Article

Neuron

Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex

Highlights

- Is hierarchical motor sequence organization mapped orderly onto human neocortex?
- Individual movements are distinctively represented in primary motor cortex
- Movement chunks and sequences are jointly represented in premotor and parietal areas
- Representational clustering revealed regions along the stimulus-to-action gradient

Authors

Atsushi Yokoi, Jörn Diedrichsen

Correspondence

ayokoi@nict.go.jp

In Brief

Hierarchical movement sequences are often assumed to mirror an anatomical hierarchy of brain regions. Using representational fMRI analysis, Yokoi and Diedrichsen challenge this common assumption. Although the primary motor cortex represents elementary movements, chunk and sequence representations coexist in the premotor and parietal cortices.

Yokoi & Diedrichsen, 2019, Neuron *103*, 1–13
September 25, 2019
https://doi.org/10.1016/j.neuron.2019.06.017

CellPress

Please cite this article in press as: Yokoi and Diedrichsen, Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex, Neuron (2019), https://doi.org/10.1016/j.neuron.2019.06.017

Neuron
Article

Neural Organization of Hierarchical Motor Sequence Representations in the Human Neocortex

Atsushi Yokoi[1,2,5,6,*] and Jörn Diedrichsen[2,3,4,5]
[1]Center for Information and Neural Networks, National Institute of Information and Communications Technology, Suita, Osaka 565-0871, Japan
[2]The Brain and Mind Institute, University of Western Ontario, London, ON N6A 5B7, Canada
[3]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON N6A 5B7, Canada
[4]Department of Computer Science, University of Western Ontario, London, ON N6A 5B7, Canada
[5]Institute of Cognitive Neuroscience, University College London, London, WC1N 3AZ, UK
[6]Lead Contact
*Correspondence: ayokoi@nict.go.jp
https://doi.org/10.1016/j.neuron.2019.06.017

SUMMARY

Although it is widely accepted that the brain represents movement sequences hierarchically, the neural implementation of this organization is still poorly understood. To address this issue, we experimentally manipulated how participants represented sequences of finger presses at the levels of individual movements, chunks, and entire sequences. Using representational fMRI analyses, we then examined how this hierarchical structure was reflected in the fine-grained brain activity patterns of the participants while they performed the 8 trained sequences. We found clear evidence of each level of the movement hierarchy at the representational level. However, anatomically, chunk and sequence representations substantially overlapped in the premotor and parietal cortices, whereas individual movements were uniquely represented in the primary motor cortex. The findings challenge the common hypothesis of an orderly anatomical separation of different levels of an action hierarchy and argue for a special status of the distinction between individual movements and sequential context.

INTRODUCTION

The ability to learn and to produce complex sequences of movements is essential to many everyday skills, such as speaking or playing an instrument. Recent evidence from human fMRI studies has suggested that a widespread network of brain regions is involved in the learning and production of complex sequences, including the prefrontal cortex (PFC), the dorsal and ventral premotor cortex (PMd/v), the supplementary motor area (SMA), the precuneus, areas along the intraparietal sulcus (IPS), basal ganglia (BGs), and the cerebellum (Doyon et al., 2002; Grafton et al., 1995; Hikosaka et al., 1999; Honda et al., 1998; Penhune and Steele, 2012; Sadato et al., 1996; Wymbs et al., 2012). It remains unclear, however, exactly how these areas contribute to sequential behavior (Hikosaka et al., 1999).

A prevalent idea is that such movement sequences are organized in a hierarchical fashion, where several elementary movements are combined into units, often called "motor chunks" (Lashley, 1951). These motor chunks can then activate the generating circuits for each elementary movement, allowing faster and more accurate execution (Rosenbaum et al., 1983) as well as computational efficiency (Ramkumar et al., 2016). Production of an entire sequence could then be achieved by sequential activation of the corresponding chunks, and new sequences could be generated by recombination of learned chunks (Sakai et al., 2003). Practice of specific sequences would eventually result in the formation of a more abstract representation that combines chunks (Figure 1A). Thus, the core idea of a multi-level action hierarchy is that each level combines elements in the hierarchically lower level, facilitating the transition between lower-level elements. Each hierarchically higher level would lose some of the fine temporal details encoded at the lower level and, hence, represent the action at a more abstract level (Cooper and Shallice, 2000; Fuster, 2008; Humphreys and Forde, 1998; Miller et al., 1960).

Although there is substantial behavioral evidence (Rosenbaum et al., 1983; Sakai et al., 2003) for such hierarchical structure, we still know very little about how these hierarchies are represented in the human brain (Krakauer et al., 2019). Electrophysiological studies in non-human primates have shown that some neurons show differential firing rates during the same elementary movement, depending on the sequential context, such as preceding or following movement (Baldauf et al., 2008; Tanji and Shima, 1994). Other neurons have been found to be active at both initiation and termination of a sequence (Fujii and Graybiel, 2003) or were selective for specific categories of sequences (Shima et al., 2007). However, the use of relatively short and simple sequences in these studies precludes further assessment of different hierarchical levels of sequence representations.

There are two extreme views of how such a hierarchical representation could be implemented in the neocortex. The first and most dominant view assumes a clear anatomical separation of different levels of the behavioral hierarchy in the brain (Cooper and Shallice, 2000; Fuster, 2008; Koechlin and Jubault, 2006;

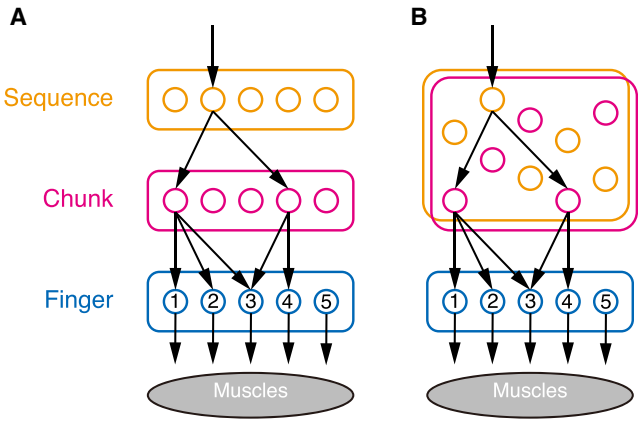

Figure 1. Two Potential Neural Architectures for Hierarchical Movement Sequence Production

(A) Spatially separable hierarchy. The example shows production of the finger sequence 12334. At the bottom of hierarchy is the generator circuit for individual finger movement. The activation of each element (circles) causes a descending command that generates the movement. The second level above specifically encodes chunks (groups of a few movements), whereas the highest level represents individual sequences, which are composed of chunks. (B) According to a spatially overlapping hierarchy, higher-level sequence representations are not separated in different brain regions but intermingled in the same areas.

Miller et al., 1960; Figure 1A). Such an organization could emerge from the fact that different regions have different intrinsic timescales on which they can learn and represent sequential dependencies. It has been suggested that basic sensory areas can represent sequences of events on the order of a few hundred milliseconds, with higher-order association areas being able to represent sequences lasting multiple seconds or even minutes (Chaudhuri et al., 2015; Chen et al., 2015; Hasson et al., 2015). Based on this idea, one would predict that the anatomical separation of motor sequence representation is determined by the differences in the timescale of individual elementary movements, chunks, and entire sequences.

An alternative view advocates that hierarchical behavior can be learned by a single recurrent network without separating different levels of representation (Botvinick and Plaut, 2004; Botvinick, 2007). Therefore, chunks and entire sequences could be represented in the same network (Figure 1B). Although this hypothesis would also predict that levels of individual movements and sequential context are intermingled, we have recently shown a clear neuroanatomical separation of these two (Yokoi et al., 2018).

To test the prediction of a clear anatomical separation between levels of representation, we designed a new behavioral paradigm in which human participants learned to produce long motor sequences that were explicitly organized on three levels; each sequence consisted of four chunks, and each chunk contained 2–3 finger presses. We then investigated the neural correlates of these three levels using representational fMRI analysis (Diedrichsen and Kriegeskorte, 2017). Rather than analyzing the increases or decreases of spatially smoothed activity, representational fMRI analysis makes inferences based on the similarity (or dissimilarity) of multivariate activity patterns across multiple experimental conditions (Kriegeskorte et al., 2008; Yokoi et al., 2018). Our paradigm therefore allowed us to experimentally manipulate the structure of motor sequence representations across three hierarchical levels and then to directly infer the representational content of each cortical area with respect to these levels.

RESULTS

Induction of a Multi-level Motor Hierarchy through Behavioral Training

Our paradigm was aimed at inducing a stable motor hierarchy by manipulating how participants built up their explicit, declarative memory of the sequence. During the fMRI experiment, we required participants to produce sequences with their right hand completely from memory, provided only with a sequence cue (Figures 2A–2C). The training taught participants to remember 8 sequences (Figure 2B; for more details, see Behavioral Training in the STAR Methods). On day 1, they practiced to produce 8 different chunks of 2 or 3 items (Figure 2C) and to associate these with a specific letter (A–H). On the second day, participants started to learn 8 different sequences, each a combination of four learned chunks. At the end of this training, participants could reliably recall most of the sequences. The number of incorrect presses they made in each sequence execution had reduced dramatically; the error rate per press was 9% ± 8% (Figure 2D). The inter-press intervals (IPIs) within each chunk quickly reduced on the first day and remained relatively stable for the following days. In contrast, IPIs for the boundaries between two successive chunks were much longer and reduced only slowly over the course of training days (Figure 2D). Even on the fifth day, the between-chunk IPIs were executed more slowly (386 ± 93 ms) than the within-chunk intervals (220 ± 62 ms, $t_{11} = -5.73$, p = 0.0001; Figure 2D). The longer between-chunk IPI is commonly taken as a behavioral indicator that the two sequence elements are represented as separate units.

Although this result provided clear evidence for chunking, it remains unclear to what degree the longer between-chunk IPIs were caused by memory retrieval and/or stable motor representation (i.e., planning ahead successive movements). To test this, we asked participants to perform a follow-up session after the fMRI scan in which the sequences were instructed not by the sequence cue but, rather, directly using digits (e.g., 13524232514; Figure 3A). This frees participants from the requirement of explicit memory recall. Nonetheless, on trained sequences, the between-chunk intervals were still longer than the within-chunk intervals (348 ± 124 ms versus 280 ± 122 ms, $t_{11} = -5.00$, p = 0.0004; Figure 3E, trained), clearly showing that the induced chunking affected motor performance over and beyond explicit recall.

Importantly, our data also show that the observed effect is not driven by differences in the biomechanical difficulty of finger transitions. The sequence design (Figure 2B) was such that 5 of the 8 sequences were the same across the two experimental groups, with the difference that they learned them using a different chunking structure. The normalized press intervals for those sequences showed a pattern that clearly reflected the original instruction (Figure 3B). The inter-subject correlation of these patterns was

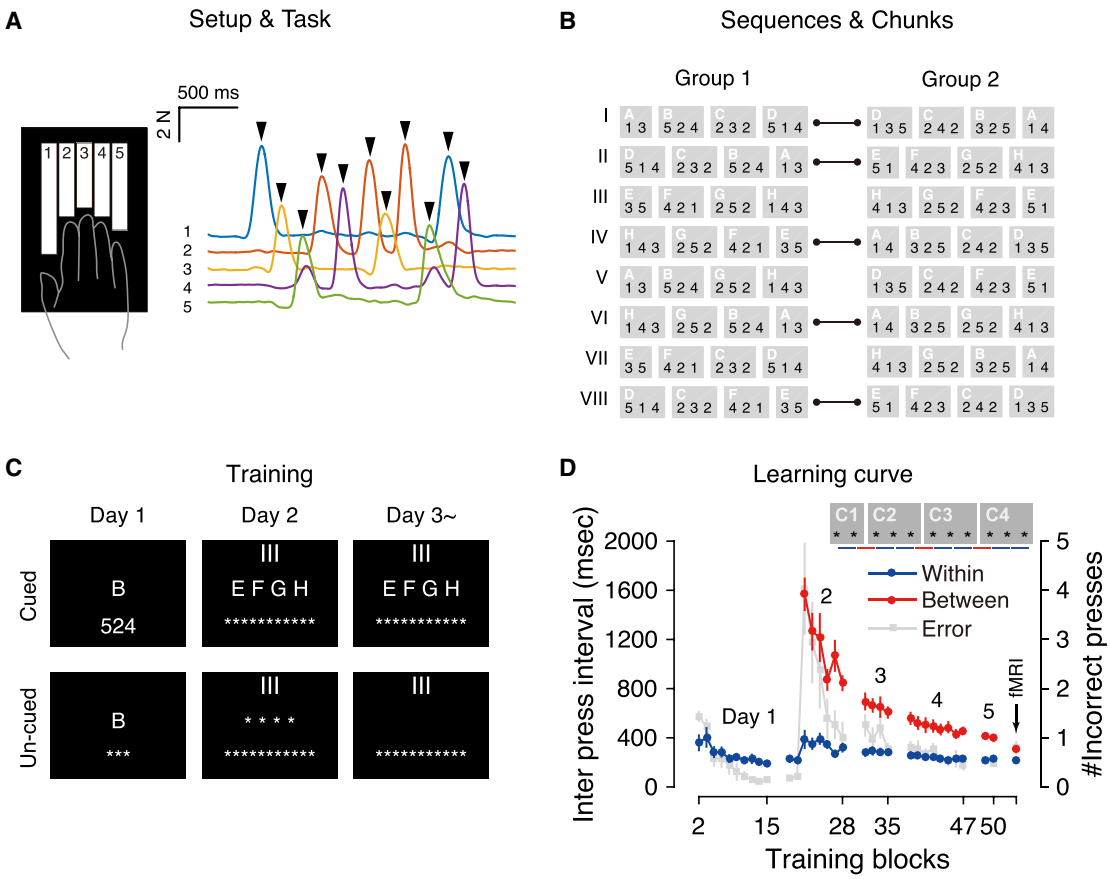

Figure 2. Explicit Induction of a Structured Sequence Memory

(A) Experimental setup. Participants practiced fast sequences of isometric finger presses on the custom-built keyboard device (left). The traces show example force data (fingers 1–5) from one sequence execution (right). Each arrowhead on peak force indicates a successful finger press.

