nature human behaviour

ARTICLES
https://doi.org/10.1038/s41562-020-00935-3

Inferring latent learning factors in large-scale cognitive training data

Mark Steyvers [1] and Robert J. Schafer [2]

The flexibility to learn diverse tasks is a hallmark of human cognition. To improve our understanding of individual differences and dynamics of learning across tasks, we analyse the latent structure of learning trajectories from 36,297 individuals as they learned 51 different tasks on the Lumosity online cognitive training platform. Through a data-driven modelling approach using probabilistic dimensionality reduction, we investigate covariation across learning trajectories with few assumptions about learning curve form or relationships between tasks. Modelling results show substantial covariation across tasks, such that an entirely unobserved learning trajectory can be predicted by observing trajectories on other tasks. The latent learning factors from the model include a general ability factor that is expressed mostly at later stages of practice and additional task-specific factors that carry information capable of accounting for manually defined task features and task domains such as attention, spatial processing, language and math.

An enduring challenge in psychology is to understand individual differences in intellectual abilities. Early theories of intelligence have explained performance on cognitive tests in terms of broad factors such as general intelligence shared among all cognitive tests as well as domain-specific factors[1,2]. Cognitive theories have been developed to explain covariation across tests of mental ability in terms of shared cognitive resources across tasks such as working memory and executive processes[3,4], the speed of information processing at various levels of processing[5,6], associative learning[7], as well as developmental models involving mutually interacting cognitive processes[8]. Recent research has started to examine neurobiological mechanisms that influence individual differences in mental ability[9].

In this research, the goal is to understand individual differences in performance on cognitive tasks when individuals practice over an extended period of time, and to understand how learning trajectories covary across tasks. To what degree is the learning trajectory in one task predictive of the learning trajectory in another task? Can we predict how well an individual will perform after extended practice on a task by observing how well this individual performs at the start of practice on the same task, or on others? We scale up the investigation of individual differences in learning trajectories by analysing large-scale data from Lumosity, an online cognitive training platform. On this platform, users perform a variety of gamified cognitive tasks that assess a wide range of cognitive abilities in areas such as memory, attention and reasoning. Some of the games on the platform correspond directly to well-known cognitive paradigms (for example, the Erikson flanker task and n-back tasks) whereas others involve complex planning and visualization tasks not typically studied in the laboratory.

The Lumosity platform offers unique advantages to study individual differences. First, the platform has a very large and diverse user base. As of 2018, over 90 million users from 182 countries signed up to participate. We analyse a sample of the Lumosity data that includes the performance of 36,297 users on 51 different cognitive games (see Methods). Unlike most laboratory studies of cognition, the users in our sample are associated with a diverse range of demographic characteristics in terms of age, gender and educational background. Second, users practice the cognitive tasks multiple times and performance on tasks typically improves with practice[10–12]. Therefore, the data are not restricted to a single snapshot of performance across tasks but instead offer a dynamic view of performance over time. In our data sample, we will analyse performance on the platform during a 6-yr period (the average user in this set spends about 2.5 yr on the platform during this time span).

The aim of this work is to analyse the latent structure of the learning trajectories across cognitive tasks. We apply a dimensionality reduction approach on the basis of a Bayesian principal component analysis (PCA)[13,14]. This approach has proved useful for predictive modelling of human preferences in large-scale datasets[15,16]. The particular Bayesian PCA approach adopted here models time in a discrete fashion and makes no assumption about the functional form of the learning curve. There are several other approaches that could be applied to the data, including continuous-time models[17] and various forms of latent growth models to learn the covariance structure of latent parameters governing the changes over time (for example, refs. [18–22]).

We use a predictive evaluation approach to assess the model's ability to account for covariation across learning trajectories and forecast future performance[23,24]. There are two types of prediction problems we use for model evaluation, as illustrated in Fig. 1. In the first prediction problem (Fig. 1a), the goal is to predict an entire learning trajectory for a particular user and a single cognitive task on the basis of the learning trajectories from other cognitive tasks the user engaged with. Because the to-be-predicted learning curve is missing entirely, the prediction has to be based on the latent structure that captures covariation across tasks as well as covariation across different stages of practice. The challenge in this prediction task goes beyond extrapolating learning curves[11,23] or predicting future errors on the same learning task[25]. In the second prediction problem (Fig. 1b,c), the goal is to predict the user performance on a specific task at a later stage of practice based on observations of learning trajectories on other tasks, as well as partially observed performance at the initial stages of practice on the target task.

[1]Department of Cognitive Sciences, University of California, Irvine, CA, USA. [2]Lumos Labs, San Francisco, CA, USA. e-mail: mark.steyvers@uci.edu; bschafer@lumoslabs.com

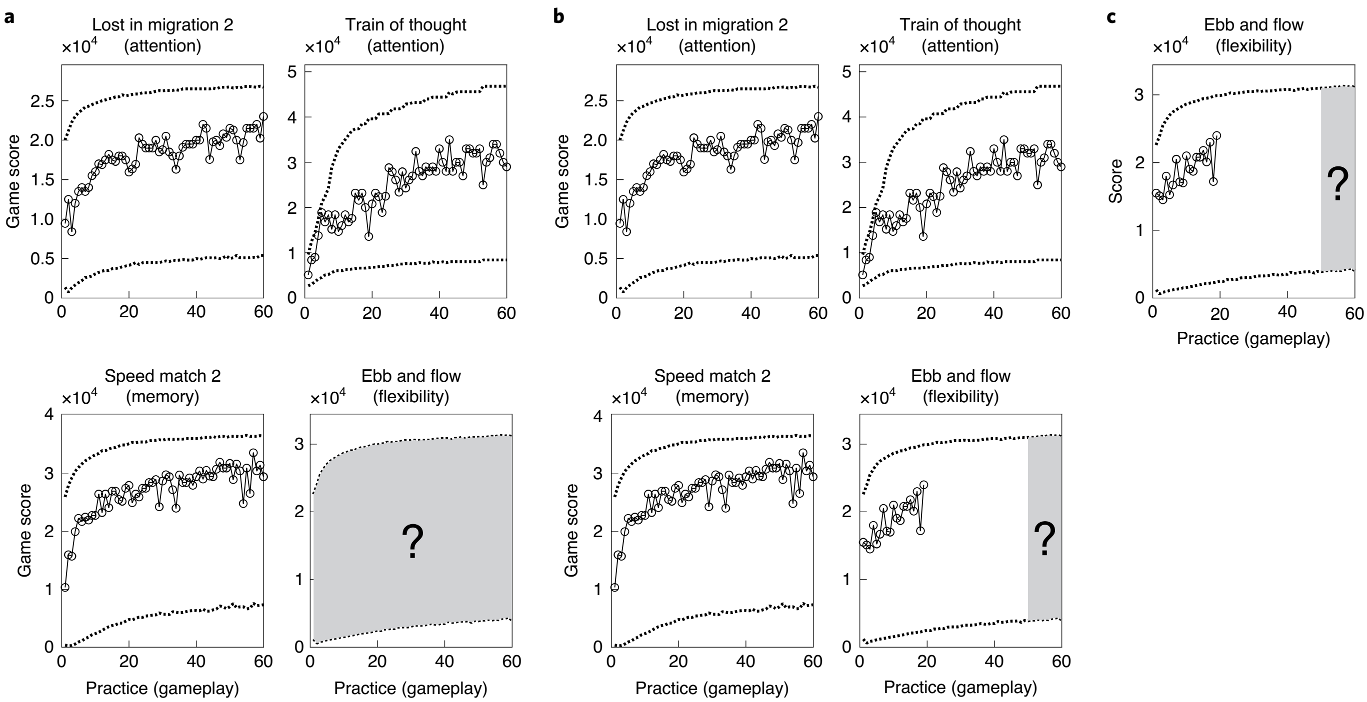


Fig. 1 | Illustration of different learning curve prediction problems for one user. a, For one task (Ebb and flow), the entire learning curve is missing and needs to be predicted from the user's performance on three other tasks (Lost in migration, Train of thought and Speed match). **b**, The performance on 'Ebb and flow' is partially observed and the goal is to predict the performance at the last stage of practice from the early performance on that task as well as the performance on other tasks. **c**, The performance on 'Ebb and flow' is partially observed and the goal is to predict the performance at the last stage of practice without any knowledge of performance on other tasks. Dotted lines show the 95% range in scores for each stage of practice.

By systematically varying the amount of observed information about early task performance and number of other tasks observed, we can assess the predictive utility of different sources of information. Overall, these prediction problems are related to real-world situations involving decisions to invest in costly training of a particular individual on the basis of prior performance on the same, or other, tasks.

We also examine whether the latent learning factors correspond to interpretable dimensions of performance. To avoid relying on subjective judgement when assessing the interpretability of the latent factors, we use a predictive approach to evaluate whether the latent structure discovered by the model can be mapped to known types of cognitive activities associated with a task, as well as the types of stimuli used in the task. For this purpose, we characterized each Lumosity game in terms of a number of predefined task features related to the particular cognitive task (for example, memory, selective attention and logical reasoning) as well as stimulus features (for example, words, single digits and objects). These task features are not mutually exclusive—a particular task can be associated with a number of cognitive task features and involve multiple types of stimulus features. If the latent structure discovered by the model carries information about mental activities involved in each task, a model-based mapping between the latent factors and predefined task features should lead to accurate predictions of the types of cognitive activities involved in a particular task.

Applying Bayesian PCA to the learning data

To explain the Bayesian PCA approach as applied to the learning trajectories, we will first ignore the temporal dimension of the data and assume that each individual is associated with a single (potentially missing) performance score for each task. We will then show how the representation of the data can be extended to apply the model to the three-way data of individuals × tasks × practice stages (time).

Let $\mathbf{x}_j$ be an $N\times 1$ vector that represents the normalized performance scores across N tasks for individual j. In probabilistic PCA[13,14], the individual performances $\mathbf{x}$ are modelled as a weighted linear combination of orthogonal latent factors (also known as components):

$$\mathbf{x}_j = W\mathbf{z}_j + \boldsymbol{\mu} + \boldsymbol{\epsilon}_j \tag{1}$$

The $N\times K$ matrix W captures the latent factors where K corresponds to the number of latent factors. The $K\times 1$ vector $\mathbf{z}_j$ represents the user-specific scores on the latent factors that determines how high or low the user scores are on the latent factor. The $N\times 1$ vector $\boldsymbol{\mu}$ is a mean offset term to capture baseline differences in the tasks. This term is needed when modelling incomplete data and cannot be removed by subtracting the data mean[13]. Finally, the $N\times 1$ vector $\boldsymbol{\epsilon}$ represents the noise in the data.

A key assumption in probabilistic PCA is that the noise is normally distributed, independent across factors and has the same standard deviation σ across factors:

$$\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \sigma^2 I) \tag{2}$$

where the symbol I represents the identity matrix. In contrast, factor analysis uses the same model as in equation (1) but has a more flexible error term that allows variances to differ across factors: $\boldsymbol{\epsilon}_j \sim \mathcal{N}(0, \boldsymbol{\psi})$. Therefore, probabilistic PCA can be considered a type of factor analysis model but with a restricted error model[13,26]. Previous commentaries on the differences between PCA and factor analysis have focused on non-probabilistic PCA, which lacks an error model[27], but these distinctions do not apply to our current modelling approach.

