nature human behaviour

Two-dimensional neural geometry underpins hierarchical organization of sequence in human working memory

Received: 22	eceived: 22 April 2024		
Accepted: 2 October 2024			
	: 7	2 @4	

Ying Fan $0^{1,2,3}$, Muzhi Wang $0^{1,2,3}$, Fang Fang $0^{1,2,3,4,5}$, Nai Ding^{6,7} \otimes & Huan Luo $0^{1,2,3,4}$

Working memory (WM) is constructive in nature. Instead of passively retaining information, WM reorganizes complex sequences into hierarchically embedded chunks to overcome capacity limits and facilitate exible behaviour. Here, to investigate the neural mechanisms underlying hierarchical reorganization in WM, we performed two electroencephalography and one magnetoencephalography experiments, wherein humans retain in WM a temporal sequence of items, that is, syllables, which are organized into chunks, that is, multisyllabic words. We demonstrate that the one-dimensional sequence is represented by two-dimensional neural representational geometry in WM arising from left prefrontal and temporoparietal regions, with separate dimensions encoding item position within a chunk and chunk position in the sequence. Critically, this two-dimensional geometry is observed consistently in di erent experimental settings, even during tasks not encouraging hierarchical reorganization in WM and correlates with WM behaviour. Overall, these ndings strongly support that complex sequences are reorganized into factorized multidimensional neural representational geometry in WM, which also speaks to general structure-based organizational principles given WM's involvement in many cognitive functions.

Working memory (WM) is constructive in nature¹⁻³. Instead of passively storing information, the WM system actively builds new representations to fulfil task-specific goals and deal with various contexts⁴⁻⁸. In fact, most high-level cognitive functions, including language, motor control, reasoning and planning, involve reorganizing pre-existing elements in ways that differ from their original formats⁹⁻¹⁴. For example, when memorizing a sequence of items such as phone numbers or words, we group them into chunks¹⁵. When forming episodic memory of a movie or a novel, a continuous stream of events tends to be

reorganized into hierarchically embedded episodes¹⁶. Reconfiguration of successive items into hierarchically embedded chunks or schemas is a major memory operation to overcome capacity limits, promote flexible goal-directed processing and facilitate generalization in novel scenarios¹⁷⁻²⁰.

Previous behavioural evidence also supports the hierarchical reorganization of sequences in WM, by revealing transposition WM errors for items occupying the same position across different chunks²¹⁻²⁵. On the basis of these findings, computational models further propose that each item in a hierarchical sequence would be represented based on two indexes: a global index indicating its affiliated chunk order and a local index denoting its position within a chunk^{26–28}. However, the neural implementation of this hierarchical reorganization in WM remains unknown. Answers to this question would shed light on the neural mechanism of WM operation and provide substantial insights into a wide range of fields given WM's involvement in almost any cognitive function.

To achieve abstract hierarchical organization, structure and content are proposed to be represented in a disentangled manner, known as factorization^{29,30}. This view has been proposed in computational models for sequence memory^{31,32} and supported by recent empirical findings^{33–36}. For instance, prefrontal cortex recordings in monkeys showed that when retaining a sequence of spatial locations in WM, each ordinal rank in the sequence occupies a subspace in multidimensional neural space, regardless of its content, that is, the memorized location³⁷. Furthermore, memorized content and sequence structure are reactivated by different triggering events during auditory WM retention, indicating their dissociated neural formats^{38,39}. In light of these findings, we postulate that WM contains factorized representations of hierarchical structure and content items, and here we aimed at examining the neural representation of abstract hierarchical structure regardless of the items being attached. Lastly, unlike most studies focusing on the encoding period when items are presented^{40,41}, our investigation centres on the maintenance period, which better elucidates the internal organizational principles of the WM system.

What types of neural representation could support the abstract hierarchical structure underlying sequence organization in WM? One straightforward possibility is a one-dimensional (1-D) clustered format, in which items within a chunk have compressed representational distance to each other, compared with items in different chunks (Fig. 1b, right middle). By contrast, however, here we propose that the hierarchical structure could also be implemented by two-dimensional (2-D) neural geometry whereby global and local ranks are separately encoded as separate dimensions spanning a 2-D space (Fig. 1b, right lower). In other words, a 1-D sequence embedded in hierarchical structures is neurally represented along two dimensions. Our hypotheses are motivated by previous findings. First, behavioural and modelling work propose that each item in a hierarchical sequence is reorganized via two indexes: a global index and a local index^{21-23,25-28}. Second, orthogonal neural geometry has been observed in various fields to minimize inference between cognitive variables^{42–44}. For instance, simple non-hierarchical sequences are encoded in near-orthogonal neural manifolds in WM, potentially reducing interference between ranks³⁷. Therefore, dissociating global and local ranks of hierarchical sequences along two neural axes would reduce their mutual interference.

To test the hypothesis, we performed two electroencephalography (EEG) and one magnetoencephalography (MEG) experiments that asked subjects to retain a series of syllable sequences, which were hierarchically organized into words and multi-word sequences, and perform a rank recalling task. We examined the neural geometry of the syllable sequence in WM using an innovative parametrical representational similarity analysis (RSA) approach. We demonstrate a 2-D factorized representation of the syllable sequence, with separate dimensions for the local rank (position of a syllable within a word) and the global rank (position of a word within a sequence). Critically, this 2-D neural geometry, originating from prefrontal and temporoparietal brain regions, is observed consistently in different stimulus settings and tasks, even when the task does not encourage hierarchical structures, and correlates with memory behaviour. Overall, these results support that the WM system can reorganize a complex linear sequence into a 2-D factorized neural representation to reveal the underlying hierarchical structure. Although only examined in linguistic contexts in the present study, the hierarchical organization finding would potentially reflect a general WM mechanism waiting to be tested in future studies.

Results

Representational geometry models of syllable sequences in WM Hierarchically organized syllable sequences were used to assess the neural representational geometry of sequence retained in WM (Fig. 1a). In Experiment 1, each trial presented 9 Chinese syllables at a constant rate of 4 Hz, which grouped into three trisyllabic words (different words in different trials). Accordingly, each syllable is associated with two ordinal ranks—a global rank (index of word in the sequence) and a local rank (index of syllable within a word). Human subjects were instructed to retain the sequence in WM. One of the nine syllables was presented during retention as the cueing syllable and subjects serially reported the global and local ranks of the cueing syllable during retrieval (balanced order across trials) (Fig. 1c).

Here we focus on the neural representational geometry of sequence in WM during the retention period and formulate three hypotheses (Fig. 1b). The 1-D chain hypothesis postulates even spacing of the neural representation of each syllable along an axis, mirroring their presentation order. The 1-D cluster hypothesis extends the 1-D chain hypothesis by additionally considering the grouping of syllables into words and postulates that syllables within a word have shorter neural representational distances than those spanning word boundaries. By contrast, the 2-D hierarchy hypothesis postulates separate neural axes for local and global ranks. The strongest form of the 2-D hierarchy hypothesis is that the two dimensions are orthogonal to each other. Under this condition, referred to as the 2-D orthogonal hypothesis, there is no interference between global-rank and local-rank representations.

We quantified the three hypotheses using a parametric representational geometry model with two free parameters, that is, the localglobal angle (LGA) and the local-global scale (LGS) (Fig. 1d, right). The LGA denotes the angle between the local and global dimensions. The 1-D chain and cluster hypotheses predict a 0° LGA, that is, local and global ranks occupy the same representational dimension. The 2-D hierarchy hypothesis predicts a non-zero LGA and the 2-D orthogonal hypothesis predicts a 90° LGA. The other parameter, LGS, denotes the ratio of representational distance between consecutive units along the local and global dimensions. The LGS is 1/3 according to the chain hypothesis and should be lower than 1/3 according to the cluster hypothesis, which assumes compressed within-word distance. The 2-D hierarchy or orthogonal hypothesis does not constrain the LGS. However, an LGS lower than 1 indicates that global rank is better discriminated by the neural response than the local rank, and vice versa. Therefore, when the neural data were fitted using the representational geometry model, the fitted parameters could reveal whether the neural representation of the syllable sequence was consistent with 1-D chain (LGA = 0° , LGS = 1/3), 1-D cluster (LGA = 0° , LGS < 1/3) or 2-D hierarchical hypothesis (LGA 0° and LGA = 90° for the 2-D orthogonal hypothesis).

