

Global determinants of navigation ability

Antoine Coutrot^{1,2}, Ricardo Silva³, Ed Manley⁴, Will de Cothi¹, Saber Sami⁵, Véronique D. Bohbot⁶, Jan M. Wiener⁷, Christoph Hölscher⁸, Ruth C. Dalton⁹, Michael Hornberger^{5,*}, Hugo J. Spiers^{1,*}

1 Institute of Behavioural Neuroscience, Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, United Kingdom.

2 Laboratoire des Sciences du Numérique de Nantes, CNRS, Université de Nantes, France.

3 Department of Statistical Science and CSML, University College London, London, United Kingdom.

4 Centre for Advanced Spatial Analysis, University College London, London, United Kingdom.

5 Norwich Medical School, University of East Anglia, Norwich, United Kingdom.

6 Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada.

7 Department of Psychology, Ageing and Dementia Research Centre, Bournemouth University, Poole, United Kingdom.

8 ETH Zürich, Swiss Federal Institute of Technology, Zürich, Switzerland.

9 Department of Architecture and Built Environment, Northumbria University, Newcastle upon Tyne, United Kingdom.

* These authors contributed equally to this work.

Lead contact: Hugo J. Spiers, h.spiers@ucl.ac.uk.

Summary

Human spatial ability is modulated by a number of factors including age [1–3] and gender [4, 5]. While a few studies showed that culture influences cognitive strategies [6–13], the interaction between these factors has never been globally assessed as this requires testing millions of people of all ages across many different countries in the world. Since countries vary in their geographical and cultural properties, we predicted that these variations give rise to an organized spatial distribution of cognition at a planetary-wide scale. To test this hypothesis we developed a mobile-app-based cognitive task, measuring non-verbal spatial navigation ability in more than 2.5 million people, sampling populations in every nation state. We focused on spatial navigation due to its universal requirement across cultures. Using a clustering approach, we find that navigation ability is clustered into five distinct, yet geographically related, groups of countries. Specifically, the economic wealth of a nation was predictive of the average navigation ability of its inhabitants, and gender inequality was predictive of the size of performance difference between males and females. Thus, cognitive abilities, at least for spatial navigation, are clustered according to economic wealth and gender inequalities globally, which has significant implications for cross-cultural studies and multi-centre clinical trials using cognitive testing.

Keywords: *spatial cognition, cross-country analysis, crowdsourcing, gender differences, ageing*

We devised a mobile video game designed to measure human spatial navigation ability through gameplay - Sea Hero Quest (SHQ). The game involves navigating a boat in search of sea creatures in order to photograph them (Figure 1 and Video S1). It features two main tasks: wayfinding and path integration. In wayfinding levels, players are initially presented with a map indicating start location and the location of several checkpoints to find in a set order (Figure 1A-C and Figure S1). Wayfinding task requires quite elaborate processing including interpretation of a map, planning a multi-stop route, memory of the route, monitoring progress along the route and updating of route plan, transformation of birds-eye perspective to an egocentric perspective needed for navigation [14]. In path integration levels, participants navigate along a river with bends to find a flare gun and then choose which three directions is the correct direction back to the starting point along the Euclidean space (Figure 1D and Figure S1). During path integration, one integrates perceived ego motion during travel to update one's position and orientation. It is a more basic (and evolutionary highly conserved) navigation mechanism, which typically only requires working memory processes [15, 16]. Together, wayfinding and path integration capture a wide range of the abilities and processes that are required for everyday successful navigation. 2,512,123 people between 18 and 99 years old from all 195 countries in the world downloaded and played the game (details in Table S1 and Figure S2 Part B). 57.6% of the participants provided demographics of their age, gender and nationality (Figure S1). To provide a reliable estimate of spatial navigation ability we examined the data only from those subjects who had completed a minimum of 9 levels of the game (see Methods). This resulted in 558,143 participants from 57 countries that were included in our analysis (Table S1).

To quantify spatial abilities, we defined Overall Performance corrected (OPcorr), a metric that captures different aspects of navigation abilities while correcting for video gaming skills (see STAR Methods). We further explain how we controlled for other potential biases such as the unavailability of SHQ in some languages, 'fake' demographics, and the virtual nature of the task in STAR Methods and Figure S2.

Results and Discussion

Across all countries we observed a similar pattern of decline in ability with age and a male advantage between 19 and 60 years old (Figure 2E, see Figure S2 Part G for plots from example nations). This result held true in all tested countries after correcting for differences in age and gender distributions.

We fit a multi-level model for OPcorr, with fixed effects for age and varying slope for gender, nested within nationality: $OP\ corr \sim age + (gender | nationality)$. Gender estimate had the same sign in every country, ranging from 0.43 to 1.49, $M = 0.97$, 95%I C = [0.90, 1.05]. We also computed Cohen's d within each country, they range from 0.09 to 0.48, $M = 0.29$, 95%I C = [0.27, 0.31], Figure S2 Part H. While a number of previous studies have examined the size of gender differences in cognitive abilities across countries, the underlying causes of such variation