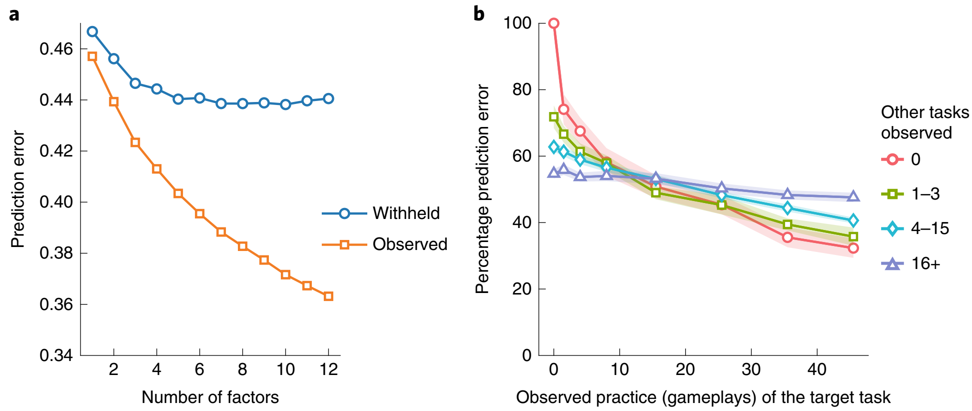


Fig. 2 | Predictive model performance. **a**, Error in predicting learning curves on a task from learning curves on other tasks as a function of the number of factors. In the prediction task, the learning curves are either observed (in-sample prediction error) or missing (out-of-sample prediction error). Prediction error is assessed by normalized mean absolute deviation. **b**, Error in predicting performance at the final practice stage of a particular task, depending on the amount of practice observed for the task (horizontal axis) and the number of learning curves observed from other tasks (colours and markers). Error is expressed in percentage of prediction error for the case where there is no historical performance data available for the user. Shaded areas indicate the s.e.m. across partitions of the data.

Common scores, different latent factors across stages of practice. There are several ways to apply probabilistic PCA to the three-way data of individuals × tasks × practice stages. Here, we make the assumption that each practice stage t is associated with its own set of latent factors W_t allowing for the covariation pattern across tasks to be dependent on the amount of practice. For example, two tasks A and B might not covary at the start of practice but as practice progresses the score in task A might start to covary with the score in task B as the two tasks begin to depend on the same latent factors. To constrain the latent factors across stages of practice, we assume that each individual has a single set of scores $\mathbf{z}_j$ that is shared among the latent factors associated with each stage of practice.

Specifically, let $\mathbf{x}_{t,j}$ be the N × 1 vector of task performance scores for user j at practice stage t. The model can be written as an extension of equation (1) by stacking the performance scores $\mathbf{x}_{t,j}$ and latent factors W_t across stages of practice:

$$\begin{bmatrix} \mathbf{x}_{1,j} \\ \vdots \\ \mathbf{x}_{T,j} \end{bmatrix} = \begin{bmatrix} W_{1,j} \\ \vdots \\ W_{T,j} \end{bmatrix} \mathbf{z}_j + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_j \quad (3)$$

Therefore, the observed performance of a user at practice stage t is based on a weighted combination of latent factors specific to t but where the scores $\mathbf{z}_j$ are shared across time. Because of the sharing of user-specific scores, the latent factors W_t cannot arbitrarily change across stages of practice. Note that the models in equations (3) and (1) are equivalent but only differ in the representation of the data $\mathbf{x}$ and the interpretation of the latent factors W.

Results

Bayesian inference. We first normalize the data such that performance scores across tasks are on the same scale. In addition, we truncate the learning curves to the first 60 gameplays and discretize the temporal dimension into eight stages of practice (see Methods). We apply a variational Bayesian inference procedure (see Methods) to infer model parameters. The main results presented here are based on the posterior mean of W which represents the latent factors and $\mathbf{z}_j$ which represents the scores on the latent factors for each user j.

Number of latent learning factors. Figure 2a shows the predictive model performance as a function of number of latent learning factors (K). The prediction problem corresponds to reconstructing an entire missing learning curve for a user on the basis of all other observed learning curves for the user (Fig. 1a). The prediction error is assessed on the original untransformed data by the normalized mean absolute deviation (see Methods). For the observed learning curves, increasing the number of latent learning factors leads to better reconstructions (lower in-sample prediction errors). However, for learning curves withheld from the model, increasing the number of factors leads to diminishing out-of-sample predictive performance beyond seven factors. The Supplementary Results show the predictive performance for three other metrics, each consistent with Fig. 2a and suggesting that predictive performance for out-of-sample learning curves reaches a maximal value for solutions in the range of five to eight factors. All our subsequent analyses will focus on the model solution with six latent learning factors, as this solution provides good predictive performance while not leading to overly complex visualizations of the factors (see the Supplementary Methods for additional description of predictive performance metrics and Extended Data Figs. 3 and 4 for inferred solutions with $K = 5$ and $K = 7$ latent factors). Note the substantial data compression that this implies: the variations in learning trajectories across all tasks for a user are captured by only six coefficients in the user scores ($\mathbf{z}_j$).

Predicting performance at the end of practice. Figure 2b shows the prediction error when the model is predicting user performance at the end of practice (between 51 and 60 gameplays, corresponding to the problem illustrated in Fig. 1b,c). The results show that predictions at the end of practice improve as a larger portion of the learning curve is observed. In addition, prediction performance increases as more performances on other tasks are observed. Of particular interest is that observing user performance on other tasks is as informative as observing about the first 10–25 performances on the learning curve for the target task itself. This shows that covariation between cognitive tasks can be used to form predictions about performance on a new task in the absence of any observations on that task.

Optimal prediction performance is obtained when observing as much of the learning curve for the target task as possible while not observing the performance on any other tasks (that is, the curves in Fig. 2b cross over). This result demonstrates that the model is not optimized to solve this particular prediction problem—the goal of the model is to reconstruct all observed learning curves as best as possible and more observed learning curves can deteriorate

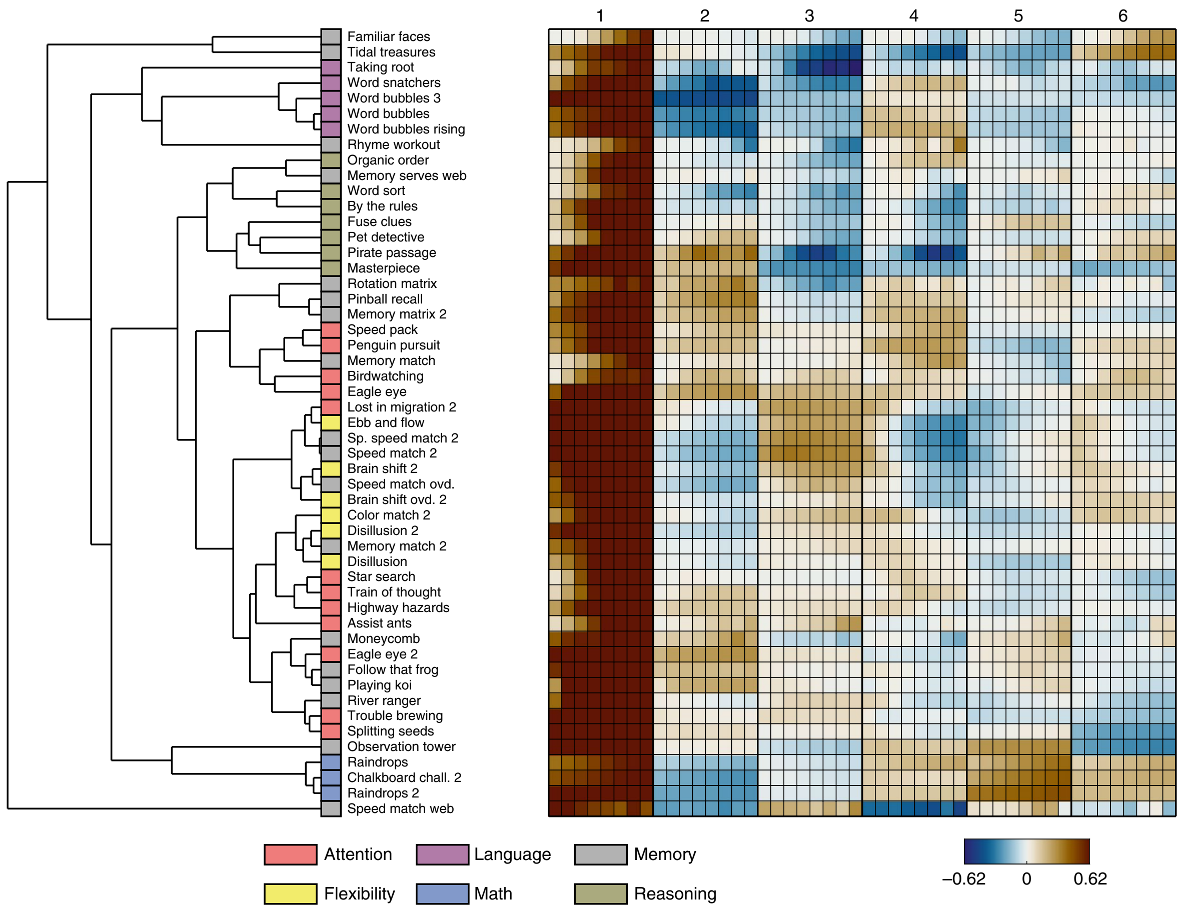


Fig. 3 | Inferred latent learning factors as a function of practice. The heatmap visualizes the six latent learning factors (columns) across games (rows). Positive (or negative) values are visualized by brown (or blue) colours. Each latent learning factor corresponds to a group of eight columns, where the columns within a group correspond to different stages of practice (practice increases from left to right). The hierarchical clustering solution to the left of the heatmap visualizes the similarity structure between cognitive tasks on the basis of the latent factors. Cognitive tasks are coloured according to primary task domain. Sp, spatial; ovd, overdrive.

predictive performance on any particular segment of any particular learning curve.

Inferred latent learning factors. Figure 3 shows the inferred latent learning factors broken down by stage of practice. The first latent learning factor is associated with positive scores across all tasks, consistent with a general ability factor often found in factor analysis of mental abilities[2]. While the first latent learning factor is positive across all tasks, there is also an increase over stages of practice (Fig. 4a). This increase in part reflects changes in the correlations in performance scores across tasks (Fig. 4b). A user who performs well in one task tends to perform well in other tasks and this correlation increases with practice.

Some of the changes in the first factor scores can be attributed to task differences related to adaptivity. In some tasks the level of difficulty adapts between gameplays (see Methods). For these between-game adaptive tasks, idiosyncratic effects might influence the performance scores in initial stages of practice, which would decrease the dependence on general ability and thus lower the first factor scores at at those stages. For example, a user who has high ability might only be able to express this ability once the game has adapted to an appropriate level. However, the first learning factor also increases with practice for tasks that maintain the same scoring system and difficulty level across stages of practice (that is, those with no between-game adaptivity).

Another potentially relevant factor is the reliability of scores as a function of practice. An increase in the first learning factor over stages of practice could be attributed to an increase in reliability (for example, smaller performance fluctuations over practice). However, Fig. 4c shows that the reliability of scores, assessed by the correlation in task scores of consecutive gameplays, ramps up quickly and remains constant after the third stage of practice (that is, after six gameplays), showing that the changes in factor scores cannot be attributed to changes in performance variability (within person) over time.

The tree diagram of Fig. 3 shows the similarity structure between tasks on the basis of the latent learning factors. The tasks cluster along the primary task domains: language tasks are similar to other language tasks, math tasks are similar to other math tasks, etc. This suggests that while the domain independent factor (the first factor) has a large influence on performance, the remaining factors modulate the learning curves for particular sets of tasks depending on the primary cognitive processing demands of the tasks. Note that, with the exception of the first factor, the values have no natural interpretation of higher-is-better. It is the sign of the user score along with the sign of the latent factor value that is important for interpretation.