We used an RSA to investigate which parameters of the representational geometry model best fit the neural data during WM retention. We focused on the neural response elicited by the cueing syllable and subsequent neutral impulse during WM retention, as both could reactivate ordinal rank information³⁹. All trials were labelled according to the continuous ordinal rank (1-9) of the cueing syllable, regardless of the identity of the syllables. Trials in which the cueing syllable was associated with identical local and global ranks (that is, the 1st, 5th, 9th syllable in the whole sequence) were excluded from the analysis, since these trials could not distinguish local and global ranks. The multivariate neural dissimilarity, that is, Mahalanobis distance between 64-channel EEG responses or 204-sensor MEG gradiometer responses in each 50 ms time window (5 values for each sensor), between neural responses to syllables of different ranks was then calculated, yielding the neural representational dissimilarity matrix (neural RDM; Fig. 1d, upper left, purple matrix).

To fit the neural RDM, a number of model RDMs (Euclidean distance) were generated based on the representational geometry model

Article



by sampling all possible pairing of LGA and LGS (Fig. 1d, upper right, black matrix). Using a regression analysis (linear mixed model), we evaluated which model RDM, that is, which parameter of the representational geometry model, best fitted the neural RDM, with fixed effects indicating across-subject effects and random effects indicating individual effects. This regression analysis was conducted at each time point, resulting in time-resolved model fitting performance, quantified by negative log-likelihood, as a function of LGA and LGS (Fig. 1d, lower panel). The model parameter leading to the lowest negative log-likelihood was selected as the best model parameter (red asterisk) quantifying the neural representational geometry of syllable sequences in WM.

Thirty-two human subjects participated in Experiment 1 with 64-channel EEG activities recorded. Subjects reported the global (96.59% ± 0.37%) and local (97.16% ± 0.40%) ranks with high accuracy. Figure 2a plots the time-resolved model fitting results (fixed effect in linear mixed model) throughout the retention period. Around 300 ms after the cueing syllable, the neural response was significantly explained by the parametric representational geometry model (300-420 ms, $P_{\text{max}} = 0.042$, false discovery rate (FDR) corrected; peak, fixed effect beta = 0.00376, t(1150) = 5.342, P < 0.001, 95% CI, 0.00238-0.00513). Critically, the best model within the time range (dashed box) supports the 2-D hierarchy hypothesis (LGA = 109°, 95% CI, 78-161°; LGS = 0.9, 95% CI, 0.57–1.48; fixed effect beta = 0.00268, t(1150) = 6.374, P < 0.001, 95% CI, 0.00186-0.00351).

To further confirm these findings, Akaike Information Criterion (AIC) values for the best model (LGA = 109°, LGS = 0.9) were compared against those of the three hypotheses: 1-D chain (LGA = 0° , LGS = 1/3), 1-D cluster (LGA = 0°, LGS = 1/4 as a representative value) and 2-D orthogonal (LGA = 90°, LGS = 1 as a representative value). Consistently, the hierarchy hypothesis demonstrated superior performance, exhibiting the smallest AIC compared with the AIC value of the best model (chain, AIC = 103.641; cluster, AIC = 90.516; orthogonal,

AIC = 5.741). Crucially, the model comparison results hold at the individual level as well (Supplementary Fig. 1a). By contrast, the neural response after the neutral impulse was not significantly explained by the parametric representational geometry model (Fig. 2a, $P_{(3)} = 0.064$, FDR corrected; see Supplementary Fig. 2a for details).

In addition, we sought to directly decode global and local rank information from neural responses. Specifically, each participant's neural RDM was regressed with two model RDMs, one based on the global rank and the other based on the local rank (Fig. 2b, upper). The cueing syllable indeed triggered neural encoding of both global and local ranks (permutation test; global rank, 300–500 ms, P < 0.001, corrected; local rank, 320-430 ms, P = 0.008, cluster-based permutation corrected), while the neutral impulse only triggered global (240-370 ms, P = 0.004,cluster-based permutation corrected) but not local rank ($P_{min} = 0.410$, cluster-based permutation corrected) (Fig. 2b, lower). This might

reflect the more excitable WM state of global rank compared with local rank, in line with the classic global precedence effect^{45,46}.

In summary, these analyses demonstrated that a 1-D syllable sequence was retained in the WM according to the 2-D hierarchy hypothesis, with separate neural dimensions for the local and global ranks.

2-D hierarchy of syllable sequences with varied word length

Experiment 1 exclusively presented trisyllabic words. In Experiment 2 (Fig. 3a), we varied the word length to introduce more variability and flexibility and test the generality of the 2-D hierarchy hypothesis. Specifically, the syllable sequence contained a random combination of three word tur syllel

2-Difference in the second sec ter was added. Thirty-two subjects participated in Experiment 2 and





not reactivate local rank information in Experiments 1 and 2. Thirty subjects participated in Experiment 3 and reported the continuous rank (96.04% \pm 0.63%) with high accuracy.

Experiment 3 replicated the main findings of Experiments 1 and 2, regardless of the critical task change (Fig. 4b). The cueing syllable elicited neural responses that could be well captured by the parametric

representational geometry model but in a more sustained manner (240–920 ms, best model fixed effect beta = 0.00141, t(2428) = 6.919, P < 0.001, 95% CI, 0.00101–0.00181). The fitted parameter was consistent with the 2-D hierarchy hypothesis (LGA = 71°, 95% CI, 74–141°; LGS = 0.87, 95% CI, 0.41–1.39), and the result was confirmed by model comparison (chain, AIC = 25.729; cluster, AIC = 25.488; orthogonal, AIC = 8.107).

https://doi.org/10.1038/s41562-024-02047-8

We further performed a multidimensional scaling (MDS) analysis on the neural RDM, which exhibits a clear hierarchical organization of the syllable sequence in WM (Fig. 4b, right dashed box, upper panel).

Moreover, we conducted a searchlight analysis (permutation test, corrected) in the sensor space and revealed one cluster within the left central and frontal channels that are engaged in the 2-D geometry representation of hierarchical structures (Fig. 4b, right dashed box, bottom-left panel). The same searchlight analysis conducted in the source space (see Methods for details) demonstrated the involvement of prefrontal cortex, temporoparietal areas and insula, with left hemispheric lateralization (Fig. 4b, right dashed box, bottom-right panel). EEG recordings in Experiments 1 and 2 showed similar spatial patterns (Supplementary Fig. 4a). These findings suggest that instead of arising from dissociated regions encoding global and local ranks separately (Supplementary Fig. 3), the 2-D geometry originates from neural operations within same brain regions. Notably, the absence of 2-D geometry during encoding (Supplementary Fig. 5) supports a constructive WM process whereby a sequence of items is gradually reorganized into a 2-D format from encoding to retention to guide subsequent behaviour.

Finally, consistent with Experiments 1 and 2, both global and local ranks were triggered by the cueing syllable (Fig. 4c, permutation test; global rank, 230–310 ms, P = 0.014, 620–690 ms, P = 0.023, 810–920 ms, P = 0.003, cluster-based permutation corrected; local rank, 340–460 ms, P = 0.003, cluster-based permutation corrected). The searchlight analysis further confirmed that global and local ranks are processed in overlapping brain areas (Supplementary Fig. 3).

Overall, even in a continuous-report task that encourages the subject to store the 1-D sequence of syllables as it is, the brain still automatically reorganizes the 1-D syllable sequence into a 2-D neural representational geometry in WM.

2-D hierarchical organization in WM behaviour

On top of MEG evidence, behavioural responses during the continuous-rank report task also supported the hierarchy hypothesis. First, the reaction time (RT) calculated for each of the 9 cueing positions showed that primacy and recency effects were observed both at the local and the global scales, that is, shorter reaction time for the first and last word (Global 1 versus Global 2, t(29) = -4.284, P < 0.001, Cohen's d = 0.800, 95% CI, - to -0.095; Global 3 versus Global 2, t(29) = -2.297, P = 0.020, Cohen's d = 0.419,95% CI, - to -0.011) and shorter reaction time for the first and last syllable within each word (Local 1 versus Local 2. t(29) = -3.707, P = 0.001. Cohen's d = 0.677. 95% CL - to -0.044: Local 3 versus Local 2, t(29) = -4.078, P < 0.001, Cohen's d = 0.745, 95% CI, - to -0.058) (Fig. 4e). Two-way repeated ANOVA analyses supported the contribution of both global and local ranks to RTs (global rank, F(1.364, 39.569) = 15.050, P < 0.001, partial $\eta^2 = 0.343$; local rank, F(1.358,39.390) = 13.098, P < 0.001, partial $\eta^2 = 0.311$; interactions, $F(3.083,89.405) = 15.758, P < 0.001, \text{ partial } \eta^2 = 0.352).$