(B) Sequences and chunks. The participants were divided into two groups, which practiced partly overlapping sequences with different chunking. Lines indicate pairs of identical sequences across the two groups.

(C) Training consisted of cued trials (top row) and un-cued trials (bottom row). On day 1, participants learned to produce single chunks from memory using a letter (A–H) cue. On the following days, they practiced sequences (indicated by Roman letters I–VIII) as combinations of learned chunks. On day 2, cued and un-cued trials were alternated. On days 3–5, cued trials and a set of three un-cued trials were alternated.

(D) Inter-press intervals over the course of the behavioral training. Within- and between-chunk intervals averaged over the sequence types are displayed as blue and red dots, respectively (axis on the left side). Average numbers of incorrect presses are indicated as gray squares (axis on the right side). An arrow indicates performance in the imaging session. Only data from un-cued trials are shown. Error bars indicate the SEM.

significantly higher for within-group than between-group comparisons (Figure 3C; $t_{11} = 9.22$, $p = 1.66 \times 10^{-6}$).

If participants acquire a motor representation of the chunks, then they should also be able to use this knowledge when producing the chunk in a novel context (Sakai et al., 2003). To test this, the participants additionally executed 3 sets of novel sequences (Figure 3D). These new sequences either consisted of trained chunks in a new order ("Chunk"), 2 trained chunks embedded in an otherwise random sequence ("Chunk+New"), or completely untrained sequences with no relation to learned chunks ("New"). As expected, IPIs in the New sequences were executed considerably slower than IPIs in trained sequences. For the other two sequence types, the intervals that lay within a learned chunk were performed significantly faster than novel intervals ($t_{11} = -4.86$, $p = 0.0005$ for Chunk; $t_9 = -4.25$, $p = 0.0021$ for Chunk+New; Figure 3D). Overall, these results suggest that the originally declaratively (i.e., cognitively) imposed chunk structure left a reliable imprint in the motor behavior and was generalized to novel contexts.

Importantly, we also observed that the between-chunk IPIs of the trained sequences were faster than the between-chunk IPIs of the Chunk sequences ($t_{11} = -6.35$, $p = 5.46 \times 10^{-5}$). This advantage may have two reasons. First, participants may have acquired a higher-order sequence representation that facilitated the transitions between chunks. Alternatively, it may be due to a form of non-hierarchical association learning (Hunt and Aslin, 2001; Reber, 1967; Stadler, 1992; Verwey and Abrahamse, 2012) at the level of the individual elements; finger transitions that had been encountered in practice would frequently become associated and, therefore, be performed faster. To disentangle these possibilities, we built a linear model for all IPIs of the follow-up session for each individual. We tested three effects

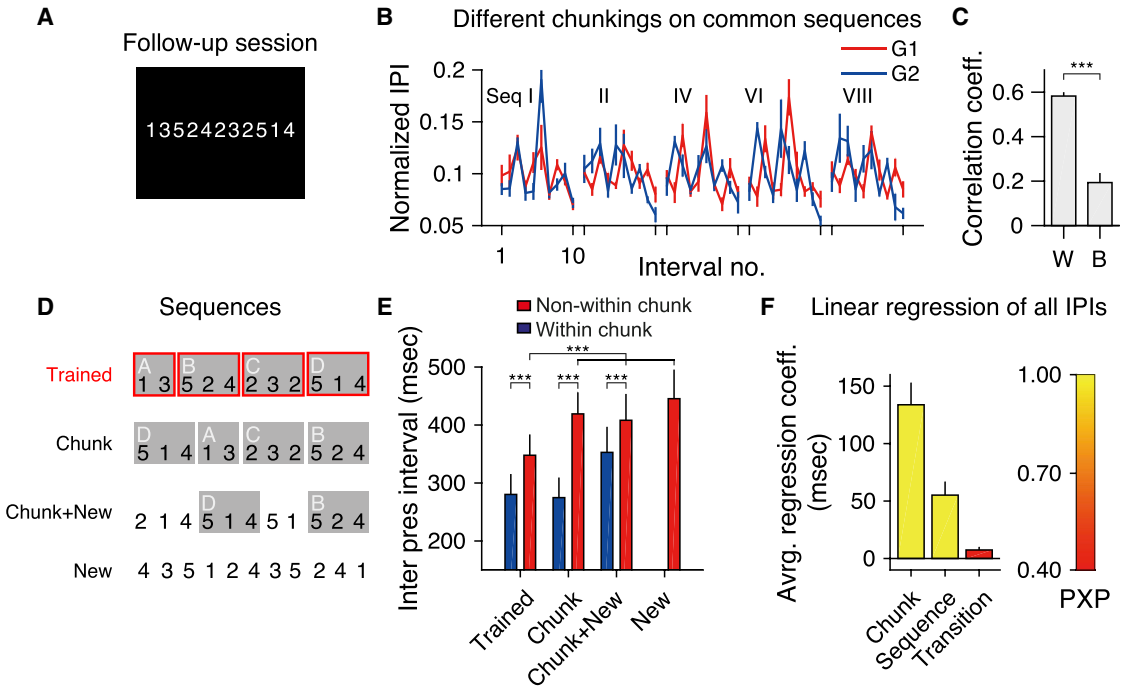

Figure 3. Cognitively Induced Chunking Structure Influences Subsequent Skilled Motor Performance

(A) In the follow-up session, the sequences were directly cued with 11 digits, removing the need for memory recall.

(B) Average IPIs, normalized to the entire sequence duration, for groups G1 and G2.

(C) Across-participant correlation (Pearson's r) of IPI profiles from (B) for within-group (W) and between-group (B) pairs.

(D) Generalization test with trained sequences, new sequences consisting of trained chunks, new sequences partly consisting of trained chunks, and completely new sequences.

(E) Within-chunk intervals were faster than between-chunk intervals for all sequence categories.

(F) Group-averaged regression weights for the 3 effects of interest (chunk, sequence, and transition). Weights were estimated using Bayesian model averaging (STAR Methods). The color of each bar represents the prevalence of the corresponding model. Protected exceedance probability (PXP) is the posterior probability that an effect was present in more than a half of the subjects (see Statistical Tests in the STAR Methods).

***$p < 0.005$; two-sided paired t test. Error bars indicate SEM across the subjects.

of interest: whether the IPI was within or between chunks, whether a chunk transition was within a known sequence, and the frequency with which a specific finger transition was practiced. The model also contained three effects of no interest (see STAR Methods for details). The results of Bayesian model averaging provided clear evidence of both a chunk and sequence effect, accompanied by poor evidence of a finger transition effect (Figure 3F; Table S1).

These results suggest that the motor system has formed a representation of the trained chunks as well as a representation of the overall sequences, which, in turn, activate these chunks. With this behavioral evidence, we next assessed where and how these different representations are implemented in the human neocortex.

Cortical Regions with Robust Sequence "Encoding"

In the fMRI session, the participants received a brief visual cue for the sequence type and then executed the sequence twice (Figure 4A). The activation associated with each sequence was estimated for each voxel by averaging the task-evoked blood oxygen level dependent (BOLD) activity over the two executions. We then applied a representational fMRI analysis to study the cortical sequence representation. Using a searchlight approach (Figure 4B), we first determined whether the activation patterns had any information about the executed sequences. For this, we computed a cross-validated estimate of the Mahalanobis distance (crossnobis distance estimator; Diedrichsen et al., 2016; Walther et al., 2016) between any possible pairs of the sequences. Systematically positive crossnobis estimates indicate reliable differences between the activity patterns for different sequences. Because all sequences consisted of the same finger presses arranged in a different order (Figure 2B), and because we averaged the activity across two sequence executions, any differences in activity patterns must reflect some dependency of the activity on the sequential context.

Figure 4C shows the resultant group searchlight map displayed on the flattened cortical surface. Consistent with other studies (Kornysheva and Diedrichsen, 2014; Wiestler and Diedrichsen, 2013; Wiestler et al., 2014; Yokoi et al., 2018), sequences were "encoded" over the wide area the cortical surface, including primary motor cortex (M1), primary somatosensory cortex (S1), the PMd, and areas around the IPS. The area also included the SMA and pre-SMA, a small part of the rostral cingulate zone (RCZ; Picard and Strick, 1996), the lateral PFC, and the

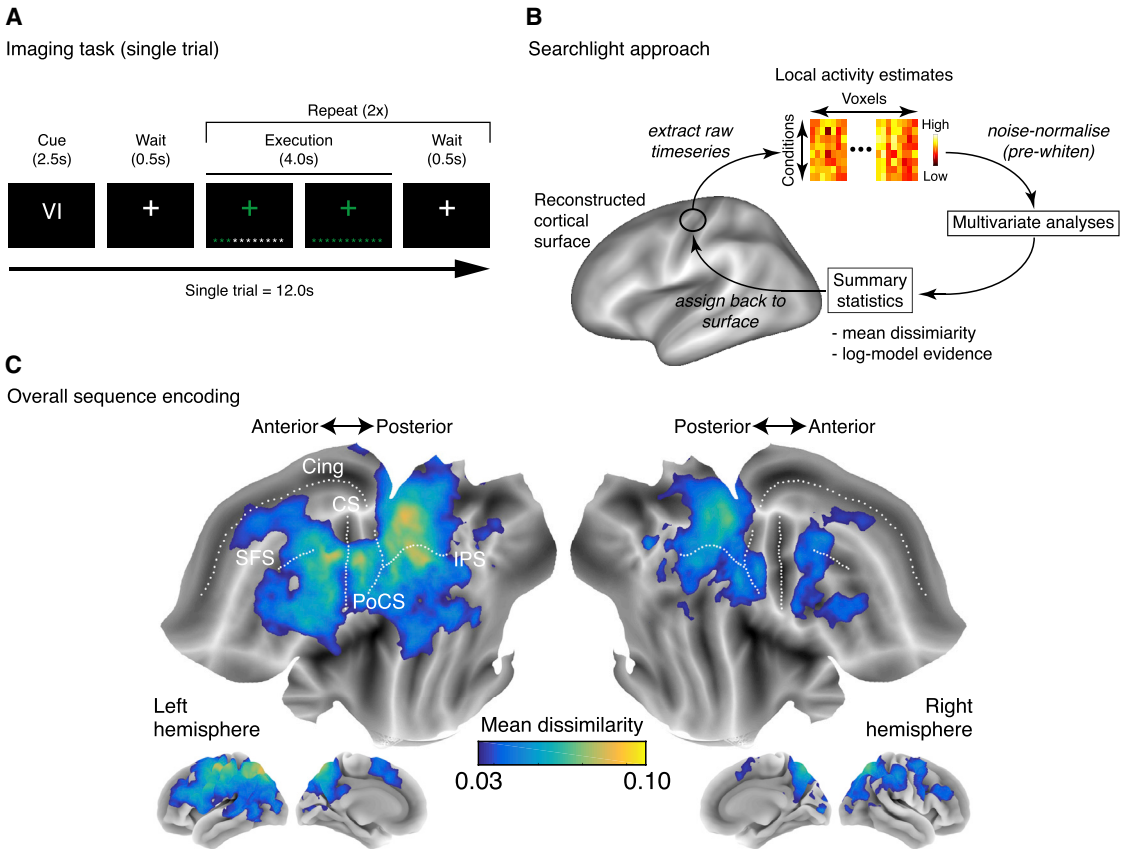

Figure 4. Overall Sequence Encoding on the Flattened Cortical Surface

(A) Time course of a single trial in the fMRI session. Shown is a presentation of the visual cue (sequences I–VIII) was followed by two executions of that sequence.

(B) Searchlight approach. We extracted the activity patterns for small circular areas (~22 mm in diameter) of the reconstructed cortical surface. The pre-whitened activity patterns were then used to calculate the crossnobis distance (pattern separability) or used for model comparison (Figure 5). The resulting statistics were then assigned back to the center node of the region.