Interpretation of latent learning factors. We examine whether the latent learning factors described by this six-factor solution can be related to known features of the tasks such as the types of cognitive

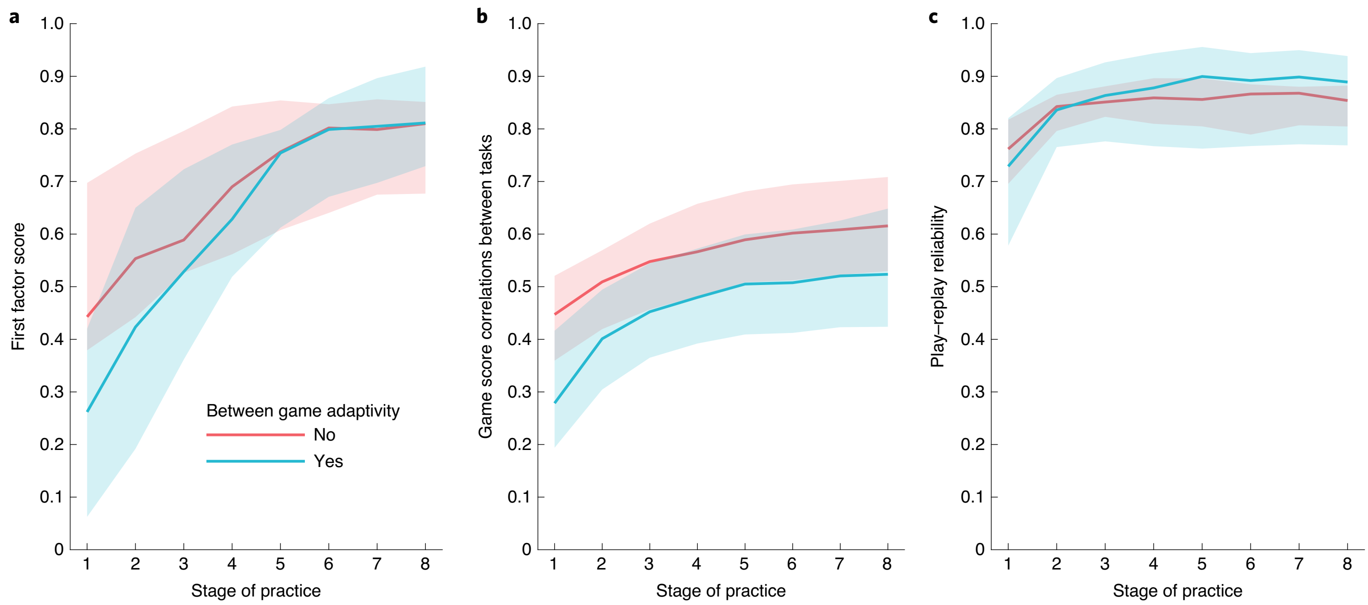


Fig. 4 | Analysis of the first learning factor. a–c, Mean score of the first learning factor (**a**), game score correlations between tasks (**b**) and play–replay reliability as a function of stage of practice (**c**). Results are broken down by tasks with and without between-game adaptivity (colours). Shaded areas represent 25–75% percentiles. Correlations were first computed for each gameplay before aggregating over gameplays corresponding to a stage of practice.

processing demands imposed by the task, the type of stimulus, method of input or aspects of game design (for example, the presence of leveling in difficulty or the presence of time pressure). Table 1 shows a list of task features we used in our analysis. Each feature is binary: a task is either associated with the feature or it is not. (Extended Data Figs. 1 and 2 show the full task × feature matrix and a definition of the task features.)

Figure 5 shows the correlations between latent learning factors and task features corresponding to cognitive processes and stimulus types. The hierarchical tree visualizes the similarity structure in the pattern of correlations. We use the primary task domains in Fig. 3 and the task features in Fig. 5 to guide an interpretation of the individual factors. The first factor, as described above, is associated with higher performance across all tasks and is consistent with a general ability factor. The second factor is related to visuospatial abilities: it is has highest scores in tasks requiring spatial recall, route planning or the integration of visual information from across the visual field. Scores are correlated with the use of locations and objects as the task stimuli. The third factor is related to executive attention, with highest scores in tasks requiring the maintenance of goal-directed information or the inhibition of inappropriate responses and activation of appropriate ones. An increased score on this factor improves performance on tasks requiring selective attention. The fourth factor is primarily associated with comprehension knowledge related to math and language tasks, that is crystallized intelligence, although it also has high scores in several tasks requiring spatial reasoning or visualization. The fifth factor is related to mathematical ability—note the high correlation with task features involving numbers and math. The sixth factor is a hybrid factor that combines executive attention and math.

While this interpretation of the factors is consistent with the task features and primary task domains, there might be others as well. In addition, the interpretation of these factors is constrained by the orthogonality of these factors, which tends to lead to tasks being associated with multiple factors. To examine the effect of orthogonality on the interpretability of the factors, we applied factor rotation methods using the promax method[28], which leads to non-orthogonal factors where the variance of each task is redistributed across factors such that tasks are associated with fewer factors (see Extended Data Fig. 5 for the resulting factors and factor correlations). As with the original orthogonal solution, the first factor retains positive scores across all tasks, consistent with a general ability factor. Though still positive, first factor scores are weakest for tasks involving language. The second factor is also similar to its counterpart before rotation but the signs have been reversed such that visuospatial tasks now correspond to negative values. The largest positive scores are for language tasks. Positive scores on the third factor are found for tasks involving spatial reasoning, with negative scores relating to tasks demanding response inhibition and other aspects of executive function. The fourth factor has strongest scores for tasks involving visualization of spatial transformations or spatial recall. The fifth factor relates to mathematical ability, similar to the orthogonal solution. The sixth factor is a hybrid that relates to tasks practicing executive attention and also to reasoning and language.

Predicting individual task features from the latent learning factors. The previous analysis focused on a qualitative interpretation of the latent learning factors. We can also perform a more fine-grained, quantitative analysis that examines whether the pattern of information contained in the latent learning factors is predictive of individual task features. We build a predictive model that has the goal of predicting (out-of-sample) the presence or absence of each task feature for any particular task on the basis of the latent learning factors associated with the task. Table 1 shows the diagnosticity of the predictive model expressed by area under the curve (AUC). This corresponds to the probability that the model assigns a higher score of 'feature present' to a new task where the feature is indeed present than a new task without the feature. For 16 of 28 task features, predictive model performance exceeds the 95% performance interval associated with chance-level performance. For example, whether a task involves cognitive processes such as selective attention, task switching, word generation, calculation, spatial reasoning and response inhibition can be determined by the latent learning factors associated with the task. Some of the other cognitive processes cannot be reliably identified in part because the feature was only associated with a few tasks (N in Table 1), which makes generalization to new tasks more challenging.

Table 1 | Performance (AUC) of predicting task features from the latent learning factors

Feature type	Task feature	AUC	Confidence interva	N
Process	Memory	0.65	(0.26, 0.71)	24
Process	Memory updating	0.71	(0.21, 0.76)	10
Process	Multiple object monitoring	0.91	(0.08, 0.91)	3
Process	Selective attention	0.82*	(0.18, 0.73)	11
Process	Divided attention	0.71	(0.23, 0.74)	13
Process	Task switching	0.86*	(0.15, 0.83)	5
Process	Vocabulary knowledge	0.93	(0.04, 0.97)	2
Process	Word generation	0.85*	(0.13, 0.80)	6
Process	Planning	0.80	(0.12, 0.84)	6
Process	Calculation	0.98*	(0.12, 0.86)	4
Process	Quantitative reasoning	0.52	(0.14, 0.85)	4
Process	Spatial reasoning	0.81*	(0.22, 0.77)	10
Process	Logical reasoning	0.77	(0.11, 0.86)	4
Process	Response inhibition	0.90*	(0.24, 0.73)	17
Stimulus	Objects	0.78*	(0.27, 0.72)	33
Stimulus	Locations	0.90*	(0.26, 0.72)	25
Stimulus	Single letters	0.79	(0.08, 0.91)	3
Stimulus	Words	0.90*	(0.20, 0.77)	10
Stimulus	Single digits	0.59	(0.14, 0.83)	6
Stimulus	Numbers	0.98*	(0.10, 0.90)	4
Input method	Cursor keys	0.76*	(0.25, 0.72)	17
Input method	Keyboard entry	0.83*	(0.20, 0.77)	9
Input method	Mouse pointer	0.89*	(0.26, 0.70)	27
Game design	Within-game adaptivity	0.95*	(0.25, 0.71)	29
Game design	Algorithmic leveling	0.73	(0.23, 0.74)	14
Game design	User-selected leveling	0.67	(0.25, 0.72)	17
Game design	Response-time pressure	0.89*	(0.25, 0.71)	31
Game design	Briefly presented stimuli	0.82*	(0.23, 0.74)	11

Values denoted with an asterisk show performances outside the 95% confidence interval given between parentheses. AUC, area under the curve; *N*, number of tasks with the feature present.

The particular Bayesian PCA approach we adopted allows for different latent factors W_t to be associated with each stage of practice (t), which raises the question whether the factors assess the same construct over time[29]. Because of the sharing of scores across stages of practice, there are constraints on the latent factors that prevent them from arbitrary changes across stages of practice. For example, a user who scores high on a particular factor will upweight the contribution of the corresponding column of all matrices W_t regardless of stage of practice t, which constrains model flexibility during inference. In model simulations (Supplementary Methods), we verified that the task feature prediction model retains high performance when the model is trained on the latent factors associated with a particular stage of practice and validated on other stages of practice.

Overall, these results show that the latent learning factors carry not only information about general learning ability but also task-specific information related to the type of cognitive process involved, the type of stimulus being processed, the input method and additional aspects of game design.

Contributions of individual latent factors to the shape of the learning curve. To better understand the contributions of individual latent learning factors to the shape of the learning curve, Fig. 6 shows the posterior predicted learning curves in the original (untransformed) data space for the set of games that maintain the same difficulty levels and scoring across all stages of practice (Extended Data Fig. 6 shows results across all games). The curves show the predictions for simulated individuals who have individual scores on a factor that score in the top 10% (red), bottom 10% (blue) or at 50% (black) while scoring at 50% on all other factors. Consistent with Fig. 3, the first factor predicts improved learning outcomes across all tasks. The remaining factors primarily shape the learning curve at later stages of practice and for subsets of tasks. For example, the fifth factor primarily differentiates performance on the math tasks and has a relatively small impact on other tasks.

Relating individual scores to demographics. Figure 7 shows the distribution of individual scores ($\mathbf{z}_j$) for each latent learning factor as a function of age and educational background. The first factor strongly differentiates between age groups. Older age groups score lower on this factor, predicting lower performance across all tasks, consistent with cognitive slowing theories of aging that affect a broad range of tasks[30]. Note that the general learning factor we found does not depend on the presence of different age groups. The same learning factor emerges when the model is applied to individual age groups, consistent with previous research that general intelligence factors can be found across different subpopulations[31]. In terms of educational differences, individuals with more advanced education score higher on the general learning factor and lower on the second factor, corresponding to higher math and language performance scores. However, overall there are no strong effects of educational background.

Discussion

Large-scale datasets offer unique opportunities to study cognition at scale[32,33]. The Lumosity dataset offers a unique multivariate and longitudinal set for studying learning across a number of cognitive tasks and users. To avoid making strong assumptions about the data, we started this investigation with a data-driven Bayesian PCA approach to discover the statistical patterns of covariation across learning trajectories. The model infers a set of latent factors that can change as a function of the stage of practice, similar to other approaches that allow for the organization of cognitive abilities to vary as a function of ability[34,35]. The modelling results showed that the learning trajectories show substantial covariation across tasks. Individuals who are good at learning one task typically are also good at learning other tasks, analogous to the standard finding of Spearman's general intelligence factor[1] and a positive manifold in tests of cognitive ability where performance across tests tends to be positively correlated.

A key finding is that the general learning factor was expressed mostly at later stages of practice. Across tasks, asymptotic performance correlates more strongly than baseline performance. The increase in the general learning factor over stages of practice was especially prevalent for tasks with adaptive difficulty, consistent with the possibility that tasks with adaptive difficulty may take some time to reach an individual's challenge level. However, the effect was also found for non-adaptive tasks, which suggests that the type of learning in early stages (that is, learning how to complete the task, not necessarily how to perform it well) may not map as well to this general factor. Overall, the temporal pattern for the general learning factor suggests that early performance on these cognitive tasks is in part influenced by idiosyncratic, task-specific factors.