Fig. 4 | Experiment 3. All results here are based on MEG recordings from

N = 30 subjects. a, Experimental procedure (identical to Experiment 1, but with the task of reporting the continuous rank, 1–9, of the cueing syllable). b, Same as Fig. 2a, with the addition of MDS results of the averaged neural RDM (upper panel, right dashed inset) and sensor-/source-level searchlight results for the 2-D orthogonal geometry model (lower panel, right dashed inset). Statistical significance was assessed using non-parametric sign-permutation tests (one-sided) for each sensor and brain parcel. Sensor-level results were corrected with cluster-based permutation (P < 0.05, 10,000 permutations, significant clusters marked by asterisks). Source-level results were FDR corrected (P < 0.01, significant parcels shown in colour). Source-level searchlight analysis includes 29 subjects (one lacked anatomical MRI data). c, Same as Fig. 2b. d, Behavioural correlates of neural RDMs with the three hypotheses (1-D chain, 1-D cluster and 2-D orthogonal). Subjects are grouped based on their HRS index. Right: continuous rank report accuracy of Higher_{HRS} and Lower_{HRS} groups (independent

The reported error profiles also support the 2-D hierarchy hypothesis. As illustrated in Fig. 4f (left), the 1-D chain and 2-D hierarchy hypotheses predict different error patterns (EPs), which were separately referred to as the continuous error (CE) and the hierarchical error (HE). Specifically, if the sequence is retained as a 1-D chain as the task required, a syllable tends to be confused with its neighbouring syllables in the sequence, for example, the 4th syllable would be reported as the 3rd or 5th syllable (Fig. 4f, left, CE, blue arrow). By contrast, if syllables are organized based on a 2-D hierarchical representation, a syllable can also be confused with its neighbours along the global rank axis, that is, the 4th syllable confused with the 1st or 7th syllable (Fig. 4f, left, HE, orange arrow). The observed errors showed a pattern consistent with HE profile (Fig. 4f, middle-left, EP). To further quantify the results, we built two RDMs based on CE and HE hypotheses (Fig. 4f, middle-right, error RDM), and regressed the overall behavioural EP by CE and HE RDMs (Fig. 4f, right). The results support the greater contribution of hierarchical structure to behaviour (bootstrapping test; $\beta_{CF} = 1.025$, 95% CI, 0.440–1.769; $\beta_{\rm HE}$ = 3.523, 95% CI, 2.507–4.998). Together with similar behavioural results in Experiments 1 and 2 (Supplementary Fig. 4c,d), the findings provide converging behavioural evidence supporting hierarchical organization in WM.

Neural representational geometry predicts WM behaviour

Finally, we examined whether the neural representational geometry was related to WM behaviour in individual subjects. Since the continuous-report task in Experiment 3 motivated subjects to maintain a 1-D chain neural representation, a 2-D hierarchical neural geometry, although observed at the population level, could in principle adversely affect the continuous-report behavioural performance. We thereby hypothesized that individuals showing 1-D chain neural geometry commensurate with task requirements demonstrated higher accuracy on the continuous-report task than those showing 2-D hierarchical neural geometry.

To test this hypothesis, we regressed individual neural RDM with the model RDMs derived from the 1-D chain and 2-D orthogonal hypotheses (Fig. 4d, left panel), based on which we computed a hierarchical representational strength (HRS) index for each participant, defined as the difference between the regression beta values of the 2-D orthogonal and 1-D chain hypotheses. Consistent with our hypothesis, subjects with higher HRS demonstrated lower report accuracy (94.64% \pm 4.22%) than subjects with lower HRS values (97.45% \pm 1.73%) on the continuous rank report task (Fig. 4d, right panel; independent *t*-test; *t*(27) = -2.368, *P* = 0.012, Cohen's *d* = 0.874, 95% CI, - to -0.55%). Interestingly, the neural-behavioural correlation in Experiments 1 and 2 showed an opposite pattern (Supplementary Fig. 4b). Specifically, participants with stronger hierarchical neural representations exhibited significantly higher reporting accuracy. The different neural-behavioural relationship is due to different tasks, such that global and local rank report in

t-test, one-sided). One outlier (beyond 3 s.d. from the mean) was excluded from the statistics ($N_{\rm HigherHES}$ = 14, $N_{\rm LowerHES}$ = 15). **e**, Hierarchical structure in the RT pattern. Left: RT by cueing syllable's continuous rank (two-way repeated measures ANOVA). Right: RT by global/local rank, marginalizing local/global rank (paired *t*-tests, one-sided, Bonferroni corrected). **f**, Hierarchical structure in error pattern. Left: illustration of CE based on the 1-D chain hypothesis (blue arrows) and HE based on the 2-D hierarchy hypothesis (orange arrows). Middle: observed behavioural error patterns (pooled across 30 subjects) for each of the 9 continuous ranks in 3 × 3 matrices, combined into a 9 × 9 matrix. Right: regression of observed behavioural error patterns with RDM_{CE} and RDM_{HE} (red line, regression coefficients; error bars, 95% bias-corrected and accelerated confidence intervals from 5,000 bootstraps; dots denote bootstrapped samples). Boxes (**d** and **e**) represent the interquartile range (IQR) with the median shown as a horizontal line. Whiskers extend to 1.5 times the IQR, and outliers are marked by crosses. Each dot represents one subject.



Experiments 1 and 2 encouraged 2-D structure, while continuous report in Experiment 3 does not, as the 2-D structure needs to be transformed to 1-D continuous sequence.

Taken together, subjects automatically reorganize syllable sequences into a 2-D hierarchical neural representational geometry in WM even when the reorganization is unfavourable to the task. The findings also further substantiate the strong link between neural representational geometry in WM and memory behaviour.

Discussion

It has long been proposed that WM organize information into hierarchically embedded chunks, yet the underlying neural implementation remains unknown. In three human EEG/MEG experiments, we provide converging evidence that hierarchically organized structures underlying a 1-D sensory sequence are represented by 2-D neural representational geometry in WM, whereby each item in the sequence is reorganized along two dimensions, separately corresponding to chunk order and the order of an item within a chunk. The 2-D geometry, arising from prefrontal and temporoparietal regions, is automatically constructed regardless of task demands and could predict behaviour.

Hierarchical organization is found in a wide range of human behaviours, including language, music, knowledge, motor control, memory, decision-making and planning^{12,13,47-49}. Humans indeed excel at tracking the hierarchical structure embedded in stimuli^{50,51}, inferring hierarchically generative rules⁵² and making hierarchically structured plans^{14,19}. This hierarchy-based organizational principle offers a compelling solution to balance capacity limits and flexible operations in WM, such as compressing information storage, decreasing interference between levels and boosting retrieval efficiency by top-down signals^{18,19,53,54}. Here, even in a task that encourages a 1-D representation by reporting continuous 1-D rank, behavioural measurements (that is, RT and reported error profile) exhibit hierarchical structures, consistent with previous findings^{21–23,25,55–57}, and global/local ranks could be separately decoded from neural activities. These findings add substantial evidence for the intrinsic hierarchical nature of WM.

Most importantly, we demonstrate that this hierarchical reorganization in WM is neurally implemented through 2-D geometry rather than 1-D clustered geometry, although both operations could achieve WM chunking. Previous modelling works propose that items in a hierarchical sequence are associated with a global and a local index $^{26-28}$. advocating a 2-D representation in mental space. Our findings support this view and provide novel insight into its neural implementations. Notably, hierarchically organized sequences do not necessarily correspond to 2-D geometry in neural space. Brain activities could track information at multiple hierarchical levels^{50,58}, yet this could not reveal how those levels are related to each other. For example, one study demonstrated neural encoding of local and global ranks during motor preparation, highlighting their relevance to WM capacity and showing that within-chunk ranks are aligned across chunks⁵⁹. Our study extends these findings by quantitatively and directly revealing the underlying neural geometry that links items within a hierarchical sequence across various settings. Specifically, the 2-D neural geometry constitutes an abstract, unified neural space that can reveal the internal chunk organization of items, regardless of the content. This neural geometry supports the representation of chunks of varying lengths (that is, Experiment 2) and in different tasks (as compared between Experiments 1 and 3). The close association with WM behaviour further confirms its central role in structure-based WM operations. Notably, the 2-D factorized encoding strategy is proposed to minimize inference between representational dimensions

their neural geometry in WM. In other words, by sampling neural representations of various items via cueing syllables as probes in different trials, we could illustrate how items in the sequence are stored in WM. Finally, the neural profile could not be attributed to action planning, for example, to-be-reported global and local ranks. Experiment 3 does not require global and local reports but both show neural representations. Moreover, typical action-related neural signals precede motor execution^{84,85}, while here the neural responses emerge immediately after the cueing syllable and do not persist into action execution.

Taken together, we leverage the innate hierarchical structure of linguistic materials and demonstrate 2-D geometrical representations of complex sequences in WM and their impact on behaviour. The findings, in a broad sense, illustrate how the brain automatically reorganizes incoming information into an appropriate structure that facilitates information storage and retrieval, and would provide potential insights into a wide range of domains given WM's involvement in almost any cognitive function.