(C) A group-averaged map of the strength of sequence encoding (average pairwise crossnobis distance between sequences) plotted on a flattened cortical map. Cing, cingulate sulcus; SFS, superior frontal sulcus; CS, central sulcus; PoCS, post-central sulcus; IPS, intraparietal sulcus.

precuneus, although the signal from these areas was weak compared with other areas such as the PMd or IPS (Kornysheva and Diedrichsen, 2014; Wiestler and Diedrichsen, 2013; Wiestler et al., 2014). Although some areas showed bilateral representations, they were consistently stronger in the left contra-lateral hemisphere.

Model-Based Representational fMRI Analysis

Although Figure 4C tells us that we can decode sequence identity from the activation patterns in this area, it does not reveal which specific representation contributed to the pattern differences. We therefore tessellated the area with significantly positive dissimilarities using a discrete set of surface patches and looked in detail into the representational structure characterized by the representational dissimilarity matrix (RDM; Diedrichsen and Kriegeskorte, 2017) within each patch (Figure 5B).

Based on our behavioral results, we considered three levels of sequence representation (sequence, chunk, and single finger). At the highest level, we propose a sequence representation with a unique neuronal activity pattern for each of the 8 trained sequences. Because we assume that all sequences are equally strongly encoded, such a representation would predict that all possible pairwise distance are equal. At the next level, we have distinct neural activity patterns for each learned chunk. A region with a pure chunk representation would therefore transition, during the sequence, through the four activity states associated with the four chunks. The resultant RDM is therefore predicted by how many chunks different sequences have in common. For instance, sequences 1 and 2 consist of the same chunks in a different order (Figure 2B) and are therefore predicted to elicit highly similar activity patterns. At the lowest hierarchical level, we considered representations of single fingers. Because all sequences consisted of exactly the same presses arranged in a different order, a single-finger representation should lead to identical activity patterns for all sequences. In a recent study, however, we found that the activity pattern for a sequence is strongly determined by the first finger (Yokoi et al., 2018). Therefore, this representation predicts that sequences starting with the same finger should be very similar to each other (Figure 5A).

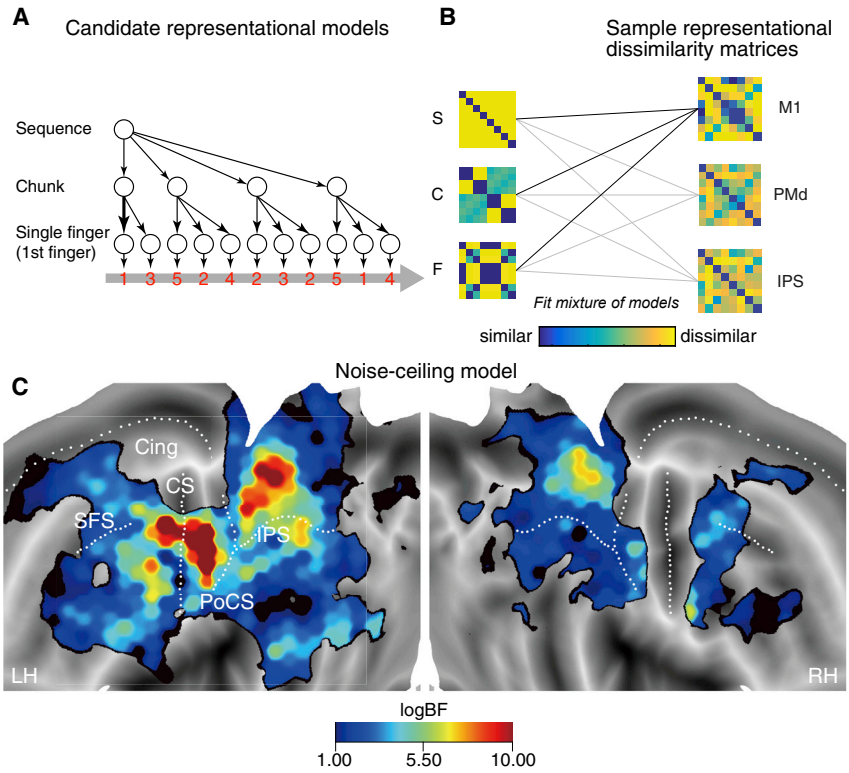

Figure 5. Fitting Candidate Representational Models onto Data

(A) Candidate representational model components of an example sequence (red numbers). Each circle represents a hypothetical neural population; the arrows between the circles represent descending commands to activate the units. Note that, for the single-finger level, a unit for each single finger is shown multiple times. Given the sequences used (Figure 2B), each representational component predicts a unique structure of the RDM. S, sequence; C, chunk; F, first finger.

(B) Empirical RDMs for representative cortical regions. We fitted various combinations of the candidate models to explain the observed representational structure at each cortical searchlight (Figures 4B and 4C).

(C) Group-level map of the strength of shared representation (log-Bayes factor for the noise ceiling versus the null model). LH, left hemisphere; RH, right hemisphere. The other abbreviations are as in Figure 4C. The log-Bayes factor (logBF) map was thresholded using a PXP of 0.75 (Rigoux et al., 2014; Rosa et al., 2010; Stephan et al., 2009) and a logBF of 1 (Kass and Raftery, 1995).

We then estimated the contribution of each candidate representational model to the observed activity patterns using pattern component modeling (PCM; Diedrichsen et al., 2011, 2018). PCM is a powerful Bayesian approach to test combinations of representational models (Diedrichsen and Kriegeskorte, 2017). We evaluated the likelihood of the data in each surface patch under each possible combination of the 3 candidate models. The relative weight of each model component was fitted. Because different combination models had different numbers of free parameters, we used "leave one subject out" cross-validation (Diedrichsen et al., 2018), fitting each model to all participants except one, and then evaluated the likelihood of the data from the left-out participant under the model (see STAR Methods for more detail). Using the resultant cross-validated (log) likelihood for each model, we could compute a log-Bayes factor (Kass and Raftery, 1995), which assesses how much more (or less) likely the observed data are under the candidate model compared with the null model (see STAR Methods).

Noise Ceiling Model

For a full model evaluation, however, it is not enough to know that a specific model can explain the data better than chance. It is also important to know how much of the systematic variation in the data the model can explain. Before testing any candidate models, we therefore fitted a fully flexible "noise ceiling" model. A positive log-Bayes factor (logBF) for the noise ceiling model (versus the null model) indicates that the structure of sequence representation was consistent across individuals. The map (Figure 5C) clearly indicates that the representational structure of the primary sensorimotor areas showed the highest inter-individual consistency, followed by parietal and frontal premotor areas. Because this model can capture any representational structure, it constitutes an estimate of how well the true model should be able to predict the representational structure (Diedrichsen et al., 2018; Nili et al., 2014).

Single Finger Movement Representations

To examine the contribution of the 3 candidate model components, we fitted all possible combination of components and then determined the model-averaged log-Bayes factor as a measure of evidence for the presence of each component in the context of the others (STAR Methods).

Replicating our previous results (Yokoi et al., 2018), the representational structure in M1 and S1 was almost fully determined by the first finger in each sequence. The model-averaged log-Bayes factor ($logBF_c$) revealed strong evidence of the first finger component (Figure 6A). Within the significant patches shown in Figure 6A, the average logBF for the first-finger model was 1.33 ± 8.12 above the noise ceiling model in M1 and S1. The protected exceedance probability (PXP), the posterior probability that the noise ceiling was the better model than the finger model for more than a half of subjects (see Statistical Tests in the STAR Methods), was 0.427, indicating that the first-finger model could predict the left-out subject's data equally well with the noise ceiling model. This implies that the systematic representational structure in this region could be fully explained by this simple model. Because the first finger press was executed with a similar force as all subsequent presses (t_{11} = 0.394, p = 0.70), this indicates that the same movement elicits more BOLD activity in M1 when it is executed in the beginning rather than in the middle of a sequence (Yokoi et al., 2018). This additional cost for sequence initiation is

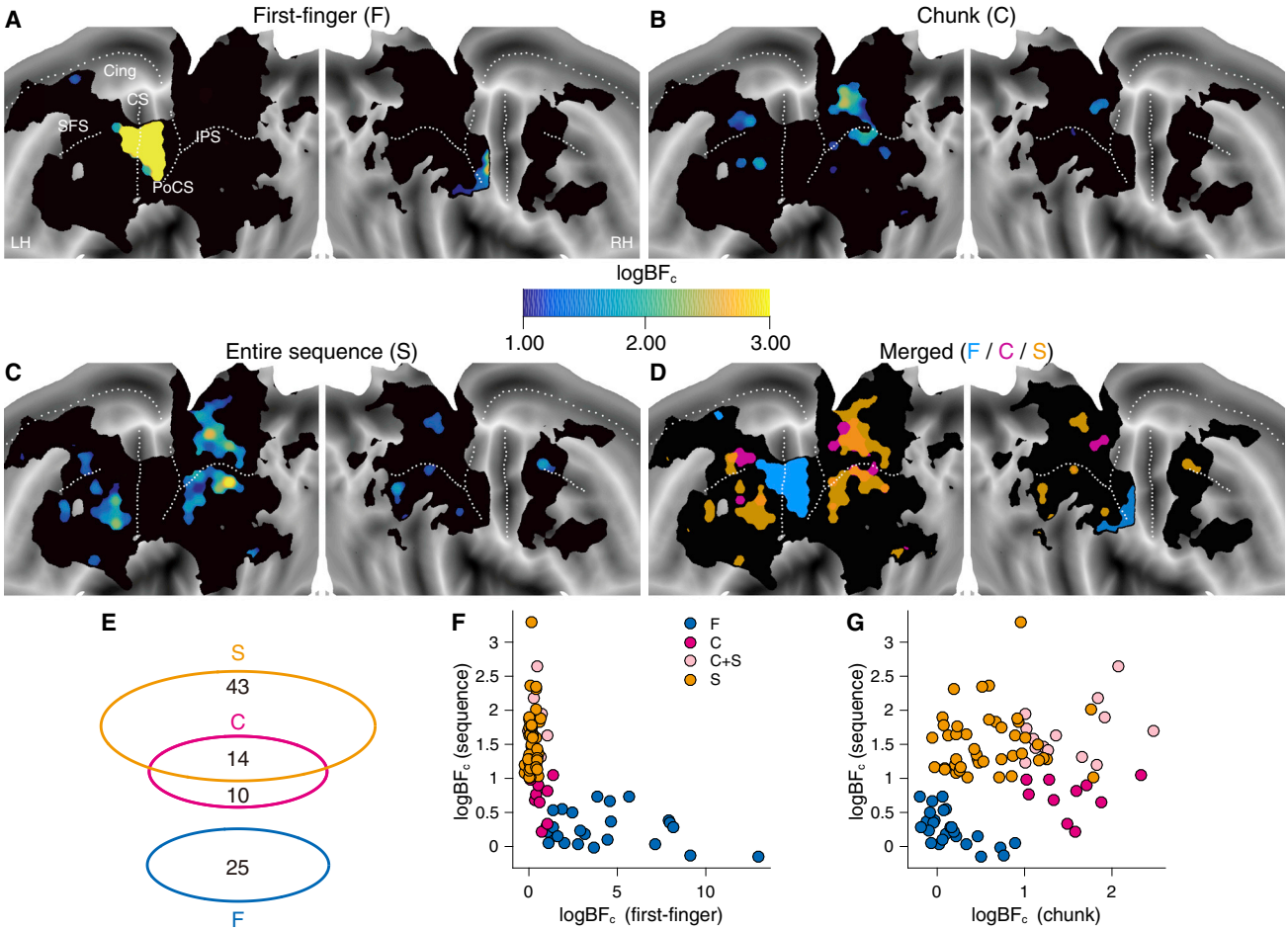

Figure 6. Elementary Movements Had Spatially Distinct Representation, whereas Chunk and Sequence Representations Are Spatially Overlapping

(A–C) Group maps for the model-averaged log-Bayes factor ($logBF_c$) for the first-finger model (A), chunk model (B), and sequence model (C). Each map was thresholded with PXP > 0.75 and $logBF_c$ > 1.

(D) Overlap of the maps for the individual models.

(E) A Venn diagram summarizing spatial overlap between the representations. Numbers indicate the counts of significant surface patches.

(F and G) Scatterplots of group-averaged, model-averaged log-Bayes factors ($logBF_c$) for first-finger versus sequence representation (F) and chunk versus sequence representation (G). Dots represent cortical patches showing logBF > 1 for the noise-ceiling model and significant model-averaged evidence (PXP > 0.75 and $logBF_c$ > 1) for the first-finger representation only (F), chunk representation only (C), chunk and sequence representations (C+S), and sequence representation only (S).

consistent with recent results from a serial reaction time task that showed longer reaction times for the first movement in a sequence (Wong et al., 2015).

We also found some positive evidence for the first-finger model in some sub-cortical regions of interest (ROIs), including the left caudate and the right putamen, although the noise ceilings were generally low in these regions (Figure S1).