The finding that general ability is expressed to a higher degree after extended practice is also consistent with cognitive and developmental theories of intelligence such as mutualism[8] and process

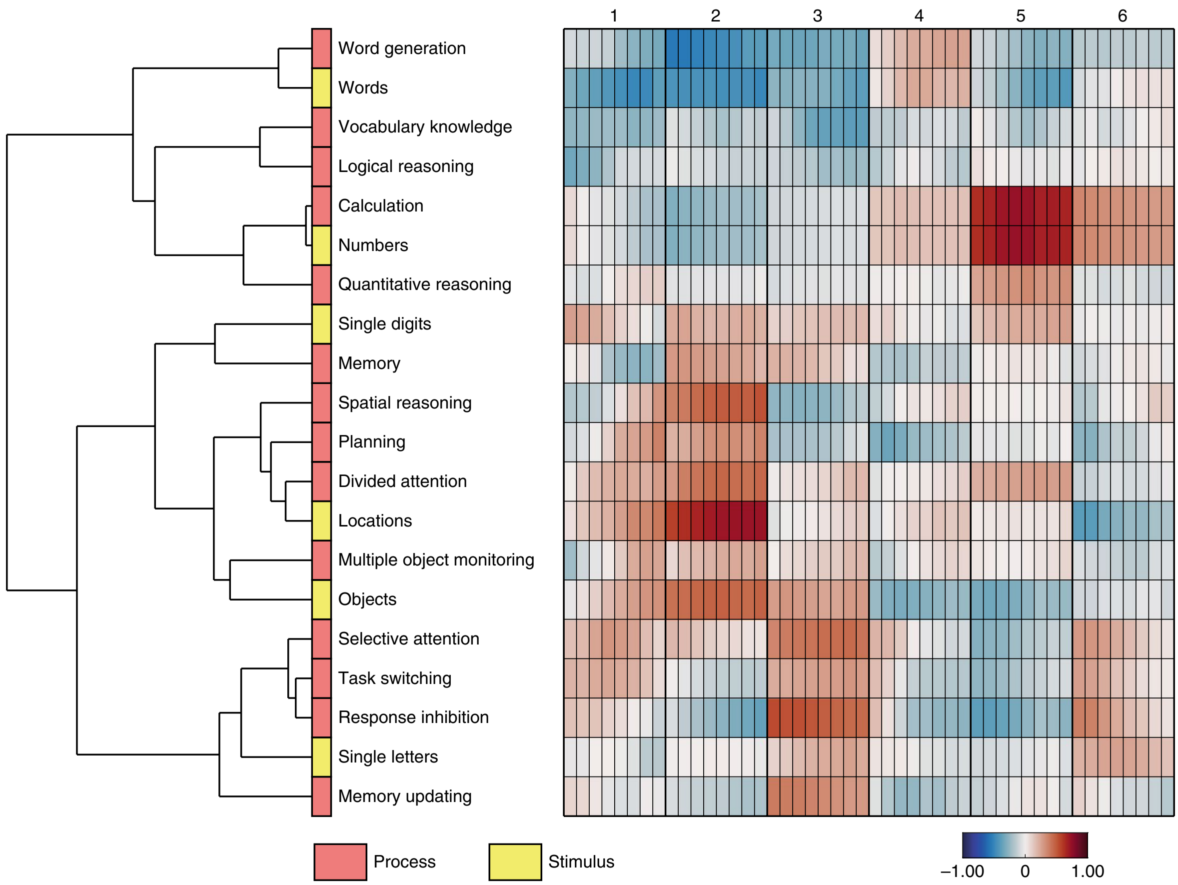


Fig. 5 | Correlations between latent learning factors and manually derived task features. Each group of columns corresponds to the different time slices of a single latent factor. Positive (or negative) correlations are illustrated by red (or blue) colours. The hierarchical tree visualizes the similarity between the task features on the basis of the pattern of correlations between task features and latent factors. Task features are grouped by process and stimulus type.

overlap theory[4]. The mutualism theory of intelligence views the cognitive system as a set of basic cognitive abilities and growth in one ability is partly based on growth in other abilities. During development, initially uncorrelated abilities become correlated as a result of the mutually beneficent interactions between these abilities. For example, there is evidence for a coupling between fluid and crystallized abilities such that higher starting point in one ability is associated with greater developmental gains in the other[36,37]. Process overlap theory, on the other hand, proposes that performance on a task depends on interactions between domain-specific as well as domain-general cognitive processes such as attentional and executive processes. Task performance is constrained by whichever required process an individual is weakest in. In the initial stages, task-specific skills need to be acquired, which limits the observed correlations between tasks. During later learning stages, more domain-general abilities may set an upper limit on learning ability.

The model also inferred additional latent factors that captured performance variations related to specific types of tasks. In fact, the latent learning factors encode information related to a number of different cognitive processes, types of stimuli, ways of assessing user responses and general design features of the tasks. Therefore, the relatively low-dimensional factor structure captures a substantial amount of information about the mental activities in each task.

We also showed that the model can be used to predict task performance from observed performance on other tasks. Specifically, we found that performance at a later stage of practice on a particular task can be predicted as accurately from observing performance on other cognitive tasks as observing performance at the very early stages of practice for that particular task. This result has implications for real-world forecasting scenarios where the most promising individuals need to be selected for expensive and time-consuming training of a particular task. Some of the future performance can be forecasted without engaging in the target task at all.

By making the Lumosity dataset publicly available, we aim to foster new research on individual differences and the development of new modelling approaches to account for covariation across learning trajectories. There are several research directions that are likely to improve the prediction results and interpretability of the latent factors. For example, we can improve the treatment of the temporal dimension of the data by building an explicit continuous-time model for the learning trajectory for each individual[17,38] and applying multivariate latent growth models to learn the covariance structure of the latent parameters governing the changes over time (for example, refs. [18–22,36,37]). In addition, future research could focus on a cognitive modelling approach to explain observed performance in terms of latent cognitive processes[39]. One Lumosity task (the task-switching game 'Ebb and flow') has already been modelled to understand the cognitive processes at the level of individual trials as well as the cognitive changes across sessions[10]. Finally, while our approach reveals changes in covariation of performance across tasks as practice progresses, it does not directly address the question of whether or how cognitive training might transfer between tasks, which has been a topic of recent debate[40,41]. Future modelling could allow for transfer effects arising from mutually coupled cognitive processes[36,37] to account for the possibility that performance is impacted by the order in which

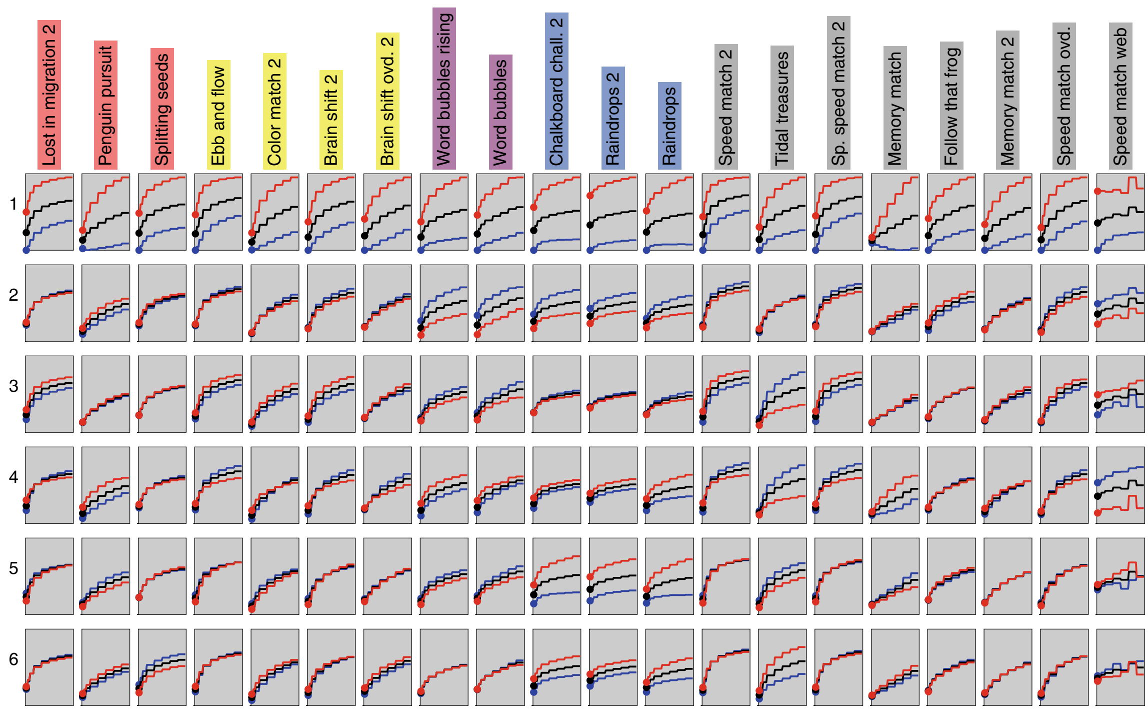


Fig. 6 | Changes in predicted learning curves based on changes in individual scores for individual factors. For each factor (row), the blue, black and red lines show the predicted learning curves for 10, 50 and 90% percentile scores on that particular factor and median scores on all other factors. Therefore, the red (blue) lines show the predicted learning curves for individuals who score high (low) on that factor while holding constant the contribution of other factors. Intercept differences are highlighted with circles at the start of the learning curve. To facilitate comparison, the vertical axis is the same for each particular task (column) but is different across tasks. Tasks are ordered by primary domain (colour) consistent with Fig. 3. Only the subset of tasks without between-game adaptivity is shown.

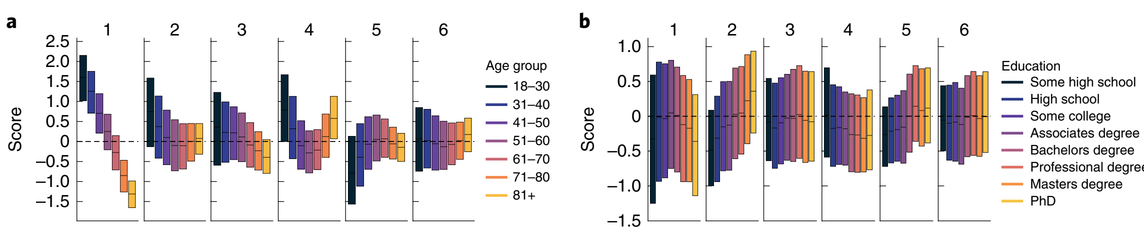


Fig. 7 | Distribution of individual scores for latent learning factors across demographic categories. a,**b**, Latent learning factors (columns) across age groups (**a**) and educational levels (**b**). Boxes represent interquartile ranges (that is, 25–75% percentile scores) and the median (horizontal line).

tasks are practiced in interleaved practice. Ultimately, the predictive evaluation approach introduced in this research will facilitate comparison between different modelling approaches.

Methods

The Lumosity platform provides several games that tap memory, attention, flexibility, speeded processing and problem solving. In the Lumosity program, individuals are given a recommended daily training session of five different cognitive training games. One five-game session takes approximately 15 min to complete. Outside of the training sessions, Lumosity users can also opt to select and play games directly from the entire library of available games.

Data sample. We performed a retrospective analysis of a sample of gameplay data on the basis of 84 distinct cognitive games. The de-identified data sample provided by Lumosity includes the gameplay event history for 36,297 individuals, 1,442,555 individual learning curves and 50,374,056 gameplay events. No statistical methods were used to predetermine sample sizes but our sample sizes are larger than those typically reported. The data sample was selected from the overall Lumosity database by focusing on active users who have at least 500 lifetime gameplays. In addition, because the focus of our research is on analysing correlations across cognitive tasks, this dataset only included users if 5% or fewer of their gameplays were repetitions of the previous game. This criterion excludes individuals who choose to repeat specific games rather than play from the broader game library. In addition, individuals were selected who signed up between 1 August 2013 and 31 December 2016, with an age at signup between 18 and 90 yr. The country of origin was restricted to the United States, Canada or Australia with English as the user's preferred language. Finally, individuals were included if at least 99% of their lifetime Lumosity gameplays were on the web product (as opposed to mobile apps) and only the web data were included in the data sample. In total, the dataset contains the full gameplay history across tasks spanning a period from 1 August 2013 to 30 June 2019. Most users in this sample (80%) signed up with Lumosity before 1 January 2015. Users in this sample spent a median of 2 yr on the platform.