Methods

Ethics

This research complied with all relevant ethical regulations and received approval from the Departmental Ethical Committee of Peking University. All subjects gave written informed consent before taking part in the experiments. Each subject received compensation for his/her participation in the form of either 60 RMB per hour or course credits.

Subjects

Thirty-two (18 females and 14 males, mean age 20.8, range 18–27 years), 32 (17 females and 15 males, mean age 21.3, range 18–27 years) and 30 (11 females and 19 males, mean age 21.2, range 18–27 years) native listeners of Mandarin Chinese with no history of neurological or psychiatric disorders were recruited in Experiments 1, 2 and 3, respectively. No subject participated in more than one experiment. All subjects had normal or corrected-to-normal vision and none of them had any known auditory disorders.

Apparatus of EEG experiments (Experiments 1 and 2)

The whole experiment was performed via Psychophysics Toolbox-3 (ref. 86) for MATLAB (MathWorks) using custom scripts. Visual stimuli were presented on a 32 inch Display++ LCD screen, operating at a resolution of $1,920 \times 1,080$ pixels and a refresh rate of 120 Hz. Auditory stimuli (that is, syllable sequences and cueing syllables) were generated in advance using custom MATLAB codes. During the experiment, these auditory stimuli were delivered to the subjects via a Sennheiser CX300S earphone that was connected to an RME Babyface Pro external sound card. Subjects' head positions were secured using a chin rest, which was positioned 75 cm away from the screen. Subjects' responses for the tasks were collected using a standard QWERTY keyboard.

Apparatus of the MEG experiment (Experiment 3)

The whole experiment was also controlled via Psychophysics Toolbox-3 (ref. 86) for MATLAB (MathWorks) using custom scripts. Visual stimuli were projected onto a 32 inch rear projection screen, positioned 75 cm away from the subjects, using a projector. The projector operated at a resolution of $1,920 \times 1,080$ pixels and a refresh rate of 60 Hz. The pre-generated auditory stimuli (that is, syllable sequences and cueing syllables) were delivered to the subjects via an air-tube earphone that was connected to an RME Babyface Pro external sound card. Subjects' responses were collected using a response pad.

Auditory stimuli

In the current research, auditory stimuli were syllable sequences and cueing syllables. All stimuli were presented in Mandarin Chinese, where each character corresponds to a single syllable and combinations of characters (that is, syllables) form words conveying specific meanings. In both Experiments 1 and 3, syllable sequences were formed by concatenating three trisvllabic words. In Experiment 2, syllable sequences were constructed by randomly selecting and concatenating three words from a pool comprising disyllabic words, trisyllabic words and quadrisyllabic idioms. We ensured that each syllable sequence did not contain any syllables with the same pronunciation. It is notable that both the fixed word-rate sequences used in Experiments 1 and 3 and the varied word-rate sequences used in Experiment 2 are naturally and effortlessly chunked into meaningful units for native Chinese speakers. For example, a sequence like 'in-cre-di-ble-buil-ding-lu-lla-by' (4-2-3 sequence) in English would be easily organized into three words by an English speaker. Similarly, there is no ambiguity in the chunking of these sequences for Chinese subjects. Although the syllable sequences could be structured into three multisyllabic words, each of the constituent syllables was independently synthesized using the Neospeech synthesizer (the male voice, Liang). The synthesized syllables had a duration range of 168-397 ms (mean 250 ms) in Experiment 1, 168-400 ms (mean 251 ms) in Experiment 2 and 155-401 ms (mean 250 ms) in Experiment 3. To achieve uniformity, all syllables in Experiments 1 and 3 were adjusted to a consistent length of 250 ms, either by truncating the syllable or adding silence at the end. Those syllables in Experiment 2 were adjusted to 230 ms. The final 25 ms of each syllable were smoothed by a cosine window. Upon the synthesis of constituent syllables, they were successively concatenated without any temporal gaps, forming the syllable sequences for both Experiments 1 and 3. For the syllable sequences in Experiment 2, a random temporal jitter ranging from 0 ms to 40 ms was introduced between two successive syllables. At present, all syllable sequences have been generated. Given the acoustic independence of constituted syllables, the hierarchical structure of syllable sequences (syllable-word-sequence) could only be extracted through semantic knowledge (Fig. 1a), not prosodic cues. In this set-up, each syllable within the syllable sequence carries two pieces of structural information: local rank information (the ordinal rank of the syllable within its multisyllabic word) and global rank information (the ordinal rank of the multisyllabic word within the entire sequence). All texts and sound files of the syllable sequence and cueing syllables are available here (https://osf.io/drzuy/#).

Experimental procedure (Experiment 1)

The initiation of a trial was signalled by the presentation of a black cross (0.9° in visual angle), positioned centrally against a grey background (RGB = (128,128,128)). This cross remained in place throughout the entirety of a trial, with the exception of during the neutral impulse presentation and the report screen (Fig. 1c). Subjects were instructed to maintain their gaze on the central cross and minimize eye blinking throughout each trial. During the encoding phase of the WM task, subjects were presented with pre-generated syllable sequences via auditory delivery. The task required subjects to memorize these sequences in the order they were presented. During the maintenance phase, two types of trigger events, an auditory cueing syllable, which was one of the syllables delivered during the encoding phase, and a visual neutral impulse (the visual angle is 18°), were used to probe the neural representation of hierarchical structure. The rationale is that each syllable is associated with specific hierarchical structure information during the encoding phase, based on the current syllable sequence. As a result, when a syllable is presented during the maintenance period, its associated hierarchical structure information could be reactivated. Meanwhile, the visual neutral impulse is designed to perturb the WM network, thereby reactivating the maintained information. These two trigger events were proved efficient in probing sequential temporal structures (1st, 2nd and 3rd) in our previous research³⁹. The subjects' task was to report both the global and local ranks of the cueing syllable. The order in which these ranks were reported was randomized for each trial and was shown on the report screen. For each report, subjects were instructed to use their index, middle or ring fingers of their right hand

to press the 'j', 'k' or 'l' keys, respectively. Notably, to prevent motor response preparation during the maintenance phase, the correspondence between the ranks (1st, 2nd and 3rd) and the response keys ('j', 'k' or 'l') was randomly assigned for each trial, as indicated on the report screen. In Experiment 1, we delivered only those syllables as cueing syllables that corresponded to inconsistent global and local ranks (that is, global 1 and local 2 (G1L2), G1L3, G2L1, G2L3, G3L1 and G3L2). This was done to avoid scenarios where the global and local ranks of a cueing syllable are identical. In such cases, subjects could complete the experimental task by representing a single rank information, without specifying whether it pertains to a global or local rank. This could potentially introduce ambiguity in discerning whether the neural representation is related to global or local hierarchical information. Experiment 1 consisted of a total of 432 trials. To mitigate fatigue, a mandatory break of at least 1 minute was instituted after every 30 trials.

Experimental procedure (Experiment 2)

Experiment 2 closely mirrors the design of Experiment 1, with the key distinction being that the syllable sequences in Experiment 2 exhibit irregular hierarchical structures (Fig. 3b; for details, see 'Auditory stimuli'). Furthermore, certain timing parameters were modified to manage the overall duration of the experiment (Fig. 3a). The response keys were also expanded to four keys, with the 'j', 'k', 'l' and ';' keys corresponding to the index, middle, ring and little fingers of the right hand, respectively. Experiment 2 comprised a total of 486 trials.

Experimental procedure (Experiment 3)

The hierarchical structure of the syllable sequences in Experiment 3 is identical to that of Experiment 1, both of which are constructed by concatenating three trisyllabic words. However, the specific syllable sequences used in Experiment 3 were newly generated. Most importantly, the behavioural task was to report the continuous rank of the cueing syllable within the syllable sequence, which ranged from 1 to 9, at the end of each trial (Fig. 4a). Contrary to tasks that involve reporting the global and local ranks of the cueing syllable in Experiments 1 and 2, the task in Experiment 3 does not encourage any hierarchical structuring of the syllable sequence explicitly. Therefore, Experiment 3 could eliminate the possibility that findings from Experiments 1 and 2 are solely confined to the hierarchical rank report task. In addition, Experiment 3 used syllables corresponding to all continuous ranks (1-9) as the cueing syllable. Meanwhile, we chose to retain only cueing syllable as the trigger event in Experiment 3 given its proven efficiency in reactivating hierarchical structures in both Experiments 1 and 2. Experiment 3 comprised a total of 459 trials.