Higher-Order Sequence Representations

In contrast to M1 and S1, the first-finger model did not provide a good explanation for the representational structure in premotor and parietal areas (Figure 6A). In these regions, we found consistent evidence of both chunk and sequence representations (Figures 6B and 6C). Evidence of chunk representation was present in the left PMd, PMv, and parietal cortex, both dorsal and ventral to the IPS. The distribution of the sequence representation was similar but extended further anterior in the PMd and PMv and encompassed larger regions in both the superior parietal lobule (SPL), inferior parietal lobule (IPL), and precuneus. In these regions, the combination of sequence and chunk models fit the data as well as or significantly better than the noise ceiling model: logBF (combination model versus noise ceiling) = 2.56 ± 2.47, PXP = 0.954 for the premotor cluster and logBF = 3.76 ± 4.81, PXP = 0.732 for the parietal cluster.

Figure 6D shows the spatial overlap of the finger, chunk, and sequence representation. Although the finger representation

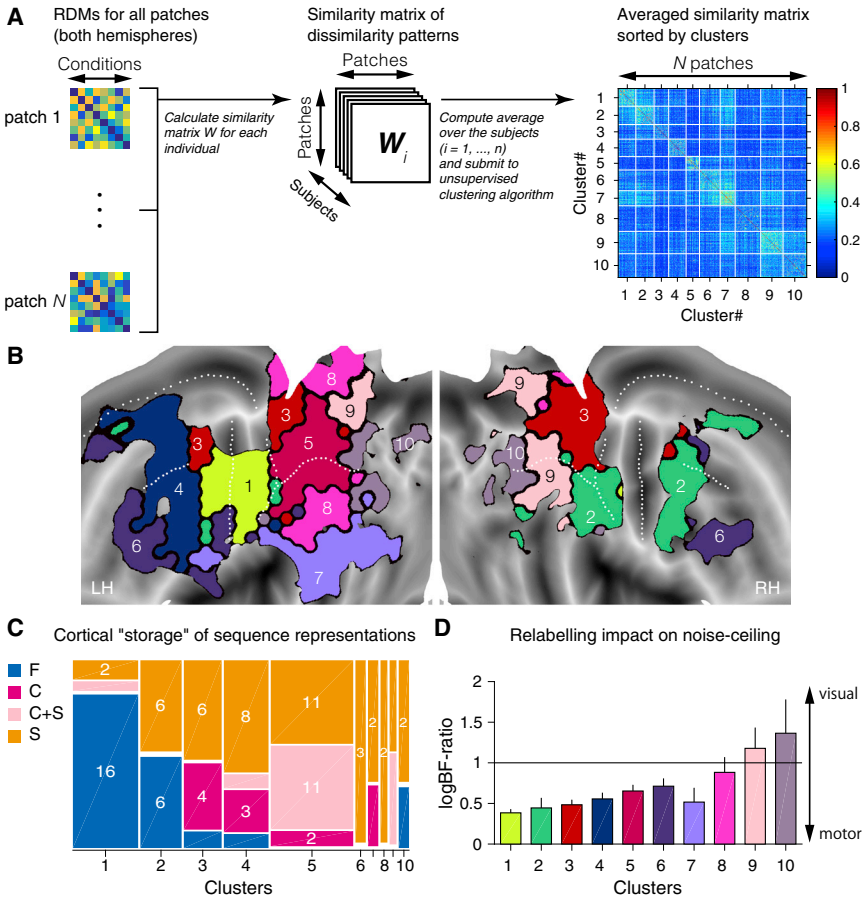

Figure 7. Model-Free Clustering of Cortical Motor Sequence Representations

(A) Within each participant, the RDM for each cortical patch was calculated. We then calculated a patch × patch correlation matrix between the RDMs, averaged the matrix across participants, and submitted them to a clustering approach.

(B) Resultant 10-cluster solution, mapped on the cortical surface.

(C) A Marimekko chart showing the distribution of different movement representations across the cortical clusters. White number indicates the number of significant cortical patches for each representation. Cells without a superimposed number contained one significant patch.

(D) A comparison between noise ceilings obtained under motor-based labeling and visual cue-based labeling for each cluster. The vertical axis represents the logBF-ratio of noise ceiling models (visual versus motor).

was clearly separated, the chunk and sequence representations overlapped substantially. To quantify this observation, we counted the cortical patches that showed evidence of each component (PXP > 0.75, $logBF_c > 1$; Figure 6E). Based on the hypothesis that the first-finger and higher-order sequence representations are independently distributed, we would expect to find, on average, 4 patches with both representations. In contrast to this prediction, we found none (likelihood ratio test, $G = 9.01$, $p = 0.0027$). This indicates a clear anatomical separation between the representation of individual movements and sequential context. The separation was also evident when plotting the component Bayes factors against each other; cortical patches with significant first-finger representation were clearly segregated from other patches (Figures 6F and 6G).

In contrast, sequence and chunk representations overlapped considerably with 14 patches. Under the assumption of independent representations, we would only have expected 3.26 patches, a highly significant deviation ($G = 28.82$, $p = 7.96e{-}8$). Thus, our results indicate a strong non-random overlap between sequence and chunk representations. Nevertheless, the mixed patches did not form one homogeneous group, but some areas showed more sequence and other more chunk representations. A Gaussian mixture model (STAR Methods) with 2 components described the data better than a single Gaussian, resulting in a logBF of 3.68 (strong evidence; Kass and Raftery, 1995).

The result indicated that, unlike the individual finger representation, chunk and sequence representations overlap considerably within the premotor and parietal cortices (Figure 6D), suggesting that the brain may not employ a strict hierarchy at this level of the neural architecture (Botvinick and Plaut, 2004). Although the overlap was clearly more than expected by chance, we also found some evidence that some regions were more heavily weighted for one representation or the other.

Representational Clustering

One shortcoming of the model-based approach is that it can only detect representations that show a consistent representational structure across participants. However, some regions showed clear evidence of differences between the activity patterns for the sequences within each subject (Figure 4) but low consistency of the representational structure across subjects (Figure 5C). This suggests that the sequences were represented in a way that was unique to the individual.

To analyze these representations, we first sought to define cortical regions that, within each subject, would show a relatively uniform representational structure. Using a model-free representational clustering approach, we first calculated the similarity (or "connectivity") of the observed RDMs across the cortical patches (Figure 7A). Because the similarity of RDMs is individually defined, this approach allowed us to group the areas with similar representational structure together without requiring a consistent shape of the RDM across participants.

The resultant clusters are shown in Figure 7B. The first cluster (cluster 1) encompassed the left M1 and S1. In these areas, most

cortical patches showed evidence of the first-finger representation (Figure 7C). In the premotor and parietal areas, the clustering was able to reveal features that could not be captured by the model-based approach alone. The approach showed that the regions with clear positive evidence of the chunk or sequence representation (Figure 6D) were sub-divided into several distinct clusters. For example, cluster 3 encompassed both Brodmann area (BA) 5 and caudal PMd, which had very similar representational structure (Figures 7B and 7C), consistent with the strong anatomical connections shown between these areas (Kurata, 1991; Tomassini et al., 2007).

Cascade of Information Processing during Sequence Production

In clusters 6–10, the inter-subject consistency was relatively poor, as indicated by the relatively low noise ceilings (Figure 5C). Although this may relate to idiosyncratic representations or strategies, it could also be driven by the perceptual similarity between the sequence cues; the associations between the visual cues (I, … VIII) and the actual sequences were randomized across participants.

To test this idea, we re-calculated the noise ceiling for each cluster after re-aligning the conditions across participants in terms of the visual cue rather than in terms of the physical sequence. The result demonstrated that the representational structure of clusters 9 and 10 was more consistent in ''visual'' than ''motor'' space (Figure 7D). As expected, cluster 1, encompassing M1 and S1, was located at the most ''motor'' position, whereas the other clusters occupied reasonable positions that suggest a mixture of motor and perceptual and/or symbolic representations of the sequences. The result also allowed us to directly infer that ''sequence'' representations located in different clusters are likely to play different roles along the continuum of the stimulus-to-output cascade (Figure 7D). Overall, the results of the model-free clustering approach provided not only a concise summary of the model-based result in terms of spatially segregated clusters but also an additional characterization of these clusters in terms of the stimulus-to-output cascade.

DISCUSSION

A Neural Hierarchy of Sequence Representations?

To produce a temporally ordered sequence of signals, the neural system requires a representation of the sequential context; i.e., the neuronal state needs to be sufficiently different for movement A when it is followed by B compared to when A is followed by C. Similar neuronal states for movement A in these two contexts would lead to ''tangling'' of the population response (Russo et al., 2018) and would require substantial input to bring the neuronal dynamics on the correct path. Consistent with our findings, recent results (Russo et al., 2019) indicate that the neural state in M1 shows high tangling on the level of movements, whereas the SMA provides an untangled signal, where the neuronal state for the same movement depends on the sequential context (Tanji and Shima, 1994).

It has been suggested that the brain develops an anatomically ordered hierarchical representation of sequences. This idea is predicated on the fact that different regions, based on differences in their connectivity and intrinsic circuit properties, learn temporal regularities on different timescales (Burt et al., 2018; Hasson et al., 2015; Honey et al., 2007; Kiebel et al., 2008). Basic sensory regions are thought to represent recent history on very short timescales, whereas increasing abstract association areas represent longer timescales (Chaudhuri et al., 2015; Chen et al., 2015; Hasson et al., 2015; Lerner et al., 2011; Murray et al., 2014). This idea provides an elegant explanation of the clear separation of finger and chunk representations; the intrinsic properties of M1 maybe ideally suited to represent individual finger movements, which are executed within 400 ms. However, it requires longer timescales of around 1 s to represent chunks of movements, and only premotor and parietal areas may sustain representations stably over this time.

Given the similar temporal relationship between fingers and chunks (3-fold) and chunks and sequences (4-fold), this idea would also have predicted a comparable anatomical separation between chunk and sequence representations (Figure 1A). This, however, was not the case. Unlike the drastic transition in the representational contents from primary to non-primary sensorimotor areas, we did not find a clear anatomical boundary between chunk and sequence representations. This is surprising, considering that our task and training paradigm had an equally clear structure across all three levels (i.e., finger, chunk, and sequence). This overlap is consistent with the idea that multiple levels of a hierarchy can be represented in a single homogeneous recurrent network (Botvinick and Plaut, 2004; Botvinick, 2007; Paine and Tani, 2005).

One advantage of such architecture is that it could, in contrast to strictly hierarchical neural architectures, deal with tasks with ''quasi-hierarchical structure'' (Botvinick and Plaut, 2004), where the detail of a lower-level abstract movement representation (e.g., ''add sugar'') can change depending on the higher-level representation (e.g., ''make coffee'' or ''bake a cake'') that recruits it. A possible implementation of such context dependency avoiding the redundancy of having many slightly different versions of the same representation would be to integrate the representations of different levels (in our case, chunks and sequences) via gain-field encoding (Andersen et al., 1985; Pouget and Sejnowski, 1997; Yokoi et al., 2011). This may explain the substantial overlap between these two in the parietal cluster 5 (Figure 7C).

At the same time, the clear dissociation between M1 and premotor areas indicates that the nervous system sometimes clearly separates representations on different levels of the hierarchy. Our data strongly argue for a special distinction between the representations of individual movements at the output level and a representation of the sequential context independent of timescale. An advantage of such a clear separation would be that it ensures that the representation of past and future movements cannot interfere with control of the currently ongoing movement.

Parcellation of Neocortical Skill Representations

The strong overlap of chunk and sequence representations, however, does not necessarily imply that all higher-order areas play the same role. Our model-free clustering approach, combined with relabeling analysis, suggested potential contributions

of each of the representationally similar regions along the stimulus-to-action gradient.

Cluster 10, encompassing visual regions along the transverse occipital sulcus, showed the highest consistency after visual relabeling and therefore represents sequences most likely in the form of the visual cue. Clusters 8 and 9 were less affected by relabeling, and we speculate that these clusters play a critical role in translating the visual cue into specific sequence representations. Consistent with this interpretation, activation of these areas has been reported to be related to sequence production and/or learning (Honda et al., 1998; Petit et al., 1996; Sakai et al., 1998) and episodic memory retrieval (Cavanna and Trimble, 2006).

Cluster 6 only showed evidence of a more abstract representation of the sequences (Figure 7C), which would be consistent with the previous electrophysiological finding that neurons in the PFC fire at the initiation of specific sequences (Shima et al., 2007).