Demographics of participants. Basic demographic information is available on the basis of information provided by users when signing up for Lumosity. In our data sample, most individuals are female (50%), with 39% males and 11% individuals who did not provide gender information. We coded the age of individuals in seven bins leading to the following breakdown of the user sample: 18–30 (5%), 31–40 (5%), 41–50 (12%), 51–60 (25%), 61–70 (32%), 71–80 (17%) and 81+ (4%). Therefore, the user sample skews towards an older demographic. Individuals reported their educational attainment according to the following categories: some high school (2%), high school (13%), some college (22%), Associates degree (4%), Bachelors degree (31%), Professional degree (6%), Masters degree (19%) and PhD (3%).

Data filtering. For some of the cognitive tasks in the dataset, the available learning data were sparse in terms of the number of individuals who performed the task or the total number of times individuals performed the task. To ensure that sufficient data were available for our analyses, we filtered out any cognitive task that had fewer than 200 individuals at the twentieth gameplay attempt. At the learning curve level, we truncated the learning curve to the first 60 gameplays. The final dataset created after these filtering steps includes the gameplay event history for 51 cognitive games, 36,297 individuals, 1,255,175 individual learning curves and 36,736,286 single gameplay events.

Data sparsity. The observed data are sparse as users do not perform all cognitive tasks. The tasks vary by popularity from 99.8% to as few as 1.5% of users in this sample performing the task. On average, users perform a median of 32 distinct tasks. An additional factor that influences sparsity is that learning curves for a particular user and a particular task might not be complete—a user might stop playing the game before our cutoff of 60 gameplays. Across tasks, users and first 60 gameplays, only 33% of the data are observed. Supplementary Fig. 3 illustrates the pattern of sparsity across games as well as stages of practice.

Performance scores. At the end of each gameplay event, users are given feedback on a variety of performance scores depending on the type of game. Some games provide feedback on mean response time per trial and/or mean accuracy. Our dataset only includes data on the total score per gameplay, which all games provide. This performance score is based on a combination of accuracy, speed of processing, as well as bonus points accumulated from a variety of factors (for example, streaks of correct responses or completing more challenging trials).

Adaptivity within and between games. Some of the Lumosity games have a level of difficulty that adapts to the user's performance. For different games, this adaptivity may occur either within or between gameplays. Between-game adaptivity is achieved either through user-selected leveling, in which users unlock more challenging levels by reaching certain performance targets, or algorithmic leveling, in which the level of difficulty is adjusted algorithmically across gameplays on the basis of performance. For within-game adaptivity, the level of difficulty can increase or decrease during gameplay depending on user performance but each gameplay begins at the same difficulty level. Extended Data Fig. 1 shows which tasks are associated with each type of adaptivity. For the purpose of our analyses, we distinguish between tasks that have no between-game adaptivity (20 total), which maintain the same level of difficulty and scoring system across gameplays, and tasks that have a form of between-game adaptivity (31 total).

Data transformations. We applied several transformation steps to prepare the data for the PCA algorithm, reduce the computational complexity and normalize the performance scores across cognitive tasks. To introduce notation, we denote the observed performance scores by a three-way array Y with elements y_{ijt} referenced by indices $i=1, \ldots, L$ (tasks), $j=1, \ldots, M$ (individuals) and $t=1, \ldots, T$ (number of gameplays). We transform the three-way array to a matrix of scores X with elements x_{ij} referenced by $i=1, \ldots, N$ (combination of task and stage of practice) and $j=1, \ldots, M$ (individual). It should be noted that the predictive performance of the PCA model is assessed on the original untransformed data (Y) to facilitate comparison to other models that might use the same data.

In the first data transformation step, we reduced the dimensionality of the temporal component of the data by discretizing time into B temporal bins. We set $B=8$ and defined the following stages of practice by the number of gameplays completed: 1 or 2 gameplays, 3–5, 6–10, 11–20, 21–30, 31–40, 41–50 and 51–60. All (observed) performance outcomes were averaged within these bins. The bins were chosen to be smaller for the early stages of the learning curve as these are the stages where performance typically changes the most.

In the second data transformation step, we normalized the performance outcomes across cognitive tasks and temporal bins. For each task, the performance scores were normalized by the standard deviation of the performance scores in the last temporal bin (gameplays 51–60). No correction was applied to the mean. After this transformation step, the data are of size $M=36{,}297$ (individuals) $\times L=51$ (cognitive tasks) $\times B=8$ (stages of practice).

In the final data transformation step, we rearrange the data to a matrix representation. For each individual, we concatenate the L performance scores across tasks and B stages of practice to create a single (column) vector with $N=L\times B$ performance scores. By combining the $\mathbf{M}$ individual vectors, we create the final $N \text{x} M$ matrix X.

Priors and model inference. The Bayesian approach to probabilistic PCA completes the model by using the following priors on the model parameters:

$$\begin{aligned} w_{ik} &\sim \mathcal{N}(0, \sigma_w^2) \\ z_{ij} &\sim \mathcal{N}(0, 1) \\ \mu_i &\sim \mathcal{N}(0, \sigma_\mu^2) \end{aligned} \quad (4)$$

In this notation, w_{ik} is an element of matrix W with indices $i=1, \ldots, N$ and $k=1, \ldots, K$ and the variable z_{ij} is the ith element of individual vector $\mathbf{z}_j$.

We use the variational Bayesian inference procedure (VBPCA) described in ref. [13] to infer model parameters. The algorithm provides estimates of the posterior mean and covariance of W, $\mathbf{z}$ and $\boldsymbol{\mu}$ as well as estimates of the hyperparameters σ, σ_w and σ_μ. Finally, the algorithm provides estimates of the variance of withheld observations of the $\mathbf{x}_j$. The VBPCA algorithm was run for 30 iterations.

Predictive tests. We assess model performance on two types of prediction problems. In the first prediction problem (illustrated in Fig. 1a), the goal is to predict the entire learning curve for a task for a particular user given the observed learning curves from other tasks that the user engaged in. We created a validation set by withholding for each user the performance scores from one cognitive task chosen at random from all tasks that the user engaged in. Because of the size of the data, we only created a single partition of the data into a training and validation set to assess predictive performance.

In the second prediction problem (illustrated in Fig. 1b,c), the goal is to predict the user performance at the last stage of practice (51–60 gameplays) from some combination of observed performance on earlier stages of practice on the task as well as practice on other tasks. We created 36 partitions of the data into validation and training sets. For each partition, we first randomly selected 20% of users and selected for each user a single target learning curve without any missing practice data. The data for the last stage of practice (gameplays 51–60) of the target learning curves were then placed in the validation set. For each of the target learning curves, the number of observed stages of practice for the training data was randomly determined, uniformly from the $B-1=7$ remaining practice stages. For each of the selected users, we selected a random number of learning curves from tasks other than the target tasks, uniformly from the number of other tasks the user engaged in. The selected learning curves were placed in the training set. For the remaining 80% of users, all observed data were placed in the training data. We repeated this procedure of partitioning data into training and validation data 36 times. The model was applied to each partition and predictions for the last practice stage for the validation learning curves were recorded.

Assessing predictive performance. We use the normalized mean absolute deviation (NMAD) metric to assess the predictive performance of the model. The mean absolute deviation (MAD) for particular cognitive task i is defined as

$$\text{MAD}_i = \frac{1}{|\Omega_i|} \sum\nolimits_{(j,t)\in\Omega_i} |y_{ijt} - \hat{y}_{ijt}| \quad (5)$$

where y_{ijt} is the withheld performance of user j on task i at time t and $\hat{y}_{ijt}$ is the corresponding model prediction. The variable Ω_i denotes the set of withheld data pairs j and t for task i and $|\Omega_i|$ is the number of withheld data pairs for task i. To combine the MAD across tasks, we normalize the scores before averaging:

$$\text{NMAD} = \frac{1}{N} \sum\nolimits_j^N \text{MAD}_i / S_i \quad (6)$$

where S_i is the standard deviation of performance scores (across all practice trials) for task i. This step normalizes the arbitrary scoring ranges associated with individual tasks. The Supplementary Methods show three additional predictive performance metrics for model evaluation.

Determining the number of factors. There are several methods to determine K, the number of latent factors in a dimensionality reduction algorithm such as PCA. We use a generalization test where we assessed the model's performance on predicting learning curves that are withheld from the model. We calculated the normalized mean absolute deviations (NMAD) for $K=1$ up to $K=12$ and chose the number of latent factors such that generalization performance was adequate and no substantial gains could be obtained by selecting solutions with a higher number of latent factors. In our simulations, about six latent factors were sufficient to maximize predictive performance. Supplementary Figs. 1 and 2 show that similar results were obtained with three other predictive performance metrics.

Determining the primary cognitive domains and task features. The tasks included in the dataset span a range of cognitive skills. For the purpose of visualizing and interpreting the results of the probabilistic PCA, we classified the games into six primary areas: attention, flexibility, language, math, memory and reasoning. These classifications closely follow the primary areas that are listed on the Lumosity platform with a few exceptions (for example, we dropped the speed area and reclassified a few games into other areas). Extended Data Fig. 1 shows the primary cognitive areas designated for each task.

While these primary cognitive domain classifications are broadly useful, it is important to note that each game might require multiple types of cognitive processes. To more fully capture the set of cognitive activities involved in each task, we enumerated a list of task features shown in Extended Data Fig. 1. The task features relate to the type of cognitive process, the type of stimulus, the method of input and other aspects of game design. For each task, the authors assessed what task features were relevant and focused mainly on features that would be applicable to a range of tasks. For definitions of task features, see Extended Data Fig. 2.

Linking latent factors to predefined task features. A (lasso) regularized logistic regression model was used to predict predefined task features from the pattern of latent factors. For each task feature, a separate model was trained using cross-validation. For each fold, the model was trained on all but two tasks and the test set consisted of the two held out tasks with the constraint that the feature was present on one task but not the other. The partitioning into training and test sets was done over 200 different combinations to create a test set with one feature-present and one feature-absent task. An AUC value was calculated from the proportion of test sets where the held out feature-present task was given a higher score by the logistic regression model than the held out feature-absent task.

In addition, we derived a distribution of AUC that can be expected from chance-level performance. In this procedure, we randomly permuted the feature pattern across tasks (losing any connection between the latent learning factors and task features but preserving the marginal distribution of variables) and calculated an AUC for the logistic regression model as described earlier. This procedure was repeated 400 times to derive a distribution of AUC values. The 2.5% and 97.5% percentiles of this distribution are reported in Table 1.

For all simulations of the regression model, a coordinate descent algorithm for model inference was used. Convergence was assessed by the relative change in the size of the estimated coefficients. All model simulations converged when the relative change dropped below 1×10^{-4}.

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

Data availability

The original and preprocessed versions of the data can be accessed at https://osf.io/g9zkf. Source data are provided with this paper.

Code availability

The code used to analyse the data, run the Bayesian PCA model and create figures and tables can be accessed at https://osf.io/g9zkf.

