ARTICLE   OPEN

Check for updates

Evidence for a unitary structure of spatial cognition beyond general intelligence

Margherita Malanchini [1,2,3,9 ✉], Kaili Rimfeld[2,9], Nicholas G. Shakeshaft[2], Andrew McMillan[2], Kerry L. Schofield[2], Maja Rodic[4], Valerio Rossi[5], Yulia Kovas [4,6], Philip S. Dale [7], Elliot M. Tucker-Drob[3,8] and Robert Plomin[2]

Performance in everyday spatial orientation tasks (e.g., map reading and navigation) has been considered functionally separate from performance on more abstract object-based spatial abilities (e.g., mental rotation and visualization). However, few studies have examined the link between spatial orientation and object-based spatial skills, and even fewer have done so including a wide range of spatial tests. To examine this issue and more generally to test the structure of spatial ability, we used a novel gamified battery to assess six tests of spatial orientation in a virtual environment and examined their association with ten object-based spatial tests, as well as their links to general cognitive ability (*g*). We further estimated the role of genetic and environmental factors in underlying variation and covariation in these spatial tests. Participants (*N* = 2660; aged 19–22) were part of the Twins Early Development Study. The six tests of spatial orientation clustered into a single '*Navigation*' factor that was 64% heritable. Examining the structure of spatial ability across all 16 tests, three, substantially correlated, factors emerged: *Navigation*, *Object Manipulation*, and *Visualization*. These, in turn, loaded strongly onto a general factor of *Spatial Ability*, which was highly heritable (84%). A large portion (45%) of this high heritability was independent of *g*. The results point towards the existence of a common genetic network that supports all spatial abilities.

INTRODUCTION

Spatial skills are fundamental for everyday life as they make it possible for us to understand and operate in the physical world around us. Studies in primates and other animals have highlighted the importance of spatial ability for evolution and survival. Food-hoarding birds rely on spatial memory to retrieve their caches, which is crucial to their subsistence, and climate harshness has been found to positively drive the evolution of spatial memory skills in black-capped chickadee, another bird species[1]. Spatial skills are also important in modern technologically oriented societies[2,3] as individual differences in spatial skills are associated with positive developmental, educational and life outcomes. Spatial ability reliably predicts scholastic and professional success and career choices, particularly in Science, Technology, Engineering and Mathematics (STEM) and related fields, even after controlling for general cognitive ability[4–6]. In spite of the increasingly fundamental role that spatial ability has for individuals and contemporary societies[3], numerous questions remain regarding the nature of spatial ability as well as its origins and structure[7].

What constitutes good spatial skills? Since its earliest conceptualization[8], spatial ability has been considered a multifaceted construct comprising several related, yet separable, skills[9]. One of the most widely adopted definitions of spatial ability describes it as the ability to generate, retain, retrieve, and transform well-structured visual images[10]. Contrary to this very broad characterization of spatial ability, however, extant research has largely focused on measuring only specific aspects of object-based spatial ability. Among the most widely studied spatial skills are

individuals' abilities to mentally rotate shapes[11], to visualize objects from different perspectives, and to find figures embedded within other shapes[12]. A much smaller body of research has considered larger-scale, practical, everyday spatial orientation abilities, such as navigation, map reading, and wayfinding.

Until recent years, studies of spatial orientation skills had been hindered by the difficulty in measuring navigation and wayfinding abilities in real-life settings utilizing rigorous approaches that are standardized across participants. In addition, assessing navigation in the real environment can be highly costly and time consuming and thus unlikely to be scalable to large samples nation wide or world wide. Technological advances in the field of virtual reality (VR) provide a novel powerful tool to study individual differences in spatial orientation skills in realistic settings that can be fully controlled and standardized across participants[13,14]. Studies assessing the validity of measuring navigation skills using VR have observed strong correlations (~0.60) with performance in real world navigation skills[13,15]. The reliability of assessing spatial abilities in VR is likely to continue increasing as accelerating technological developments provide progressively immersive and realistic tools.

Likely due, at least in part, to such difficulties in assessing multiple spatial orientation skills reliably in large, representative samples, few studies have examined the structure of spatial orientation ability and its association with other spatial skills. More broadly, evidence concerning the nature and factor structure of spatial ability remains mixed, with most studies focusing on differentiating between relatively few measures rather than examining the communalities across a broad range of spatial

[1]Department of Biological and Experimental Psychology, Queen Mary university of London, London, UK. [2]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [3]Population Research Center, The University of Texas at Austin, Austin, USA. [4]InLab, Department of Psychology, Goldsmiths University of London, London, UK. [5]ETT Solutions LTD, London, UK. [6]Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk, Russia. [7]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, USA. [8]Department of Psychology, The University of Texas at Austin, Austin, USA. [9]These authors contributed equally: Margherita Malanchini, Kaili Rimfeld. ✉email: m.malanchini@qmul.ac.uk

skills[7,10,16,17]. In our previous work[18], we have shown that a general factor of spatial ability captures a substantial proportion of variance across numerous tests of spatial skills, and that communalities across tests are largely explained by shared genetic variance[18]. However, one major limitation characterized our previous study: although we considered ten object-based spatial abilities, including tests of rotation, visualization and scanning abilities, we did not include measures of spatial orientation, such as navigation, map reading, and wayfinding.

The omission of spatial orientation measures has special theoretical relevance because evolutionary and cognitive theories have pointed to a distinction between the ability to mentally manipulate objects on a small scale (object-based spatial skills) and the ability to orient in large-scale environments (spatial orientation ability)[19–21]. This proposition is partly supported by psychological studies suggesting that the two abilities are influenced by separate cognitive processes and brain structures. For example, in a study of the association between performance in object-based psychometric spatial tests and large-scale spatial learning, partial support was found for a differentiation between these skills. Individual differences in measures of spatial learning (measuring skills such as placing landmarks on a map, intra-route distance estimates and route reversal) were unrelated to variation in object-based spatial tests. However, the ability to learn maze and maze reversal, was found to be related to both object-based tests and spatial learning[22]. Other studies in the field of cognitive psychology have found evidence for a partial dissociation between object-based tests and large-scale spatial orientation skills[23–25].

Neuroimaging studies have also provided preliminary converging evidence for the distinction between object-based abilities and spatial orientation skills, suggesting that the two are supported by separate brain networks. Object-based spatial skills, and particularly mental rotation ability, were found to be primarily associated with activation of the parietal lobes[26]. Conversely, variation in learning and remembering the layout of large-scale spaces has been found to be related to processing in the hippocampus and the medial temporal lobes[27].

Other theoretical accounts and studies, however, have suggested that object-based and spatial orientation skills might be closely related. For example, theories concerning the evolution of sex differences have argued that individual variation in object-based spatial skills, such as mental rotation, are the product of different selection pressures for large-scale spatial orientation abilities between males and females over evolutionary history[28,29]. Therefore, these theories suggest that spatial orientation and object-based spatial skills largely reflect a common set of abilities. Empirical evidence also supports the idea of a largely unitary set of abilities. A study of the association between object-based spatial abilities, measured with a limited battery of three psychometric tests, and large-scale spatial orientation skills, measured both in realistic settings and a virtual environment, found a substantial correlation ($r = 0.60$) between the two[15].

The proposition of a unitary set of cognitive processes underlying object-based and spatial orientation skills is consistent with the idea that these are aspects of a more general set of cognitive abilities. It is plausible that at the heart of individual differences in all spatial skills is general cognitive ability, or general intelligence (g). G is a psychometric construct that emerged at the beginning of the twentieth century from observations that almost all cognitive tests correlate moderately and positively[30]. Individuals performing highly on one cognitive test are also likely to show good performance on other tests of cognitive abilities, and g indexes this covariance observed between cognitive measures. Therefore, g is thought to represent individual differences in the domain-general abilities to plan, learn, think abstractly, and solve problems that are necessary for successfully completing cognitive tests[31].

In our previous work on the factor structure of object-based spatial tests, we have shown that individual differences in spatial abilities cluster into a unitary factor, at both the observed and genetic levels, even after accounting for g[18]. Along the same lines, another study found that the association between object-based and spatial orientation abilities was largely independent of verbal ability[15]. These studies suggest that the coherence of spatial abilities is not simply due to their being part of g, but rather inherent in the spatial domain itself. However, neuropsychological evidence contradicts this view. Case studies of patients with neuropsychological impairments suggest that damage to navigation-related structures in humans typically leads to broad memory deficits that are not limited to the spatial domain[10].

Extant literature is therefore characterized by contrasting theories and evidence with respect to the factor structure and associations between object-based spatial abilities, assessed mostly through psychometric tests, and large-scale spatial orientation skills, assessed both in real settings and VR. The lack of a cohesive account is likely due to a paucity of studies that have investigated the association between object-based and large-scale spatial orientation skills with a sufficiently diverse battery of tests. In addition, to our knowledge, no study to date has investigated their links within a genetically informative framework, testing the hypothesis that a common genetic network, independent of g, supports performance in all spatial skills.