Experimental materials (Experiment 1)

In Experiment 1, a total of 81 unique 3-syllable words were used in Experiment 1, with each word presenting an average of 16 times throughout the experiment (s.e. = 0.28). All 81 words were used as targets (where the cued syllable belonged), with each being cued an average of 5.3 times (s.e. = 0.079). Across all 432 trials, the cued syllable consisted of 241 different pronunciations, with each unique syllable being cued an average of 1.79 times (s.e. = 0.026).

Experimental materials (Experiment 2)

In Experiment 2, 248 different words were used, each appearing an average of 5.88 times throughout the experiment (s.e. = 0.066). Of these, 237 words served as targets, with each being cued an average of 2.051 times (s.e. = 0.073). Across 486 trials, the cued syllable included 338 different pronunciations, with each unique syllable being cued an average of 1.44 times (s.e. = 0.047).

Experimental materials (Experiment 3)

Experiment 3 used 301 different words, each appearing an average of 4.57 times throughout the experiment (s.e. = 0.029). All 301

words served as targets, with each being cued an average of 1.52 times (s.e. = 0.033). Across 459 trials, the cued syllable comprised 413 different pronunciations, with each unique syllable being cued an average of 1.11 times (s.e. = 0.016).

EEG acquisition and pre-processing (Experiment 1)

The EEG signals were acquired using a 64-channel EasyCap (Brain Products). The data were recorded using two BrainAmp amplifiers (Brain Products) and the BrainVision Recorder software (Brain Products) at a frequency of 500 Hz. Throughout the entire EEG recording process, the impedance of all electrodes was maintained below 10 k . The recorded EEG data were subsequently pre-processed offline utilizing FieldTrip⁸⁷ in MATLAB 2022a. The data were segmented into epochs extending from 200 ms before the onset of the syllable sequence to 500 ms following the onset of the report screen. These epochs were then baseline-corrected using the mean activity from 50 ms to 100 ms before the onset of the syllable sequence as the baseline for subtraction. Following this, the data were re-referenced to the average value across all channels, downsampled to a frequency of 100 Hz and subjected to a band-pass filter within the 1-30 Hz range. Independent component analysis using FastICA algorithm was conducted to eliminate components associated with eye movement and other artefacts, such as bad channels and heartbeat. The remaining components were back-projected into the EEG channel space for subsequent analysis.

EEG acquisition and pre-processing (Experiment 2)

The acquisition of EEG data in Experiment 2 was carried out in the same manner as in Experiment 1. The pre-processing procedure was also largely identical, with the exception that epochs were extracted from 2500 ms before the onset of the cueing syllable to 700 ms following the offset of the neutral impulse in Experiment 2.

MEG acquisition and pre-processing (Experiment 3)

Neuromagnetic data were acquired using a whole-head MEG system with 204 planar gradiometers and 102 magnetometers (Elekta Neuromag system) in a magnetically shielded room. The MEG experiment was divided into ten blocks, each followed by a brief break. Before the commencement of each block, the position of the subject's head was estimated using index coils placed at four points on the head. This position was then compared with the initial position recorded at the start of the experiment to ensure that any head movement did not exceed 4 mm throughout the experiment. Magnetic field strength was sampled at a frequency of 1,000 Hz. The recorded MEG data were subsequently pre-processed offline utilizing MNE-Python tools⁸⁸. Initially, the data underwent de-noising and motion correction using the Maxfilter Signal Space Separation method. Subsequently, the data were band-pass filtered to a range of 1-40 Hz and downsampled to a frequency of 100 Hz. Epochs were then extracted, specifically from 600 ms before the onset of the syllable sequence to 600 ms following the onset of the report screen. Finally, ICA using FastICA algorithm was conducted to eliminate components associated with eye movement and heartbeat.

Source reconstruction (MEG)

Among the 30 participants who completed the MEG data recording, 29 underwent individual anatomical imaging using a 3 T GE Discovery MR750 MRI scanner (GE Healthcare). Only these 29 participants with individual anatomical images were included in the source reconstruction analysis. Each participant's MEG data were co-registered with their structural MRI, and boundary element model (BEM) models were generated using the FreeSurfer watershed algorithm. The forward model was computed using a surface-based source space with 4,096 vn, time window of -800 ms to -200 ms relative to cueing syllable onset, and the data interval was defined as -90 ms to 1,000 ms.

Source reconstruction (EEG)

The main processes are the same as MEG source reconstruction. However, as no anatomical images were collected in the EEG experiments, we used the standard template MRI subject 'fsaverage' provided by MNE-Python instead.

RSA based on the parametric representational geometry model

Neural RDM. Each entry in the neural RDM represents the Mahalanobis distance between neural representations of two hierarchical ranks, which were reactivated by cueing syllable (Fig. 1d, upper left). In the computation of neural RDM, each trial was labelled based on the hierarchical rank associated with the cueing syllable. For instance, consider a three trisyllabic word sequence to be memorized as 'Zhi-NanZhenYaoLanQuTuJiDui', where the cueing syllable is 'Zhen'. Given that the 'Zhen' is the third syllable of the first word, this trial would be labelled as global 1 local 3, abbreviated as G1L3. Furthermore, a subsampling of trials was performed. Specifically, for each cueing syllable with a unique pronunciation, we randomly selected trials corresponding to its appearance at one specific position in one subsampling. This procedure ensured that within all sampled trials, no cueing syllables shared the same pronunciation. Consequently, the computed neural dissimilarity across different hierarchical ranks was not influenced by the recurrence of the same cueing syllable pronunciation across various ranks, which further ensured that the results reflected a genuine representation of the hierarchical structure, dissociable from the content. Meanwhile, the trial subsampling process ensured an approximately equal number of trials for each hierarchical rank to preserve comparability across conditions. For subjects with the fewest trials, each hierarchical rank included 28 trials. Finally, the calculation of the neural RDM incorporated both spatial and temporal features⁸⁹. Specifically, at each time point, signals from all sensors (64 EEG channels for Experiments 1 and 2, and 204 gradiometers for Experiment 3) within a forward 50 ms time window (5 values, one for each of the five 10 ms windows) were included as spatiotemporal features. This resulted in 320 features (that is, 64×5) for the EEG experiments and 1,020 features (that is, 204 × 5) for the MEG experiment at each time point.

We utilized an eight-fold cross-validated approach to compute the neural RDM^{38,90}. The previously subsampled trials were partitioned into eight folds, ensuring an approximately equal distribution of trials for each hierarchical rank within each fold. Subsequently, one fold was taken as testing dataset, while the remaining seven folds constituted training dataset. The condition-specific spatiotemporal neural pattern was computed by averaging the trials that corresponded to identical hierarchical ranks within the training dataset. Subsequently, the Mahalanobis distance was computed between the spatiotemporal neural activity of each individual trial in the testing dataset and each condition-specific spatiotemporal neural pattern that was derived from the training dataset. Finally, the original Mahalanobis distances were adjusted by subtracting the mean distance between each testing trial and all training conditions for each testing trial (that is, mean-centred). The above procedure was iteratively performed until each of the eight folds had served as testing dataset. Upon completion, we obtained a neural RDM of dimensions corresponding to the number of trials by the number of conditions (that is, hierarchical ranks). Finally, distances of trials corresponding to identical hierarchical ranks were averaged, resulting in a condition-by-condition neural RDM. The entire procedure, from trial subsampling to the generation of the final condition-by-condition neural RDM, was performed 300 times. The resulting 300 neural RDMs were then averaged to yield the final neural RDM at each time point.

Parametric representational geometry model. It is proposed that the representational geometry for abstract hierarchical structure can be projected onto a 2-D plane (Fig. 1d, upper right). This plane is defined by two axes, each representing a distinct level of rank information: one corresponds to high-level global ranks, and the other to low-level local ranks. The specific geometry can be quantified using two parameters: LGA and LGS. The LGA represents the angle between representational

es14.6(ulting in a modee RDM. Tfr)14.6(ughd the s)4.699999809(s14.6is)4.6(s)

rm t15.50e 2inOfinreents) we were(a).6bol t15.50ect15.50estu tas seris of15

the neural RD viah apliear mixFed modeetck asrG nts the neural RDM,(as adlcaet)10.1(ed b)10.0(y thesm lle s)5.199999809t ne

n th o(t).5(eb).5(s)10.399999618(e).5(t)-4.6(w)15.399999618(o..5(a-5.4(a-5.4))) archical ranky can beicacul(at)10(ed b)10tased on tias defined geometry,

G = 1/yand hierarcss fome fom; e G = 90 G = 1 G prn

The aboveRcA bed on temparametric representational geosy modee was performed at each time poin d(uring te WMeemaiat)1(es)]TJ O T

 $o illumindateg \ temparametric \ mode e(f) - 9.899999618 (ntlinspr) 15 (es) 5 (ulterative and the second second$

are suitable inputs for the classical MDS method. Finally, by applying classical MDS, we projected the averaged symmetric RDM across all subjects and significant time points into a three-dimensional space, ultimately deriving the final neural geometry.