The representations found in clusters 3, 4, and 5 suggest that these premotor and parietal areas serve as intermediate layers in the network of sequence production. In these areas, we found substantial evidence of both sequence and chunk representations, with a large degree of overlap between the two. The presence of abstract movement representations in the IPL is consistent with the fact that lesions in the left supramarginal gyrus can cause apraxia (Haaland et al., 2000) and the presence of abstract goal representations in the anterior IPL (Hamilton and Grafton, 2006). Cluster 4 consisted of a large complex of lateral and medial premotor areas (rostral PMd, PMv, SMA, and pre-SMA). Although the SMA and pre-SMA has been extensively discussed as a crucial region for sequence production (Shima and Tanji, 1998; Tanji and Shima, 1994), we found no clear evidence of chunk or sequence representations in this area. It may be that movement representations in medial motor areas are spatially more intermingled than movement representations in lateral regions, making them harder to detect with multivariate fMRI.

According to the relabeling analysis, cluster 3 was closer to the output end of the stimulus-to-response gradient. The cluster includes both the left caudal PMd and bilateral BA 5, areas with dense anatomical interconnections to both M1 and higher-level areas (Kurata, 1991; Lu et al., 1994; Tomassini et al., 2007). This is consistent with a critical role of the caudal PMd as a bottleneck that transforms higher-order movement representations into activity that can directly drive the adjacent generating circuit (M1) (Dum and Strick, 2005; Ohbayashi et al., 2003, 2016).

Advances over Previous Studies

Our study provides several important advances over previous studies. Behaviorally, previous work has failed to cleanly discern whether the consistent temporal regularities of sequential behaviors reflect a hierarchical organization (Sakai et al., 2003), associative learning of transition statistics (Verwey and Abrahamse, 2012), or merely the biomechanical requirement at the specific finger transitions (Jiménez, 2008). Our experimental design with two groups of participants physically acquiring the same sequences through two cognitive routes provides clear behavioral evidence of chunking independent of the biomechanical property of those sequences (Verwey and Dronkert, 1996). The follow-up experiment also provided evidence of the formation of a higher-order representation that facilitated transition between trained chunks within a known sequence. In contrast, the associative learning of transition between fingers did not seem to contribute to the pattern of responding.

Second, our model-based multivariate fMRI approach allowed direct assessment of sequence representations. Many prior univariate studies on motor sequence learning have revealed experience-dependent activity changes in multiple cortical regions, including the dorsolateral prefrontal cortex (DLPFC), M1, premotor cortex (PM), SMA, IPS, and precuneus (Doyon et al., 2002; Grafton et al., 1995; Honda et al., 1998; Kawashima et al., 1998; Penhune and Steele, 2012; Sadato et al., 1996; Sakai et al., 1998). A recent line of multivariate fMRI studies (Kornysheva and Diedrichsen, 2014; Nambu et al., 2015; Wiestler and Diedrichsen, 2013) has provided direct evidence that these previously reported regions represent some information about motor sequences. The current study now characterizes these representations in detail.

Finally, many existing methods for cortical parcellation rely on correlations between time series across different brain regions, mostly during rest (i.e., functional connectivity; Margulies et al., 2016; Yeo et al., 2011). In contrast, our model-free clustering approach is unique in using similarity between representational structures. The specific advantage of the current task-based representational parcellation approach is that the identified regions show a relatively homogeneous representation that can then be analyzed using model-based representational fMRI approaches to draw better inference about the functional role for each network or cluster.

Limitations and Open Issues

Our study was restricted to analysis of neocortical representation, and we did not address how sequences are represented in sub-cortical structures (Doyon et al., 2002; Graybiel, 1998; Graybiel and Grafton, 2015; Wymbs et al., 2012). Although we found some evidence of sequence representations in subcortical ROIs (Figure S1), conclusions of the model-based analyses here are limited because of the relatively low signal-to-noise ratio.

Another important limitation of the current study is that training was limited to 1 week. Thus, it should be noted that the cortical map of sequence representations presented in the current study is one snapshot in the dynamic process of skill acquisition, and extended training may provide a different map. Specifically, longer training may lead to emergence of a sequence representation in M1 and S1. Although we found no evidence of such a representation after a week of training, animal studies with training of more than 6 months report sequence-related neuronal activity in M1 (Matsuzaka et al., 2007; Picard et al., 2013). One caveat is, however, that these studies focus on the difference in neuronal firing associated with a certain action between trained and untrained sequential contexts, which is insufficient to conclude that sequences are truly encoded in M1 (see 'Testing for a Neural Hierarchy of Sequence Production). Nonetheless, a recent human fMRI study using repetition suppression also suggests that sequence representation in M1 may emerge slowly over a month of training (Wymbs and Grafton, 2015). Therefore, the next important challenge is to use the multivariate

approach introduced here to understand how these representations are formed and reorganized over the course of prolonged training (Berlot et al., 2017).

Summary and Conclusion

Is there an anatomical hierarchy of motor sequence representations in the human neocortex? Our data provide evidence that each level of the behavioral movement hierarchy is represented. However, although we found a clear anatomical separation of individual movements, higher-order sequence representations were represented in an intermingled fashion. These finding argue for a special status of execution-level representations as opposed to representations of the sequential context, independent of timescale or level of abstraction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
- METHOD DETAILS
  - Apparatus
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral data Analysis
  - Imaging data analysis
  - Statistical tests
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2019.06.017.

ACKNOWLEDGMENTS

This study was supported by JSPS postdoctoral fellowship 15J03233 (to A.Y.), a James S. McDonnell Foundation scholar award, NSERC discovery grant RGPIN-2016-04890 (to J.D.), and the Canada First Research Excellence Fund (BrainsCAN). We thank E. Bamber and E. Thomas for assistance with data collection; N. Ejaz, U. Hertz, and S. Nishimoto for helpful discussions; and E. Berlot and N. Hagura for comments on the manuscript.

AUTHOR CONTRIBUTIONS

A.Y. and J.D. designed the experiment. A.Y. collected and analyzed the data. A.Y. and J.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 17, 2018
Revised: March 18, 2019
Accepted: June 21, 2019
Published: July 22, 2019

REFERENCES

Andersen, R.A., Essick, G.K., and Siegel, R.M. (1985). Encoding of spatial location by posterior parietal neurons. Science *230*, 456–458.

Baldauf, D., Cui, H., and Andersen, R.A. (2008). The posterior parietal cortex encodes in parallel both goals for double-reach sequences. J. Neurosci. *28*, 10081–10089.

Berlot, E., Popp, N., and Diedrichsen, J. (2017). Neural correlates of long-term motor sequence learning (Open Science Framework).

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (Springer).

Botvinick, M.M. (2007). Multilevel structure in behaviour and in the brain: a model of Fuster's hierarchy. Philos. Trans. R. Soc. Lond. B Biol. Sci. *362*, 1615–1626.

Botvinick, M., and Plaut, D.C. (2004). Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. Psychol. Rev. *111*, 395–429.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., and Cohen, J.D. (2001). Conflict monitoring and cognitive control. Psychol. Rev. *108*, 624–652.

Burnham, K.P., and Anderson, D.R. (2004). Multimodel inference - understanding AIC and BIC in model selection. Sociol. Methods Res. *33*, 261–304.

Burt, J.B., Demirtaş, M., Eckner, W.J., Navejar, N.M., Ji, J.L., Martin, W.J., Bernacchia, A., Anticevic, A., and Murray, J.D. (2018). Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat. Neurosci. *21*, 1251–1259.

Cavanna, A.E., and Trimble, M.R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. Brain *129*, 564–583.

Chaudhuri, R., Knoblauch, K., Gariel, M.A., Kennedy, H., and Wang, X.J. (2015). A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. Neuron *88*, 419–431.

Chen, J., Hasson, U., and Honey, C.J. (2015). Processing Timescales as an Organizing Principle for Primate Cortex. Neuron *88*, 244–246.

Cooper, R., and Shallice, T. (2000). Contention scheduling and the control of routine activities. Cogn. Neuropsychol. *17*, 297–338.

Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS Comput. Biol. *13*, e1005508.

Diedrichsen, J., and Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. Neuroimage *27*, 624–634.

Diedrichsen, J., Ridgway, G.R., Friston, K.J., and Wiestler, T. (2011). Comparing the similarity and spatial structure of neural representations: a pattern-component model. Neuroimage *55*, 1665–1678.

Diedrichsen, J., Wiestler, T., and Krakauer, J.W. (2013). Two distinct ipsilateral cortical representations for individuated finger movements. Cereb. Cortex *23*, 1362–1377.

Diedrichsen, J., Zareamoghaddam, H., and Provost, S. (2016). On the distribution of crossvalidated mahalanobis distances. arXiv, arXiv:1607.01371 https://arxiv.org/abs/1607.01371.

Diedrichsen, J., Yokoi, A., and Arbuckle, S.A. (2018). Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. Neuroimage *180* (Pt A), 119–133.

Doyon, J., Song, A.W., Karni, A., Lalonde, F., Adams, M.M., and Ungerleider, L.G. (2002). Experience-dependent changes in cerebellar contributions to motor sequence learning. Proc. Natl. Acad. Sci. USA *99*, 1017–1022.

Dum, R.P., and Strick, P.L. (2005). Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. J. Neurosci. *25*, 1375–1386.

Ejaz, N., Hamada, M., and Diedrichsen, J. (2015). Hand use predicts the structure of representations in sensorimotor cortex. Nat. Neurosci. *18*, 1034–1040.

Fischl, B., Sereno, M.I., Tootell, R.B.H., and Dale, A.M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. *8*, 272–284.

Fujii, N., and Graybiel, A.M. (2003). Representation of action sequence boundaries by macaque prefrontal cortical neurons. Science *301*, 1246–1249.

Fuster, J. (2008). The Prefrontal Cortex (Elsevier Science).

Grafton, S.T., Hazeltine, E., and Ivry, R. (1995). Functional mapping of sequence learning in normal humans. J. Cogn. Neurosci. *7*, 497–510.

Graybiel, A.M. (1998). The basal ganglia and chunking of action repertoires. Neurobiol. Learn. Mem. *70*, 119–136.

Graybiel, A.M., and Grafton, S.T. (2015). The striatum: where skills and habits meet. Cold Spring Harb. Perspect. Biol. *7*, a021691.

Greve, D.N., Van der Haegen, L., Cai, Q., Stufflebeam, S., Sabuncu, M.R., Fischl, B., and Brysbaert, M. (2013). A surface-based analysis of language lateralization and cortical asymmetry. J. Cogn. Neurosci. *25*, 1477–1492.

Haaland, K.Y., Harrington, D.L., and Knight, R.T. (2000). Neural representations of skilled movement. Brain *123*, 2306–2313.

Hamilton, A.F., and Grafton, S.T. (2006). Goal representation in human anterior intraparietal sulcus. J. Neurosci. *26*, 1133–1137.

Hasson, U., Chen, J., and Honey, C.J. (2015). Hierarchical process memory: memory as an integral component of information processing. Trends Cogn. Sci. *19*, 304–313.

Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. Trends Neurosci. *22*, 464–471.

Honda, M., Deiber, M.P., Ibáñez, V., Pascual-Leone, A., Zhuang, P., and Hallett, M. (1998). Dynamic cortical involvement in implicit and explicit motor sequence learning. A PET study. Brain *121*, 2159–2173.

Honey, C.J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc. Natl. Acad. Sci. USA *104*, 10240–10245.

Humphreys, G.W., and Forde, E.M.E. (1998). Disordered action schema and action disorganisation syndrome. Cogn. Neuropsychol. *15*, 771–811.

Hunt, R.H., and Aslin, R.N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. J. Exp. Psychol. Gen. *130*, 658–680.

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., and Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. Neuroimage *16*, 217–240.

Jiménez, L. (2008). Taking patterns for chunks: is there any evidence of chunk learning in continuous serial reaction-time tasks? Psychol. Res. *72*, 387–396.

Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. J. Am. Stat. Assoc. *90*, 773–795.

Kawashima, R., Matsumura, M., Sadato, N., Naito, E., Waki, A., Nakamura, S., Matsunami, K., Fukuda, H., and Yonekura, Y. (1998). Regional cerebral blood flow changes in human brain related to ipsilateral and contralateral complex hand movements–a PET study. Eur. J. Neurosci. *10*, 2254–2260.

Kiebel, S.J., Daunizeau, J., and Friston, K.J. (2008). A hierarchy of time-scales and the brain. PLoS Comput. Biol. *4*, e1000209.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. Neuron *50*, 963–974.

Kornysheva, K., and Diedrichsen, J. (2014). Human premotor areas parse sequences into their spatial and temporal features. eLife *3*, e03043.

Krakauer, J.W., Hadjiosif, A.M., Xu, J., Wong, A.L., and Haith, A.M. (2019). Motor Learning. Compr. Physiol. *9*, 613–663.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.

Kurata, K. (1991). Corticocortical inputs to the dorsal and ventral aspects of the premotor cortex of macaque monkeys. Neurosci. Res. *12*, 263–280.

Lashley, K.S. (1951). The problem of serial order in behavior. In. In Cerebral Mechanisms in Behavior, L.A. Jeffress, ed. (John Wiley & Sons), pp. 112–146.

Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J. Neurosci. *31*, 2906–2915.

Lu, M.T., Preston, J.B., and Strick, P.L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J. Comp. Neurol. *341*, 375–392.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. USA *113*, 12574–12579.