The current study addresses these limitations by investigating the structure of spatial ability using two comprehensive online batteries of object-based and spatial orientation skills, administered to a large genetically informative sample of twins aged 19–22. Importantly, we assessed spatial orientation abilities with an innovative gamified battery of six tests measuring navigation, map reading, wayfinding and large-scale scanning and perspective-taking skills set in a virtual environment. The current work has three main aims: first, we examined the factor structure and origins of spatial orientation skills; second, we investigated the structure and genetic and environmental origins of spatial ability across sixteen tests of object-based and spatial orientation skills; third, we explored the role that g has in unifying individual differences in performance across tests of spatial abilities.
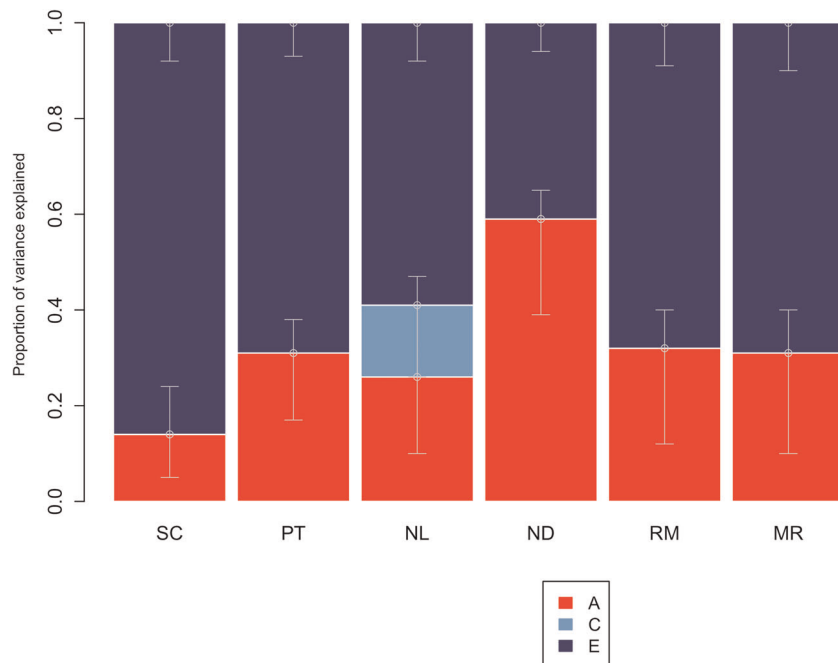
RESULTS

Individual differences in spatial orientation can be measured reliably in a virtual environment and are moderately heritable

We first assessed whether our newly developed gamified battery set in a virtual environment could effectively capture individual differences in spatial orientation skills in our large sample of twins. Beyond showing good test–retest reliability (average $r = 0.74$, ranging from 0.60 to 0.89; see Method for information on the reliability of each tests), the six tests—scanning, perspective taking, navigation based on landmarks, navigation following directions, route memorizing and map reading—showed normal distributions, with acceptable values for skewness and kurtosis ($< +/-2$; Fig. 1 and Supplementary Table 1). Therefore, our gamified battery was able to discriminate and reliably capture variation in spatial orientation abilities.

Because sex differences are often found for spatial abilities (though not always in the same direction)[32,33], we examined whether performance differed between males and females. We found significant differences in performance between males and females across all tests (Supplementary Table 2). Males outperformed females in all tests, effect sizes ranged between small and moderate. The biggest effect size was observed for map reading ($R^2 = 0.17$) and the smallest effect size was observed for scanning ($R^2 = 0.03$).
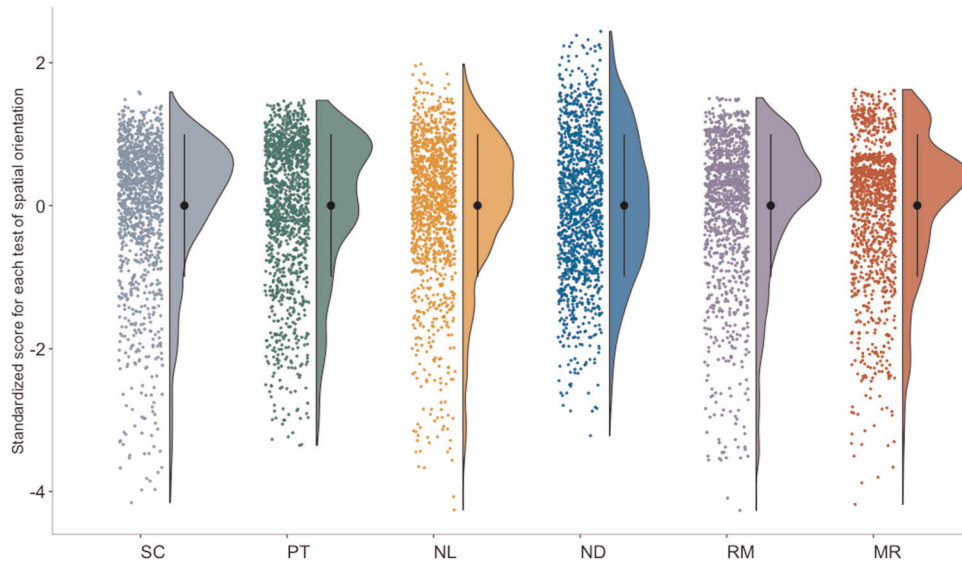
Fig. 1 Individual differences and distributions for the six tests included in our novel gamified battery of spatial orientation set in a virtual environment. All variables were residualized for age and sex, and standardized in one randomly selected half of the sample (only one twin within each pair was randomly selected for descriptive and phenotypic analyses in order to account for the non-independence of observations); full descriptive statistics for both randomly selected halves of the sample are presented in Supplementary Table 1. The black dots represent means and the error bars standard deviations. SC scanning, PT perspective taking, NL navigation landmarks, ND navigation directions (cardinal points), RM route memory, MR map reading.


Fig. 2 Genetic and environmental estimates for navigation tests: univariate model-fitting results. A additive genetic; C shared environmental; E nonshared environmental components of variance. Error bars show 95% confidence intervals. SC scanning, PT perspective taking, NL navigation landmarks, ND navigation directions (cardinal points), RM route memory, MR map reading.

We applied the twin method (see Method section) to calculate heritability estimates for the six measures of spatial orientation; these are presented in Fig. 2. Heritability estimates, the extent to which variation in a trait is accounted for by genetic differences[34], ranged from 14 to 57%. The remaining variance in all tests was accounted for by nonshared environmental factors, environmental factors that do not contribute to similarities between siblings[34], with the only exception being the test of orientation ability using landmarks, which showed a significant proportion of shared

environmental variance (15%). These substantial nonshared environmental estimates in part reflect measurement error.

Given the significant sex differences observed at the phenotypic level, we conducted univariate full sex-limitation model fitting (see Method section) to examine whether these estimates of heritability differed between males and females. We found significant quantitative sex differences for navigation ability and several of the subtests (Supplementary Table 3), that is, differences in the magnitude of genetic and environmental influences were
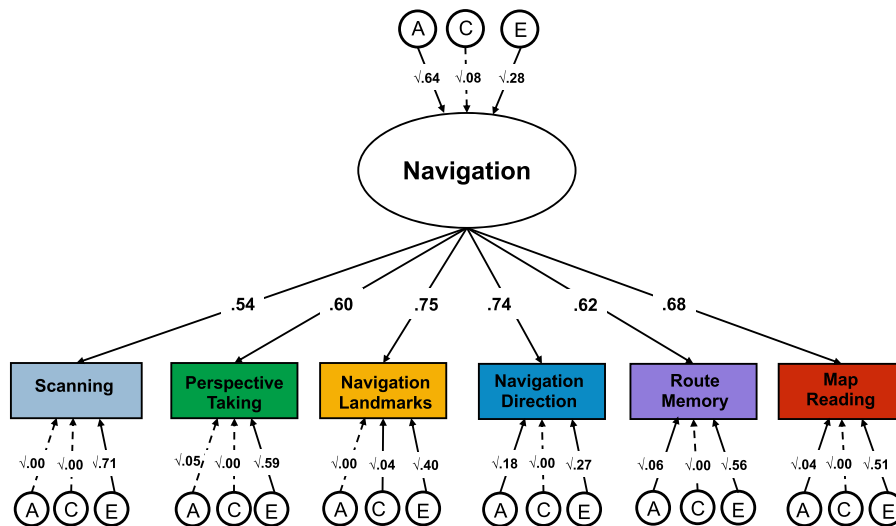
Fig. 3 Factor structure and genetic and environmental variance common across the six tests of spatial orientation. We applied the common pathway model to parse the genetic (A), shared environmental (C) and nonshared environmental (E) variance that is shared across all the tests (represented by the A, C and E paths leaving from the common *Navigation* factor) from the genetic and environmental variance that is specific to each test (indexed by the individual A, C, and E latent factors from each rectangle). Each individual test loaded substantially onto a common factor, which we named *Navigation factor* (loadings ranging from $\lambda = 0.54$ for scanning ability to $\lambda = 0.75$ for navigation based on landmarks). All A, C, and E paths are standardized and squared.

observed for males and females. However, the effect sizes of these differences were small to moderate. For example, for for an overall composite measure of navigation ability, heritability was 52% (95% CI: 0.31; 0.70) for males and 54% for females (95% CI: 0.29; 0.62). For these reasons, and to increase power, the full sample was used in subsequent analyses, combining males and females, and same- and opposite-sex twin pairs.

A single 'navigation' factor captured the variance common across all tests of spatial orientation
We applied factor analysis (see Method section) to examine the covariance structure across the six tests in the spatial orientation battery. The results showed that the six tests correlated substantially and clustered into one common factor, capturing between 32 and 57% of the variance in each individual test, which we named 'Navigation', as it indexed abilities that are generally described in the literature as spatial navigation skills (see Supplementary Table 4—factor structure, and Fig. 4—intercorrelations between tests). This unifactorial model provided a good fit for the data ($\chi^2 = 269.937$ (148), $p = <0.001$; comparative fit index (CFI) = 0.968; Tucker-Lewis index (TLI) = 0.971; root mean square error of approximation (RMSEA) = 0.030; standardized root mean square residual (SMRR) = 0.049).

We used the Common Pathway model (see Method section) to examine the extent to which genetic (A), shared environmental (C) and nonshared environmental (E) effects were common or specific across the six tests (Fig. 3). We found that the heritability of the common navigation factor was 64% (95% CIs = 0.41–0.91); shared environmental and nonshared environmental factors accounted for smaller proportions of variance, 8% (95% CIs 0.00–0.43) and 28% (95% CIs 0.21–0.36), respectively. The largest part of the genetic variance in navigation ability was shared across all tests; between 66 and 100% of the heritability of each test was captured by the common factor of navigation. Consequently, test-specific genetic effects were found to account for between 0 and 34% of the genetic variance in each test of spatial orientation (Supplementary Table 5).