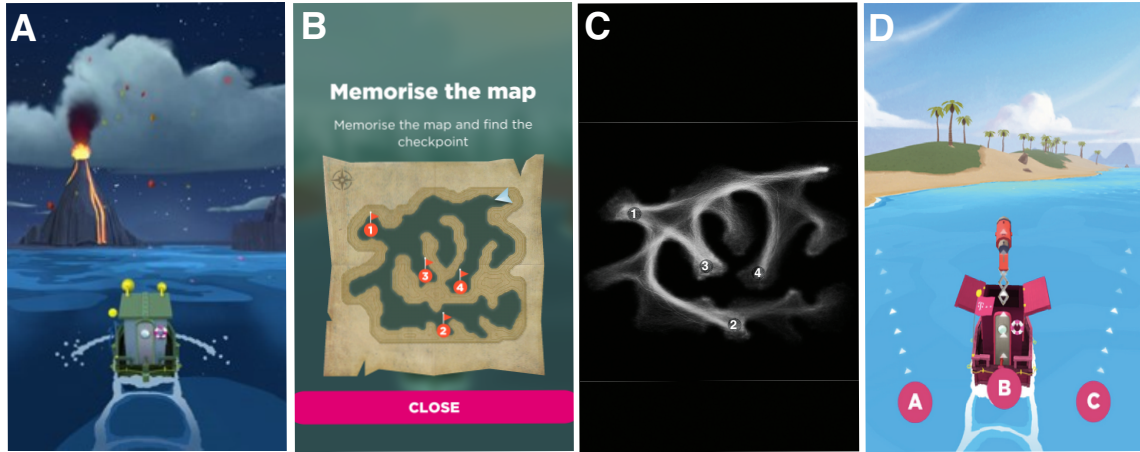


Figure 1. Tasks design. (A-B) Wayfinding task: a map of the level featuring the ordered set of checkpoints to reach is presented and disappears when the game starts. (C) Superposition of 1000 individual trajectories randomly sampled from level 32. (D) Path integration task: after navigating the level, participants must shoot a flare back to the starting point. See also Figure S1.

are still debated. Advocates of the gender stratification hypothesis argue that gender differences are more pronounced in countries with less equity [17, 18]. Data from Programme for International Student Assessment (PISA) that reports on more than 250,000 15-year-old students from 40 countries show that the gender gap in math scores disappears in countries with a more gender-equal culture [19, 20]. By contrast, other studies link gender differences more to evolved sex-linked dispositions and environmental affordances [21]. For example, difference in mental rotation and line angle judgment performance in more than 200,000 men and women from 53 nations remained even when controlling for gender equality [22]. Here, we report a positive correlation between the magnitude of gender differences measured in the aforementioned multi-level model and gender inequalities assessed by the World Economic Forum's Gender Gap Index (GGI), which reflects economic and political opportunities, education, and well-being for women (Figure 2D, $r = 0.62$, $p < .001$). We computed a multiple linear regression to predict gender estimates based on GDP and GGI. Both GGI ($t(52) = -2.93$, $p = 0.005$) and GDP per capita ($t(52) = -3.08$, $p = 0.003$) significantly predicted Countries' gender estimates. This suggests that the gender effect is not just related to countries' wealth, but also to the improvement of the role of women in society.

The age related decline in navigation abilities - OPcorr decreases in a linear fashion between 19 and 60 years-old (Figure 2E) - held true in all tested countries, age estimates ranging from -0.092 to -0.022 per year, $M = -0.059$, 95%IC = [-0.063, -0.055], Figure S2 Part G. Our observed early decline in performance mirrors the decline in 'fluid intelligence' components, which generally occurs in healthy adults [1-3]. Fluid intelligence refers to the capacity to reason independently of any knowledge from the past and is often linked to working memory. Our observed performance increment after 70 years of age was not predicted from the past literature and is consistent with a selection bias that those older participants willing to participate in online testing are likely to have greater cognitive skills. At the individual level performance should continue to decline, as