Matsuzaka, Y., Picard, N., and Strick, P.L. (2007). Skill representation in the primary motor cortex after long-term practice. J. Neurophysiol. *97*, 1819–1832.

Miller, G.A., Galanter, E., and Pribram, K.H. (1960). Plans and the Structure of Behavior (Holt, Rinehart and Winston).

Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., and Wang, X.J. (2014). A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci. *17*, 1661–1663.

Nambu, I., Hagura, N., Hirose, S., Wada, Y., Kawato, M., and Naito, E. (2015). Decoding sequential finger movements from preparatory activity in higher-order motor regions: a functional magnetic resonance imaging multi-voxel pattern analysis. Eur. J. Neurosci. *42*, 2851–2859.

Ng, A.Y., Jordan, M.I., and Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. https://papers.nips.cc/paper/2092-on-spectral-clustering-analysis-and-an-algorithm.pdf.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. PLoS Comput. Biol. *10*, e1003553.

Ohbayashi, M., Ohki, K., and Miyashita, Y. (2003). Conversion of working memory to motor sequence in the monkey premotor cortex. Science *301*, 233–236.

Ohbayashi, M., Picard, N., and Strick, P.L. (2016). Inactivation of the Dorsal Premotor Area Disrupts Internally Generated, But Not Visually Guided, Sequential Movements. J. Neurosci. *36*, 1971–1976.

Oosterhof, N.N., Wiestler, T., Downing, P.E., and Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage *56*, 593–600.

Paine, R.W., and Tani, J. (2005). How hierarchical control self-organizes in artificial adaptive systems. Adapt. Behav. *13*, 211–225.

Patenaude, B., Smith, S.M., Kennedy, D.N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage *56*, 907–922.

Penhune, V.B., and Steele, C.J. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. Behav. Brain Res. *226*, 579–591.

Petit, L., Orssaud, C., Tzourio, N., Crivello, F., Berthoz, A., and Mazoyer, B. (1996). Functional anatomy of a prelearned sequence of horizontal saccades in humans. J. Neurosci. *16*, 3714–3726.

Picard, N., and Strick, P.L. (1996). Motor areas of the medial wall: a review of their location and functional activation. Cereb. Cortex *6*, 342–353.

Picard, N., Matsuzaka, Y., and Strick, P.L. (2013). Extended practice of a motor skill is associated with reduced metabolic activity in M1. Nat. Neurosci. *16*, 1340–1347.

Pouget, A., and Sejnowski, T.J. (1997). Spatial transformations in the parietal cortex using basis functions. J. Cogn. Neurosci. *9*, 222–237.

Ramkumar, P., Acuna, D.E., Berniker, M., Grafton, S.T., Turner, R.S., and Kording, K.P. (2016). Chunking as the result of an efficiency computation trade-off. Nat. Commun. *7*, 12176.

Reber, A.S. (1967). Implicit learning of artificial grammars. J. Verbal Learn. Verbal Behav. *6*, 855–863.

Rigoux, L., Stephan, K.E., Friston, K.J., and Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. Neuroimage *84*, 971–985.

Rosa, M.J., Bestmann, S., Harrison, L., and Penny, W. (2010). Bayesian model selection maps for group studies. Neuroimage *49*, 217–224.

Rosenbaum, D.A., Kenny, S.B., and Derr, M.A. (1983). Hierarchical control of rapid movement sequences. J. Exp. Psychol. Hum. Percept. Perform. *9*, 86–102.

Russo, A.A., Bittner, S.R., Perkins, S.M., Seely, J.S., London, B.M., Lara, A.H., Miri, A., Marshall, N.J., Kohn, A., and Jessell, T.M. (2018). Motor cortex embeds muscle-like commands in an untangled population response. Neuron *97*, 953–966.e8.

Russo, A.A., Khajeh, R., Bittner, S.R., Perkins, S.M., Cunningham, J.P., Abbott, L.F., and Churchland, M.M. (2019). Neural trajectories in the supplementary motor area and primary motor cortex exhibit distinct geometries, compatible with different classes of computation. bioRxiv. https://doi.org/10.1101/650002.

Sadato, N., Campbell, G., Ibáñez, V., Deiber, M., and Hallett, M. (1996). Complexity affects regional cerebral blood flow change during sequential finger movements. J. Neurosci. *16*, 2691–2700.

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Sasaki, Y., and Pütz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. J. Neurosci. *18*, 1827–1840.

Sakai, K., Kitaguchi, K., and Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. Exp. Brain Res. *152*, 229–242.

Shen, S.M., and Ma, W.J. (2019). Variable precision in visual perception. Psychol. Rev. *126*, 89–132.

Shima, K., and Tanji, J. (1998). Both supplementary and presupplementary motor areas are crucial for the temporal organization of multiple movements. J. Neurophysiol. *80*, 3247–3260.

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. Nature *445*, 315–318.

Sokal, R.R., and Rohlf, F.J. (1995). Biometry: The Principles and Practice of Statistics in Biological Research (W. H. Freeman).

Stadler, M.A. (1992). Statistical Structure and implicit serial learning. J. Exp. Psychol. Learn. Mem. Cogn. *18*, 318–327.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., and Friston, K.J. (2009). Bayesian model selection for group studies. Neuroimage *46*, 1004–1017.

Tanji, J., and Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. Nature *371*, 413–416.

Tomassini, V., Jbabdi, S., Klein, J.C., Behrens, T.E.J., Pozzilli, C., Matthews, P.M., Rushworth, M.F.S., and Johansen-Berg, H. (2007). Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. J. Neurosci. *27*, 10259–10269.

Verwey, W.B., and Abrahamse, E.L. (2012). Distinct modes of executing movement sequences: reacting, associating, and chunking. Acta Psychol. (Amst.) *140*, 274–282.

Verwey, W.B., and Dronkert, Y. (1996). Practicing a structured continuous key-pressing task: Motor chunking or rhythm consolidation? J. Mot. Behav. *28*, 71–79.

Verwey, W.B., Abrahamse, E.L., and de Kleine, E. (2010). Cognitive processing in new and practiced discrete keying sequences. Front. Psychol. *1*, 32.

Von Luxburg, U. (2007). A tutorial on spectral clustering. Stat. Comput. *17*, 395–416.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage *137*, 188–200.

Wiestler, T., and Diedrichsen, J. (2013). Skill learning strengthens cortical representations of motor sequences. eLife *2*, e00801.

Wiestler, T., Waters-Metenier, S., and Diedrichsen, J. (2014). Effector-independent motor sequence representations exist in extrinsic and intrinsic reference frames. J. Neurosci. *34*, 5054–5064.

Wong, A.L., Lindquist, M.A., Haith, A.M., and Krakauer, J.W. (2015). Explicit knowledge enhances motor vigor and performance: motivation versus practice in sequence tasks. J. Neurophysiol. *114*, 219–232.

Wymbs, N.F., and Grafton, S.T. (2015). The Human Motor System Supports Sequence-Specific Representations over Multiple Training-Dependent Timescales. Cereb. Cortex *25*, 4213–4225.

Wymbs, N.F., Bassett, D.S., Mucha, P.J., Porter, M.A., and Grafton, S.T. (2012). Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. Neuron *74*, 936–946.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.

Yokoi, A., Hirashima, M., and Nozaki, D. (2011). Gain field encoding of the kinematics of both arms in the internal model enables flexible bimanual action. J. Neurosci. *31*, 17058–17068.

Yokoi, A., Bai, W., and Diedrichsen, J. (2017). Restricted transfer of learning between unimanual and bimanual finger sequences. J. Neurophysiol. *117*, 1043–1051.

Yokoi, A., Arbuckle, S.A., and Diedrichsen, J. (2018). The Role of Human Primary Motor Cortex in the Production of Skilled Finger Sequences. J. Neurosci. *38*, 1430–1442.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB R2015b	Mathworks	https://www.mathworks.com
SPM 8	FIL Method group	https://www.fil.ion.ucl.ac.uk/spm
SPM 12	FIL Method group	https://www.fil.ion.ucl.ac.uk/spm
PCM toolbox	Diedrichsen et al., 2018	https://github.com/jdiedrichsen/pcm_toolbox; RRID:SCR_015891
RSA toolbox	RSA group	https://github.com/rsagroup/rsatoolbox
Dataframe toolbox	Jörn Diedrichsen	https://github.com/jdiedrichsen/dataframe
Freesurfer	Martinos Center for Biomedical Imaging	https://surfer.nmr.mgh.harvard.edu/
FSL	FMRIB Analysis Group	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
Caret	Van Essen Laboratory (Department of Anatomy and Neurobiology at the Washington University School of Medicine)	http://brainvis.wustl.edu/wiki/index.php/Caret:About
Other		
MATLAB code for reproducing all figures and related data	This paper	https://github.com/ayokoi/sh1

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Atsushi Yokoi (ayokoi@nict.go.jp).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

All experimental procedures were approved by local ethics committee at the University College London (London, UK). We recruited 23 healthy, right-handed, neurologically healthy volunteers, who participated in the study after providing written informed consent. None of the participants was a professional musician. Of these, 8 participants were excluded, as they did not meet the performance criterion necessary to go on to the imaging and follow-up sessions. The remaining 15 participants went through the imaging session, and the subsequent follow-up session. Of these 15 remaining participants, the data from 3 participants were excluded, as they failed to achieve sufficient behavioral performance during scanning (57% correct versus 81% correct for all other subjects). As a result, only the data from the remaining 12 participants (5 females, 7 males, age: 23 ± 4) was submitted to analysis. These participants reported 5.8 ± 3.8 years of practice with musical instruments (e.g., piano, guitar, violin, etc.).

METHOD DETAILS

Apparatus

We used a custom-built five-finger keyboard device (Figure 2A). The keys of the device were immobile and equipped with force transducers that could measure isometric finger forces (Wiestler and Diedrichsen, 2013; Yokoi et al., 2017). The analog signals were passed through a penetration panel in the magnet room to avoid radio-frequency leakage. The signals were then low-pass filtered, amplified, digitized, and sent to PC for online task control and data recording. The forces were recorded at 200 Hz. During the training sessions, the participants were seated in front of an LCD monitor and placed only their right hand on the keyboard to perform the task, while in the scanner the fingers of their left hand were placed on a mirror-symmetric device to monitor potential implicit mirror movement (Diedrichsen et al., 2013).

Procedure

Sequences production task

We employed a discrete sequence production (DSP) task, in which participants were asked to produce a specific sequence of key presses as fast and accurate as possible. Over the course of 5-6 days (~2 hours per each day), participants learned to produce 8

different sequences consisting of 11 presses from the memory as quickly as possible. All the sequences were matched with the number of finger presses used; 2 presses with thumb, middle, ring, and little fingers, and 3 presses with index finger, respectively. A finger press was detected when the force crossed a threshold of 3 N and a release was detected when it fell below the threshold. To successfully complete a finger press, the pressed finger needed to be pressed, while all other finger needed to be released. In the display, either 11 digits or asterisks representing the finger presses were presented on the screen, as well as other visual cues related to the sequences (Figure 2C). Each correct/incorrect press was informed by turning corresponding digit (or asterisk) into green/red. The participants were rewarded by points for their speed and accuracy (see Behavioral training). After completion of each experimental block, the average movement time (MT; the time between the beginning of the first press and the end of the last press) for the correct trials, the earned points, and the error rate, were presented on the screen.

During the imaging task, a central fixation cross was presented. Each trial started with a 2.5 s of the presentation of a visual cue (roman numerals I-VIII) that indicated the sequence to be executed. The cue presentation was followed by 0.5 s of interval. Then, the fixation cross turned green, and 11 asterisks were presented, triggering the subject to produce the sequence (Figure 4A). For each correct press the corresponding asterisks turned green - for each incorrect press red. Participants were instructed to complete the sequence even if an error has occurred. After each execution, feedback was given (during the ITI) by the color of fixation cross (white: correct, red: one or more presses were incorrect, and blue: unfinished, but presses were correct).

Behavioral training

In order to ensure consistent and stable chunk structure across individuals, we deliberately imposed chunk structure by manipulating how the participants build up their explicit memory of the sequences. In brief, we initially trained the participants with single "chunks" (Figure 2B) until they could produce these chunks from memory, and then trained them with sequences that consisted of these chunks. Eight different chunks of 2 or 3 presses (cued by alphabets) were organized into 8 different sequences, every one of which consists of combination of 4 chunks (Figure 2B). The associations between chunks and chunk cues (A, B, ∼, H), and between sequences and sequence cues (I, II, ∼, VIII) were randomized across participants. To dissociate the influence of the explicit training from subsequent biomechanical optimization of the sequence, we assigned participants randomly to one of two groups, which were trained with different set of chunks. Five sequences were physically identical across the two groups (i.e., the same order of finger presses chunked differently, Figure 2B). The sequences were designed to maximize the difference in the prediction of the different representational models (see Multivariate fMRI analysis: Model-based approach).