Environmental factors were largely specific to each test, as indicated by the considerable size of the specific E paths (bottom of Fig. 3); between 64 and 90% of the nonshared environmental variance was found to be specific to each test. The common navigation factor only captured between 10 and 36% of nonshared environmental variance in each test of spatial orientation (Supplementary Table 5).
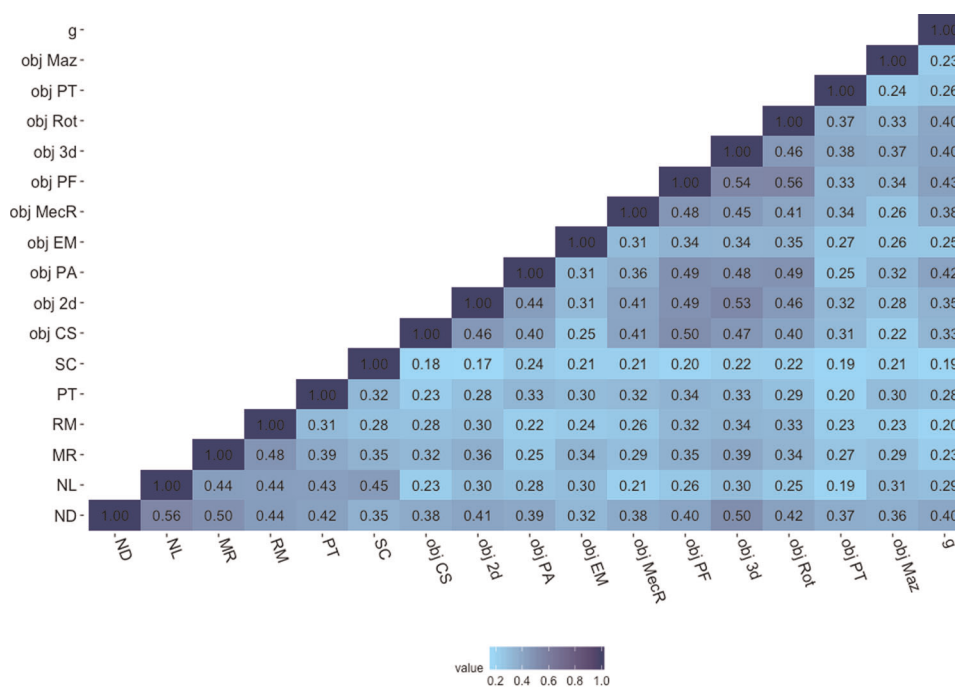
Substantial associations between measures of spatial orientation and object-based spatial tests
We investigated the structure of spatial ability across a greater diversity of spatial tests. To this end, we extended our analyses beyond the six tests of spatial orientation to incorporate 10 additional tests of object-based spatial skills[18]. This additional battery of spatial tasks included measures that very closely align with traditional psychometric tests of spatial ability, including mental rotation, visualization, 2D and 3D drawing ability, and mechanical reasoning. Figure 4 presents phenotypic correlations between the 16 spatial tests included in the two batteries (spatial orientation and object-based) and their correlations with g.

Correlations between spatial tests were positive and ranged from modest (between 0.10 and 0.30) to strong (>0.5) with r ranging between 0.17 and 0.56. Stronger links were observed between certain tests within each battery. For example, the four tests assessing navigation and map reading skills in the spatial orientation battery clustered more strongly together (r ranging from 0.44 to 0.56). The same was observed for measures of 2D and 3D drawing, pattern assembly, paper folding, and mental rotation in the object-based battery (r ranging from 0.34 to 0.54).

We conducted a series of confirmatory factor analyses to formally evaluate the covariance structure between the 16 spatial tests. We tested different theoretical models about the structure of spatial skills, starting from the simplest model and progressing to increasingly complex representations of the structure of spatial skills. The first model we tested was a one-factor model (Fig. 5a), positing that variation in spatial orientation and object-based skills could be largely considered a unitary ability. Although all tests loaded substantially onto a single factor (Fig. 5a), model fit indices ($\chi^2 = 692.730$ (104), $p < 0.001$, CFI = 0.890, TLI = 0.873, RMSEA = 0.061, SMRS = 0.059) suggested that this structure did not provide a good fit for the data.

Secondly, we tested whether including two factors of spatial ability (one for each battery, Supplementary Fig. 1) would provide

	ND	NL	MR	RM	PT	SC	obj CS	obj 2d	obj PA	obj EM	obj MecR	obj PF	obj 3d	obj Rot	obj PT	obj Maz	g
g																	1.00
obj Maz																1.00	0.23
obj PT															1.00	0.24	0.26
obj Rot														1.00	0.37	0.33	0.40
obj 3d													1.00	0.46	0.38	0.37	0.40
obj PF												1.00	0.54	0.56	0.33	0.34	0.43
obj MecR											1.00	0.48	0.45	0.41	0.34	0.26	0.38
obj EM										1.00	0.31	0.34	0.34	0.35	0.27	0.26	0.25
obj PA									1.00	0.31	0.36	0.49	0.48	0.49	0.25	0.32	0.42
obj 2d								1.00	0.44	0.31	0.41	0.49	0.53	0.46	0.32	0.28	0.35
obj CS							1.00	0.46	0.40	0.25	0.41	0.50	0.47	0.40	0.31	0.22	0.33
SC						1.00	0.18	0.17	0.24	0.21	0.21	0.20	0.22	0.22	0.19	0.21	0.19
PT					1.00	0.32	0.23	0.28	0.33	0.30	0.32	0.34	0.33	0.29	0.20	0.30	0.28
RM				1.00	0.31	0.28	0.28	0.30	0.22	0.24	0.26	0.32	0.34	0.33	0.23	0.23	0.20
MR			1.00	0.48	0.39	0.35	0.32	0.36	0.25	0.34	0.29	0.35	0.39	0.34	0.27	0.29	0.23
NL		1.00	0.44	0.44	0.43	0.45	0.23	0.30	0.28	0.30	0.21	0.26	0.30	0.25	0.19	0.31	0.29
ND	1.00	0.56	0.50	0.44	0.42	0.35	0.38	0.41	0.39	0.32	0.38	0.40	0.50	0.42	0.37	0.36	0.40

value 0.2 0.4 0.6 0.8 1.0

Fig. 4 Correlations between the 16 tests of spatial ability and _g_. Starting from the bottom left of the matrix, the first six tests are part of the spatial orientation battery. ND navigation according to directions, NL navigation according to landmarks, MR map reading, RM route memory, PT perspective taking, SC scanning. The following 10 tests were part of the other battery assessing object-based spatial skills: obj CS cross-section, obj 2d 2d drawing, obj PA pattern assembly, obj EM Elithorn Maze, obj MecR mechanical reasoning, obj PF paper folding, obj 3d 3d drawing, obj Rot mental rotation, obj PT perspective taking, obj Maz mazes, _g_ general cognitive ability. All correlations were significant at _p_ < 0.001; variables were residualized for age and sex and standardized prior to analyses.

a more accurate description of the structure of spatial skills. This model provided a good fit ($\chi^2 = 316.000$ (103), $p < 0.001$, CFI = 0.958, TLI = 0.951, RMSEA = 0.037, SMRS = 0.040). However, it also presented one major limitation: due to the substantial difference in test administration and properties of the two batteries, we could not exclude the possibility that the two separate factors emerging from this analysis were a product of differences between the two batteries, rather than underlying a real set of separate, although substantially correlated, abilities. In addition, the two batteries included some cases of parallel measures, so that specific skills were tested in both batteries using different methods (e.g., scanning and perspective taking).

In order to overcome this limitation, we tested another two-factor model, but this time we constructed the two factors based on theoretically driven differences between the constructs. The first factor included all those tests that are described in the literature as tapping spatial orientation abilities (navigation, wayfinding, and map reading) available across the two batteries. This resulted in six tests loading onto a first factor of 'Spatial Orientation': navigation according to directions, navigation according to landmarks, map reading, route memory and two tests originally part of the object-based battery, Elithorne maze and mazes. The second factor of 'Object Manipulation' included the eight remaining tests included in the object-based battery along with the scanning and perspective-taking measures included in the spatial orientation battery (Supplementary Fig. 2). However, this model did not provide a good fit for the data (Supplementary Table 6a).

We then tested whether a bifactor model would present a more accurate reflection of the structure of spatial ability. The bifactor model allowed us to examine how each test of spatial skills loaded onto a general factor of spatial ability, after removing the variance specific to each battery. The two specific factors are likely to include a mixture of true battery-specific underlying abilities and methodological artefacts common to every test within each battery. This bifactor model, presented in Supplementary Fig. 3, provided an adequate yet not excellent fit for the data (Supplementary Table 6a).

The last model we examined was based on the structure of the correlations observed between the 16 spatial tests (Fig. 4), which clustered into three main components. Consequently, this fourth model included three factors representing individual differences in: (1) Object Manipulation, (2) Navigation and (3) Visualization abilities (Fig. 5b). This model provided a good fit for the data ($\chi^2 = 351.870$ (101), $p < 0.001$, CFI = 0.953, TLI = 0.944, RMSEA = 0.041, SMRS = 0.041). However, the three factors were strongly correlated (r ranging from 0.73 to 0.95). These strong correlations are reflective of an underlying common set of abilities across the three factors. Consequently, we re-specified the model as a hierarchically structured model of spatial skills: The 16 tests of spatial skills clustered onto three separate abilities (object manipulation, navigation and visualization), which in turn loaded onto a common factor of Spatial Ability (Fig. 6). AIC indices and Akaike weights[35] indicated that this model provided a better fit as compared with the other one and two-factor models (Supplementary Tables 6a and 6b). This three-factor hierarchical model provided a better fit for the data, when compared with a simpler hierarchical model including two first order factors: Object manipulation and Navigation/Visualization (see Supplementary Fig. 4).