Model-based neural decoding analyses

To directly investigate the reactivation patterns of both global and local rank information for cueing syllables during the maintenance period, we implemented model-based neural decoding analyses (Fig. 2b, Fig. 3c and Fig. 4c, left). Specifically, for each subject individually, the time-resolved neural RDM was regressed using two predictors: the global rank RDM and the local rank RDM. Each regression coefficient's () time courses were smoothed with a Gaussian-weighted window (window length = 40 ms). A regression coefficient () significantly exceeding zero signifies that corresponding information is represented in the neural activities. We used a non-parametric sign-permutation test⁹¹ to conduct statistical analyses on the regression coefficients. Specifically, the sign of regression coefficients for each subject at each time point was subjected to 100,000 random flips, which process allowed the generation of null distribution of the population mean . The P value of the observed population mean was then estimated from this null distribution (one-sided). To adjust for multiple comparisons over time, a cluster-based permutation test was conducted, with a cluster-forming threshold of P < 0.05. We performed cluster-based permutation tests within specific time windows following each triggering event: from -40 ms to 1,000 ms after the cueing syllable (Experiments 1-3) and from -40 ms to 900 ms after the neutral impulse (Experiments 1 and 2). These window lengths were chosen based on previous studies indicating that various types of WM representations reactivated by a triggering event consistently last less than 1 second (refs. 39,76,77). However, the window for the neutral impulse was shortened to 900 ms, the maximum duration possible, because the report screen appeared 900 ms after the neutral impulse in both Experiments 1 and 2.

Searchlight analyses (sensor level)

A searchlight analysis was performed to identify MEG sensors (gradiometers) involved in representing the 2-D orthogonal geometry, global rank and local rank. For each sensor, we computed a neural multivariate RDM incorporating neural activities from this sensor and its neighbours (averaging 8.33 neighbours, range 4-11) within a 50 ms time window as features at each time point. Next, the neural RDMs across significant time windows identified for 2-D hierarchy (240-920 ms; Fig. 4b, left) were averaged and regressed with the orthogonal model RDM (LGA = 90° and LGS = 1). This was performed for each sensor, yielding the spatial distribution of neural representation strength. Similar procedures were applied to both global and local ranks, within the corresponding significant time windows (global 230-920 ms, local 340-460 ms; Fig. 4c, left). Model-based neural decoding analyses (Fig. 2b, upper) were then conducted on the averaged neural RDMs to derive the spatial topography for each rank. Finally, cluster-based permutation tests (Monte Carlo method for cluster-based permutation; cluster-forming threshold P < 0.05) were conducted to identify sensor with significantly higher regression coefficients compared with the group median value (one-sided).

We applied the same analyses to the EEG data in Experiments 1 and 2, using each EEG sensor and its neighbouring sensors (averaging 6.5 neighbours, range 3–8) as features to investigate the spatial topography of 2-D orthogonal geometry representational strength. Statistics (cluster-based permutation tests) were applied to the aggregated results of Experiments 1 and 2.

Searchlight analyses (source level)

To investigate the spatial topography of 2-D orthogonal geometry representational strength at the source level, we conducted whole-brain

analyses for both MEG and EEG experiments. We parcellated the 8,192 vertices into 400 parcels according to the Schaefer atlas⁹². For each parcel, we conducted RSA as in the sensor level, using all vertices within each parcel as features. Owing to the computational complexity at the source level, temporal information was not included in the features. Non-parametric sign-permutation test was conducted to identify parcels with significantly stronger 2-D orthogonal representation compared with the group median value (one-sided). We applied FDR correction at P < 0.01 to account for multiple comparisons across the 400 parcels. Owing to the lower signal-to-noise ratio in EEG and the absence of individual MRI structural images for co-registration, the results for Experiments 1 and 2 were combined, and the statistical analyses were restricted to the brain parcels where MEG showed significant findings (P < 0.05).

Correlations between brain and behaviour

To quantify the predictive power of neural representational strengths on behavioural responses, we conducted a median split analysis for three experiments. Specifically, for each subject, we first averaged the neural RDM within the identified significant time window (Fig. 2a for Experiment 1; Fig. 3d for Experiment 2; Fig. 4b for Experiment 3). This averaged RDM was then individually regressed against the three predefined RDMs corresponding to the chain model, cluster model and orthogonal model. Subsequently, we computed the HRS by calculating the difference in regression coefficients between the orthogonal (task irrelevant) and chain (task relevant) models. Finally, subjects were divided into two groups based on a median split of their HRS values. To assess the relationship between HRS and behaviour, we compared the behavioural response accuracy of the higher HRS group with the lower HRS group using an independent t-test (one-sided). Extreme data points (beyond 3 s.d. from the mean) were excluded from the statistics (one subject from the combined Experiments 1 and 2; one subject from Experiment 3).

Model-based behavioural decoding analyses

To examine whether the hierarchical organization of the syllable sequence is reflected in the patterns of error responses, we carried out model-based behavioural decoding analyses (Fig. 4f and Supplementary Fig. 4d). Erroneous responses from all subjects were collectively aggregated. For each continuous rank (1-9), we calculated the distribution of erroneous responses, which we denote as the EP. Regression analysis was applied to the EP using two predictors. The first, CE RDM, represents a scenario where a syllable's rank could be mistaken as its adjacent ranks in a 1-D chain (for example, the fourth syllable could be mistaken as the third or fifth). The second predictor, HE RDM, encapsulates instances where a syllable's rank could be confused with its adjacent ranks along the global rank dimension in a 2-D hierarchical form (for example, the fourth syllable could be mistaken as the first or seventh). A significant regression coefficient () implies the existence of a corresponding organization within the EP. We used a bootstrapping method (N = 5,000) for the statistical analysis of the regression coefficients, ensuring that the current results were not disproportionately influenced by a single subject. Meanwhile, given our focus on incorrect responses, the diagonal elements of the matrices were omitted from the analyses. What is more, it is worth noting that mistaking the fourth item for the fifth could also be a condition where diffusion occurs along the local rank dimension in a 2-D hierarchical form. However, we currently classify this scenario as a CE, which results in an underestimation of HE strength and an overestimation of CE strength.

Reporting summary

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

Data availability

Data supporting main findings of the study are available at https://osf.io/drzuy/#.

Code availability

The code illustrating key analyses of the study can be found here https://osf.io/drzuy/#.

References

- Schacter, D. L., Norman, K. A. & Koutstaal, W. The cognitive neuroscience of constructive memory. *Annu. Rev. Psychol.* 49, 289–318 (1998).
- Brady, T. F. & Alvarez, G. A. Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychol. Sci.* 22, 384–392 (2011).
- Prabhakaran, V., Narayanan, K., Zhao, Z. & Gabrieli, J. D. E. Integration of diverse information in working memory within the frontal lobe. *Nat. Neurosci.* 3, 85–90 (2000).
- van Ede, F. & Nobre, A. C. Turning attention inside out: how working memory serves behavior. *Annu. Rev. Psychol.* 74, 137–165 (2023).
- 5. Kwak, Y. & Curtis, C. E. Unveiling the abstract format of mnemonic representations. *Neuron* **110**, 1822–1828 (2022).
- Lee, S.-H., Kravitz, D. J. & Baker, C. I. Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat. Neurosci.* 16, 997–999 (2013).
- Ehrlich, D. B. & Murray, J. D. Geometry of neural computation unifies working memory and planning. *Proc. Natl Acad. Sci. USA* 119, e2115610119 (2022).
- Gelastopoulos, A., Whittington, M. A. & Kopell, N. J. Parietal low beta rhythm provides a dynamical substrate for a working memory bu er. *Proc. Natl Acad. Sci. USA* 116, 16613–16620 (2019).
- Caucheteux, C., Gramfort, A. & King, J.-R. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nat. Hum. Behav.* 7, 430–441 (2023).
- 10. de Saussure, F. *Course in General Linguistics* (Columbia Univ. Press, 2011).
- 11. Georgopoulos, A. P. Higher order motor control. *Annu. Rev. Neurosci.* **14**, 361–377 (1991).
- 12. Sarafyazd, M. & Jazayeri, M. Hierarchical reasoning by neural circuits in the frontal cortex. *Science* **364**, eaav8911 (2019).
- Balaguer, J., Spiers, H., Hassabis, D. & Summerfield, C. Neural mechanisms of hierarchical planning in a virtual subway network. *Neuron* 90, 893–903 (2016).
- 14. Koechlin, E. & Jubault, T. Broca's area and the hierarchical organization of human behavior. *Neuron* **50**, 963–974 (2006).
- Thalmann, M., Souza, A. S. & Oberauer, K. How does chunking help working memory? *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 37–55 (2019).
- 16. Farrell, S. Temporal clustering and sequencing in short-term memory and episodic memory. *Psychol. Rev.* **119**, 223–271 (2012).
- 17. Brady, T. F., Konkle, T. & Alvarez, G. A. A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* **11**, 4 (2011).
- Geddes, C. E., Li, H. & Jin, X. Optogenetic editing reveals the hierarchical organization of learned action sequences. *Cell* 174, 32–43.e15 (2018).
- Tomov, M. S., Yagati, S., Kumar, A., Yang, W. & Gershman, S. J. Discovery of hierarchical representations for e icient planning. *PLoS Comput. Biol.* 16, e1007594 (2020).
- Wu, S., Élteto, N., Dasgupta, I. & Schulz, E. Learning structure from the ground up—hierarchical representation learning by chunking. *Adv. Neural Inf. Process. Syst.* 35, 36706–36721 (2022).
- Ng, H. L. H. & Maybery, M. T. Grouping in short-term verbal memory: is position coded temporally? *Q. J. Exp. Psychol. A* 55, 391–424 (2002).