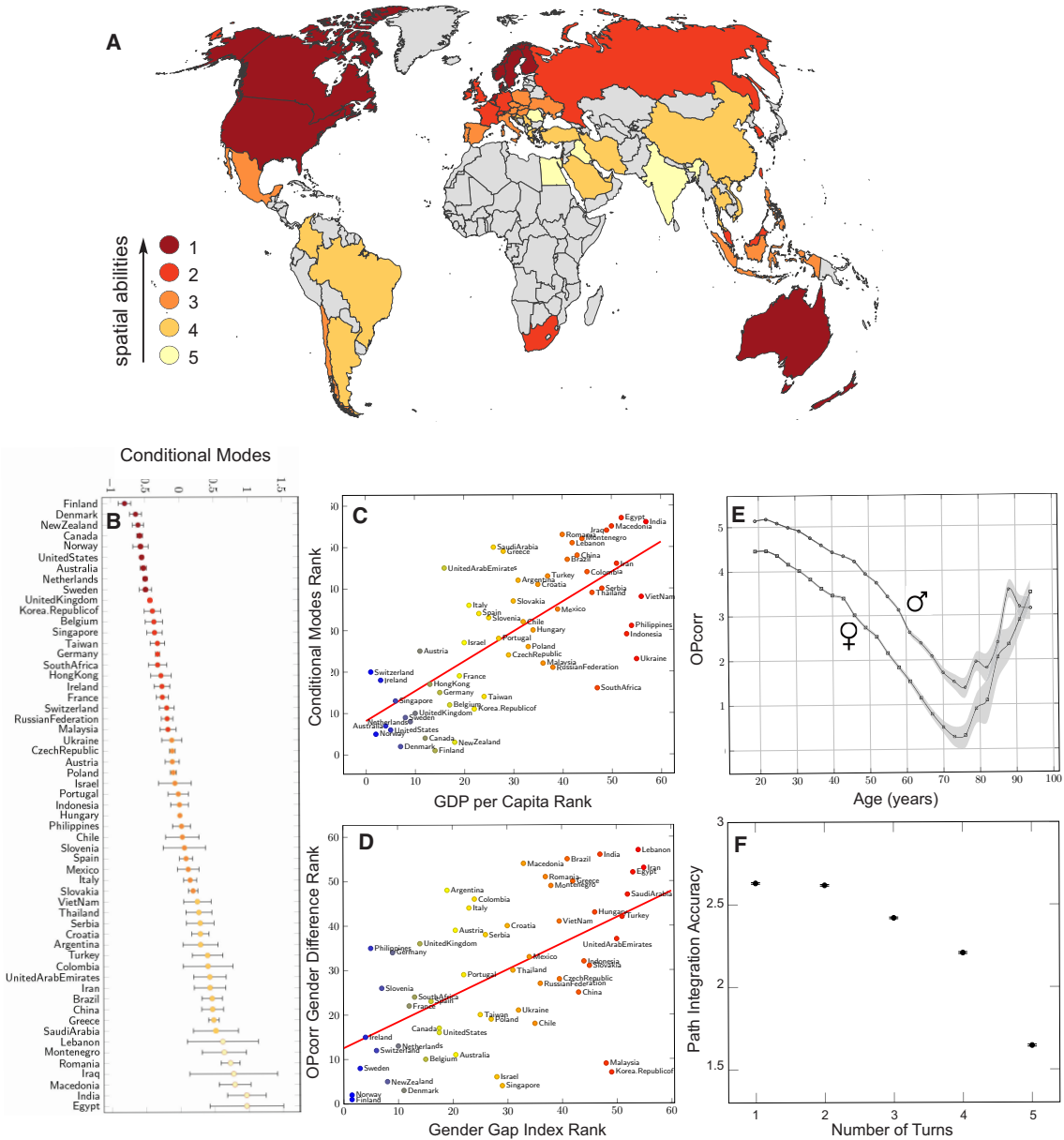


Figure 2. Spatial ability distribution across age, gender, and nations. (A-B) Five world clusters of people with similar overall performance corrected for video gaming skill (OPcorr). We used a multilevel model to predict OPcorr with fixed effect for age and gender and random effect for nationality. Conditional Modes (CM) represent the country-level performance (the lower the better). (C) Correlation between country performance (CM) and GDP per capita ($r=0.69$, $p < 0.001$). (D) Correlation between gender estimates and Gender Gap Index ($r=0.62$, $p < 0.001$). (E) Evolution of OPcorr across age and gender. Data points correspond to the average OPcorr within 3-year windows. Error bars correspond to standard errors. (F) Path integration accuracy (number of stars) vs. path complexity (number of turns). This plot includes participants that completed all five levels ($N = 19,038$). For more details see Figure S2 Part D. Error bars correspond to standard errors.

demonstrated in prior studies of navigation in elderly humans [23–26]. Increasing data shows that the first pathophysiological changes in dementia occur up to 20 years before diagnosis. Hence, the data of participants beyond 75 y.o., when we see the change of data trajectory, is not as

important as the data of participants up to 75 y.o. for future dementia screening based on the SHQ benchmark [27]. To our knowledge no prior large-scale studies have quantified the impact of nationality on a cognitive task. To assess the impact of nationality on spatial navigation we fit a multi-level model for OPcorr, with fixed effects for age and gender and random effect for nationality: $OP\ corr \sim age + gender + (1 | nationality)$. When we compare our multi-level model with a single-level model including only age and gender, nationality has a significant impact on OPcorr ($\chi^2(1) = 6413.8, p < 0.001$). The variance partition coefficient (VPC) indicates that 1.7% of the variance in performance can be attributed to differences between nationalities. Figure 2B represents countries ranked according to their conditional modes (CM), that is the difference between the global average predicted response in performance and the response predicted for a particular country. A reasonable assumption is that while OPcorr differs around the world, it follows a relatively smooth uniform distribution with some countries populations performing well and other performing less well on average. An alternative possibility is that countries are grouped by similar cognitive strategies, and that some countries will tend to behave more similarly. Indeed, country-level might not be the optimal scale to work at, since many social and geographical traits know no borders. To the best of our knowledge this hypothesis has never been tested with data from cognitive tests. To address this we pooled countries with similar CM into k clusters via the optimal 1D k-means algorithm [28]. We defined the optimal k as the one maximizing VPC. This was achieved for $k = 5$, $VPC = 2.6\%$, Figure 2A and Figure S2 Part B. Thus, spatial navigation ability appears to be clustered. Importantly, this clustering is distinct from GDP per Capita and video gaming skill distributions across countries (Figure S2 Part B). We downsampled the data to equate video gaming skill in our population and found a ranking and clustering nearly identical to the one with the full dataset (Pearson's correlation $r = 0.99, p < 0.001$, Figure S2 Part A and Part C).

The clustering of navigation abilities is not geographically random. Indeed, countries' CM were correlated with Gross Domestic Product (GDP) per capita (Figure 2C, Pearson's correlation $r = 0.69, p < 0.001$). This can be explained by different variables highly correlated with GDP associated with better spatial abilities, such as level of education [29] - particularly in science [30, 31] - or ability to travel [32]. Figure S2 Part I shows a positive correlation between countries' CM and average scores at PISA 2015 (Pearson's correlation $r = 0.73, p < 0.001$).