This hierarchical characterization of spatial skills describes the complexity of the structure of individual differences in spatial abilities, while highlighting the strong interconnection between all abilities at a higher level of analysis. The higher order factor of spatial ability accounted for a large portion of individual differences in the navigation ($R^2 = 0.79$), object manipulation ($R^2 = 0.69$) and visualization ($R^2 = 1.00$) factors. We adopted this
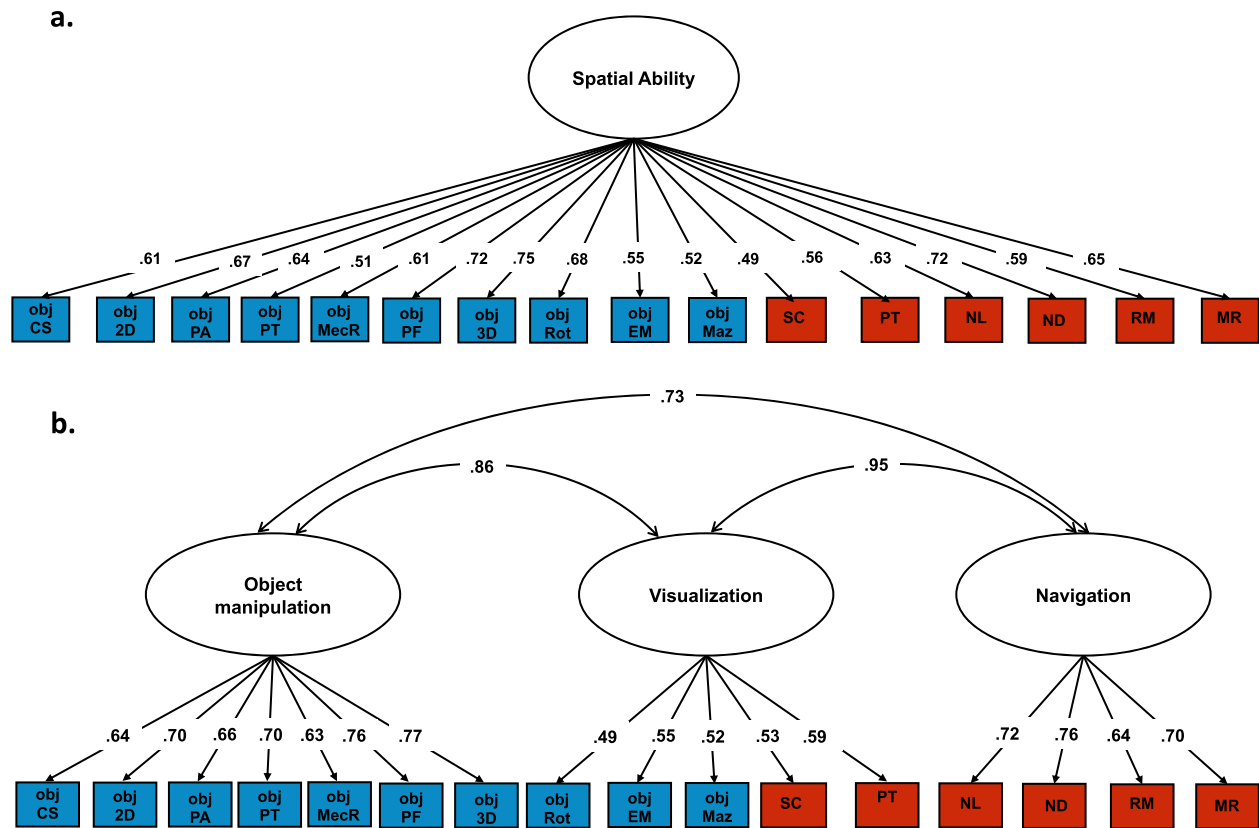
Fig. 5 Factor structure of spatial ability across all 16 tests. a Unifactorial model of spatial ability; **b** Three-factor model of spatial ability. Obj CS cross-section, obj 2D 2D drawing, obj PA pattern assembly, obj SR shapes rotation, obj MecR mechanical reasoning, obj PF paper folding, obj 3D 3D drawing, obj PT perspective taking, obj EM Elithorn Maze, obj Maz Mazes, SC scanning, PT perspective taking, NL navigation according to landmarks, ND navigation according to directions, RM route memory, MR map reading.

hierarchical characterization of individual differences in spatial skills in subsequent analyses.

A common genetic network underlies performance in all spatial tests

We used the multivariate twin method to analyze the genetic and environmental origins of the hierarchical structure of spatial abilities. First, we found that a model decomposing variation in spatial abilities into additive genetic (A) and nonshared environmental (E) factors provided a good fit for the data ($\chi^2 = 1681.128$ (1040), $p < 0.001$, CFI = 0.941, TLI = 0.944, RMSEA = 0.026, SMRS = 0.056). That is, there was no evidence that shared environmental variance, which encompasses those experiences that make children growing up in the same family more similar to one another beyond their genetic similarity, played a meaningful role in accounting for individual differences in spatial skills.

This hierarchical AE model (Fig. 6) showed that spatial skills clustered together largely due to shared genetic variance. The common spatial ability factor was in fact highly heritable (84%) and subsumed 67% of the genetic variance in object manipulation. This is calculated, based on path tracing, as the standardized squared genetic variance in the general factor of spatial ability (0.84) multiplied by twice the path estimate for object manipulation (0.81) divided by the total genetic variance ($0.84 \times 0.81^2 + 0.27$), resulting in $(0.84 \times 0.81^2)/(0.84 \times 0.81^2 + 0.27)$. The common factor of spatial ability accounted for 93% of the genetic variance in the navigation factor and for the entirety of the genetic variance in the visualization factor (see Supplementary Table 7 for the full model including 95% confidence intervals). Nonshared environmental variance accounted for a much smaller proportion

of individual differences in the common spatial ability factor (16%). Nonshared environmental factors, which at the test-specific level include measurement error, were the major source of test-specific variance.

General cognitive ability (g) measured with several tests of verbal and nonverbal ability over development, only partly accounts for the genetic clustering of spatial skills

It is well established that cognitive skills correlate with each other, and that a substantial portion of variation in different abilities can be accounted for by a general factor of cognitive ability (g), both at the observed and genetic level[17,36,37]. We applied a Cholesky decomposition (Method) to examine to what extent the genetic and environmental variance in spatial ability could be captured by g. The Cholesky approach, similar to hierarchical regression, parses the genetic and environmental variation in each trait into that which is accounted for by traits that have previously been entered into the model and the variance which is unique to a newly entered trait. We applied this method to examine the extent to which the clustering of spatial tests into a common factor of spatial ability could be accounted for by the broader g factor. The results presented in Fig. 7 (see Supplementary Fig. 5 for the full model) showed that g accounted for 55% of the genetic variance in the second-order common spatial ability factor. In other words, 45% of the genetic variance in spatial ability was independent of g. As described in greater detail in the Method section, our measure of g was obtained combining multiple tests of verbal and nonverbal cognitive ability administered during development (from age 7 to 16). Tests of spatial ability were not included in our measure of g over development.
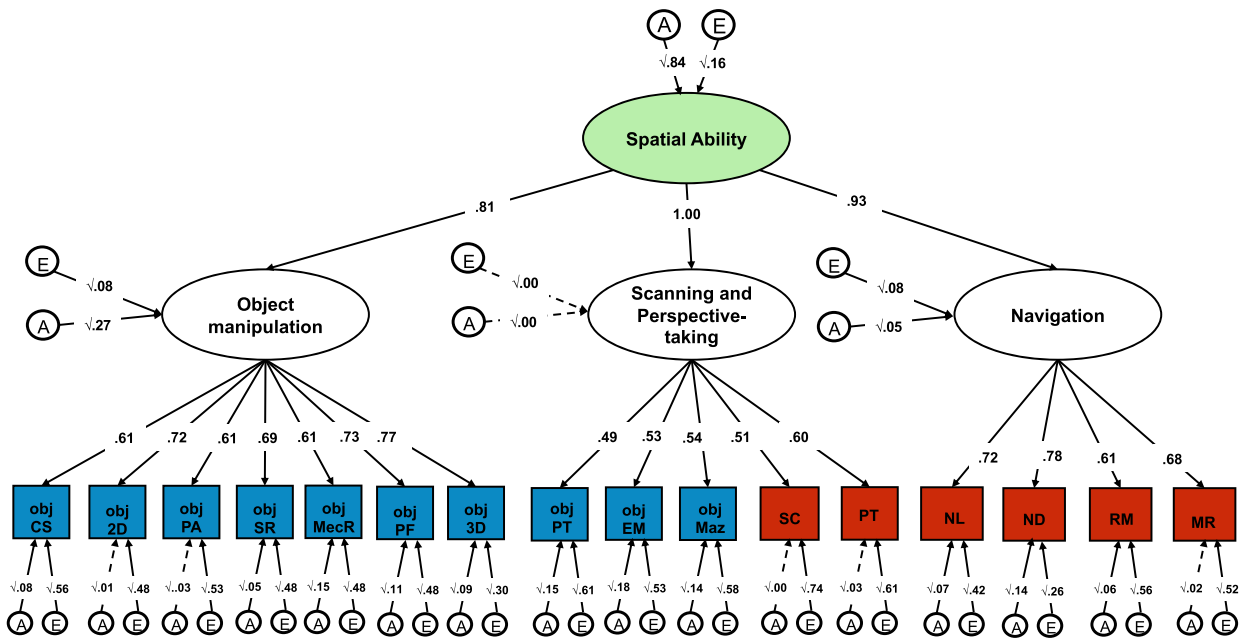
Fig. 6 Genetic and environmental variance characterizing the hierarchical structure of spatial ability. Within each blue rectangle are the ten tests that were included in the object-based spatial battery, while shaded in red are the six tests that are part of the spatial orientation battery set in a naturalistic virtual environment. Obj CS cross-section, obj 2D 2D drawing, obj PA pattern assembly, obj SR shapes rotation, obj MecR mechanical reasoning, obj PF paper folding, obj 3D 3D drawing, obj PT perspective taking, obj EM Elithorn Maze, obj Maz Mazes, SC scanning, PT perspective taking, NL navigation according to landmarks, ND navigation according to directions, RM route memory, MR map reading. Both the phenotypic ($\chi^2 = 351.870$ (101), $p < 0.001$, CFI = 0.953, TLI = 0.944, RMSEA = 0.041, SMRS = 0.041) and genetic ($\chi^2 = 1681.128$ (1040), $p = < 0.001$; CFI = 0.941; TLI = 0.944; RMSEA = 0.026; SRMR = 0.056) model provided good fit for the data.
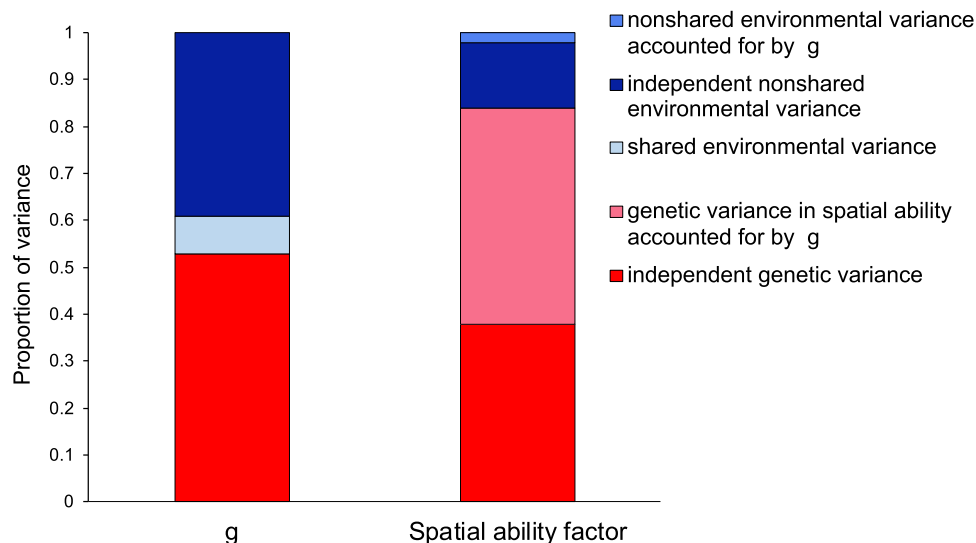
Fig. 7 Genetic and environmental variance in a measure of _g_ over development (age 7–16) and in the common spatial ability factor. For the common spatial ability factor the bar is divided into the genetic and environmental contributions independent of _g_ and those that are accounted for by the genetic and environmental variance in _g_. Results are from a Cholesky decomposition (see Supplementary Fig. 5 for the full model).