Received: 9 January 2020; Accepted: 15 July 2020;
Published online: 31 August 2020

References

1. Spearman, C. 'General intelligence' objectively determined and measured. *Am. J. Psychol.* **15**, 201–293 (1904).
2. Carroll, J. B. et al. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (Cambridge Univ. Press, 1993).
3. Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J. & Minkoff, S. R. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* **30**, 163–183 (2002).
4. Kovacs, K. & Conway, A. R. Process overlap theory: a unified account of the general factor of intelligence. *Psychol. Inq.* **27**, 151–177 (2016).
5. Jensen, A. R. *Clocking the Mind: Mental Chronometry and Individual Differences* (Elsevier, 2006).
6. Schubert, A.-L., Hagemann, D. & Frischkorn, G. T. Is general intelligence little more than the speed of higher-order processing?. *J. Exp. Psychol.* **146**, 1498–1512 (2017).
7. Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J. & Mackintosh, N. Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence* **37**, 374–382 (2009).
8. Van Der Maas, H. L. et al. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* **113**, 842–861 (2006).
9. Barbey, A. K. Network neuroscience theory of human intelligence. *Trends Cogn. Sci.* **22**, 8–20 (2018).
10. Steyvers, M., Hawkins, G. E., Karayanidis, F. & Brown, S. D. A large-scale analysis of task switching practice effects across the lifespan. *Proc. Natl Acad. Sci. USA* **116**, 17735–17740 (2019).
11. Steyvers, M. & Benjamin, A. S. The joint contribution of participation and performance to learning functions: exploring the effects of age in large-scale data sets. *Behav. Res. Methods* **51**, 1531–1543 (2019).
12. Donner, Y. & Hardy, J. L. Piecewise power laws in individual learning curves. *Psychon. Bull. Rev.* **22**, 1308–1319 (2015).
13. Ilin, A. & Raiko, T. Practical approaches to principal component analysis in the presence of missing values. *J. Machine Learning Res.* **11**, 1957–2000 (2010).
14. Tipping, M. E. & Bishop, C. M. Probabilistic principal component analysis. *J. R. Stat. Soc. B* **61**, 611–622 (1999).
15. Lim, Y.J. & Teh, Y.W. Variational Bayesian approach to movie rating prediction. In *Proc. International Conference on Knowledge Discovery and Data Mining* 15–21 (ACM, 2007).
16. Bell, R. M. & Koren, Y. Scalable collaborative filtering with jointly derived neighborhood interpolation weights. In *Proc. Seventh IEEE International Conference on Data Mining* 43–52 (IEEE Computer Society, 2007).
17. Driver, C. C. & Voelkle, M. C. Hierarchical Bayesian continuous time dynamic modeling. *Psychol. Methods* **23**, 774–799 (2018).
18. Kievit, R. A. et al. Developmental cognitive neuroscience using latent change score models: a tutorial and applications. *Dev. Cogn. Neurosci.* **33**, 99–117 (2018).
19. Isiordia, M. & Ferrer, E. Curve of factors model: a latent growth modeling approach for educational research. *Educ. Psychol. Meas.* **78**, 203–231 (2018).
20. Ram, N. & Grimm, K. J. in *Handbook of Child Psychology and Developmental Science* (ed. Lerner, R. M.) 1–31 (Wiley, 2015).
21. McArdle, J. J., Ferrer-Caja, E., Hamagami, F. & Woodcock, R. W. Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Dev. Psychol.* **38**, 115–142 (2002).
22. Preacher, K. J., Wichman, A. L., MacCallum, R. C. & Briggs, N. E. *Latent Growth Curve Modeling* (Sage, 2008).
23. McNeish, D., Dumas, D. G. & Grimm, K. J. Estimating new quantities from longitudinal test scores to improve forecasts of future performance. *Multivariate Behav. Res.* https://doi.org/10.1080/00273171.2019.1691484 (2019).
24. Rosenberg, M. D., Casey, B. & Holmes, A. J. Prediction complements explanation in understanding the developing brain. *Nat. Commun.* **9**, 589 (2018).
25. Settles B., Brust, C., Gustafson, E., Hagiwara, M. & Madnani, N. Second language acquisition modeling. In *Proc. Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications* (eds Tetreault, J., Burstein, J., Kochmar, E., Leacock, C. & Yannakoudakis, H.) 56–65 (ACL, 2018).
26. Luttinen, J. & Ilin, A. Transformations in variational Bayesian factor analysis to speed up learning. *Neurocomputing* **73**, 1093–1102 (2010).
27. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods* **4**, 272–299 (1999).
28. Abdi, H. in *Encyclopedia for Research Methods for the Social Sciences* (ed. Lewis-Beck, M. S. et al.) 792–795 (Sage, 2004).
29. Widaman, K. F., Ferrer, E. & Conger, R. D. Factorial invariance within longitudinal structural equation models: measuring the same construct across time. *Child Dev. Perspect.* **4**, 10–18 (2010).
30. Salthouse, T. A. The processing-speed theory of adult age differences in cognition. *Psychol. Rev.* **103**, 403–428 (1996).
31. Jensen, A. R. Regularities in Spearman's law of diminishing returns. *Intelligence* **31**, 95–105 (2003).
32. Griffiths, T. L. Manifesto for a new cognitive revolution. *Cognition* **135**, 21–23 (2015).
33. Goldstone, R. L. & Lupyan, G. Discovering psychological principles by mining naturally occurring data sets. *Topics Cogn. Sci.* **8**, 548–568 (2016).
34. Molenaar, D., Dolan, C. V., Wicherts, J. M. & van der Maas, H. L. Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence* **38**, 611–624 (2010).
35. Tucker-Drob, E. M. Differentiation of cognitive abilities across the life span. *Dev. Psychol.* **45**, 1097–1118 (2009).
36. Kievit, R. A. et al. Mutualistic coupling between vocabulary and reasoning supports cognitive development during late adolescence and early adulthood. *Psychol. Sci.* **28**, 1419–1431 (2017).
37. Kievit, R. A., Hofman, A. D. & Nation, K. Mutualistic coupling between vocabulary and reasoning in young children: a replication and extension of the study by Kievit et al.(2017). *Psychol. Sci.* **30**, 1245–1252 (2019).
38. Evans, N. J., Brown, S. D., Mewhort, D. J. & Heathcote, A. Refining the law of practice. *Psychol. Rev.* **125**, 592–605 (2018).
39. Frischkorn, G. & Schubert, A.-L. Cognitive models in intelligence research: advantages and recommendations for their application. *J. Intell.* **6**, 34 (2018).
40. Melby-LervÅg, M., Redick, T. S. & Hulme, C. Working memory training does not improve performance on measures of intelligence or other measures of 'far transfer' evidence from a meta-analytic review. *Perspect. Psychol. Sci.* **11**, 512–534 (2016).
41. Simons, D. J. et al. Do 'brain-training' programs work? *Psychol. Sci. Public Interest* **17**, 103–186 (2016).

Acknowledgements

Feedback on the task feature categorization framework was provided by A. Kaluszka, O. Claflin, A. Osman and N. Ng. The authors received no specific funding for this work.

Author contributions

M.S. and R.J.S. planned the research. R.J.S. provided the data. M.S. analysed the data. M.S. and R.J.S. interpreted the results and wrote the paper.

Competing interests

R.J.S. is an employee of Lumos Labs and owns stock in the company. M.S. has no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-00935-3.

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00935-3.

Correspondence and requests for materials should be addressed to M.S. or R.J.S.

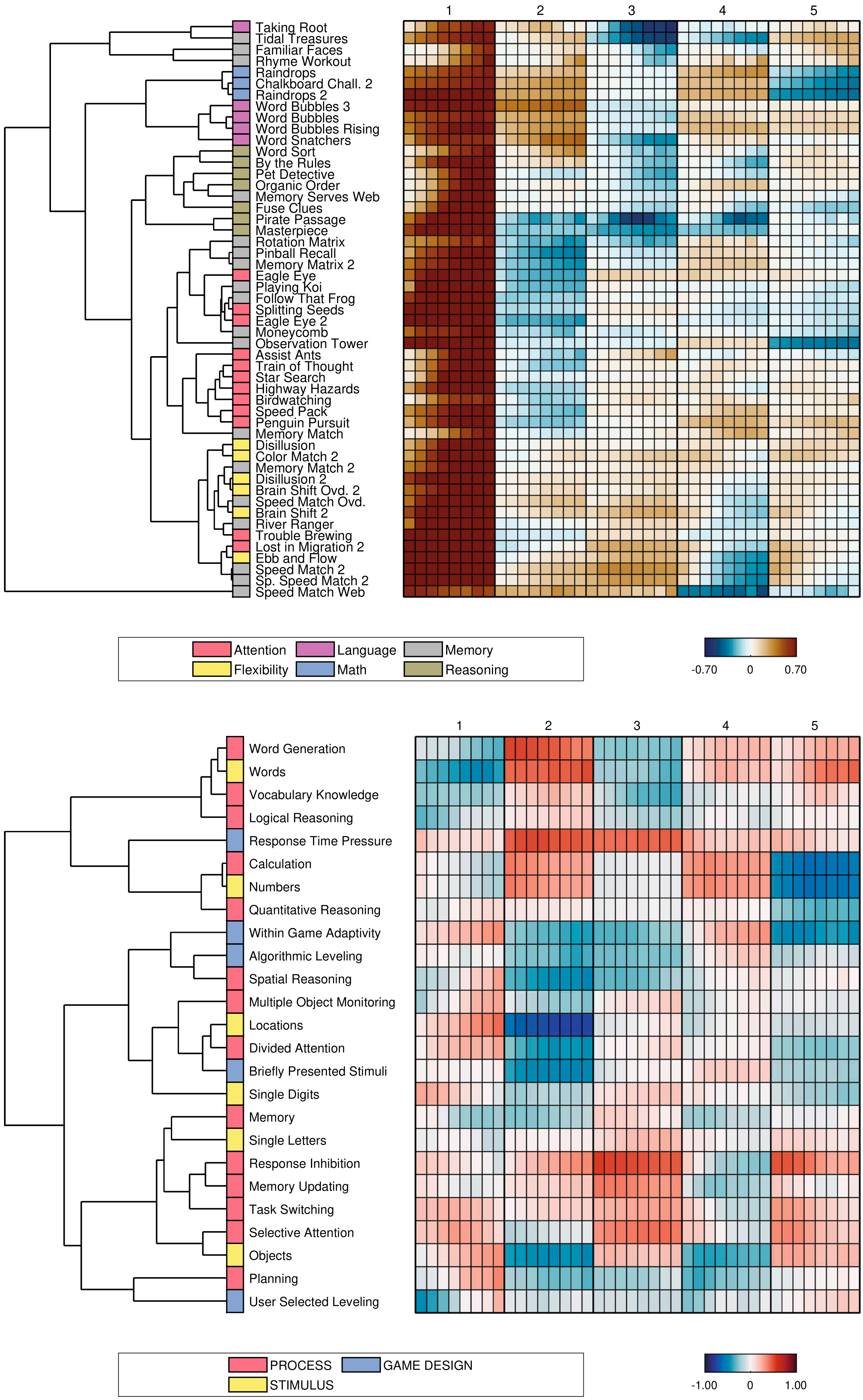
Peer review information Primary Handling Editor: Marike Schiffer.