While GDP and GGI have a strong predictive influence on navigation ability, other country-level factors might influence navigation ability. Evidence suggests that driving rather than taking public transport has a positive effect on spatial knowledge [33, 34]. While this might explain why North Americans and Australians are particularly successful as populations compared to equivalent (GDP) European countries that rely more on public transport [35, 36], it fails to explain why the Nordic countries perform so well as a group. Many factors likely conspire to drive the superior performance in Nordic countries and it is impossible with the current data to precisely determine what the key factors are. One factor might be cultural activities that enhance navigation skill. Notably, the Nordic countries share a culture of participating in a sport related to navigation: orienteering. Nordic countries have a tradition of teaching orienteering in schools [37] and winning medals in the Orienteering World Championship (OWC). Across the 19 nations represented in the top 100 OWC ranking, performance in orienteering is significantly correlated with countries' CM in SHQ (Pearson's correlation $r = 0.55, p = 0.01$), even after correcting for

GDP per capita (Figure S2-N). This preliminary observation shows the potential for future targeted research to evaluate the impact of cultural activities on cognitive performance.

Assessing human cognition through video games constitutes a profound shift in behavioral sciences as it gives access to very large sample size and to populations difficult to reach through a lab setting (e.g. from remote countries). When considering such data it is important to be mindful of selection biases, such as the requirement to access a mobile device to participate. For example, our observation that performance improves with age after 75 y.o. (Figure 2E) is likely due to such a bias, as detailed above. Lab-based studies have their own selection biases, since people must choose to attend to participate and the testing pool is often very stereotyped [8]. Understanding these different biases will be crucial in future research. For our task it is important to also consider that familiarity with video games will influence navigation performance, such as via the expertise involved in controlling virtual movement. As noted previously, to account for these differences, in all our analyses we used a corrected performance score (OPcorr) which involved dividing our navigation performance measure by our measure of video game skill. To provide further scrutiny we down-sampled our participants to a population where performance on the first two tutorial levels was equivalent across the population and we found similar results to our main analyses (see Figure S2-B and the section *Controlling for familiarity with technology* in the STAR Methods). In future research it will be important to explore in greater detail the impact of video games proficiency on navigation skill in virtual environments and the real-world.

Navigation performance in the real-world has been found to correlate with navigation performance in virtual environments, both for healthy individuals [38, 39] and patients with early stage Alzheimer's disease [40]. Thus, virtual tests appear to provide some predictive value for real-world performance. However, because past studies have used desktop or immersive displays rather than a mobile version it remains unclear if SHQ performance would be predictive of real-world behaviour. To assess this, we tested participants with SHQ and with a real-world navigation task in the streets of London (UK) and Paris (France) [41]. In agreement with previous studies [38, 39] we find a significant correlation between real-world navigation performance and performance on SHQ levels that specifically test navigation [41]. No such correlation was found when performance on the tutorial levels was compared to real-world navigation. Thus, at least for certain real-world contexts, SHQ appears to predict real-world navigation performance. Further research exploring the construct validity of SHQ by testing it with other tests relating to navigation skill will be important to gain a more precise understanding of the behaviour it captures. For instance, studies examining path integration ability have shown that performance decreases with complexity of the path travelled [42, 43]. We replicated this pattern in every country examined in our data (Figure 2F and Figure S2 Part D).

Conclusion

Through the use of a mobile app-based approach, we have been able to reveal for the first time a global benchmark for spatial navigation. This approach enables predictions to be made about an individual's spatial navigation performance based on their age, gender and nationality. Specifically, GDP per capita and GGI of countries are predictive of their inhabitants' average spatial navigation performance. Thus, the collected dataset embodies a unique resource, which not only informs our understanding of global cognitive abilities but also provides a stepping stone

towards spatial navigation diagnostics and treatment in patient populations with navigation deficits, such as incipient Alzheimer's pathophysiology [26, 44, 45]. Use of online technology for the assessment of cognitive abilities has a promising future in particular with an ever-increasing world population making use online mobile technology.

Acknowledgments

The authors wish to thank Deutsche Telekom for supporting and funding this research, Alzheimer's Research UK for funding the analysis, the Glitchers Limited for the game production, Saatchi and Saatchi London for project management and creative input, Dara Mohammadi, Rogier Kievit and Charlotte Malcolm for their advice.

Author Contributions

HS and MH supervised the project, HS, MH, SS, JW, RD, VB and CH designed research; AC, RS, WdC, and EM analyzed the data; AC, MH and HS wrote the paper.

Declaration of Interests

The authors declare no competing interests.

References

1. Ghisletta, P., Rabbitt, P., Lunn, M., and Lindenberger, U. (2012). Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. *Intelligence* 40, 260–268.
2. Anguera, J.A., Boccanfuso, J., Rintoul, J.L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101.
3. Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science* 346, 572–578.
4. Linn, M.C. and Petersen, A.C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development* , 1479–1498.
5. Reilly, D. and Neumann, D.L. (2013). Gender-role differences in spatial ability: A meta analytic review. *Sex Roles* 68, 521–535.
6. Darwin, C.R. (1873). Origin of Certain Instincts. *Nature* 7, 417–418.
7. Dehaene, S., Izard, V., Spelke, E., and Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science* 320, 1217–1220.