When we accounted for _g_ at different levels in the models (Supplementary Figs. 6–11), results remained consistent with the existence of a general genetic network of spatial skills that covaries independently of _g_. Results remained highly consistent when we considered a measure of g constructed from one verbal and one nonverbal test collected when participants were 16-years-old (Supplementary Fig. 12).

DISCUSSION

The current study provides new knowledge on the structure and nature of spatial ability by addressing three outstanding issues in the field of spatial cognition. First, we examined the structure of spatial orientation abilities, measured with a novel gamified battery set in a virtual environment that included a broad range of measures tapping putatively different aspects of spatial

orientation ability. Second, we explored the structure of the associations between spatial orientation skills and object-based spatial tests, a topic that remains mostly unexplored in the cognitive psychology literature and is characterized by strong, contrasting theoretical views[7,15,22]. Third, we investigated the extent to which an index of the developmentally stable component of g accounted for the shared variance observed across spatial skills. Across these three broad aims, we leveraged the genetically informative quality of our twin sample to address parallel questions related to the genetic and environmental structure of spatial ability and of its association with g. At every level of analysis our results highlighted communalities rather than differences across tests of spatial ability, largely supporting a unitary structure of spatial cognition.

Support for the unitary structure of spatial cognition first emerged from phenotypic analyses of our battery of spatial orientation tasks. This finding of a strong general component of variation was remarkable given the breadth of spatial orientation skills covered by our newly developed battery. In fact, the development of this novel gamified, battery set in a virtual environment was guided by a careful process of literature review aimed at covering all the main domains of spatial orientation described in the existing literature. This resulted in six broad domains that ranged from navigation according to directions and large-scale perspective taking, which, based on Newcombe and Shipley's (2015) taxonomy, could be categorized as extrinsic-dynamic spatial abilities, to route memory and large-scale scanning, which, based on the same taxonomy, could be described as extrinsic-static spatial abilities[7]. Although extrinsic-static and extrinsic-dynamic abilities have been proposed to be separate skills[7], and a meta-analysis of the effects of training spatial ability partly supported this distinction for a few selected tests[38], our results contradict this largely theoretical taxonomy.

We found support for a unitary structure of spatial orientation skills not only at an observed (phenotypic) level, but also in terms of the genetic and environmental factors supporting spatial orientation skills. We found that a common factor of 'navigation ability' was 64% heritable and captured between 66 and 100% of the heritability of the six individual tests of spatial orientation, and to a lesser extent their nonshared environmental variance (between 10 and 36%). This suggests that, to the extent that measures of spatial orientation covary, they do so largely due to their shared genetic variance. These results push our knowledge of the nature of spatial orientation skills further, providing support for a unitary structure of spatial orientation skills at the genetic level.

Further support for a unitary structure of spatial cognition emerged when we considered an even greater breadth of spatial tests, including, in addition to our six measures of spatial orientation, ten psychometric tests of object-based spatial skills, administered in the same sample as part of another gamified spatial battery. These sixteen tests of spatial skills were specifically selected to cover all the main areas of spatial cognition identified in extant literature, making the current work, to our knowledge, the most comprehensive investigation of spatial abilities to date. We approached the examination of the structure of associations between such a broad umbrella of spatial measures by moving through increasing levels of complexity.

A simple unitary account of spatial ability, represented by a general factor common to all measures, did not provide an accurate description of the foundations of spatial skills. At first glance, the results could have been interpreted as supporting the existence of three factors of spatial ability. These three factors described individual differences in navigation, object-based abilities and visualization. Existing taxonomies of spatial ability[7], differentiate not only between static and dynamic spatial skills, but also between intrinsic and extrinsic abilities. Consistent with this account we observed a partial differentiation between object-based spatial tests, such as mental rotation, that are largely concerned with the intrinsic properties of objects, and visualization tests, such as perspective taking and scanning, which are largely concerned with extrinsic relations among objects[7,39]. However, the very strong correlations, from 0.73 to 0.95, observed between the object-based, navigation and visualization factors contradicted this putative distinction, and opened the possibility that a coherent, underlying set of abilities held these three factors together.

Factor analytic evidence supported this hierarchical account of spatial cognition: All sixteen tests loaded onto three, substantially correlated, factors (navigation ability, object-based ability and visualization ability), which in turn loaded strongly on a common factor of spatial ability. Particularly striking was the clustering of tests of scanning and perspective taking abilities into a single factor, *Visualization*. In fact, this visualization factor included measures belonging to both spatial ability batteries. This is significant since the tests were administered in very different formats and, according to proposed accounts of spatial skills[19–21], might reflect separate spatial abilities (large vs. small-scale spatial skills). Despite these theoretical and practical differences in task administration, we found that our measures of scanning and perspective taking across the two batteries reflected a common set of visualization abilities.

A hierarchical structure, which highlights both communalities and differences between cognitive tests, has also been found to provide the most accurate characterization in other domains of cognition, most notably executive functions[40–43]. Also consistent with what has been observed for individual differences in executive functions, we found that genetic factors were largely shared across all tests of spatial abilities. These results point to the existence of a common genetic network at the basis of individual differences in spatial ability, therefore providing additional support for a unitary account of spatial cognition.

A further line of evidence supporting the existence of a unitary account of spatial cognition was provided by our analyses examining the role of g in the clustering of spatial ability at the genetic and environmental levels. We found that individual differences in g correlated moderately with all individual tests of spatial skills and substantially with the common spatial ability factor. However, nearly half of the substantial genetic variance in spatial ability was found to be independent of the genetic variance in g, measured aggregating multiple cognitive tests over development. Taken together, our results indicate that spatial skills cluster together phenotypically and genetically beyond the simple fact that they are all tests reflecting a general, developmentally stable, capacity for planning, thinking abstractly and solving problems, all skills that are indexed by g[36]. It should be noted that, since the genetic and environmental components of cognitive abilities have differential longitudinal stabilities, aggregating across waves might have resulted in 'cancelling out' environmental variance that is specific to each developmental stage, in favour of aggregating stable genetic variance in g over development[37].

In summary, our current work provides a threefold line of support for the unitary nature of spatial cognition, partly independent of other measures of cognitive skills. Interestingly, this unitary account of abilities is at odds with individuals' perceptions of their own ability and feelings towards spatial activities. In our previous work examining the structure of spatial and mathematics anxiety, we found evidence for a separation between the anxiety people feel towards spatial navigation and the anxiety towards object-based skills, such as completing difficult jigsaw puzzles and building flat-pack furniture from instructions[44]. This observed difference in perceptions and feelings towards different spatial activities might contribute to explaining why ideas, theories and taxonomies of spatial cognition have mostly favoured a multifaceted account of spatial skills.

Although our study provides a highly comprehensive investigation of the structure of spatial ability in a large sample and addresses several outstanding research questions concerning spatial cognition, it was limited by the technology available to us at the time. Although we developed a new gamified battery set in a virtual environment to reliably examine individual differences in spatial orientation skills, it is possible that assessment in a computer-simulated environment might not be able to capture individual differences in spatial orientation and navigation as well as does assessment in real-life settings. It has been proposed that spatial orientation in computer-simulated environments might reflect an allocentric (object-to-object) approximation of the abilities involved in egocentric (self-to-object) real-life spatial orientation[45]. However, studies that have examined the reliability of measuring navigation skills in VR, as compared with real-life settings, have found good concordance between the two[13]. While we leveraged the newest technological developments to create a realistic virtual environment to host our gamified test, future studies might explore navigation in VR by applying even more immersive tools such as, for example, head-mounted displays (e.g., oculus technology).