Training consisted of 5 days before the imaging session (Figure 2C). On day 1 and early blocks of day 2, subjects were specifically trained with individual chunks. We alternated cued trials in which the chunk cue (A-H) was presented together with the required digits, and uncued trials where only the chunk cue was presented (Figure 2C). In each block, each chunk type was repeated for three times. Participants received a total of 720 trials of chunk training. Starting on day 2, participants practiced entire movement sequences. On cued trials, the sequence cue (I-VIII) and chunk cues were presented (Figure 2C), but no longer with finger cues. On uncued trials, participants needed to retrieve the entire sequence from memory. In each block, each sequence type was repeated for three times. They received a total of 1512 trials of sequence training.

On the fifth day of the training session, after the practice session, the participants practiced the task in a supine position on a mock MRI scanner bed. For the half of the participants, we added the 6th day of additional familiarization session to ensure that they could correctly produce the sequences within 4 s. They were familiarized with the actual task in the scanner during on average 10 ± 8 blocks of the familiarization session.

To keep motivation, the participants were rewarded by points for their performance in the following way. Participants earned 1 point for each correct completion of a chunk. When all the presses were correct, they earned additional points according to their MT compared with the median MT of the previous block (mMT); 3 points (MT < 0.8 × mMT), 1 point (otherwise). For each session, the median MT (mMT) was initialized with [#press] × 600 (ms). Participants lost 1 point when any one of the presses was incorrect. Therefore, at each trial, they could earn maximally 7 points (= 1+1+1+1+3). At the end of each trial, an incremental value of the points for the block was displayed on the screen. The acquired points were later translated into additional monetary reward at the end of the whole experiment (1000 points = 1£). Points were not added for the familiarization session. For the follow-up session (see below), points were considered only for presses. This additional "bonus" was paid for only those who completed the whole experiment, including the follow-up session.

Follow-up session

To confirm that the cognitively imposed chunk structures actually influenced participants' motor representations, we conducted a follow-up session within 1 week after the imaging session (1 ± 0.6 days). In the session, all the digits were presented on the screen to release the participants from the necessity to recall any sequence from memory (Figure 3A). No sequence cues were provided. Additionally, we assessed the generalization of learned chunks to unlearned sequences by additionally introducing 3 new sets of sequences (Figure 3D): New: completely novel sequences which did not contain any of the trained chunks; Chunk: sequences composed of trained chunks in untrained order; and Chunk + New: novel sequences that contained two learned chunks at random positions in the sequence. Each category except for Chunk, which had 16 different sequences, had 8 different sequences, resulting in totally 40 sequences. Sequences were executed 4 times in a row. The order of sequences was randomized and all sequences were repeated for 4 times (16 executions per a sequence). The resultant 640 executions were divided into 16 blocks. Two participants received slightly different protocol with Trained, Chunk, and New sequences.

Imaging session

During the imaging session, the participants were placed on the scanner bed with their knees slightly bent and supported by a wedge-shaped cushion. The two keyboards were tied together with plastic screws and stabilized on the participants' lap with foam pads. Visual stimuli were presented on a back-projecting screen, and participants viewed the screen through the mirror mounted above the head coil. The presentation of the sequence cue (2.5 s), was followed by two execution phases for the same sequence (4 s for each). After this, the next trial started after an ITI of 0.5 s (Figure 4A). The order of 8 sequences was randomized and each sequence had 3 trials (with 2 executions each) within each functional imaging run. Each run also contained 4 randomly interspersed rest phases (12 s). The number of correct trials for the run was presented in the screen at the end of the run. Each functional run lasted about 7 min and 9 runs per participant were conducted. Short breaks (up to a few minutes) were interleaved on the participants' request.

Imaging data acquisition

Imaging sessions were conducted on a Siemens Trio 3T scanner system with a 32-channel head coil at the Welcome Trust Centre for Neuroimaging (London, United Kingdom). B0-field maps were acquired at the beginning of the session to correct for inhomogeneities of the magnetic field (Hutton et al., 2002). Functional images were acquired for 9 runs of 135 volumes each, using a 2-D echo-planer imaging sequence (TR = 2.72 s, in-plane acceleration factor = 2, resolution = 2.3mm isotropic with 0.3 mm gap between each slice, and 32 slices interleaved). The slices were acquired in an axial orientation and covered the dorsal aspects of the brain, including most of the frontal, parietal, occipital lobes, and basal ganglia. The ventral aspects of the frontal and temporal lobes, brainstem, and the cerebellum were not scanned. The first 5 volumes of each run were discarded to ensure stable magnetization. A T1-weighted anatomical image was obtained using MPRAGE sequence (1mm isotropic resolution).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral data Analysis

Recorded force data were analyzed offline. The inter-press-intervals (IPIs) were calculated as the time between the moments of peak force for two consecutive finger presses (Figure 2A). We obtained similar result when IPIs were calculated as based on the time of press onset.

Linear IPI modeling

To assess the contribution of imposed chunk structure and other effects, we ran a linear regression analysis on the IPI data of the follow-up session. We treated the individual IPIs for all trials as a single data vector. We then built linear models to explain the variation in the IPIs. These models consisted of all possible combinations of the following components: 1) chunk: whether the IPI was within a chunk (coded as -1) or not (0); 2) sequence: whether the between-chunk transition was trained (-1) or not (0); 3) transition: the frequency of each specific digit transition (25 total) in training, the sign of the regressor was negative, so that higher frequencies would predict lower IPIs; 4) explicit knowledge (Verwey et al., 2010): whether the interval belonged to trained sequence (-1) or not (0); 5) biomechanical difficulty: the mean IPI of that particular finger transition in a control experiment that tested for the execution speed of all possible 2- and 3-finger transitions, as will be reported in a separate paper, and 6) post-error slowing (Botvinick et al., 2001): whether the proceeding press was incorrect (1) or correct (0). All models also included an intercept. Importantly, the effect of explicit knowledge (4) was modeled to dissociate non-representational account of performance gain, such as increased movement vigor for familiar sequences (Wong et al., 2015).

For each model (m), we calculated cross-validated log-likelihood $(\log \mathcal{L}_m^{cv})$ from the sum of squared errors. As each trial in the follow-up session had N repetitive executions per one sequence-type (5 times for 2 participants, and 4 for the remaining participants), we fitted a model using IPIs from N-1 executions and then tested the model with the left-out IPI data. We further applied Bayesian model-averaging to estimate regression weight and corresponding contribution of each model component taking the presence of all other components into account.

The regression coefficients for the models were averaged by using the following variant of Akaike-weight:

$$w_m = \frac{\exp\left(\log \mathcal{L}_m^{cv}\right)}{\sum \exp\left(\log \mathcal{L}_i^{cv}\right)},$$

where w_m is the Akaike-weight for the model m, and $\log \mathcal{L}_m^{cv}$ the cross-validated log-likelihood. We obtained similar results by using the AIC (Akaike information criterion) instead of cross-validated log-likelihood to control for model complexity. Model averaging results in better prediction accuracy than using the best model alone (Burnham and Anderson, 2004). For parameter averaging, the regression weights were treated as zero when a model combination did not contain those terms.

The evidence for the presence of each model component was estimated from the cross-validated model evidence $(\log \mathcal{L}_m^{cv})$ of all possible model combination. The log-Bayes factor for each model component $(\log BF_c)$ was calculated as the log of the ratio between averaged likelihood for the models that contained the component (c = 1) versus the averaged likelihood for the models that did not (c = 0) (Burnham and Anderson, 2004; Shen and Ma, 2019)

$$\log BF_c = \log \frac{\frac{1}{N_{m:c=1}} \sum_{m:c=1} \exp\left(\log \mathcal{L}_m^{cv}\right)}{\frac{1}{N_{m:c=0}} \sum_{m:c=0} \exp\left(\log \mathcal{L}_m^{cv}\right)},$$

where $N_{m:c\ =\ 1}$ ($N_{m:c\ =\ 0}$) denotes the number of models (not) containing the component.

Imaging data analysis

Preprocessing and first-level model

Functional imaging data were pre-processed using SPM 8 (https://www.fil.ion.ucl.ac.uk/spm/). Functional images were first slice-time-corrected, motion corrected, and then co-registered to the individual anatomical image. We also corrected for B0 inhomogeneity by using field map images when correcting the head motion. The data were then submitted to a 1st-level GLM to estimate the size of the evoked activity for each sequence in each run. We used the standard high-pass filtering with a cut-off frequency of 128 s before GLM estimation. We applied robust-weighted least square estimation (Diedrichsen and Shadmehr, 2005) to reduce the effect of any motion-induced artifact.

Each trial was modeled as a boxcar function, starting at the presentation of the go-cue with a length of 7.5 s for each execution. The boxcar function was then convolved with a standard hemodynamic response function. In the GLM used for subsequent analysis we included the activity of both correct and incorrect trials in the analysis. This was justified by two reasons. First, even if participants made a mistake, they were instructed to complete the sequence. This often happened automatically, as a substantial number of errors arose from omissions, in which the participants did not apply enough force to have the press registered. Second, given that each trial consisted of two executions of the sequences, many trials consisted of a correct and incorrect execution. Given the low temporal resolution of fMRI, we had little power to resolve this. An alternative analysis in which we excluded error trials from the activity estimation yielded similar, albeit noisier results.

Surface-based analyses

Our primary focus of analysis was cortical surface. We first reconstructed individual cortical surfaces (i.e., the pial and white-gray matter surfaces) from the anatomical image by using Freesurfer software (Fischl et al., 1999). The reconstructed cortical surfaces were then registered to a common symmetrical template (“fsaverage_sym”) (Greve et al., 2013). Subsequently, we defined the surface-based searchlight (Oosterhof et al., 2011) as small circular patches that contains 160 voxels (approximately 11 mm radius) centerd on each node which was defined on the reconstructed cortical surface. The activity patterns of these 160 voxels for each center were submitted to the multivariate analysis, and the result from each searchlight was re-assigned to the center (see Multivariate fMRI analyses: Overview). The overall result was then used to restrict the regions of interest for further analysis.

For detailed testing of representational models, we defined a discrete searchlight parcellation. In contrast to the continuous searchlight map, we aimed to define a reduced set of surface patches. This was done for computational efficiency for model testing, as well as for the model-free clustering approach. The centers of these patches were selected within the regions in which the continuous searchlight analysis showed an average pattern distance across the sequence conditions greater than 0.03 (Figure 4C). Within this region, we defined hexagonally-arranged patch centers on the flattened cortical surface coordinate for both hemispheres (using a 7 mm of spacing). Each patch contained 150 voxels. This definition resulted in 465 (302 for the left hemisphere) partly overlapping patches on the cortical surface.

Multivariate fMRI analyses: Overview

For each of the defined searchlight, the beta-weights for each sequence type for each imaging run were extracted. The resultant beta-weights across voxels were then spatially pre-whitened by using multivariate noise-normalization with a regularized estimate of the spatial noise-covariance matrix (Walther et al., 2016). As a result, the activity estimates across voxels became approximately uncorrelated with respect to the noise (Diedrichsen and Kriegeskorte, 2017). We first took an RSA-approach to restrict the regions where the sequences are encoded. We then applied PCM to the regions with substantial sequence encoding to assess detailed content of sequence encoding.

The key quantity for both of these representational analysis techniques is the second moment (**G**) of the patterns for the sequences. The second-moment matrix is a covariance matrix, where the mean activity for each condition (across voxels) is not subtracted out (Diedrichsen and Kriegeskorte, 2017; Diedrichsen et al., 2018). Thus, when two patterns for sequence *i* and sequence *j* are similar to each other, the corresponding i,j^{th} element of **G** has a high value. For RSA, we computed the cross-validated estimate of **G** and derived from this cross-validated pattern distances. For PCM, we explicitly modeled the structure of **G** and then evaluated the likelihood of the data under these different models.

Overall sequence encoding:

To define the cortical region which reliably encode different sequences, we assessed the discriminability of the elicited activity patterns on the surface-based searchlight. For this purpose, we first calculated a cross-validated estimate of the second moment matrix **G** as,

$$\hat{\mathbf{G}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\mathbf{U}}_m \hat{\mathbf{U}}_{\sim m}^{\mathsf{T}} \Big/ P,$$

where M is the total number of imaging runs, P is the number of voxels within a searchlight, $\hat{\mathbf{U}}_m$ is estimated pre-whitened activity pattern for the m-th imaging run, and $\hat{\mathbf{U}}_{\sim m}$ is the estimate of the pattern independent of the m-th imaging run. Both of $\hat{\mathbf{U}}_m$ and

$\hat{\mathbf{U}}_{\sim m}$ have size of 8 × P. We then computed a cross-validated distance estimate from $\hat{\mathbf{G}}$. The squared cross-validated Mahalanobis distance estimator (crossnobis for short; Diedrichsen et al., 2016), between activity estimates for sequence 1, $\hat{\mathbf{u}}_1$, and for sequence 2, $\hat{\mathbf{u}}_2$, can be calculated as,

$$\hat{d}^2_{1,2} = \hat{\mathbf{G}}_{1,1} - 2\hat{\mathbf{G}}_{1,2} + \hat{\mathbf{G}}_{2,2},$$

where $\hat{\mathbf{G}}_{i,j}$ is the i,j^{th} element of $\hat{\mathbf{G}}$. We calculated the mean of all pairwise crossnobis distance estimators across the sequences at each searchlight (Figure 5). The crossnobis estimator is unbiased – meaning it can be used to directly test whether a distance is larger than zero. Finding consistently positive distance estimates therefor implies that the two condition activity patterns differ from each other more than expected by chance.