8. Henrich, J., Heine, S.J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences* 33, 61–135.
9. Chiu, M.M. and Klassen, R.M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction* 20, 2–17.
10. Hanushek, E.A. and Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth* 17, 267–321.
11. Haun, D.B.M., Rapold, C.J., Janzen, G., and Levinson, S.C. (2011). Plasticity of human spatial cognition: Spatial language and cognition covary across cultures. *Cognition* 119, 70–80.
12. Lovett, A. and Forbus, K. (2011). Cultural commonalities and differences in spatial problem- solving: A computational analysis. *Cognition* 121, 281–287.
13. Goeke, C., Kornpetpanee, S., Köster, M., Fernandez-Revelles, A.B., Gramann, K., and König, P. (2015). Cultural background shapes spatial reference frame proclivity. *Scientific Reports* 5, 1–13.
14. Wiener, J.M., Büchner, S.J., and Hölscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation* 9, 152–165.
15. Mittelstaedt, M.L. and Mittelstaedt, H. (1980). Homing by path integration in a mammal. *Naturwissenschaften* 67, 566–567.
16. Etienne, A.S. and Jeffery, K.J. (2004). Path integration in mammals. *Hippocampus* 14, 180–192.
17. Baker, D.P. and Jones, D.P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education* 66, 91–103.
18. Else-Quest, N.M., Hyde, J.S., and Linn, M.C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin* 136, 103–127.
19. Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, Gender, and Math. *Science* 320, 1164–1165. 7
20. Hyde, J.S. and Mertz, J.E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences* 106, 8801–8807.
21. Reilly, D. (2012). Gender, Culture, and Sex-Typed Cognitive Abilities. *PLoS ONE* 7,

22. Lippa, R.A., Collaer, M.L., and Peters, M. (2010). Sex Differences in Mental Rotation and Line Angle Judgments Are Positively Associated with Gender Equality and Economic Development Across 53 Nations. *Archives of Sexual Behavior* 39, 990–997.
23. Lindenberger, U., Singer, T., and Baltes, P.B. (2002). Longitudinal Selectivity in Aging Populations: Separating Mortality-Associated Versus Experimental Components in the Berlin Aging Study (BASE). *Journal of Gerontology: Psychological Sciences* 57B, 474–482.
24. Moffat, S.D. (2009). Aging and Spatial Navigation: What Do We Know and Where Do We Go? *Neuropsychology Review* 19, 478–489.
25. Klencklen, G., Després, O., and Dufour, A. (2012). What do we know about aging and spatial cognition? Reviews and perspectives. *Ageing Research Reviews* 11, 123–135.
26. Lester, A.W., Moffat, S.D., Wiener, J.M., Barnes, C.A., and Wolbers, T. (2017). The Aging Navigational System. *Neuron* 95, 1019–1035.
27. Gordon, B.A., Blazey, T.M., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., Christensen, J., McDade, E., Wang, G., Xiong, C., et al. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology* 17, 241–250.
28. Wang, H. and Song, M. (2011). Ckmeans. 1d. dp: optimal k-means clustering in one dimension by dynamic programming. *The R Journal* 3, 29.
29. Skirbekk, V. and Loichinger, E. (2012). Variation in cognitive functioning as a refined approach to comparing aging across countries. *Proceedings of the National Academy of Sciences* 109.
30. Gunderson, E.A., Ramirez, G., Beilock, S.L., and Levine, S.C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Developmental Psychology* 48, 1229.
31. Uttal, D.H., Meadow, N.G., Hand, E.T.L.L., Warren, C., and Newcombe, N.S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin* 139, 352–402.
32. Poumanyvong, P., Kaneko, S., and Dhakal, S. (2012). Impacts of urbanization on national transport and road energy use: Evidence from low, middle and high income countries. *Energy Policy* 46, 268–277.

33. Maguire, E.A., Woollett, K., and Spiers, H.J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. *Hippocampus* 16, 1091–1101.
34. Sandamas, G. and Foreman, N. (2015). Active Versus Passive Acquisition of Spatial Knowledge While Controlling a Vehicle in a Virtual Urban Space in Drivers and Non-Drivers. *SAGE Open* 5, 1–9. 8
35. Pucher, J. (1988). Urban travel behavior as the outcome of public policy: the example of modal-split in Western Europe and North America. *Journal of the American Planning Association* 54, 509–520.
36. Bassett, D.R., Pucher Jr, J., Buehler, R., Thompson, D.L., and Crouter, S.E. (2008). Walking, cycling, and obesity rates in Europe, North America, and Australia. *Journal of Physical Activity and Health* 5, 795–814.
37. Annerstedt, C. (2008). Physical education in Scandinavia with a focus on Sweden: a comparative perspective. *Physical Education and Sport Pedagogy* 13, 303–318.
38. Moffat, S.D., Zonderman, A.B., and Resnick, S.M. (2001). Age differences in spatial memory in a virtual environment navigation task. *Neurobiology of Aging* 22, 787–796.
39. Conroy, R.A. (2001). Spatial navigation in immersive virtual environments. Ph.D. thesis, University College London.
40. Cushman, L.A., Stein, K., and Duffy, C.J. (2008). Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. *Neurology* 71, 888–895.
41. Coutrot, A., Schmidt, S., Pittman, J., Hong, L., Wiener, J.M., Hölscher, C., Dalton, R.C., Hornberger, M., and Spiers, H.J. (2018). Virtual navigation tested on a mobile app (Sea Hero Quest) is predictive of real-world navigation performance: preliminary data. *bioRxiv* URL <https://www.biorxiv.org/content/early/2018/04/22/305433>.
42. Benhamou, S. and Séguinot, V. (1995). How to find one's way in the labyrinth of path integration models. *Journal of Theoretical Biology* 174, 463–466.
43. Wiener, J.M., Berthoz, A., and Wolbers, T. (2011). Dissociable cognitive mechanisms underlying human path integration. *Experimental Brain Research* 208, 61–71.
44. Kunz, L., Schröder, T.N., Lee, H., Montag, C., Lachmann, B., Sariyska, R., Reuter, M., Stirnberg, R., Stöcker, T., Messing-Floeter, P.C., et al. (2015). Reduced grid-cell-like representations in adults at genetic risk for Alzheimer's disease. *Science* 350, 430–433.
45. Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.M., Spiers, H., and Hornberger, M. (2018). Impact of Sex and APOE Status on Spatial Navigation in Pre-symptomatic

Alzheimer's disease. bioRxiv URL
<https://www.biorxiv.org/content/early/2018/03/23/287722>.