Our finding of a unitary structure of spatial cognition across sixteen diverse tests of spatial skills is likely to inform several disciplines beyond cognitive psychology. Investigations on the nature and structure of spatial ability have concerned researchers in a wide range of scientific disciplines, from evolutionary biology to neuroscience, ecology and molecular genetics. Our evidence for a largely unitary phenotypic and genetic network supporting individual differences in spatial cognition can serve as a basis for future research on the nature of spatial ability across all these disciplines and suggests a shift in our consideration of the architecture of human cognitive abilities. These findings are also likely to inform the development of programs aimed at advancing STEM learning through training spatial skills[46].

METHODS

Sample

Participants were part of the Twins Early Development Study (TEDS), a longitudinal study of twins born in the United Kingdom between 1994 and 1996. The families in TEDS are representative of the British population in their socio-economic distribution, ethnicity and parental occupation. See Rimfeld et al. for additional information on the TEDS sample[47,48]. The present study focuses on data collected in a sample of 2,660 TEDS twins aged 19–22 ($M = 21.23$, $SD = 0.53$, range = 2.29). All individuals with major medical, genetic or neurodevelopmental disorders were excluded from the dataset. These included twins with ASD, cerebral palsy, Downs syndrome, chromosomal or single-gene disorders, organic brain problems, e.g. hydrocephalus, profound deafness and developmental delay. TEDS twins completed two online batteries assessing multiple aspects of spatial abilities. One was set in a virtual environment and assessed six aspects of large-scale spatial navigation and orientation skills. 2660 twins (178 pairs were MZ males, 169 pairs were MZ females, 325 pairs were DZ males, 260 pairs were DZ females and 398 pairs were opposite sex) took part in this battery (868 complete twin pairs). 74.3% of participants who completed the spatial orientation battery ($N = 1978$; 740 complete pairs) also completed an online battery of tests developed to assess ten aspects of object-based spatial abilities. At least five months passed between the administration of the two batteries (median time lag = 265.00 days). The time lag was not associated with spatial measures (Supplementary Table 8). The object-based spatial battery was administered first, starting from May 2015, while the data collection for the spatial orientation battery started in September 2015. The Institute of Psychiatry, Psychology and Neuroscience ethics committee at King's College London approved the study. Informed consent was obtained from all participants prior to data collection.

Measures

Spatial orientation battery. Putatively different facets of spatial orientation skills were assessed through a novel gamified battery called 'Spatial Spy'.

Participants were invited to solve a mystery by collecting clues while orienting and navigating around the streets of a virtual environment (Fig. 8). The online battery was developed in Unity (https://unity3d.com) by ETT Solutions. After a comprehensive literature review, we identified four core aspects of spatial orientation and navigation skills: (1) navigating when reading a map; (2) navigating based on a previously memorized map or route; (3) navigating following directions (e.g., cardinal points), and (4) navigating using reference landmarks. In addition to these four abilities, the spatial orientation battery included two tests based on paradigms that have been frequently used in the object manipulation spatial literature: perspective-taking and scanning. Two research aims motivated the decision to include these two tests in the battery. First, we aimed to explore how perspective taking and scanning measured within a large-scale spatial framework (i.e., within a more naturalistic context approximating a virtual environment) related to the same abilities assessed within a smaller-scale, object manipulation framework (i.e., psychometric tests collected as part of another online battery). Second, due to the innovative and experimental nature of the spatial orientation battery, we included measures of scanning and perspective taking, for which we had corresponding data from more traditional psychometric tests, in order to explore the external validity of assessing spatial skills within this new virtual environment. The measures included in this spatial orientation battery are described in detail below. The statistical properties (distribution characteristics and test–retest reliability) of each measure were assessed through two pilot studies.

The final battery started with a training session that helped participants become acquainted with using the cursor or mouse for navigating around the virtual environment, as well as with the requirements and mechanics of each of the six tests. The battery was administered online, with participants taking the tests in web browsers on their own desktop or laptop computers, using a mouse or trackpad to 'look' around the virtual environment, and the keyboard to move. The battery took between 35 and 60 min (median time 43 min) to complete and participants could pause at any time by pressing the key 'P' on their keyboard and could resume the game at any given time. Prior to the testing session, participants were provided with practice trials for every test. A two-minute video providing examples of how each subtest was implemented within the Spatial Spy virtual environment is available at the following link https://www.youtube.com/watch?v=wHj0–19rbil. Following is a detailed description of each test included in the spatial orientation battery. Test–retest reliability for each measure was calculated as part of a pilot study including a sample of 100 participants who completed the battery twice over the space of two months.

Map Reading (Fig. 8a), assessed individual differences in the ability to efficiently read a map to travel from one location to another. Once a map had appeared on the top-right corner of the screen, a flashing yellow dot on the map indicated participants' starting location (A), while a red pointer designated the end-point location on the map (B). Participants were instructed to get from A to B by finding the fastest route and notified that they had 1 min to complete their mission. If participants could not reach their destination within 60 s, they were 'teleported' back to the initial location and allowed a second opportunity to complete the task. The ability was assessed though five non-consecutive iterations of increasing difficulty. Each iteration was allocated a score of 2 if participants had successfully travelled from A to B through the quickest (most direct) route, a score of 1 if participants had successful completed the mission but had not selected the fastest route, and a score of 0 if participants had failed to complete the mission. This created a final maximum score of 10. The final score was calculated by combining this accuracy score with participants' reaction time (time taken to successfully complete the mission), equally weighted. The test showed good test–retest reliability ($r = 0.69$, $p < 0.001$) and distribution (Fig. 1).

Route Memory (Fig. 8b), assessed individual differences in the ability to travel from one location to another by remembering the content of a map. As for the map reading condition, a map appeared on the top-right corner of the screen, with a flashing yellow dot indicating participant's starting location (A), and a red pointer designating the end-point location (B). However, the route memorizing test asked participants to memorize the content of the map before the map disappeared from the screen. Participants were given 20 s to memorize the map and plan the route before travelling from A to B and were allowed 120 s to reach the target location. The number of increasingly difficult iterations, procedure and scoring were the same as those for the previously described map-reading without memory task. Test–retest reliability was acceptable ($r = 0.60$, $p < 0.001$), and distribution (Fig. 1).

Fig. 8 The virtual city where the spatial orientation battery takes place and examples of the six tasks included. a Map Reading; **b** route memory; **c** navigation directions; **d** navigation landmarks; **e** scanning; **f** perspective taking.

Navigation Directions (Fig. 8c) assessed participants' skills in navigating around a virtual environment following instructions based on directions. At the start of the task, participants were 'teleported' to one location of the virtual environment and given instructions to navigate around the virtual city in terms of compass points (north, south, east and west). The test included five non-consecutive iterations of increasing difficulty and each iteration comprised 4–6 tasks. Each task that was solved correctly was assigned a score of 1. Participants were allowed a maximum of three attempts to respond correctly to each task and consequently proceed to the next set of instructions. After three consecutive failed attempts, the iteration was discontinued and the remaining tasks in that iteration (if any) were assigned a score of 0. Each iteration had a time limit of 180 s, if the time limit expired before participants had completed all the tasks, the remaining tasks for that iteration were discontinued and assigned a score of 0. There was no progress bar or timer on screen to help participants keep track of time; however, "hurry up" prompts appeared on screen as the time limit approached. At the end of each iteration (either successfully completed or discontinued) participants were teleported to another part of the virtual environment to complete the subsequent iteration. For the first two iterations the image of a compass providing cardinal directions was available on the top-left corner of the screen, but the compass was not available for the last three iterations, making them more difficult to complete. Examples of instructions were: 'Now turn east' and 'You are

facing southwest. Go north and immediately turn west'. The final score was calculated by combining the accuracy score with participants' reaction time (time taken to successfully complete each iteration), equally weighted. The test showed excellent test–retest reliability ($r = 0.89$, $p < 0.001$) and distribution of the scores (Fig. 1).

Navigation Landmarks (Fig. 8d) measured the ability to navigate following instructions based on the descriptive features of the destination or other nearby landmarks. The test included five non-consecutive iterations each comprising 4 or 5 tasks. Each task lasted for a maximum of 60 s, so participants had 60 s to reach a certain landmark within the virtual environment. If the time limit expired before participants had reached the required landmark, they were discontinued, teleported to the landmark in question, and were able to proceed to the next task. Each task solved correctly, meaning that participants were able to reach the described landmark within the time limit, was assigned a score of 1, while for each trial when participants were not able to reach the location in 60 s, they were assigned a score of 0. Neither a map nor a compass was provided to help participants navigate around the environment. Examples of instructions are: 'Now reach the tall white pyramid skyscraper', and 'The message instructs you to go to the park near the old clock tower'. The target landmark was visible at the start of the session, but it was not always in plain sight as participants were navigating throughout the city to reach the target landmark. The final score was calculated by combining this accuracy

score with participants' reaction time (time taken to successfully complete each iteration), equally weighted. The test showed excellent test–retest reliability ($r = 0.80$, $p < 0.001$) and distribution of the scores (Fig. 3a).