- 22. Ryan, J. Grouping and short-term memory: di erent means and patterns of grouping. *Q. J. Exp. Psychol.* **21**, 137–147 (1969).
- 23. Ryan, J. Temporal grouping, rehearsal and short-term memory. *Q. J. Exp. Psychol.* **21**, 148–155 (1969).
- 24. Brown, G. D. A., Preece, T. & Hulme, C. Oscillator-based memory for serial order. *Psychol. Rev.* **107**, 127–181 (2000).
- 25. Hartley, T., Hurlstone, M. J. & Hitch, G. J. E ects of rhythm on memory for spoken sequences: a model and tests of its stimulus-driven mechanism. *Cogn. Psychol.* **87**, 135–178 (2016).
- 26. Henson, R. N. A. Short-term memory for serial order: the start-end model. *Cogn. Psychol.* **36**, 73–137 (1998).
- 27. Lee, C. L. & Estes, W. K. Item and order information in short-term memory: evidence for multilevel perturbation processes. *J. Exp. Psychol. Hum. Learn. Mem.* **7**, 149–169 (1981).
- 28. Hurlstone, M. J. Functional similarities and di erences between the coding of positional information in verbal and spatial short-term order memory. *Memory* **27**, 147–162 (2019).
- 29. Behrens, T. E. J. et al. What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
- 30. Whittington, J. C. R. et al. The Tolman-Eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249–1263.e23 (2020).
- Hitch, G. J. Temporal grouping e ects in immediate recall: a working memory analysis. *Q. J. Exp. Psychol. A* 49, 116–139 (1996).
- Burgess, N. & Hitch, G. J. A revised model of short-term memory and long-term learning of verbal sequences. *J. Mem. Lang.* 55, 627–652 (2006).
- Liu, Y., Dolan, R. J., Kurth-Nelson, Z. & Behrens, T. E. J. Human replay spontaneously reorganizes experience. *Cell* 178, 640–652. e14 (2019).
- Kornysheva, K. et al. Neural competitive queuing of ordinal structure underlies skilled sequential action. *Neuron* **101**, 1166–1180 (2019).
- Guidali, G., Pisoni, A., Bolognini, N. & Papagno, C. Keeping order in the brain: the supramarginal gyrus and serial order in short-term memory. *Cortex* 119, 89–99 (2019).
- 36. Zhou, J. et al. Evolving schema representations in orbitofrontal ensembles during learning. *Nature* **590**, 606–611 (2021).
- 37. Xie, Y. et al. Geometry of sequence working memory in macaque prefrontal cortex. *Science* **375**, 632–639 (2022).
- Fan, Y., Han, Q., Guo, S. & Luo, H. Distinct neural representations of content and ordinal structure in auditory sequence memory. *J. Neurosci.* 41, 6290–6303 (2021).
- 39. Fan, Y. & Luo, H. Reactivating ordinal position information from auditory sequence memory in human brains. *Cereb. Cortex* **33**, 5924–5936 (2023).
- 40. Gwilliams, L., King, J.-R., Marantz, A. & Poeppel, D. Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nat. Commun.* **13**, 6606 (2022).
- Goldstein, A. et al. Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* 25, 369–380 (2022).
- 42. Panichello, M. F. & Buschman, T. J. Shared mechanisms underlie the control of working memory and attention. *Nature* **592**, 601–605 (2021).
- Libby, A. & Buschman, T. J. Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* 24, 715–726 (2021).
- Weber, J. et al. Subspace partitioning in the human prefrontal cortex resolves cognitive interference. *Proc. Natl Acad. Sci. USA* 120, e2220523120 (2023).
- 45. Navon, D. Forest before trees: the precedence of global features in visual perception. *Cogn. Psychol.* **9**, 353–383 (1977).

Article

- Sanders, L. D. & Poeppel, D. Local and global auditory processing: behavioral and ERP evidence. *Neuropsychologia* 45, 1172–1186 (2007).
- Fitch, W. T. & Martins, M. D. Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N. Y. Acad. Sci.* 1316, 87–104 (2014).
- Hasson, U., Chen, J. & Honey, C. J. Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* 19, 304–313 (2015).
- 49. Khanna, A. R. et al. Single-neuronal elements of speech production in humans. *Nature* **626**, 603–610 (2024).
- Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164 (2016).
- Henin, S. et al. Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* 7, eabc4530 (2021).
- 52. Badre, D. Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends Cogn. Sci.* **12**, 193–200 (2008).
- 53. Planton, S. et al. A theory of memory for binary sequences: evidence for a mental compression algorithm in humans. *PLoS Comput. Biol.* **17**, e1008598 (2021).
- D'Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142 (2015).
- 55. Frankish, C. Modality-specific grouping e ects in short-term memory. *J. Mem. Lang.* **24**, 200–209 (1985).
- 56. Frankish, C. Perceptual organization and precategorical acoustic storage. J. Exp. Psychol. Learn. Mem. Cogn. **15**, 469–479 (1989).
- Henson, R. N. A. Positional information in short-term memory: relative or absolute? *Mem. Cogn.* 27, 915–927 (1999).
- Yokoi, A. & Diedrichsen, J. Neural organization of hierarchical motor sequence representations in the human neocortex. *Neuron* 103, 1178–1190 (2019).
- Kikumoto, A. & Mayr, U. Decoding hierarchical control of sequential behavior in oscillatory EEG activity. *eLife* 7, e38550 (2018).
- Flesch, T., Juechems, K., Dumbalska, T., Saxe, A. & Summerfield, C. Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* 110, 1258–1270 (2022).
- 61. Riggall, A. C. & Postle, B. R. The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *J. Neurosci.* **32**, 12990–12998 (2012).
- 62. Ólafsdóttir, H. F., Bush, D. & Barry, C. The role of hippocampal replay in memory and planning. *Curr. Biol.* **28**, R37–R50 (2018).
- Grosmark, A. D., Sparks, F. T., Davis, M. J. & Losonczy, A. Reactivation predicts the consolidation of unbiased long-term cognitive maps. *Nat. Neurosci.* 24, 1574–1585 (2021).
- 64. Liu, Y., Nour, M. M., Schuck, N. W., Behrens, T. E. J. & Dolan, R. J. Decoding cognition from spontaneous neural activity. *Nat. Rev. Neurosci.* 23, 204–214 (2022).
- 65. Son, J.-Y., Vives, M.-L., Bhandari, A. & FeldmanHall, O. Replay shapes abstract cognitive maps for e icient social navigation. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-024-01990-w (2024).
- Nassar, M. R., Helmers, J. C. & Frank, M. J. Chunking as a rational strategy for lossy data compression in visual working memory. *Psychol. Rev.* 125, 486–511 (2018).
- 67. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
- Servan-Schreiber, E. & Anderson, J. R. Learning artificial grammars with competitive chunking. J. Exp. Psychol. Learn. Mem. Cogn. 16, 592–608 (1990).