46. Spiers, H.J., Burgess, N., Maguire, E.A., Baxendale, S.A., Hartley, T., Thompson, P.J., and O'Keefe, J. (2001). Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. *Brain* 124, 2476–2489.
47. (2015). The Global Gender Gap Report (Geneva, Switzerland: The World Economic Forum).
48. (2015). PISA Results in Focus (Paris, France: Organisation for Economic Co-operation and Development).

STAR METHODS

CONTACT FOR RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Hugo J. Spiers (h.spiers@ucl.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study has been approved by UCL Ethics Research Committee. The ethics project ID number is CPB/2013/015. Between May 2016 and July 2017, 2,512,123 participants from 255 countries and dependent territories downloaded and completed at least the first level of the game, see Table S1. Amongst them, 1,446,954 (57.6%) entered their age, gender and nationality. Examining the age distribution (Figure S1-Q), it is evident that age groups 18 and 99 years old contain more participants than would be predicted from the distribution. This is likely due to these numbers being the two extremes of the age range. It is likely that players under the age of 18 may have adopted these age bands. Since it is impossible to separate 'real' from 'fake' 18 and 99 years old players, for the current analysis we removed these two age groups from the dataset, leaving 926,456 (36.9%) participants. SHQ level progression is linear, i.e. one needs to complete level N in order to unlock level N+1 (Figure S1-C). Hence, the number of participants decreases with the progression through the game, as shown in Figure S1-T,U. To ensure a good tradeoff between sample size and amount of data per player, we included in the analysis participants who played at least the first 6 wayfinding levels (level numbers 1, 2, 3, 6, 7 and 8) and the first 2 path integration levels (levels 4 and 9). Level 5 is a creature chase level. This represents 625,626 (24.9%) valid participants. To reduce selection bias and ensure stable cross-country comparisons, we only included participants from countries with at least 500 valid participants. As a result of this sampling process, 558,143 (22.2%) participants from 57 countries were included in the analysis. Amongst them, 312,886 males (age: 34.97 ± 14.39 years old) and 245,257 females (age: 35.98 ± 15.50 years old), cf. Table S1 for country by country information.

METHOD DETAILS

Game Design

To test the global population on their navigation ability we worked with the independent video games design company Glitchers Ltd to produce a video game using Unity 3D (Unity Technologies, Copenhagen Denmark) for smart phones and tablets (apple and android devices). We were supported in this design process by staff at Deutsche Telekom (Germany) and Saatchi and Saatchi London (UK). ‘Sea Hero Quest’ (SHQ) was released on 4 May 2016 on the App Store for iOS and on Google Play for Androids. It is available in 17 languages: English, French, German, Spanish, Macedonian, Greek, Croatian, Dutch, Albanian, Hungarian, Romanian, Slovak, Czech, Polish, Portuguese, Italian, and Serbian, Figure S1A. The game is manipulated through three controls, designed to be intuitive; specifically, these were tap left to turn left, tap right to turn right, swipe up to speed up. Alongside tasks and levels, players were also asked a set of optional questions, which included their age, gender and nationality (Figure S1).

Tasks

Sea Hero Quest has been designed to reproduce as closely as possible classic navigation tasks from the literature. SHQ wayfinding task belongs to the path planning category defined in Wiener et al’s taxonomy of human wayfinding tasks (2009) [14]. We chose this type of task because, as the authors of this paper put it, “path planning is probably based on the most elaborate reasoning processes.” SHQ path integration task taps into a different yet classic type of navigation task: uninformed search coupled with the computation of momentary changes in location and orientation and their summation into a resultant vector allowing the navigator to come back home [16]. The experimental tasks in SHQ were accessed by unlocking levels sequentially (Figure S1C). These levels were comprised of 5 themed areas, each containing 15 levels. Through the game, participants followed a sea captain as he tries to recover his father’s lost memories (Figure S1B). There were three types of task. Wayfinding levels: at the beginning of each level, participants were given locations to visit from a map. The map disappeared, and they had to navigate a boat through a virtual environment to find different checkpoints. After initial levels 1-3 checkpoints are not encountered in the order of passage, but rather have to be navigated to by returning from one checkpoint to another (Figure S1). Path Integration levels: participants had to find a flare and shoot it back toward the starting point (Figure S1). Chase levels: participants chased a sea creature to take a picture of it. Chase levels were purely for motivational purposes and allow participants the capacity to share their game progress via social media in the form of a ‘photograph’ of the sea creature found. No data was collected from chase levels. Participants are encouraged to collect as many ‘stars’ as possible across the levels: the faster (Wayfinding task) or the more accurate (Path Integration task), the more stars were obtained. These stars unlocked the capacity to modify the boat in the game.