Large-scale Scanning (Fig. 8e) measured participants' ability to quickly process visual information and identify a target object, a black briefcase, located somewhere nearby within the virtual city. The target object remained the same across the five non-consecutive iterations of increasing difficulty. When looking for the target, participants' perspective could be rotated freely in any direction, but could not be moved vertically or horizontally around the virtual environment. Participants could identify the target object by clicking on the mouse or trackpad within 60s. Within the time limit, participants were allowed four attempts to correctly spot the target object and, as for all other tasks, they were encouraged to do it as quickly as possible. Feedback was provided after each attempt, and as soon as participants had identified the target object correctly, they were 'teleported' to the next task. It was not possible to pause half-way through the 60-s iteration. The final score was calculated by combining this accuracy score with participants' reaction time (time taken to successfully complete each iteration). The test showed excellent test–retest reliability ($r = 0.80$, $p < 0.001$) and wide distribution of the scores (Fig. 1).

Large-scale Perspective-taking (Fig. 8f) measured participants' ability to identify objects from a different perspective in large-scale 'naturalistic' settings. The test comprised five iterations of increasing difficulty that followed the same test rules. Each iteration started with a CCTV-like image showing an aerial shot of a location within the virtual world, and within this location one target object was depicted flashing on screen for ten seconds. During this initial stimulus presentation, participants could not move within the virtual environment, so all participants were exposed to the same image of the flashing target object. After the ten seconds had elapsed, the CCTV image disappeared and participants were teleported back to the target location within the virtual environment, which shifted their perspective back to ground level; they were then instructed to identify the target object as quickly as possible. When looking for the target object (the one that was flashing when presented from the CCTV perspective), participants' perspective could be freely rotated but could not be moved vertically or horizontally around the virtual environment. Participants could identify the target object by clicking on it with their mouse or trackpad within 60 s. Within the time limit, participants were allowed four attempts to correctly spot the target object and they were encouraged to do it as quickly as possible. A message would appear on the screen after each attempt (either 'Yes' or 'Try again') to provide participants with feedback on their performance, and each iteration terminated either after a successful attempt, or after participants had used up their four attempts, or if they timed out. A 'Hurry up' message was displayed on the screen a few seconds before the time for each iteration elapsed. The test showed good distribution (Fig. 1) and acceptable test–retest reliability ($r = 0.67$, $p < 0.001$).

Object manipulation. Object manipulation was tested using an online, gamified, battery called 'The King's Challenge'[18]. This test battery measures the major putative dimensions of spatial ability, and is comprised of 10 tests: (1) a mazes task (searching for a way through a 2D maze in a speeded task); (2) 2D drawing (sketching a 2D layout of a 3D object from a specified viewpoint); (3) Elithorn mazes (joining together as many dots as possible from an array); (4) pattern assembly (visually combining pieces of objects together to make a whole); (5) mechanical reasoning (multiple-choice naïve physics questions); (6) paper folding (visualizing where the holes are situated after a piece of paper is folded and a hole is punched through it); (7) 3D drawing (sketching a 3D drawing from a 2D diagram); (8) mental rotation (mentally rotating objects); (9) perspective-taking (visualizing objects from a different perspective), and (10) cross-sections (visualizing cross-sections of objects). The development of the battery is described in detail elsewhere[18]. A brief demonstration of the battery can be accessed here: https://www.youtube.com/watch?v=awnfeiAPmQc

General cognitive ability (g) over development. General cognitive ability (g; intelligence) was assessed in TEDS at ages 7, 9, 10, 12, 14, and 16. For the present analyses we created a longitudinal composite measure of g as a mean of these six assessments. At age 7, 'g' was calculated as a mean of conceptual grouping[49], a WISC similarities test[50], a WISC vocabulary test[50], and a WISC picture completion test[50] all collected over telephone testing. At age 9, 'g' was calculated as a mean of a shapes test[51], a WISC vocabulary test[52], a WISC general knowledge task[52], and a puzzle test[51]; all tests were administered in booklets sent to the twins by post. At age 10, 'g' was calculated as a mean of the Raven's Standard Progressive Matrices[53], a

WISC vocabulary[52], WISC picture completion[50], and a WISC general knowledge test[52]; at age 10 and subsequent assessments, all 'g' data were obtained by internet testing. At age 12, 'g' was calculated as a mean of the Raven's Progressive Matrices[53], a WISC picture completion[50], a WISC vocabulary[52], and a WISC general knowledge test[52]. At age 14, 'g' was calculated as a mean of the Raven's Progressive Matrices[53] and a WISC vocabulary test[52]. At age 16, 'g' was calculated as a mean of Mill Hill Vocabulary test[54] and Raven's Progressive Matrices[53].

Analytic strategies
The R package 'psych'[55] was used to obtain descriptive statistics and correlations, and the R package 'ggplot2'[56] was used for data visualization purposes. For all phenotypic analyses, including all model fitting, one twin was selected randomly from each pair to ensure independence of data. The random selection of one twin out of each pair also provided the opportunity of replicating all phenotypic analyses in the other, randomly selected, half of the sample, although we acknowledge that the replication sample is not independent (due to relatedness) and does not provide a full replication. Similar results were obtained when the analyses were conducted on the second half of the sample (see Supplementary Figs. 13, 14, and 15). Structural Equation modelling (SEM) was conducted in Mplus version 8[57] and Open Mx version 2.0[58]. Full Information Maximum Likelihood (FIML) was applied to account for missingness in the data.

Confirmatory factor analyses. Confirmatory Factor Analysis (CFA) is a data reduction technique whereby latent factors are constructed from observed (measured) indicators based on a pre-imposed structure which is hypothesized to underlie the data. CFA is, in most instances, theory-driven and allows for testing a hypothesis on the associations between variables and their underlying latent constructs. Alternative theoretical models were compared examining multiple model fit indices. Model fit indices include (a) the chi-square test, which indicates the correspondence between the expected and the observed covariance matrices, a chi-square value close to zero indicates greater correspondence between them; (b) the Akaike information criterion that allows us to compare model fit between competing models, with smaller values indicating better fit; (c) Akaike weights[35] which provide an index of the conditional probability for each model, Akaike weights range between 0 and 1 and a larger value indicates better fit; (d) the CFI is an incremental fit index that is based on the non-centrality measure. The CFI ranges from 0 to 1.00, with values closer to 1.00 indicating better fit (acceptable values > 0.90); (e) the RMSEA is related to residual in the model. RMSEA values range from to 1 with a smaller RMSEA value indicating better model fit. Acceptable model fit is indicated by an RMSEA value of 0.08 or less[59].

Genetic analyses: univariate and multivariate twin modelling. The twin method allows for the decomposition of individual differences in a trait into genetic and environmental sources of variance by capitalizing on the genetic relatedness between monozygotic twins (MZ), who share 100% of their genetic makeup, and dizygotic twins (DZ), who share on average 50% of the genes that differ between individuals. The method is further grounded in the assumption that both types of twins who are raised in the same family share their rearing environments to approximately the same extent[60]. By comparing how similar MZ and DZ twins are for a given trait (intraclass correlations), it is possible to estimate the relative contribution of genetic factors and environments to variation in that trait. Heritability, the amount of variance in a trait that can be attributed to genetic variance (A), can be roughly estimated as double the difference between the MZ and DZ twin intraclass correlations[61]. The ACE model further partitions the variance into shared environment (C), which describes the extent to which twins raised in the same family resemble each other beyond their shared genetic variance, and nonshared environment (E), which describes environmental variance that does not contribute to similarities between twin pairs (and also includes measurement error). It also provides confidence intervals for all estimates.

When data are available from opposite-sex and same-sex DZ twin pairs, the standard univariate ACE model can be extended to a sex-limitation model to test for the differences in the aetiologies of sex differences by comparing five sex and zygosity groups: MZ females, DZ females, MZ males, DZ females, and DZ opposite-sex twin pairs[34]. This method allows for estimating quantitative and qualitative sex differences (i.e., the same factors affecting males and females to a different extent). Differences in the magnitude of ACE estimates for males and females are referred to as quantitative sex differences; qualitative sex differences indicate whether

different genetic or environmental factors influence males and females. The sex limitation model is described in detail elsewhere[62]. Here we conducted sex-limitation model-fitting by fitting a series of nested models and then testing the relative drop of the fit between the models when the parameters for the sexes are forced to be equal[58].

The twin method can also be extended to the exploration of the covariance between two or more traits (multivariate genetic analysis). Multivariate genetic analysis allows for the decomposition of the covariance between multiple traits into genetic and environmental sources of variance, by modelling the cross-twin cross-trait covariances. Cross-twin cross-trait covariances describe the association between two variables, with twin 1's score on variable 1 correlated with twin 2's score on variable 2, which are calculated separately for MZ and DZ twins. The examination of shared variance between traits can be further extended to test the aetiology of the variance that is common between traits and of the residual variance that is specific to individual traits. Here we used the common pathway model which is a multivariate genetic model in which the variance common to all measures included in the analysis can be reduced to a common latent factor, for which the A, C, and E components are estimated. As well as estimating the aetiology of the common latent factor, the model allows for the estimation of the A, C, and E components of the residual variance in each measure that is not captured by the latent construct[63]. The common pathway model estimates the extent to which the general factor of spatial ability is explained by A, C, and E. The common pathway model is illustrated in Fig. 3. Based on factor analytic evidence, the common pathway model can be extended to include multiple common factors and, consequently to the examination of the genetic and environmental associations between the multiple latent factors. This extension of the common pathway model is presented in Fig. 6.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY
The Twins Early Development Study's data access policy can be found at the following link https://www.teds.ac.uk/researchers/teds-data-access-policy.

CODE AVAILABILITY
The corresponding author will share the code upon request.

