neurology* 69(10):986–997.
11. Mokrisova I, et al. (2016) Real-space path integration is impaired in Alzheimer's disease and mild cognitive impairment. *Behav Brain Res* 307:150–158.
12. Weston PSJ, et al. (2016) Presymptomatic cortical thinning in familial Alzheimer disease: A longitudinal MRI study. *Neurology* 87(19):2050–2057.
13. Braak H, Del Tredici K (2015) The preclinical phase of the pathological process underlying sporadic Alzheimer's disease. *Brain* 138(10):2814–2833.

14. Dubois B, et al. (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* 13(6):614–629.
15. Tu S, et al. (2015) Lost in spatial translation - A novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia. *Cortex* 67:83–94.
16. Husain M (2017) Alzheimer's disease: Time to focus on the brain, not just molecules. *Brain* 140(2):251–253.
17. Ferretti M, et al. (2018) Sex differences in Alzheimer disease — the gateway to precision medicine. *Nat Rev Neurol*. doi:10.1038/s41582-018-0032-9.
18. Doody RS, et al. (2014) Phase 3 Trials of Solanezumab for Mild-to-Moderate Alzheimer's Disease. *N Engl J Med* 370(4):311–321.
19. Sevigny J, et al. (2016) The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* 537(7618):50–56.
20. Nelson PT, et al. (2012) Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol* 71(5):362–381.
21. Snyder HM, et al. (2016) Sex biology contributions to vulnerability to Alzheimer's disease: A think tank convened by the Women's Alzheimer's Research Initiative. *Alzheimer's Dement* 12(11):1186–1196.
22. Pettigrew C, et al. (2013) Relationship of cognitive reserve and APOE status to the emergence of clinical symptoms in preclinical Alzheimer's disease. *Cogn Neurosci* 4(3–4):136–142.
23. Chan D, et al. (2018) Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation and late-life activities. *bioRxiv* 70:267831.
24. O'Keefe, John & Nadel L (1978) *The Hippocampus as a Cognitive Map* doi:10.1017/CBO9781107415324.004.
25. Hartley T, Lever C, Burgess N, O'Keefe J (2013) Space in the brain: how the hippocampal formation supports spatial cognition. *Philos Trans R Soc B Biol Sci* 369(1635):20120510–20120510.
26. Lester AW, Moffat SD, Wiener JM, Barnes CA, Wolbers T (2017) The Aging Navigational System. *Neuron* 95(5):1019–1035.
27. Byrne P, Becker S, Burgess N (2007) Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychol Rev* 114(2):340–375.
28. Epstein RA, Patai EZ, Julian JB, Spiers HJ (2017) The cognitive map in humans: Spatial navigation and beyond. *Nat Neurosci* 20(11):1504–1513.
29. Spiers HJ, Barry C (2015) Neural systems supporting navigation. *Curr Opin Behav Sci* 1:47–55.
30. Hartshorne JK, Germine LT (2015) When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychol Sci* 26(4):433–443.
31. Malek-Ahmadi M, et al. (2015) Age- and education-adjusted normative data for the Montreal Cognitive Assessment (MoCA) in older adults age 70–99. *Aging, Neuropsychol Cogn* 22(6):755–761.
32. Kong XZ, Huang Y, Hao X, Hu S, Liu J (2017) Sex-linked association between cortical scene selectivity and navigational ability. *Neuroimage* 158(February 2016):397–405.
33. Driscoll I, Hamilton DA, Yeo RA, Brooks WM, Sutherland RJ (2005) Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Horm Behav* 47(3):326–335.
34. Mosconi L, et al. (2017) Sex differences in Alzheimer risk: Brain imaging of endocrine vs chronologic aging. *Neurology* 89(13):1382–1390.
35. Press D (2014) Clinical epidemiology of Alzheimer ' s disease : assessing sex and gender differences. 37–48.
36. Perrot T, Perrot-sinal TS, Kostenuik MA, Ossenkopp K, Kavaliers M (1997) Sex differences in performance in the Morris water maze and the effects of initial nonstationary hidden platform training Sex Differences in Performance in the Morris Water Maze and the Effects of Initial Nonstationary Hidden Platform Training. ll(January):1309–1320.
37. Berteau-Pavy F, Park B, Raber J (2007) Effects of sex and APOE epsilon 4 on object recognition and spatial navigation in the elderly. *Neuroscience* 147(1):6–17.
38. Yasen AL, Raber J, Miller JK, Piper BJ (2015) Sex, but not Apolipoprotein E Polymorphism,