Data Collection

Within the opening screen and the Journal menu, participants were made aware of the purpose of

the game. They were asked whether they were willing to share their data with us and were guided to where they can opt out. The opt out was always available in the settings. The website for the game (www.seaheroquest.com) was linked to from the About menu and provided full information about the study and what the data was going to be used for. If the participant agreed, their data (boat trajectory, flare accuracy and demographics) were anonymously stored in a secure T-Systems server in Germany. The application is managed by T-Systems' scalable Docker offering called 'AppAgile' which is operated out of T-Systems' datacenter to ensure data integrity and data privacy according to German data security law. The data are owned by Deutsche Telekom and then licensed to University College London for analysis. Each participant was identified by a universally unique identifier (user-uuid), a 128-bit number commonly used to identify information in computer systems. Participant's sessions were identified by another universally unique identifier (instance-uuid). Only completed levels were stored and analysed. During Wayfinding levels, the coordinates of participants' trajectories were sampled at $F_s = 2$ Hz. During Path Integration levels, flare accuracy was quantified in term of stars obtained by the participant. Stars were awarded based on participant's choice between 3 proposed directions: 3 stars for the correct answer (their starting point), 2 stars for the second closest direction, and 1 star for the third closest direction.

QUANTIFICATION AND STATISTICAL ANALYSIS

Metrics

To quantify spatial abilities, three measures were computed.

For wayfinding levels, we computed:

- Trajectory length in pixels, defined as the Euclidean distance

$$\sqrt{\sum_{i=1}^{N-1} (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \text{ with } (x_i, y_i)_{i \in [1..N]} \text{ a N-points trajectory.}$$

- Duration to complete the game in seconds. Since $F_s = 2$ Hz, Duration = $N/2$.

While duration is a direct measure of the performance at the task (find all checkpoints in the correct order as fast as possible), it may be biased toward participants who have the reflex to accelerate. Trajectory length gets around this bias and has been used in clinical settings to measure error while participants were rewarded for duration [46]. Taken together, these metrics allow to efficiently characterise participant's profile: slow/fast and accurate/inaccurate, see Figure S2 Part E and Part F.

For path integration levels, we computed flare accuracy in number of stars: 1, 2 or 3. We considered that video games experience might bias performance, with players familiar with similar games having an advantage. Therefore we normalized durations and trajectory lengths by dividing them by the sum of their values at the first two levels, where no sense of direction is needed (Figure S1F,G: goals are visible from the starting point). We found all three measures correlated: Pearson's correlation between trajectory length and duration $r = 0.75$, $p < 0.001$; trajectory length and path integration accuracy $r = -0.21$, $p < 0.001$; duration and path integration accuracy $r = -0.20$, $p < 0.001$. While path integration relied on the integration of perceived ego

motion information over time, wayfinding required the planning of multi-stop routes, memory of survey/map, transformation of survey representation into egocentric reference frame. It is therefore not surprising that trajectory lengths and duration - both wayfinding-related metrics - were more correlated with each other than with flare accuracy. We defined an overall performance metric corrected for video gaming skill (OPcorr) summarizing normalized durations, trajectory lengths and flare accuracies. OPcorr was the 1st component of a Principal Component Analysis across the normalized durations and trajectory lengths of levels 6, 7, 8 and the flare accuracies of levels 4 and 9 (66.9% of the variance explained). The first unrotated loadings respectively corresponding to (distance, duration, flare accuracy) are (0.99, 0.12, -0.035), showing that distance is preponderant while flare accuracy is the least contributing factor to OPcorr. Flare accuracy loading is negative for arbitrary reasons: performance increases with the number of stars, while it decreases with increasing duration and distance.

Controlling for familiarity with technology

A common potential bias when assessing cognitive abilities with virtual tasks is the influence of participants' computer experience on their performance. We controlled that Sea Hero Quest tasks capture participants' spatial ability and not only their familiarity with technology through different approaches.

Normalization - We normalized durations and trajectory lengths by dividing them by the sum of their values at the first two levels. The first two levels only reflect video gaming skill (motor dexterity with the game controls) as no sense of direction is required to complete them, see Figure S1F,G.

Participants with homogeneous video gaming skills - We re-ran the analyses presented in the main body of the manuscript with a subset of participants with similar performance at levels 1 and 2. Performance at levels 1 and 2 was defined as the first component of a Principal Component Analysis across trajectory lengths and durations for levels 1 and 2 (70.3% of variance explained). We included in this subset all participants within the [0.25 0.75] quantile interval of the distribution of performance at levels 1 and 2 (N = 280,885). As shown in Figure S2-B, country ranking remained very similar to the ranking based on the full dataset (Pearson's correlation $r = .99$).

Video gaming skill ranking - We also re-ran the analyses presented in the main body of the manuscript based only on performance at levels 1 and 2. Country ranking based on levels 1 and 2 was correlated to the Overall Performance corrected for video gaming skill (OPcorr) ranking based on all levels (Pearson's correlation $r = 0.73$). However, many differences appeared between rankings, see Figure S2-C. For instance, the United Kingdom jumped from the tenth place in OPcorr ranking to the first place in video gaming skill ranking, and Finland dropped from the first to the eight place.

Geographical clustering - As described in the manuscript, we fit a multi-level model for OPcorr, with fixed effects for age and gender and random effect for nationality. Nationality levels were

either individual countries (granularity = 0, 57 levels) or cluster numbers for country partitions with 2, 3, 4, 5, 6 or 7 clusters. Variance Partition Coefficient (VPC) represents the percentage of variance explained by the random effect. Clusters have been computed with the optimal 1D K-mean algorithm. As stated in the manuscript, when the country clustering was based on OPcorr, VPC was maximal for 5 clusters (Figure S2-D). However, when the country clustering was based on performance at level 1 and 2 (reflecting video gaming skill) VPC was maximal for 2 clusters (Figure S2-E), and when the country clustering was based on GDP per capita VPC was maximal without country clustering (i.e. with all 57 levels, see Figure S2-F).