REFERENCES
1. Morand-Ferron, J., Cole, E. F. & Quinn, J. L. Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biol. Rev.* **91**, 367–389 (2016).
2. Lubinski, D. Spatial ability and STEM: a sleeping giant for talent identification and development. *Pers. Individ. Dif.* **49**, 344–351 (2010).
3. Newcombe, N. Harnessing spatial thinking to support STEM learning. *OECD Educ. Work. Pap.* https://doi.org/10.1787/7d5dcae6-en (2017).
4. Webb, R. M., Lubinski, D. & Benbow, C. P. Spatial ability: a neglected dimension in talent searches for intellectually precocious youth. *J. Educ. Psychol.* **99**, 397–420 (2007).
5. Wai, J., Lubinski, D. & Benbow, C. P. Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* **101**, 817–835 (2009).
6. Kell, H. J., Lubinski, D., Benbow, C. P. & Steiger, J. H. Creativity and technical innovation. *Psychol. Sci.* **24**, 1831–1836 (2013).
7. Newcombe, N. S. & Shipley, T. F. Thinking about spatial thinking: new typology, new assessments. In *Studying visual and spatial reasoning for design creativity* (ed. Gero, J. S.). 179–192 (Springer, New York, NY, 2015).
8. Galton, F. Generic images. *Ninet. Century* **6**, 157–169 (1879).
9. Buckley, J., Seery, N. & Canty, D. A heuristic framework of spatial ability: a review and synthesis of spatial factor literature to support its translation into STEM education. *Educ. Psychol. Rev.* **30**, 947–972 (2018).
10. Lohman, D. In *Human Abilities: Their Nature and Measurement* (eds. Dennis, I. & Tapsfield, P.) 97–116 (Lawrence Erlbaum, 1996).
11. Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects abstract. The time required to recognize that two perspective drawings portray. *Science* **171**, 701–703 (1971).
12. Linn, M. C. & Petersen, A. C. Emergence and characterization of sex differences in spatial ability: a meta-analysis author (s): Marcia C. Linn and Anne C. Petersen Published by: Wiley on behalf of the Society for Research in Child Development Stable. *Child Dev.* **56**, 1479–1498 (1985).
13. Coutrot, A. et al. Virtual navigation tested on a mobile app (Sea Hero Quest) is predictive of real-world navigation performance: preliminary data. *bioRxiv* 1–10 (2018).
14. Newcombe, N. S. Finding our way: Book Review. *Curr. Biol.* **29**, R108–R109 (2019).
15. Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T. & Lovelace, K. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* **34**, 151–176 (2006).
16. Mackintosh, N. & Mackintosh, N. *IQ and Human Intelligence.* (Oxford University Press, UK, 2011).
17. Carroll, J. B. (John B. *Human Cognitive Abilities: A Survey of Factor-analytic Studies.* (Cambridge University Press, 1993).
18. Rimfeld, K. et al. Phenotypic and genetic evidence for a unifactorial structure of spatial abilities. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.1607883114 (2017).
19. Ittelson, W. H. *Environment and Cognition.* (Seminar Press, 1973).
20. Cutting, J. & Vishton, P. In *Handbook of Perception and Cognition (2nd edn.). Perception of Space and Motion* (eds. Epstein, W. & Rogers, S. J.) 69–117 (Academic Press, 1995).
21. Silverman, I. & Eals, M. In *He Adapted Mind: Evolutionary Psychology and the Generation of Culture* (eds. Barkow, J. H., Cosmides, L. & Tooby, J.) 533–549 (Oxford University Press, 1992).
22. Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G. & Beck, S. Predicting environmental learning from spatial abilities: an indirect route. *Intelligence* **22**, 327–355 (1996).
23. Tversky, B., Bauer Morrison, J., Franklin, N. & Bryant, D. J. Three spaces of spatial cognition. *Prof. Geogr.* **51**, 516–524 (1999).
24. Zacks, J. M., Mires, J. O. N., Tversky, B. & Hazeltine, E. Mental spatial transformations of objects and perspective DISCUSSION (INTRODUCTION). *Spat. Cogn. Comput.* **2**, 315–332 (2002).
25. Hegarty, M. & Waller, D. A. In *The Cambridge Handbook of Visuospatial Thinking* (eds. Shah, P. & Miyake, A.) 121–169 (Cambridge University Press, 2005).
26. Kosslyn, S. M. & Thompson, W. L. When is early visual cortex activated during visual mental imagery? *Psychol. Bull.* **129**, 723–746 (2003).
27. Shelton, A. L. & Yamamoto, N. Visual memory, spatial representation, and navigation in. In Current issues in memory series. The visual world in memory (ed. Brockmole, J. R.). 140–177 (Psychology Press, 2009).
28. Gaulin, S. J. C. In *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 1211–1225 (The MIT Press, 1995).
29. Jones, C. M., Braithwaite, V. A. & Healy, S. D. The evolution of sex differences in spatial ability. *Behav. Neurosci.* **117**, 403–11 (2003).
30. Spearman, C. 'General Intelligence,' Objectively Determined and Measured Author (s): C. Spearman Source: The American Journal of Psychology, Vol. 15, No. 2 (Apr., 1904), pp. 201-292 Published by: University of Illinois Press Stable URL: http://www.jsto. *Am. J. Psychol.* **15**, 201–292 (1904).
31. Deary, I. J. Intelligence. *Curr. Biol.* **23**, R673–R676 (2013).
32. Toivainen, T. et al. Prenatal testosterone does not explain sex differences in spatial ability. *Sci. Rep.* **8**, 1–8 (2018).
33. Newcombe, N., Bandura, M. M. & Taylor, D. G. Sex differences in spatial ability and spatial activities. *Sex. Roles* **9**, 377–386 (1983).
34. Knopik, V. S., Neiderhiser, J. M., Defries, J. C. & Plomin, R. *Behavioral Genetics.* (Macmillan Higher Education, 2016).
35. Wagenmakers, E. J. & Farrell, S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11**, 192–196 (2004).
36. Plomin, R. & Deary, I. J. Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).
37. Tucker-Drob, E. M. & Briley, D. A. Continuity of genetic and environmental influences on cognition across the life span: a meta-analysis of longitudinal twin and adoption studies. *Psychol. Bull.* **140**, 949–979 (2014).
38. Uttal, D. H. et al. The malleability of spatial skills: a meta-analysis of training studies. *Psychol. Bull.* **139**, 352–402 (2013).
39. Kozhevnikov, M. & Hegarty, M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cogn.* **29**, 745–756 (2001).
40. Friedman, N. P. et al. Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol.* **137**, 201–225 (2008).
41. Engelhardt, L. E. et al. Strong genetic overlap between executive functions and intelligence. *J. Exp. Psychol. Gen.* **145**, 1141–1159 (2016).
42. Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P. & Tucker-drob, E. M. "Same But Different": Associations Between Multiple Aspects of Self-Regulation,

Cognition and Academic Abilities. *J. Pers. Soc. Psychol.* **117**, 1164–1188 (2019).

43. Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P. & Tucker-Drob, E. M. Genes unite executive functions in childhood. *Psychol. Sci.* https://doi.org/10.1177/0956797615577209 (2015).

44. Malanchini, M. et al. The genetic and environmental aetiology of spatial, mathematics and general anxiety. *Sci. Rep.* **7**, 42218 (2017).

45. Tu, S., Spiers, H. J., Hodges, J. R., Piguet, O. & Hornberger, M. Egocentric versus allocentric spatial memory in behavioral variant frontotemporal dementia and Alzheimer's disease. *J. Alzheimer's Dis.* **59**, 883–892 (2017).

46. Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S. & Ratliff, K. Sex differences in spatial cognition: advancing the conversation. *Wiley Interdiscip. Rev. Cogn. Sci.* **7**, 127–155 (2016).

47. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res. Hum. Genet.* **16**, 117–125 (2013).

48. Rimfeld, K. et al. Twins Early Development Study: a genetically sensitive investigation into behavioral and cognitive development from infancy to emerging adulthood. *Twin Res. Hum. Genet.* https://doi.org/10.1017/thg.2019.56 (2019).

49. McCarthy, D. *McCarthy Scales of Children's Abilities.* (The Psychological Corporation, New York, 1972).

50. Wechsler, D. *Wechsler Intelligence Scale for Children (3rd Edn. UK).* (The Psychological Corporation, 1992).

51. Smith, P., Fernandes, C. & Strand, S. *Cognitive Abilities Test 3 (CAT3).* (nferNELSON, Windsor, 2001).

52. Kaplan, E., Fein, D., Kramer, J., Delis, D. & Morris, R. WISC-III as a process instrument (WISC-III-PI) (1999).

53. Raven, J., Raven, J. C. & Court, J. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Raven Manual* (Oxford University Press, Oxford, 1996).

54. Raven, J. C., Raven, J. & Court, J. *Mill Hill Vocabulary Scale.* (OOP, Oxford, 1998).

55. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 1.9.12, https://CRAN.R-project.org/package=psych (2019).

56. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* https://ggplot2.tidyverse.org (Springer-Verlag, New York, 2016).

57. Muthén, L. K. & Muthén, B. O. *Mplus User's Guide.* 8th edn (Muthén & Muthén, Los Angeles, CA, 2017).

58. Boker, S. et al. OpenMx: an open source extended structural equation modeling framework. *Psychometrika* **76**, 306–317 (2011).

59. Hu, L. T. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* **6**, 1–55 (1999).

60. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. A twin study of recent life events and difficulties. *Arch. Gen. Psychiatry* **50**, 789–796 (1993).

61. Martin, N. G. & Eaves, L. J. Stages; the first to determine the genetical and environmental model. *Most* **38**, 79–95 (1977).

62. Medland, S. E. Alternate parameterization for scalar and non-scalar sex-limitation models in Mx. *Twin Res.* **7**, 299–305 (2004).

63. Rijsdijk, F. V. In *Encyclopedia of Statistics in Behavioral Science* (eds. Everitt, B. S. & Howell, D. C.) 330–331 (John Wiley & Sons Ltd., 2005).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceived and designed the experiment: M.M., K.R., N.G.S., K.L.S., M.R., Y.K., V.R., R.P. Analyzed the data: M.M., K.R., N.G.S. Drafted the paper: M.M., K.R., R.P. Revised the work critically: M.M., K.R., N.G.S., K.L.S., Y.K., V.R., P.S.D., E.T.D., R.P. All authors approved the final draft. M.M. and K.R. contributed equally to the work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41539-020-0067-8.

Correspondence and requests for materials should be addressed to M.M.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.