- 69. Norris, D. & Kalm, K. Chunking and data compression in verbal short-term memory. *Cognition* **208**, 104534 (2021).
- Chase, W. G. & Simon, H. A. Perception in chess. Cogn. Psychol. 4, 55–81 (1973).
- Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97 (1956).
- 72. Heusser, A. C., Poeppel, D., Ezzyat, Y. & Davachi, L. Episodic sequence memory is supported by a theta–gamma phase code. *Nat. Neurosci.* **19**, 1374–1380 (2016).
- 73. Oberauer, K. Removing irrelevant information from working memory: a cognitive aging study with the modified Sternberg task. *J. Exp. Psychol. Learn. Mem. Cogn.* **27**, 948–957 (2001).
- Oberauer, K. Access to information in working memory: exploring the focus of attention. J. Exp. Psychol. Learn. Mem. Cogn. 28, 411–421 (2002).
- 75. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015).
- Kandemir, G. & Akyürek, E. G. Impulse perturbation reveals cross-modal access to sensory working memory through learned associations. *NeuroImage* 274, 120156 (2023).
- Wol , M. J., Jochim, J., Akyürek, E. G. & Stokes, M. G. Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871 (2017).
- Sreenivasan, K. K., Curtis, C. E. & D'Esposito, M. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89 (2014).
- Lundqvist, M., Herman, P. & Miller, E. K. Working memory: delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* 38, 7013–7019 (2018).
- Nelli, S., Braun, L., Dumbalska, T., Saxe, A. & Summerfield, C. Neural knowledge assembly in humans and neural networks. *Neuron* 111, 1504–1516 (2023).
- 81. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299 (2002).
- 82. Polyn, S. M. & Kahana, M. J. Memory search and the neural representation of context. *Trends Cogn. Sci.* **12**, 24–30 (2008).
- Averbeck, B. B., Chafee, M. V., Crowe, D. A. & Georgopoulos, A. P. Parallel processing of serial movements in prefrontal cortex. *Proc. Natl Acad. Sci. USA* 99, 13172–13177 (2002).
- 84. Watanabe, K. & Funahashi, S. Prefrontal delay-period activity reflects the decision process of a saccade direction during a free-choice ODR task. *Cereb. Cortex* **17**, i88–i100 (2007).
- Van Ede, F., Chekroud, S. R., Stokes, M. G. & Nobre, A. C. Concurrent visual and motor selection during visual working memory guided action. *Nat. Neurosci.* 22, 477–483 (2019).
- Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? Perception 36, 1–16 (2007).
- 87. Oostenveld, R., Fries, P., Maris, E. & Scho elen, J.-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 15686

Acknowledgements

This work was supported by the National Science and Technology Innovation 2030 Major Project 2021ZD0204100 (2021ZD0204103 to H.L. and 2021ZD0204105 to N.D.), the National Natural Science Foundation of China (31930052 to H.L.), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (T2421004 to F.F.), the China Postdoctoral Science Foundation (2023M740124 to Y.F.), the National Natural Science Foundation of China (32222035 to N.D.), the National Science and Technology Innovation 2030 Major Project (2022ZD0204802 to F.F.) and the National Natural Science Foundation of China (31930053 to F.F.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper. We thank the National Center for Protein Sciences and Center for MRI Research at Peking University in Beijing, China, for assistance with data acquisition. We would like to thank D. Liu, Q. Han and J. Gao for their help during the data collection, and M. Luo and J. Li for helpful support on data analysis.

Author contributions

Y.F. and H.L. originally conceived and designed the experiments. Y.F. performed the experiments. Y.F. and M.W. analysed the data. F.F. and N.D. contributed to the experimental materials. Y.F., F.F., N.D. and H.L. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-024-02047-8.

Correspondence and requests for materials should be addressed to Nai Ding or Huan Luo.

Peer review information *Nature Human Behaviour* thanks Atsushi Kikumoto, Freek van Ede and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

 $\ensuremath{\textcircled{\sc line to springer}}$ Nature Limited 2024

nature portfolio

Corresponding author(s): Huan Luo

Last updated by author(s): Sep 26, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	nfirmed
		The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	\square	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
	\boxtimes	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
	\boxtimes	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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted Give <i>P</i> values as exact values whenever suitable.
\boxtimes		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
\boxtimes		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
 We used the 64-electrode EasyCap system and two BrainAmp amplifiers with BrainVision Recorder software (Brain Products) to acquire EEG signals. The neuromagnetic signals were acquired using a 306-sensor MEG system (102 magnetometers and 204 planar gradiometers, Elekta Neuromag TRIUX system, Helsinki, Finland). The MIR structural images were collected using a 3T GE Discovery MR750 MRI scanner (GE Healthcare, Milwaukee, WI, USA). All stimuli were controlled by custom Matlab (R2019a) code using Psychophysics Toolbox (v3.0.17.0).
 Custom Matlab code (R2022a) using FieldTrip (v20201229) and Matlab built-in Statistics and Machine Learning Toolbox. Python code using MNE-Python (1.4.2). Data supporting main findings and codes illustrating key analyses of the research are available at https://osf.io/drzuy/#.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data files supporting main findings are provided here: https://osf.io/drzuy/#

Research involving human participants, their data, or biological material

Policy information about studies with <u>human participants or human data</u>. See also policy information about <u>sex, gender (identity/presentation)</u>, and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We collected information about the (biological) sex of the participants (self-reported), and kept the gender balance of the subjects roughly Thirty-two (18 females and 14 males), thirty-two (17 females and 15 males), and thirty (11 females and 19 males) native listeners of Mandarin Chinese with no history of neurological or psychiatric disorders were recruited in Experiment 1, Experiment 2, and Experiment 3, respectively. No subject participated in more than one experiment. We did not make any analysis of the gender of the subjects since this was beyond the scope of our research.
Reporting on race, ethnicity, or other socially relevant groupings	Participants were not classified into different race, ethnicity of other social categories.
Population characteristics	Thirty-two (18 females and 14 males, mean age 20.8, range 18-27 years), thirty-two (17 females and 15 males, mean age 21.3, range 18-27 years), and thirty (11 females and 19 males, mean age 21.2, range 18-27 years) native listeners of Mandarin Chinese with no history of neurological or psychiatric disorders were recruited in Experiment 1, Experiment 2, and Experiment 3, respectively. No subject participated in more than one experiment. All subjects had normal or corrected-to-normal vision and none of them had any known auditory disorders.
Recruitment	We recruited participants through an online advertisement placed on a university campus bulletin board system (BBS). Participants were healthy adult university students who volunteered their time in exchange for compensation (money or course credits). There was no self-selection bias.
Ethics oversight	All three experiments were approved by the Departmental Ethics Committee of Peking University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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 m

Behavioural & social sciences study design

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

where trials were randomized.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

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

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the

Specimen	denosition
Speciment	ucposition

Dating methods

Indicate where the specimens have been deposited to permit free access by other researchers.

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment

Experiments of concern

1

Does the work involve any of these experiments of concern:

No	Yes
	Demonstrate how to render a vaccine ineffective
	Confer resistance to therapeutically useful antibiotics or antiviral agents
	Enhance the virulence of a pathogen or render a nonpathogen virulent
	Increase transmissibility of a pathogen
	Alter the host range of a pathogen
	Enable evasion of diagnostic/detection modalities
	Enable the weaponization of a biological agent or toxin
\square	Any other potentially harmful combination of experiments and agents

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Flow Cytometry

Plots

Confirm that:

The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

All plots are contour plots with outliers or pseudocolor plots.

A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.
Instrument	Identify the instrument used for data collection, specifying make and model number.
Software	Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.
Cell population abundance	Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.
Gating strategy	Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	resting state	
Design specifications	n/a	
Behavioral performance measure	s n/a	
Acquisition		
Imaging type(s)	structural images for MEG source-level analysis	
Field strength	ЗТ	
Sequence & imaging parameters	A three-dimensional fast spoiled gradient echo (FSPGR) sequence was applied to obtain anatomical T1-weighted MRI images (256 × 256 matrix, 192 slices, 1.00 × 1.00 × 1.00 mm3 voxel; TR = 6.70 ms, TE Min Full, FOV = 25.6 × 25.6 cm, FA = 12°) for each participant on a 3T GE Discovery MR750 MRI scanner (GE Healthcare, Milwaukee, WI, USA).	
Area of acquisition	whole brain	
Diffusion MRI Used	🔀 Not used	
Preprocessing		
Preprocessing software Freesurfer (v7.3.2)		
Normalization	n/a	
Normalization template	n/a	
Noise and artifact removal	n/a	
Volume censoring	n/a	

Statistical modeling & inference

statistical modeling & interence			
Model type and settings	n/a		
Effect(s) tested	n/a		
Specify type of analysis: 🛛 Wh	ole brain 🗌 ROI-based 🗌 Both		
Statistic type for inference	n/a		
(See <u>Eklund et al. 2016</u>)			
Correction	n/a		
Models & analysis			
n/a Involved in the study Image: Structional and/or effective connectivity Image: Structional analysis Image: Structional analysis			
Functional and/or effective conne	ectivity n/a		
Graph analysis	n/a		

Multivariate modeling and predictive analysis n/a