Figure S2 Part C represents the geographical clustering of 5 variables: OPcorr, OPcorr computed from participants with similar video gaming skill, video gaming skill, GDP per capita and Gender Gap Index. Even though we just saw that 5 clusters was only optimal for OPcorr, we chose to represent each geographical distribution with the same number of clusters to facilitate the comparison with OPcorr. We see that while the 5 clusters of OPcorr and OPcorr computed from participants with similar video gaming skill remained unchanged (Figure S2-G,H), many differences appeared between clusters of OPcorr and the other variables (Figure S2-I-K). This shows that the geographical clustering presented in Figure 2B is distinct from the distributions of familiarity with technology or of video gaming skill.

Validation from the spatial cognition literature - Another way to control whether Sea Hero Quest captured spatial ability is to quantify how our data fit with results from the spatial cognition literature. For instance, path integration models predict error accumulation over travelled distance and increasing turning angle [42, 43]. We compared this prediction with participants' performance in the path integration task. We chose path integration levels with increasing complexity: level 14 (1 turn), level 34 (2 turns), level 54 (3 turns), level 44 (4 turns) and level 74 (five turns), Figure S2 Part D. Note that here, the number of turns and overall turning angle were highly correlated. To avoid selection bias, we included participants who completed all five levels. In all tested countries, path integration accuracy decreased with complexity. This shows that real world path integration models do predict performance in Sea Hero Quest path integration task.

Controlling for fake demographics

Figure S1-Q suggests that many participants did not enter their real age, as the lower and upper bound of the age range are clearly over-represented. For ethical reasons, we set the youngest available age to 18 years-old. This might explain in part the spike at 18 y.o., younger participants being unable to select their real age. To limit the amount of incorrect information, we removed 18 and 99 y.o. from all analyses. Checking the authenticity of other age groups is more difficult. To do so, we investigated the coherence of age and gender distributions in different clusters of performance. We projected each participant in a 2-dimensional space representing the first PCA component of normalized duration and trajectory length (Figure S2 Part5). We fitted to this cloud of points a Gaussian Mixture Model (GMM) with 4 bivariate normal components (2D Gaussians). We tried different numbers of components but the structure of the model remained similar, i.e. the center of the Gaussians aligned on the space diagonal. We assigned each participant to exactly one cluster (hard clustering) based on a cluster membership score. Each cluster membership

scores is the estimated posterior probability that the data point came from the corresponding Gaussian. Each participant is assigned to the Gaussian corresponding to the highest posterior probability. In Figure S2 Part E, the blue cluster is closest to the space origin, hence composed of the most efficient participants. The further away from the origin, the less efficient. For each cluster, we computed the age and gender distribution of their participants. The mean age increases when performance decreases, in accordance with Figure 1e. We performed the exact same analysis only with older participants: if younger participants pretended to be older (likely more than 90 years old) we should see anomalous age distribution, with very old participants clearly over-performing. This analysis is represented in Figure S2 Part F, where a GMM is fitted on 8,653 participants above 70 y.o. Looking at the blue cluster, representing the most efficient older participants, we see that its distributions of age and gender do not particularly show spikes above 90 y.o., suggesting that most entered ages are real. Age distributions have similar shapes for males and females, which suggests that gender information is mostly reliable. Country information is more difficult to validate, this issue is not addressed in this manuscript.

Controlling for language availability

We designed very intuitive tasks and game controls to enable anyone to play, even without understanding the few textual information. To evaluate the influence of language on performance, we defined L as a coding variable for the existence of Sea Hero Quest in one of the official languages of the countries: $L_i = 1$ if SHQ is translated in language of country i , $L_i = 0$ otherwise. We fitted a logistic binomial model of L as a function of countries' CM, and found no significant effect ($t(55) = -1.01$, $p = 0.31$).

DATA AND SOFTWARE AVAILABILITY

Given the magnitude of the dataset, the authors made it available from a dedicated server. Access will be granted upon reasonable request. Analyses and figures were made using Python, JavaScript, R, and Matlab. Scripts are available from the Lead Contacts upon request.

Gross Domestic Product per Capita information were provided by the World Bank in 2015 (<https://data.worldbank.org/>).

Gender Gap Index information were provided by the World Economic Forum in 2015 [47] (<http://www3.weforum.org/docs/GGGR2015/cover.pdf>).

Programme for International Student Assessment scores were provided by the Organisation for Economic Co-operation and Development (OECD) in 2015 [48] (<https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>). Orienteering championships scores were provided by the International Orienteering Federation in 2017 (<http://ranking.orienteering.org/>).

SUPPLEMENTAL LEGENDS

Table S1 - Information about participants in each country, related to Figure 2. Number, female/male ratio, female mean age and male mean age. Columns 2 to 5: all participants who entered their demographics. Columns 6 to 9: participants included in the analysis. Countries with less than 500 participants have been excluded.

Video S1 - Wayfinding task example, related to Figure 1 and 2E. First part: one participant completing level 8. Second part: 1000 randomly sampled older participants (in blue, between 70 and 95 y.o.) and 1000 randomly sampled younger participants (in red between 19 and 25 y.o.) completing level 16.