Differences in Spatial Performance in Young Adults. *Arch Sex Behav* 44(8):2219–2226.
39. Acevedo SF, Piper BJ, Craytor MJ, Benice TS, Raber J (2010) Apolipoprotein E4 and sex affect neurobehavioral performance in primary school children. *Pediatr Res*. doi:10.1203/PDR.0b013e3181cb8e68.
40. Astur RS, Ortiz ML, Sutherland RJ (1998) A characterization of performance by men and women in a virtual Morris water task: *Behav Brain Res* 93(1–2):185–190.
41. Laczó J, Andel R, Vyhnalek M, Vlcek K (2014) APOE and spatial navigation in amnestic MCI: Results from a computer-based test. *Neuropsychology* 28(5):676–684.
42. Reiman EM, et al. (1996) Preclinical Evidence of Alzheimer's Disease in Persons Homozygous for the ε4 Allele for Apolipoprotein E. *N Engl J Med* 334(12):752–758.
43. Corder E, et al. (1993) E Type 4 Allele Gene Dose of Apolipoprotein and the Risk of Alzheimer ' s Disease in Late Onset Families. *Science (80-)* 261(5123):921–923.
44. Farrer LA (1997) Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease. *Jama* 278(16):1349.
45. Greenwood PM, Sunderland T, Friz JL, Parasuraman R (2000) Genetics and visual attention: Selective deficits in healthy adult carriers of the varepsilon 4 allele of the apolipoprotein E gene. *Proc Natl Acad Sci* 97(21):11661–11666.
46. Luck T, et al. (2015) Association of the apolipoprotein E genotype with memory performance and executive functioning in cognitively intact elderly. *Neuropsychology* 29(3):382–387.
47. Rajah MN, et al. (2017) Family history and APOE4 risk for Alzheimer's disease impact the neural correlates of episodic memory by early midlife. *NeuroImage Clin* 14:760–774.
48. Hardcastle K, Ganguli S, Giocomo LM (2015) Environmental Boundaries as an Error Correction Mechanism for Grid Cells. *Neuron*. doi:10.1016/j.neuron.2015.03.039.
49. Morgan J (2016) Gaming for dementia research: a quest to save the brain. *Lancet Neurol* 15(13):1313.
50. Coutrot A, et al. (2018) Global Determinants of Navigation Ability. *Curr Biol* 28(17):2861–2866.e4.
51. Banino A, et al. (2018) Vector-based navigation using grid-like representations in artificial agents. *Nature* 557(7705):429–433.
52. Matias-Guiu JA, et al. (2017) Addenbrooke's cognitive examination III: Diagnostic utility for mild cognitive impairment and dementia and correlation with standardized neuropsychological tests. *Int Psychogeriatrics* 29(1):105–113.
53. Shin M-S, Park S-Y, Park S-R, Seol S-H, Kwon JS (2006) Clinical and empirical applications of the Rey–Osterrieth Complex Fig. Test. *Nat Protoc* 1(2):892–899.
54. Neu SC, et al. (2017) Apolipoprotein E genotype and sex risk factors for Alzheimer disease: A meta-analysis. *JAMA Neurol* 74(10):1178–1189.
55. Cothi D (2017) Global determinants of navigation ability. *Curr Biol*:0–7.
56. Mielke M, Vemuri P, Rocca W (2014) Clinical epidemiology of Alzheimer ' s disease : assessing sex and gender differences. *Clin Epidemiol* (6):37–48.
57. Coutrot A, et al. (2018) Virtual navigation tested on a mobile app (Sea Hero Quest) is predictive of real-world navigation performance: preliminary data. *bioRxiv*:1–10.

Figure Legends

Fig. 1. SHQ Goal-orientated Wayfinding levels (A) 6, (B) 8 and (C) 11. Players initially see a map featuring a start location and several checkpoints (in red) to find in a set order. Checkpoints are buoys with flags marking the checkpoint number. Participants study a map of the level for a recorded number of seconds. When participants exit the map view, they are asked to immediately find the checkpoints (or goals) in the order indicated on the map under timed conditions. As participants navigate the boat through the level, they must keep track of their location using selfmotion and environmental landscape cues such as water-land separation. The initiation time is zero as the boat accelerates immediately after the map disappears. If the participant takes more than a set time, an arrow appears pointing in the direction along the Euclidean line to the goal to aid

navigation. (D) In flare accuracy levels (here level 9 and 14), participants are not provided with an allocentric map. Instead, they immediately navigated along a river to find a flare gun. Once they find the flare gun at the end of the river, the boat rotates by 180°, and participants are asked to choose one of three possible directions (right, front, left) that they believe points to the starting point. This level requires participants to i) form an accurate representation of the starting point relative to their position and ii) integrate this representation with a representation of the direction they are facing after the rotation. Depending on their accuracy, players receive either one, two or three stars.

Fig. 2. Mixed effects models, with subject level random effects, adjusted for age, sex and baseline cognitive ability show **A** Main effect of genotype (b=0.22; *p*=0.004) on Wayfinding distance; e3e4 carriers participants deviate from the more Euclidean trajectory leading to an overall greater distance travelled to complete the wayfinding levels relative to the e3e3 carriers**. B** No main effect of genotype on wayfinding duration (i.e. time taken to complete wayfinding levels); both groups used the same boat acceleration during wayfinding. **C** No main effect of genotype on flare accuracy which required participants to integrate newly acquired allocentric information with egocentric-viewpoint based cues presented at the end of the level. The spatial trajectory of each participant (colours red and green was used to differentiate the trajectories by the genetic groups) on wayfinding level 6 **D** level 8 **E** and level 11 **F** using x and y coordinates generated during gameplay. The maps generated illustrated a drift like navigation tendency in the e3e4 group that can be characterised as navigational preference to deviate from the most Euclidean path and travel toward the border of the environment compared to the e3e3 who demonstrated a preference to navigate more along the direct path to the checkpoint goal. A by level analysis on Wayfinding distance in the three levels showed that the e4 allele increased Wayfinding distance on level 6 (F=5.48, p=0.023) and level 8 (F=4.08, p=0.04).

Fig. 3. ROC curves for SHQ distance (pink line [lab-cohort]; dark pink line [lab – benchmark combined]) and non-verbal episodic memory (gold line [lab-cohort]) predicting APOE genotype. SHQ (lab-cohort) AUC .714, SE .068, 95% CI .555 - .822 | SHQ distance (lab – benchmark combined) AUC .701 SE .031 95% CI .639 - .759 | Non-verbal episodic memory (lab-cohort): AUC .541, SE .074, 95% CI .286 - .578.

Fig. 4 Each e3e4 carrier score (red line) on SHQ distance plotted against the normal distribution of scores from an age-sex-education matched sub-population of the benchmark dataset (green histogram). Wayfinding distance scores are on the x axis and frequency of the benchmark population on the y axis. Sex is represented by M = male, F = female sex. Age is illustrated under each distribution right of sex.