Type/Feature	Assist Ants (Attention)	Birdwatching (Attention)	Eagle Eye (Attention)	Eagle Eye 2 (Attention)	Highway Hazards (Attention)	Lost in Migration 2 (Attention)	Penguin Pursuit (Attention)	Speed Pack (Attention)	Splitting Seeds (Attention)	Star Search (Attention)	Train of Thought (Attention)	Trouble Brewing (Attention)	Brain Shift 2 (Flexibility)	Brain Shift Overdrive 2 (Flexibility)	Color Match 2 (Flexibility)	Disillusion (Flexibility)	Disillusion 2 (Flexibility)	Ebb and Flow (Flexibility)	Taking Root (Language)	Word Bubbles (Language)	Word Bubbles 3 (Language)	Word Bubbles Rising (Language)	Word Snatchers (Language)	Chalkboard Challenge 2 (Math)	Raindrops (Math)	Raindrops 2 (Math)	Familiar Faces (Memory)	Follow That Frog (Memory)	Memory Match (Memory)	Memory Match 2 (Memory)	Memory Matrix 2 (Memory)	Memory Serves Web (Memory)	Moneycomb (Memory)	Observation Tower (Memory)	Pinball Recall (Memory)	Playing Koi (Memory)	Rhyme Workout (Memory)	River Ranger (Memory)	Rotation Matrix (Memory)	Spatial Speed Match 2 (Memory)	Speed Match 2 (Memory)	Speed Match Overdrive (Memory)	Speed Match Web (Memory)	Tidal Treasures (Memory)	By the Rules (Reasoning)	Fuse Clues (Reasoning)	Masterpiece (Reasoning)	Organic Order (Reasoning)	Pet Detective (Reasoning)	Pirate Passage (Reasoning)	Word Sort (Reasoning)
PROCESS																																																			
Memory	○	●	●	●	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	○	○	○	○	●
Memory Updating	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	○	●	○	○	○	○	●	●	○	●	●	●	●	○	○	○	○	○	○	○	○
Multiple Object Monitoring	●	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Selective Attention	○	●	●	●	○	●	○	○	○	●	○	○	●	●	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Divided Attention	●	●	●	●	●	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	●	●	●	○	●	○	○	○	○	○	○	○	○	○	○	○	●	○
Task Switching	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Vocabulary Knowledge	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●
Word Generation	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Planning	●	○	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	●	●	○
Calculation	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○
Quantitative Reasoning	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○
Spatial Reasoning	●	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	●	○	○	○	●	○	○	○	○	○	○	○	●	●	●	●	○
Logical Reasoning	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	●	○	○	●
Response Inhibition	○	○	○	○	○	●	●	○	○	○	○	○	●	●	●	●	●	●	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	●	○	○	●	●	●	●	○	●	○	○	○	○	○	●
STIMULUS																																																		
Objects	●	●	●	●	●	●	○	●	●	●	●	●	○	○	○	●	●	●	○	○	○	○	○	○	○	○	●	○	●	●	●	●	○	○	●	●	○	●	●	●	●	●	●	●	●	○	●	●	●	●	○
Locations	●	●	●	●	●	○	●	●	●	○	●	●	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	●	○	○	●	○	●	●	●	●	○	●	●	●	○	○	○	○	○	○	●	●	●	●	○
Single Letters	○	●	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Words	○	●	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	●	●	●	●	●	○	○	○	●	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	●
Single Digits	○	○	●	●	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Numbers	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○	○	○
INPUT METHOD																																																		
Cursor Keys	○	○	○	○	●	●	●	○	○	○	○	○	●	●	●	○	○	●	○	○	○	○	○	●	○	○	○	○	●	●	○	○	○	○	○	○	●	○	○	●	●	●	●	○	●	○	○	○	○	○	●
Keyboard Entry	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	○	●	●	●	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Mouse Pointer	●	●	●	●	○	○	○	●	●	●	●	●	○	○	○	●	●	○	●	○	○	○	○	○	○	○	●	●	○	○	●	○	●	●	●	●	○	●	●	○	○	○	○	●	○	●	●	●	●	●	○
GAME DESIGN																																																		
Within Game Adaptivity	○	○	●	●	●	○	●	●	●	●	●	●	○	○	○	○	●	○	●	○	●	○	●	●	●	●	○	●	○	○	●	●	●	●	●	●	○	●	●	○	○	○	○	○	○	●	●	●	○	●	○
Algorithmic Leveling	○	○	●	●	●	○	○	●	○	○	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	○	○	○	●	○	●	●	●	○	○	○	●	○	○	○	○	○	●	○	●	●	○	○	○
User Selected Leveling	●	●	○	○	○	○	○	○	○	●	●	●	○	○	○	●	●	○	○	○	○	○	●	○	○	○	●	○	○	○	○	●	○	○	○	●	●	●	○	○	○	○	○	○	○	●	○	○	●	●	●
Response Time Pressure	●	○	○	○	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	○	○	●	●	○	●	○	○	○	○	●	○	○	●	●	●	●	○	○	○	○	○	○	○	○
Briefly Presented Stimuli	○	●	●	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	●	●	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○

Extended Data Fig. 1 | Task features for cognitive tasks grouped by process, stimulus, input method, and game design. Solid circles indicate the presence of the task feature. The primary cognitive area associated with each task is given between parentheses.

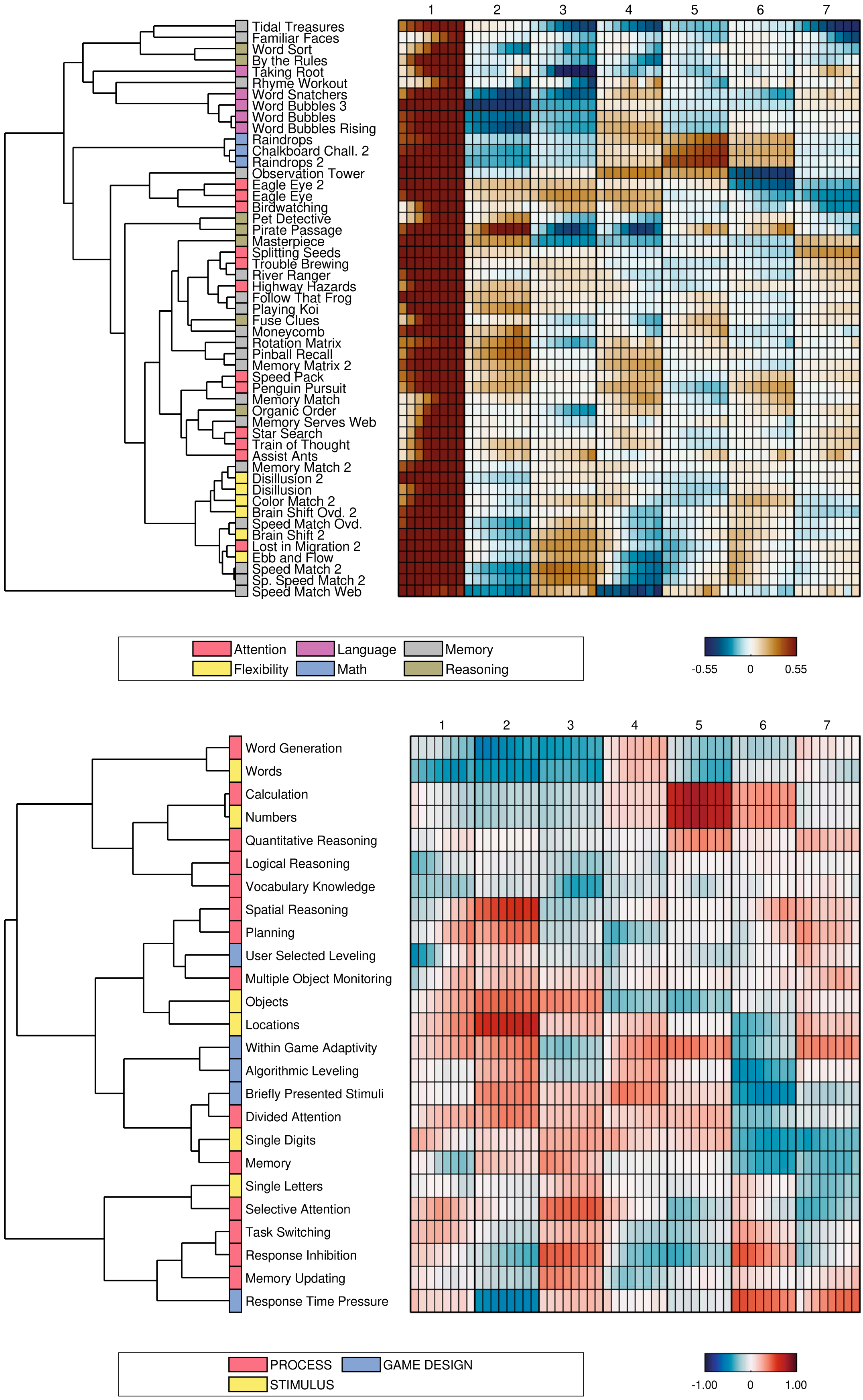
Task Feature	Feature Type	Description
Memory	Process	The task requires recall, recognition, and/or transformation of information that is no longer shown
Memory Updating	Process	The task involves updating a representation (or representations) of information as additional stimuli are presented
Multiple Object Monitoring	Process	The task requires monitoring multiple, independently moving, task-relevant objects over time
Selective Attention	Process	The task requires attending to some sources of information while ignoring others at the same time
Divided Attention	Process	The task requires integration of information from multiple sources presented concurrently
Task Switching	Process	The task requires switching between different rules for mapping from the stimulus to the response
Vocabulary Knowledge	Process	The task requires knowledge of the meaning of words or semantic relationships between words
Word Generation	Process	The task requires producing or identifying words
Planning	Process	The task requires selecting and sequencing a set of actions to reach a goal
Calculation	Process	The task requires the application of elementary mathematical operations
Quantitative Reasoning	Process	The task requires evaluating the mathematical relationship between two or more numbers
Spatial Reasoning	Process	The task requires perception of relationships between objects in space and reasoning about spatial transformations
Logical Reasoning	Process	The task requires recognizing patterns in symbolic information to arrive at unstated conclusions
Response Inhibition	Process	The task requires suppression of previously learned mappings between stimulus and response
Objects	Stimulus	Stimulus involves depictions of objects (other than words or numbers)
Locations	Stimulus	Spatial positions are the targets of cognitive processes
Single Letters	Stimulus	Stimulus involves letters of the alphabet as symbols or for property evaluation (e.g. vowel/consonant)
Words	Stimulus	Stimulus involves words
Single Digits	Stimulus	Stimulus involves single digits (0-9) as symbols with no numerical value or for property evaluation (e.g. odd/even)
Numbers	Stimulus	Stimulus involves numerical values
Cursor Keys	Input Method	Input method requires cursor keys (up/down/left/right)
Keyboard Entry	Input Method	Input method requires keyboard
Mouse Pointer	Input Method	Input method requires mouse pointer movements
Within-Game Adaptivity	Game Design	Task difficulty can change within a gameplay according to an adaptivity algorithm
Algorithmic Leveling	Game Design	Task difficulty can change between gameplays and difficulty to start a gameplay is determined by an adaptivity algorithm
User-Selected Leveling	Game Design	Task difficulty can change between gameplays. Users can unlock more difficult levels and can play at any unlocked level of difficulty
Response Time Pressure	Game Design	The game session and/or individual trials are limited to a fixed amount of time
Briefly Presented Stimuli	Game Design	Stimulus presentations are limited to a fixed amount of time

Extended Data Fig. 2 | Glossary of task features. Definitions are provided for each process, stimulus, input method and game design feature used to describe the cognitive tasks.



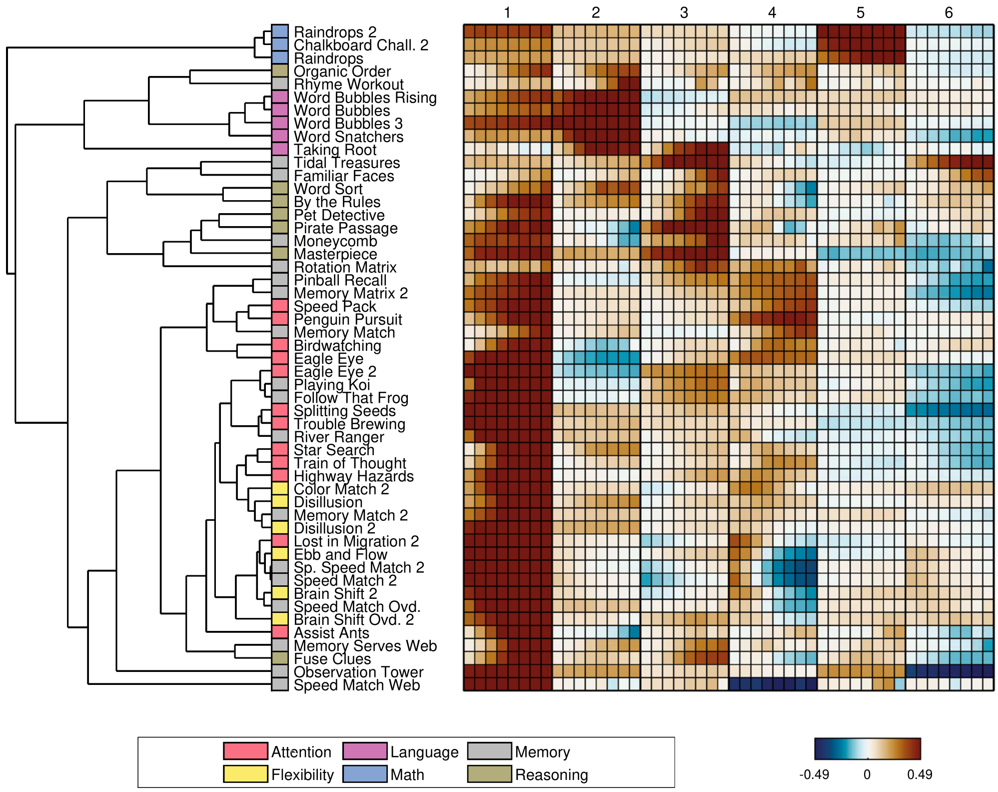
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Inferred latent learning factors with five factors. Top panel: the heatmap visualizes the latent learning factors (columns) across games (rows). Positive (negative) values are visualized by brown (blue) colours. Each latent learning factor corresponds to a group of 8 columns, where the columns within a group correspond to different stages of practice (practice increases from left to right). Cognitive tasks are coloured according to primary task domain. Bottom panel: correlations between latent learning factors and manually derived task features. Positive (negative) correlations are illustrated by red (blue) colours. The hierarchical tree visualizes the similarity between the task features on the basis of the pattern of correlations between task feature and latent factors. Task features are grouped by process, stimulus, and game design features.