Multivariate fMRI analyses: Model-based approach

The above analysis is sensitive to any possible differences between the patterns associated with the different sequences. To dissect different forms of sequence representation, we used pattern component modeling (PCM, RRID:SCR_015891) that allows to model the covariance structure (second moment matrix) across the activity patterns according to different representational hypotheses (Diedrichsen and Kriegeskorte, 2017; Diedrichsen et al., 2011, 2018; Yokoi et al., 2018). For our experiment, we defined the following five representational model components, including a null-model which predicts no difference across the sequences.

Sequence model

This model component assumes that each sequence is associated with unique activity pattern with common variance,

$$\mathbf{U} = \mathbf{M}_{seq}\mathbf{U}_{seq},$$

where the weighting matrix $\mathbf{M}_{seq}$ is an identity matrix (i.e., $\mathbf{M}_{seq} = \mathbf{I}_8$) and the pattern $\mathbf{U}_{seq}$ is uncorrelated (i.e., $\mathbf{U}_{seq}\mathbf{U}_{seq}^T = \sigma^2_{seq}\mathbf{I}_8$). Therefore, the predicted second moment matrix has the simple form of,

$$\mathbf{G}_{seq} = \sigma^2_{seq}\mathbf{I}.$$

Chunk model

The chunk model assumes that the activity for each sequence is a combination of activities associated with the chunks it contains,

$$\mathbf{U} = \mathbf{M}_{chunk}\mathbf{U}_{chunk},$$

where the weighting matrix $\mathbf{M}_{chunk}$ specifies the membership of chunks used in each sequence (e.g., $\mathbf{M}^1_{chunk} = [1,1,1,1,0,0,0,0]$, Figure 1b), and the pattern $\mathbf{U}_{chunk}$ is also assumed to be uncorrelated (i.e., $\mathbf{U}_{chunk}\mathbf{U}^T_{chunk} = \sigma^2_{chunk}\mathbf{I}_8$). Therefore, the predicted second moment matrix has the specific form that reflects the composition of sequences in terms of chunks,

$$\mathbf{G}_{seq} = \sigma^2_{chunk}\mathbf{M}_{chunk}\mathbf{M}^T_{chunk}.$$

1st-finger model

We have previously shown that differences in the sequence-specific activity pattern of M1 and S1 can we well explained by the fact that the first finger press shows a particularly strong activation compared to the subsequent finger presses (Yokoi et al., 2018). Because each sequence contained each finger equally often (and because the peak force of all finger presses was approximately the same), we can assume that the only thing that would differentiate these sequences in a region that only encodes single-finger movements, is which finger starts the sequence. The first-finger model therefore characterizes the part of the activity pattern that is different between sequences as a scaled version of the pattern associated with the first finger,

$$\mathbf{U} = \mathbf{M}_{1f}\mathbf{U}_{sf},$$

where the 8 × 5 weighting matrix $\mathbf{M}_{1f}$ has a scaler σ_{1f} at the column corresponds to the starting finger of each sequence (row), and the $\mathbf{U}_{sf}$ is the activity patterns associated with single finger presses. As we did not measure the single finger activity $\mathbf{U}_{sf}$ for the current experiment, we utilized the fact that the second moment of single finger patterns $\mathbf{G}_{sf}$ in M1 and S1 is well-characterized by the natural statistics of hand usage (Ejaz et al., 2015). The predicted second moment matrix is therefore,

$$\mathbf{G}_{1f} = \mathbf{M}_{1f}\left(\mathbf{U}_{sf}\mathbf{U}^T_{sf}\right)\mathbf{M}^T_{1f} = \mathbf{M}_{1f}\mathbf{G}_{sf}\mathbf{M}^T_{1f} = \mathbf{M}_{1f}\mathbf{G}_{ns}\mathbf{M}^T_{1f},$$

where $\mathbf{G}_{ns}$ is a second moment matrix predicted from the natural statistic of hand movement (Ejaz et al., 2015). Replacing $\mathbf{G}_{ns}$ with the second moment matrices for single finger representation that was derived from an independent experiment (Yokoi et al., 2018) did not affect the results. The second moment matrix predicted by this model therefore reflects which sequences share the first finger, and how similar the respective finger representations are to each other.

Null-model

As a baseline to evaluate each model, we defined a null-model that hypothesized no difference between any of the sequence patterns. For this, the hypothesized second moment matrix was

$$\mathbf{G}_0 = \mathbf{0}_8,$$

where $\mathbf{0}_8$ is a 8 × 8 matrix whose elements are all 0.

Noise-ceiling model

In addition to the above models, we also fitted a "noise-ceiling" model to assess the maximally explainable information shared across individual. Here, we used a naive noise-ceiling model (Diedrichsen et al., 2018; Nili et al., 2014) that uses the empirical, cross-validated estimate of the second moment matrix as the predicted second moment matrix;

$$\mathbf{G}_{NC} = \widehat{\mathbf{G}}.$$

Because it simply uses empirical mean of the observed representational structure (i.e., second moment matrix) over subjects as a model, the resultant cross-validated log-likelihood value approximates the best-achievable performance by any of the models, serving as a positive control. Therefore, significant advantage of the noise-ceiling model over the best model considered indicates the existence of better models. We fitted this noise-ceiling model separately for the two groups, as these groups of participants practiced partly different set of sequences with different chunking structure. We then combined the evidence for both groups for evaluation.

Model design

The prediction of each model was determined for each group separately, as they differed in both sequences and chunks. Because the two groups had 5 sequences that were identical, but were chunked differently, difference in the representational structure across these shared sequences could be specifically attributed to the difference at the level of the chunk representation.

Importantly, the activity estimates for each trial contained trials that either contained an error, or were not completed (see Preprocessing and first-level modeling). We included these trials, because even for these incorrect trials, most of the sequence was produced correctly. To account for the fact that the beginning of the sequence was more often executed than the end of the sequence, we weighted each element of the representation (e.g., the first, second, third and fourth chunks or the first to tenth transitions) by the relative %-correct with which this element was produced. For example, the %-correct of chunks were on average 96, 94, 86, and 82%, hence each chunk element in the PCM model was weighted accordingly, which slightly changed the structure of corresponding weighting matrix.

Model evaluation

First, we fitted all possible combinations of the candidate models (i.e., first-finger, chunk, and entire sequence models). As each of the combination models had different number of free parameters, i.e., combination weights, we evaluated the models using leave-one-subject-out cross-validation to prevent overfitting (Diedrichsen et al., 2018). Because the combination weights should be the same for the same cortical patch across the two groups of participants with different chunking structures, they were constrained to be the same across the two groups. We then used the resultant cross-validated log-likelihood $(\log \mathcal{L}_m^{cv})$ for each model m as an estimate of the model evidence. Superiority of one model (e.g., some model combination) over the other (e.g., null or noise-ceiling model) was evaluated by directly comparing model evidence $(\log \mathcal{L}_m^{cv})$ between the two models as;

$$\log \mathrm{BF}_{AB} = \log \frac{\mathcal{L}_A^{cv}}{\mathcal{L}_B^{cv}} = \log \mathcal{L}_A^{cv} - \log \mathcal{L}_B^{cv},$$

where $\log \mathrm{BF}_{AB}$ represents the log of Bayes factor (BF; Kass and Raftery, 1995) of model A over model B. The evidence for each of the model components (i.e., first-finger, chunk, and sequence) in the context of other components $(\log \mathrm{BF}_c)$ was calculated using Bayesian model-averaging in the same manner described in behavioral data analysis (see Linear IPI modeling).

Model-based representational fMRI analysis for sub-cortical ROIs

We applied the same model-based analysis described above on selected sub-cortical ROIs, which included caudate nucleus, globus pallidum, putamen, and thalamus bilaterally. The ventral aspects the brainstem and the cerebellum were not included in our imaging volumes (see Imaging data acquisition). These sub-cortical ROIs were segmented by using FSL *FIRST* algorithm (Patenaude et al., 2011), and then transformed into individual space. It should be noted that the spatial resolution of the current study (2.3 mm isotropic) may not be optimal for applying MVPA to these structures, as indicated by the relatively weak sensitivity of our result (Figure S1).

Multivariate fMRI analyses: Model-free clustering

As a complement of our model-based approach, we also applied a model-free clustering of cortical surface regions based on their representational structure. As an input data, we calculated the crossnobis distance estimator (see Multivariate fMRI analysis: Overview) for all 28 sequence pairs for each surface patch. The input data for clustering was a N-by-28 matrix, where N is the number of surface patches. We calculated an N-by-N similarity matrix across the patches for each participant by first calculating the correlation distance $d = (1\text{-}r)$ between the representational structures. We chose the correlation distance in order to emphasize the profiles of the crossnobis distances, rather than their magnitude, which is determined by the signal-to-noise ratio in the region. The resultant correlation distances were then transformed into similarity using a Gaussian similarity transformation;

$$w = \exp\left(-\frac{d^2}{2\sigma^2}\right),$$

where w is the similarity, d the correlation distance, and σ the width of Gaussian. The width was determined individually as the lower 5-percentile value of the correlation distances. We then applied the spectral clustering algorithm (Von Luxburg, 2007) to the group-averaged-similarity matrix with Jordan-Weiss normalization (Ng et al., 2002).

Statistical tests

All the statistical analyses were performed on MATLAB R2015b (Mathworks, Inc.). We set the significance level to $p = 0.05$. When a test involved multiple comparisons (i.e., Figure 3E), it was divided by the number of comparisons (i.e., 4).

Behavioral data analysis

To compare within- and between-chunk intervals, we used paired t test. To test group-specificity of press-interval patterns, pairwise correlation coefficients (Pearson's *r*) across the subjects (within- and across-group) were first *z*-transformed and then submitted to a two-sample t test.

For the regression analysis of IPI data, the model-averaged log-Bayes factors $(\log BF_c)$ were separately calculated for each participant and then submitted to the Bayesian group analysis, which estimates the posterior model probability (Rigoux et al., 2014; Rosa et al., 2010; Stephan et al., 2009) (spm_BMS() function implemented in the SPM 12). Significance of model contribution was assessed by the protected exceedance probability (PXP), which is the posterior probability of a model being more likely than any other competing model including chance-level (i.e., model probabilities are equal across all the models considered) (Rosa et al., 2010). In our case of testing two models (i.e., presence versus absence of an effect), the value of PXP simply indicates the posterior probability that an effect is present in more than half of the participants, above and beyond chance. We deemed a model contribution is significant when PXP is larger than 0.75.

Imaging data analysis

The resultant log-Bayes factors both for the simple comparison between two models $(\log BF)$ and overall contribution of a specific model component $(\log BF_c)$ were submitted to the group-level Bayesian analysis, as described above. We again used the PXP of 0.75 as a threshold of group-level significance of model evidence. Group-log-Bayes factor maps were further thresholded with the log-Bayes factor of one (Kass and Raftery, 1995).

To test significant overlap between two representations A and B than chance-level, we used the likelihood-ratio *G*-test (Sokal and Rohlf, 1995). We counted the number of cortical patches with significant evidence ($\log BF_c > 1$ and $PXP > 0.75$) for first-finger, chunk, or sequence representations that also satisfied $\log BF > 1$ for the noise-ceiling model. The observed counts were then compared to the expected number of counts if the presence/absence of the representation was independent of each other.

To test whether observed surface patches with significant chunk and sequence representation (Figure 6G) represent two separate groups, we compared Gaussian mixture models with one component and two components using leave-one-out cross validation. Subsequently, the log-Bayes factor for the mixture model over the single Gaussian model ($\log BF_{DS}$) was calculated as,

$$\log BF_{DS} = \log \mathcal{L}_D^{cv} - \log \mathcal{L}_S^{cv},$$

where $\log \mathcal{L}_D^{cv}$ and $\log \mathcal{L}_S^{cv}$ represent the cross-validated log-likelihood of the double- and single-Gaussian mixture models, respectively. Parameters for the mixture models (mean, covariance, and mixing ratio) were estimated using standard EM algorithm (Bishop, 2006). At each cross-validation fold, convergence of the mixture model to global minimum was ensured by iterating the estimation from 100 randomly chosen starting values. Strength of evidence was assessed according to Kass and Raftery (1995).

DATA AND CODE AVAILABILITY

The MATLAB code used for the multivariate fMRI analysis (Pattern Component Modeling Toolbox, RRID:SCR_015891) are available online (https://github.com/jdiedrichsen/pcm_toolbox). Upon publication data and other custom-written codes to reproduce all figures will be made available online (https://github.com/ayokoi/sh1).