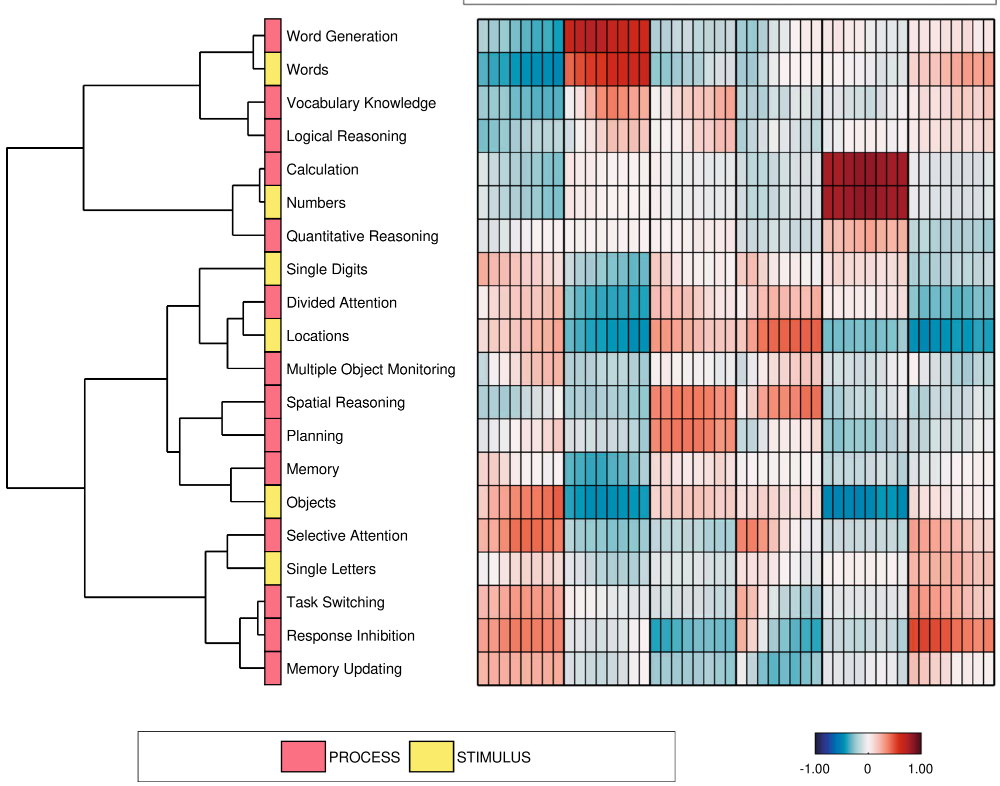


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Inferred latent learning factors with seven factors. Top panel: the heatmap visualizes the latent learning factors (columns) across games (rows). Positive (negative) values are visualized by brown (blue) colours. Each latent learning factor corresponds to a group of 8 columns, where the columns within a group correspond to different stages of practice (practice increases from left to right). Cognitive tasks are coloured according to primary task domain. Bottom panel: correlations between latent learning factors and manually derived task features. Positive (negative) correlations are illustrated by red (blue) colours. The hierarchical tree visualizes the similarity between the task features on the basis of the pattern of correlations between task feature and latent factors. Task features are grouped by process, stimulus, and game design features.

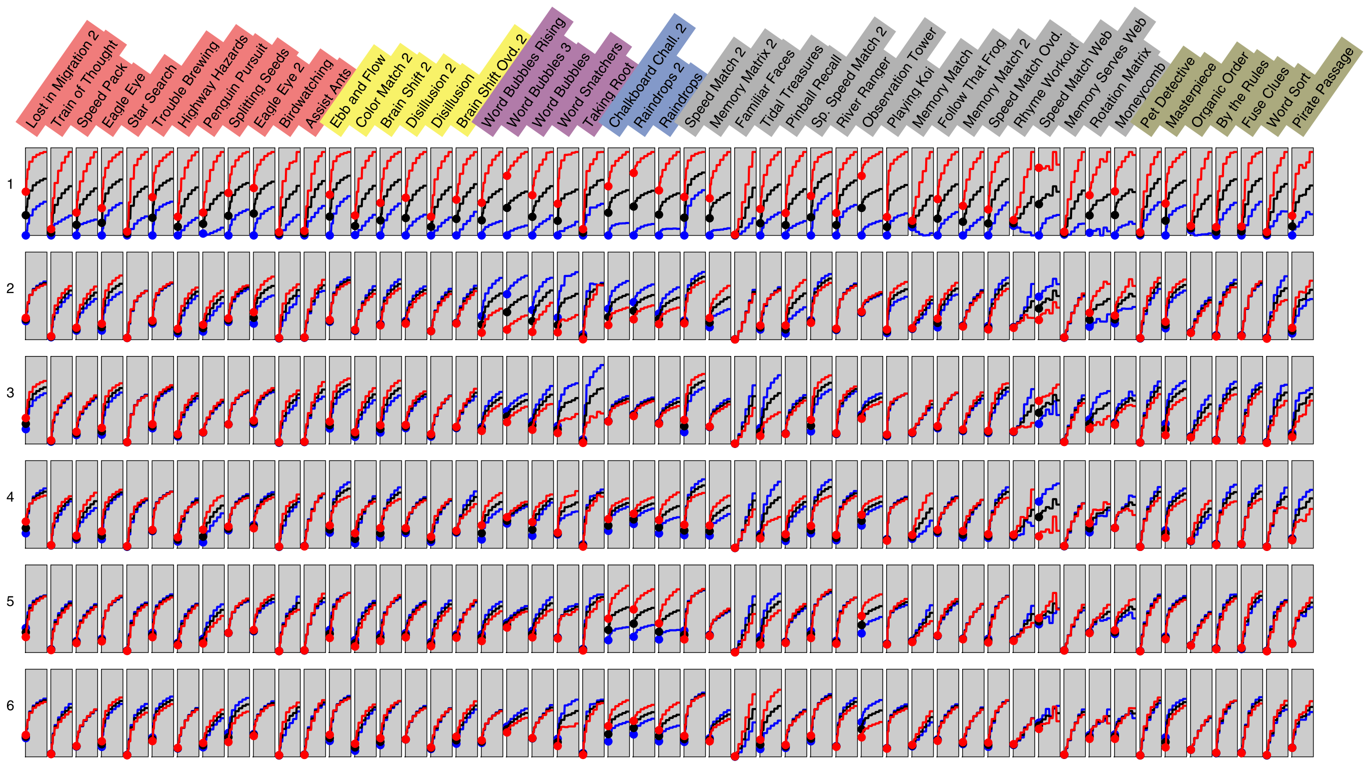


	1	2	3	4	5	6
1	1.00	0.41	0.40	0.33	0.36	0.01
2	0.41	1.00	0.28	0.17	0.26	0.05
3	0.40	0.28	1.00	0.25	0.14	-0.24
4	0.33	0.17	0.25	1.00	0.16	-0.13
5	0.36	0.26	0.14	0.16	1.00	0.04
6	0.01	0.05	-0.24	-0.13	0.04	1.00



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Inferred latent learning factors after oblique rotation (promax) that results in non-orthogonal factors. Top panel: the heatmap visualizes the latent learning factors (columns) across games (rows). Positive (negative) values are visualized by brown (blue) colours. Each latent learning factor corresponds to a group of 8 columns, where the columns within a group correspond to different stages of practice (practice increases from left to right). Cognitive tasks are coloured according to primary task domain. The matrix shows the correlations between factors. Bottom panel: correlations between latent learning factors and manually derived task features. Positive (negative) correlations are illustrated by red (blue) colours. The hierarchical tree visualizes the similarity between the task features on the basis of the pattern of correlations between task feature and latent factors. Task features are grouped by process and stimulus features.



Extended Data Fig. 6 | Changes in predicted learning curves based on changes in individual scores for individual factors. For each factor (row), the blue, black, and red lines show the predicted learning curves for 10, 50, and 90 percentile scores on that particular factor and median scores on all other factors. Therefore, the red (blue) lines show the predicted learning curves for individuals who score high (low) on that factor while holding constant the contribution of other factors. Intercept differences are highlighted with circles at the start of the learning curve. To facilitate comparison, the vertical axis is the same for each particular task (column) but is different across tasks. Tasks are ordered by primary domain (colour).

nature research

Corresponding author(s): Mark Steyvers and Robert J. Schafer

Last updated by author(s): 6/11/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	We used code for Bayesian PCA based on the paper as described by Ilin & Andraiko (2010) in the Journal of Machine Learning Research (vol 11, pp. 1957–2000). The code is included in repository we make available at https://osf.io/g9zkf

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The original and preprocessed versions of the data can be accessed at: https://osf.io/g9zkf

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We analyze the latent structure of learning trajectories from 36,297 individuals as they learned 51 different tasks on the Lumosity online cognitive training platform. Our study is a retrospective data analysis of pre-existing data. The data is longitudinal and tracks user performance in each task over time. The data is cross-sectional as it allows comparisons of different age groups and educational backgrounds. Finally, the data is multivariate in the sense that for each user, the data includes performance scores across 51 different cognitive tasks.
Research sample	We performed a retrospective analysis of a sample of gameplay data based on 84 distinct cognitive games. The de-identified data sample provided by Lumosity includes the gameplay event history for 36,297 individuals, 1,442,555 individual learning curves, and 50,374,056 gameplay events.
Sampling strategy	This data sample was selected from the overall Lumosity database by focusing on users who have at least 500 lifetime gameplays. In addition, because the focus of our research is on analyzing correlations across cognitive tasks, this data set only included users if 5% or fewer of their gameplays were repetitions of the previous game. This criterion excludes individuals who choose to repeat specific games rather than play from the broader game library. In addition, individuals were selected who signed up between Aug 1, 2013 and December 31st, 2016, with an age at signup between 18 and 90. The country of origin was restricted to the US, Canada, or Australia with English as the user's preferred language. Finally, individuals were included if at least 99% of their lifetime Lumosity gameplays were on the web product (as opposed to mobile apps) and only the web data was included in the data sample. In total, the data set contains the full gameplay history across tasks spanning a period from Aug 1, 2013 and June 30, 2019. The majority of users (80%) signed up with Lumosity before Jan 1 2015. Users in this sample spent a median of 2 years on the platform.
Data collection	The Lumosity platform provides a number of games that tap memory, attention, flexibility, speeded processing, and problem solving. In the Lumosity program, individuals are given a recommended daily training session of five different cognitive training games. One five-game session takes approximately 15 minutes to complete. Outside of the training sessions, Lumosity users can also opt to select and play games directly from the entire library of available games. For the purpose of our analysis, we only analyzed the performance score at the end of each gameplay and not the raw data of individual decisions/button presses within an individual gameplay.
Timing	The data is based on a sample of user activity between Aug 1, 2013 and June 30, 2019.
Data exclusions	As described in the Method section, for some of the cognitive tasks in the dataset, the available learning data was sparse in terms of the number of individuals who performed the task or the total number of times individuals performed the task. To ensure that sufficient data was available for our analyses, we filtered out any cognitive task that had fewer than 200 individuals at the 20th gameplay attempt. At the learning curve level, we truncated the learning curve to the first 60 gameplays. The final dataset created after these filtering steps includes the gameplay event history for 51 cognitive games, 36,297 individuals, 1,255,175 individual learning curves, and 36,736,286 single gameplay events.
Non-participation	No users declined participation. This study is a retrospective data analysis of pre-existing data.
Randomization	Users were not allocated into experimental groups (this was not a randomized study).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging