

CORTICAL AND BEHAVIORAL SIGNATURES OF
RESPONSE BINDING DURING SEQUENTIAL
SKILL LEARNING

by

Patrick Leininger Beukema

BA, McGill University, 2009

MS, Carnegie Mellon University, 2011

Submitted to the Graduate Faculty of
the Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Patrick Leininger Beukema

It was defended on

April 20, 2018

and approved by

Timothy Verstynen, Associate Professor, Psychology

Beatriz Luna, Professor, Psychiatry

Robert Kass, Professor, Statistics

Steven Chase, Associate Professor, Biomedical Engineering

Robert Turner, Professor, Neurobiology

Jörn Diedrichsen, Professor, Computer Science

Dissertation Director: Timothy Verstynen, Associate Professor, Psychology

Copyright © by Patrick Leininger Beukema
2018

CORTICAL AND BEHAVIORAL SIGNATURES OF RESPONSE BINDING DURING SEQUENTIAL SKILL LEARNING

Patrick Leininger Beukema, PhD

University of Pittsburgh, 2018

Sequence learning plays a central role in motor skill acquisition. One of the algorithms that is essential to the consolidation process is binding, where multiple movements are driven by a single motor command. Here we show that this binding process originates from learning the transitional probabilities between sequential movements. To determine how binding impacted the neural population patterns which control the movements, we generated unbiased encoding networks of the movement representations at multiple levels of the motor hierarchy. The representational geometries associated with individual movements and movement sets were readily dissociable. In addition, the set network comprised a subset of the individual movement network. However, extensive training by human participants on a sequence learning task did not result in any shift of the representational geometries. This suggests that the representations associated with expert performance are rapidly constructed and highly stable.

TABLE OF CONTENTS

PREFACE	x
1.0 PREDICTING AND BINDING: INTERACTING ALGORITHMS	1
1.1 INTRODUCTION	1
1.1.1 PREDICTING SERIAL ORDERS OF EVENTS	3
1.1.2 BINDING	4
1.2 OPEN QUESTIONS	8
1.3 SPECIFIC AIMS	12
1.3.1 Aim 1	12
1.3.2 Aim 2	12
2.0 BINDING AND INDIVIDUAL ACTIONS	15
2.1 INTRODUCTION	15
2.2 RESULTS	17
2.2.1 Behavior	17
2.2.2 Movement and cue set representations	22
2.2.3 Stability of representational geometries	23
2.3 DISCUSSION	29
2.4 MATERIALS AND METHODS	31
2.4.1 Participants	31
2.4.2 Serial reaction time task	31
2.4.3 Analysis of training data	32
2.4.4 Imaging acquisition	33
2.4.5 Neuroimaging tasks	34

2.4.6 Imaging analysis	34
3.0 BINDING AND SETS OF ACTIONS	37
3.1 INTRODUCTION	37
3.2 RESULTS	38
3.2.1 Dissociating visual and motor set encoding	38
3.2.2 Movement and cue sets are distinctly represented	40
3.2.3 Region of interest analysis	42
3.2.4 Representational dissimilarity matrices of sets	42
3.2.5 Distinguishing first item effects	45
3.2.6 Cue and movement sequence learning	46
3.2.7 Effects of training on the set representations	51
3.3 DISCUSSION	58
3.4 MATERIALS AND METHODS	60
3.4.1 Participants	60
3.4.2 Scanner Task	60
3.4.3 Serial reaction time task	61
3.4.4 Analysis of training data	62
3.4.5 Imaging acquisition	62
3.4.6 Imaging data analysis	63
3.4.7 Representational Similarity Analysis	63
3.4.8 Whole brain searchlight	64
3.4.9 Regions of interest	64
3.4.10 Statistical testing	65
3.4.10.1 Software Accessibility	65
3.4.10.2 Data Availability	65
4.0 SUMMARY AND CONCLUSIONS	66
4.1 FINAL SUMMARY	66
4.2 METHODOLOGICAL LIMITATIONS	67
4.3 INSIGHTS	69
4.4 CONCLUSION	73

BIBLIOGRAPHY	74
-------------------------------	----

LIST OF TABLES

1	Unbiasedness of individual movement encoding	25
2	Stability of individual movement representations	27
3	Stability of frequently paired movements	28
4	Stability of infrequently paired movements	28
5	Movement and cue mappings used to disassociate responses	41
6	Statistics for comparing plasticity across groups	57

LIST OF FIGURES

1	Continuum of sensorimotor sequence learning	2
2	Dual algorithm model of sequence learning	6
3	Illustration of representational similarity analysis (RSA)	11
4	Visualization of the impact of binding on representational distances	14
5	The process of response binding	18
6	Task design and behavioral performance	19
7	Binding originates from transitional probabilities	21
8	Multivariate activity patterns during cued finger movements	24
9	Stable representational distances after training	26
10	fMRI task used to dissociate movement and cue set representations	39
11	Whole brain surface searchlight showing movement and cue set encoding	43
12	Absence of overlap in encoding networks	44
13	Representational structures of movement sets and cue sets	47
14	Representational structures of first item effects	48
15	Representative reaction time plot	50
16	Learning measures for cue and motor set learning	52
17	Correlation between movements after binding	53
18	Binding similarity across groups	54
19	Stability of set representational geometries	56
20	Summary networks showing individual movement and set encoding	70

PREFACE

Thank you to my brothers in this endeavor, Kevin Jarbo and Kyle Dunovan. I will never forget our many discussions and profound insights - such as the number of degrees of freedom in a cooked spaghetti noodle. Thank you to my advisor, Tim Verstynen, for guidance and endless encouragement, and for encouraging more open and reproducible science. Thank you to Ran Liu, for being a brilliant scientist, partner, and friend. This dissertation is dedicated to my parents, Michael Beukema and Jean Anne Leininger.

1.0 PREDICTING AND BINDING: INTERACTING ALGORITHMS¹

1.1 INTRODUCTION

Many everyday behaviors depend on our ability to effortlessly produce complex, serially ordered actions. For example, typing the word “brain” on a keyboard requires serially pressing the “b”, “r”, “a”, “i”, and “n” keys as quickly and accurately as possible. Novices execute each key press slowly, planning each successive movement independently. In contrast, experts can perform the same series of keystrokes in rapid succession, executed as a fluid, unified action. This ability to execute a unified set of serially ordered actions represents one example of a more general form of skill learning known as sensorimotor sequence learning.

Sensorimotor sequence learning has fascinated cognitive science for over sixty years [2] and is thought to play a central role in a wide range of intelligent behaviors, including language learning (see [3] for a review). Initial research sought to characterize this form of learning as a singular process [4] or as a single learning system operating at different representational levels [5]. Now, overwhelming evidence supports the hypothesis that the consolidation of sequential motor skills relies on multiple interacting systems that learn different parts of the serial ordering problem at different timescales. Here we show how these systems can be segregated into two general categories of learning processes, based on the computational goals they serve: fast prediction of serial ordered events and slow binding of responses into sets of unified actions (Fig. 1). These two interacting computational mechanisms operate along a learning continuum, between sensory and motor levels, working together to sculpt behaviors over time to maximize the complexity of produced actions while minimizing the computational costs of planning and executing them.

¹This chapter is reproduced from previously published work [1].

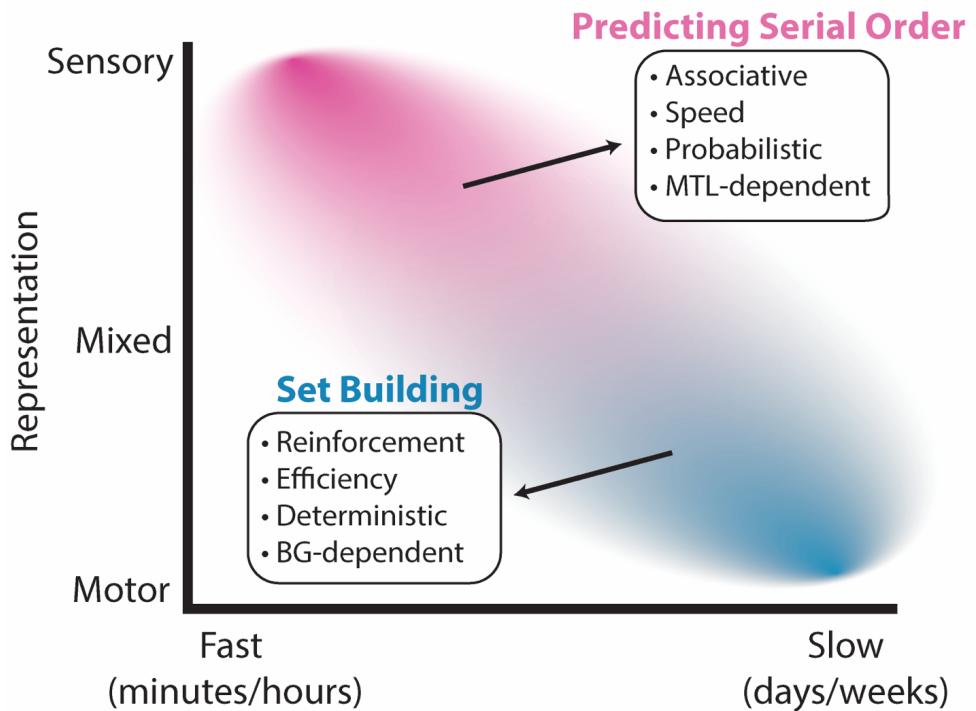


Figure 1: Continuum of sensorimotor sequence learning. Schematic of the temporal and representational spaces occupied by processes that learn to predict transition probabilities between serially ordered events (magenta) and processes that learn to bind actions together into unified sets (blue). Insets indicate summary features of each learning process that are described in detail in text. MTL, medial temporal lobe; BG, basal ganglia.

1.1.1 PREDICTING SERIAL ORDERS OF EVENTS

Early in learning, a novice typer will show faster and more accurate responses to frequently paired serial actions than to infrequently paired actions. For example, repeatedly typing the word “brain” leads to faster “r” key presses when they follow a “b” key press, but pressing the “r” key would be slower if the preceding letter was something unusual, such as “q”. In this way, the brain associatively learns the transition probabilities between sequential stimulus-response events over the course of only a few minutes of practice [4, 6]. In the context of language development, this same process is referred to as statistical learning [7]. Classically, statistical learning refers to the phenomenon whereby neural and behavioral responses become more efficient to serially repeated sensory events than to unexpected events. This learning happens very quickly and can be detected within the course of a single training session (for a review of statistical learning, see [8]). In this way, over the course of several minutes of repeated exposure, the brain learns to estimate the conditional probability of an upcoming sensorimotor plan, $X_{(t+1)}$, given the immediately preceding plans, $P(X_{(t+1)}|X_t, X_{(t-1)}, \dots, X_{(t-n)})$, to make faster and more accurate responses.

This predictive process is evident in very early sensory processing [9, 10]. For example, neurons in the inferotemporal cortex (IT), a visual processing area, learn to modulate their responses to serially presented visual stimuli depending on the transition probability between cues [10]. Yet IT neurons do not simply track the co-occurrence of stimuli, but also appear to track the conditional probabilities of events, as their activity is attenuated when the conditional probabilities are modified [11]. Of course, learning transition probabilities may not be restricted to predicting sensory cues. For example, in a classic study, Mushiake and colleagues [12] showed that cells in the macaque dorsal premotor cortex (PMd) were tuned to the transition probability between sequentially cued movements [12], suggesting that motor systems also track the transitions between serial actions. Although given the extensive training required to get the animals to learn the task, this may reflect a probabilistic variant of chunking, rather than true statistical learning (see 1.1).

Mechanistically, recent evidence suggests that this fast detection of serial ordering may rely, at least in part, on the medial temporal lobe, particularly the hippocampus [13, 14, 15,

[16, 17, 18, 19]. A sub-population of hippocampal cells shows tuning for the temporal associations between sequences of events [20, 21, 22], suggesting that the hippocampus may track serial probabilities and bias cortical sensory and motor processing via top-down signals. This hypothesis is bolstered by several other lines of evidence. For example, the consolidation of complex response sequences is improved following a normal sleep cycle [16, 23, 24, 25], a classic signature of hippocampal-dependent learning. Co-activation of hippocampal and striatal networks is observed during sequence learning [26, 27], particularly when learning the temporal structure of sequential events [28]. Finally, patients with damage to the hippocampus show impairments in single-session sensorimotor sequence learning [29, 30], particularly during the initial acquisition phases of learning when declarative mechanisms are crucial for picking up transition probabilities between stimuli [31].

Taken together, the emerging evidence suggests that fast associative mechanisms learn first-order transition probabilities between both sensory cues and actions early in learning (see also [32]). This ability to reliably predict upcoming events speeds up the ability to resolve a stimulus-response mapping and thus results in faster responses, likely through adjusting the threshold for evidence needed to initiate a response (for review see [33]).

1.1.2 BINDING

Relying solely on learning first order transition probabilities limits the capacity of producing complex sequential actions. This is because the number of events, n , that can be included in the estimate of the conditional probability, $P(X_{(t+1)}|X_t, \dots, X_{(t-n)})$, is constrained by working memory capacity [34]. One way to overcome this memory limitation is to learn the hierarchical organization of movements and bind sequences of actions into sets or “chunks” [35]. Returning to the typing example, after extensive practice the plan to execute the set of key presses “b”, “r”, “a”, “i”, and “n”, can be represented internally as a single action decision “b-r-a-i-n”, where actions are coarticulated together in a unified manner (Fig. 2B). While each action within the bound set carries its own execution noise, the action initiation decision shifts from waiting for individual sensory cues to automatically triggering one item after another without reliance on sensory cues. Using an optimal control theory

framework, Ramkumar et. al. [36] demonstrated that as animals become more experienced at a sequential skill, efficiency increases as the number of chunks decreases, reflecting more bound elements in each chunk, while simultaneously minimizing the overall computational cost during learning [36]. Therefore, binding actions into sets or chunks may represent a critical step in the optimal solution to the problem of complex sequential movements, where action sets can themselves be formed at multiple levels of movement hierarchy (e.g., goals, plans, execution).

The organization of serially ordered behaviors into unified sets is often controlled experimentally by making the transition probabilities between cued actions completely deterministic. This can be done by explicitly presenting the sequential order before production [37] or through extensive practice on short action sequences [38]. Behaviorally both approaches lead to slower responses to the first item in the set than to subsequent items in the series [38, 39, 40, 41, 42, 43]. This slowing could be due to the fact that the first item in the set has no preceding event with which to estimate the transition probability or the result of the increased time associated with loading the motor buffer [38]. While this first item slowing has classically been used as a behavioral signature for sensorimotor “chunking”, it is not sensitive to detecting whether the responses within the set are bound together under a shared motor decision, nor does it easily allow for looking at the natural evolution of sensorimotor sets during learning.

More recent research has focused on the concept of binding by looking at correlations between temporally adjacent movements within a common set. Several studies in both human and non-human primates show that it takes days of practice or longer to detect the emergence of binding between serial actions under a shared motor command [36, 44, 45]. For example, the and colleagues [45] found a dissociation between simple decreases in response time during sequence production, observed during a single session of training, versus correlations in response times between serially ordered actions, that did not emerge until after several days of training [45]. The binding hypothesis predicts this correlation as the result of multiple bound movements originating from a common generative motor command.

The ability to detect binding in sequential responses, as opposed to demarcation of chunk

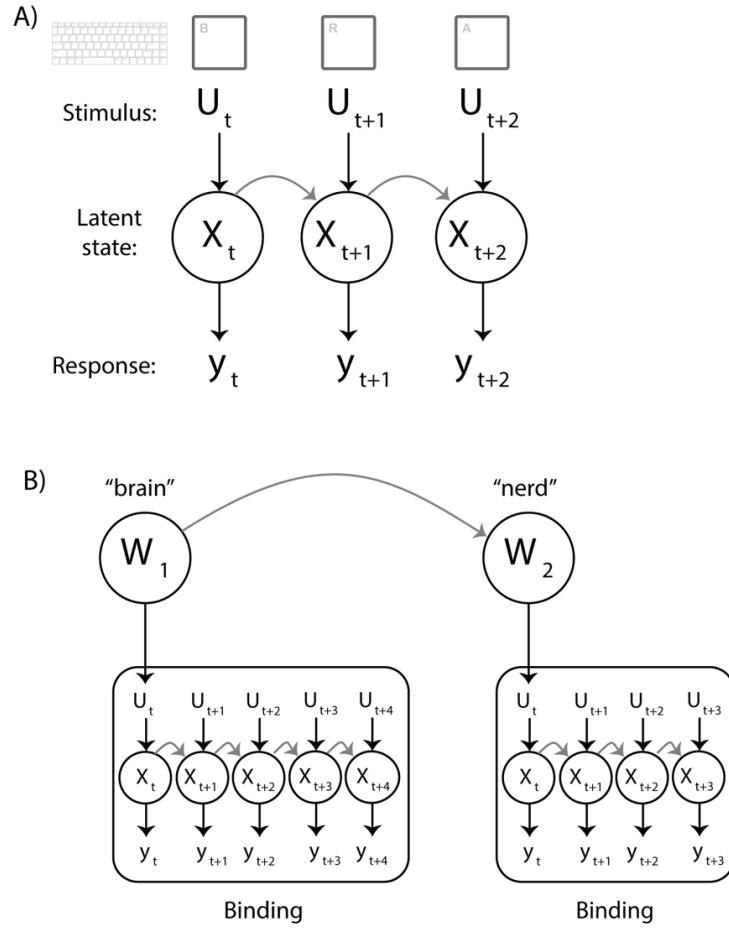


Figure 2: Dual algorithm model of sequence learning. A) Learning transitional probabilities. Typing the word "brain" on a keyboard is a temporally organized series of responses. Each letter key (e.g. "b") is a stimulus, U , that corresponds to a particular internal sensorimotor plan represented as a latent state (X), that initiates a response, Y , at time t . With training, learning transition probabilities between events increases the efficiency of sensorimotor processing for subsequent events, leading to increased speed and more accurate selections when the expected event occurs. B) Building action sets. Over time, once the transition probabilities between serially ordered items (e.g. "b", "r", "a", "i", "n" in "brain") becomes deterministic, it is computationally efficient to group sensorimotor decisions into unified sets, e.g., a "chunk", where the entire set is represented as a single, generative sensorimotor decision, W , and sequences of bound sets of actions can then be the target of learning, i.e., $P(W_2|W_1)$.

boundaries with the first item slowing, opens the door to ask questions about where in the sensorimotor hierarchy this binding occurs over the course of learning. Recently, Lynch and colleagues [46] adopted a novel remapping paradigm that dissociates learning ordered sets of visual cues, across days of training, from learning ordered sets of finger movements [46]. Relying on the same correlation measure as Verstynen et al. (2012), we found that action binding was stronger when sequences were learned in the sensory domain than when they were learned motorically. Importantly, the level of explicit awareness of the sequence, and thus reliance on declarative processes, was not affected by whether a visual or motor sequence was learned. Of course this does not necessarily exclude the possibility that binding occurs in motor representations [47]. Multivariate pattern analysis approaches have recently opened the door to exploring the nature of action representations with neuroimaging tools like fMRI. Using this approach, Wiestler and Diedrichsen (2013) showed that cued sequence sets, akin to explicitly cued chunks, can be reliably decoded from population-level activity in higher order motor cortical areas, such as the supplementary motor area (SMA) [48]. Later work by the same group showed that, while the patterns of activity for individual fingers are organized in the primary motor cortex (M1) according to the natural statistics of everyday hand use [49], the population level activity of M1 itself does not appear to distinguish well-learned sets of actions [50]. Instead, the patterns of task-related activity in upstream premotor regions, such as the dorsal premotor cortex, more reliably distinguished between learned sets of actions [50] (see also [51]).

Mechanistically the implementation of this binding process appears to rely, in part, on basal ganglia (BG) pathways (but see [52]). For example, patients with Parkinson’s Disease (PD) have deficits in chunking ability when they are in low dopamine states [53]. At the neural level, cells in the striatum, the main input nucleus to the BG, become tuned to bracketing segments of sequential actions over time, particularly as action sequences become habitual [54, 55]. Based on the variety of action sequence-linked cell types in the striatum, Jin and colleagues [56] proposed that during learning, the striatum facilitates concatenating, or binding actions together. As this binding process unfolds and action sets become established, a subset of cells in the striatum, likely in more executive regions, become sensitive to the initiation of the bound set of actions [57, 58, 59, 60]. Since the BG are thought to gate motor

responses, this onset sensitivity of striatal cells is consistent with the notion that the entire sequence set becomes a unique action decision that gets triggered by BG pathways.

The emerging behavioral and neuroscientific evidence points to an interacting set of algorithms that contribute to the long-term consolidation of sequential skills. Fast associative processes estimate the transition probabilities between serially ordered events so as to improve the speed and efficiency of stimulus-response gating. These associative mechanisms appear to primarily target the processing of sensory signals, but may also impact downstream motor processes. As the transition probabilities between actions become deterministic, computational complexity is reduced by having reinforcement learning processes unify sets of actions and initiate the bound set as a single decision. Thus the initiation of subsequent actions no longer depends on sensory cues, but on the state of the preceding actions in the set. Signatures of this binding mechanism are primarily found to impact motor computations, but can be moderated by upstream sensory processing as well (e.g., [46]). Rather than reflect serial stages of processing, these associative and binding mechanisms appear to interact during the consolidation process to support balancing goals of making fast and accurate responses while also reducing computational complexity.

1.2 OPEN QUESTIONS

Thus far, the field has focused on understanding sensorimotor sequencing learning by focusing mostly on the fast, predictive associative mechanisms, but less so on the process of binding. One challenge is that the longer time scales necessary to detect the emergence of bound movements together are more costly to study in laboratory environments compared to prediction and associative mechanisms that operate on much faster time scales. Therefore, less is known about the behavioral markers and neural mechanisms associated with the prolonged refinement of sequential motor skills that results from binding. Our review of the current literature suggests two potential avenues. First, as the brain learns transition probabilities so that series of actions become deterministic, it can begin to bind actions together. This means that binding should reflect the learned statistical regularities. This

predicts correlated responses according to the transition probabilities between elements of a sequence. Second, binding should coincide with changes to the relations between movement representations in cortical motor areas.

To better understand neural mechanisms associated with sequential skill learning, neuroimaging studies have usually examined what regions of the brain show more or less activation after training. Greater activation is typically inferred to reflect increased recruitment [61, 62, 63], while reduced activation is often thought to reflect increased efficiency [64, 65, 66, 67]. However, these inferences are tenuous at best. It is equally plausible that greater activation could be the result of non-skill-specific increases in performance (e.g., response speed) while reduced activation could be the result of reduced involvement rather than increased efficiency [68]. It is also possible for metabolic activity to reduce without any change in neuronal firing [69]. These observations make it difficult to interpret changes in the magnitude of the BOLD signal after sequence learning, insofar as understanding the consequences of plasticity on neural representations.

Any successful account of the binding mechanism during sequential skill learning will necessitate understanding how sequential information is represented, and how the underlying representations change from before to after learning. Therefore, rather than examining the average magnitude of activation, which ignores fine grained spatial information, it is necessary to examine the representational patterns themselves. One emerging method of examining neural activation patterns is representational similarity analysis (RSA) [70, 71, 72, 73], illustrated in Fig. 3. RSA compares different conditions, such as individual finger movements (Fig. 3A), in terms of their representational similarity. These representational patterns are the measurement of the brain activity during each of these conditions (Fig. 3B). The multivariate pattern of activity across measurement channels (Fig. 3C) are the representations of each condition within a particular area. The similarity between these representation is computed using one of several distance metrics [71]. The pairwise similarities between all of the conditions are then summarized in a representational dissimilarity matrix (RDM) (Fig. 3D). In the context of RSA, each representation can be thought of as a point in a high dimensional space spanned by the measurement channels (e.g. voxels). The relationship between all of these points specifies a representational geometry [74, 75]. Because represen-

tational geometries are thought to remain stable across a wide range of BOLD amplitudes [76], they are especially useful in studies of motor skill acquisition where movement speed can confound BOLD interpretations. Therefore, representational geometries are a promising tool to investigate the neural plasticity associated with binding.

How should sequence learning, and specifically response binding, impact representational geometries? Since patterns are inherently of much higher dimensionality than BOLD signal magnitude, it is not clear how plasticity should be characterized in terms of representational change. In visual cortex, plasticity is characterized by a shift of the representational geometry due to certain patterns of activation becoming more or less similar to one another [17]. Representational geometries have also been shown to shift in somatosensory cortex, for example, following a 24 hr intervention where two fingers were yoked together [77]. These findings in sensory regions suggest that long term consolidation of a sensorimotor sequence may also be reflected in the representational geometries of actions in sensorimotor areas.

A movement sequence is composed of individual sub-movements that each generate some pattern of activity, or representation, across various motor cortical regions, in addition to subcortical structures. Therefore, the representational geometry associated with a sequence describes the similarity between sub-movements that comprise the sequence. Plasticity associated with representational geometries would entail a reorganization of the elemental movements that make up a sequence. To quantify the change in representational geometry, we summarized the RDM in terms of the average distance across all conditions (Fig. 3D). This measure was used as a neural plasticity marker for all the neuroimaging analyses. We examined the representational geometries associated with sequential movements, before and after training, using two different experimental tasks.

In the first task (Aim 1), we explore the impact of binding during sequence learning on the representations of individual elements of the sequence (i.e., individual cue-finger pairs). We hypothesized that the representational distances between more frequently paired actions would decrease due to the fact that they become bound in a common motor command (Fig. 4A). Since this would decrease behavioral flexibility in primary motor areas [2], we suggest that this process would only result in representational changes in higher order planning regions.

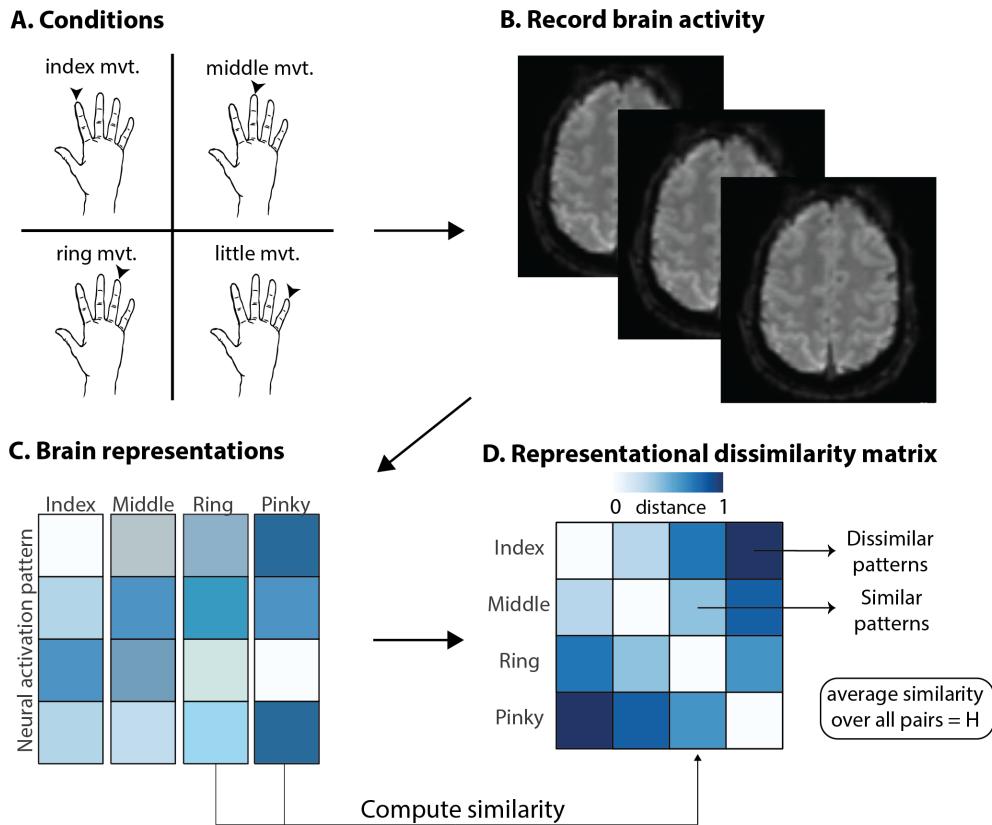


Figure 3: Illustration of representational similarity analysis (RSA). A. RSA is used to compare how different conditions are represented and related to one another in the brain. The illustrated conditions are four different finger movements: index, middle, ring, and little finger movements (mvt.). B. Images of the brain are recorded during the different conditions. C. Each condition will generate a neural activation pattern, or representation. These representations are activity measurements at different measurement channels in the brain (e.g. voxels in fMRI). D. Computing the similarity between all the conditions results in a representational dissimilarity matrix which summarizes the pairwise distances between all conditions. The average similarity over all pairs (H) can be used as a summary statistic for tests of encoding and plasticity. In ch.2 and ch.3, this statistic was examined throughout the entire cortical surface, using a searchlight, to test for encoding of the conditions.

We have previously shown that sequences can be learned in different modalities, such as cue and motoric [46]. Therefore, in the second task (Aim 2), we explore the impact of binding during sequence learning on the representations of sets of actions within specific modalities (i.e. visual vs. motoric sequences). If binding of actions within a set refines the relative representation of that set, then this should lead to an increase of distances between sets of actions. Importantly, this should only happen within the modality that the set was trained in (Fig. 4B).

1.3 SPECIFIC AIMS

1.3.1 Aim 1

Aim 1 examines the nature of binding and its influence on cortical motor representations of individual actions. We trained human subjects for 25 days on a complex finger movement sequence to determine if responses were bound after training and whether binding reflected the transitional probabilities of the sequence. Then, using RSA, we examined the plasticity of the population representations of single finger movements. Hypothesis 1.1 The nature of binding in movements will reflect the temporal structure of the learned sequence. Hypothesis 1.2 Cortical movement representations of individual movements will remain stable in primary motor cortex but will change in higher order motor planning areas. Aim 1 is addressed in Chapter 2.

1.3.2 Aim 2

Aim 2 examines the influence of binding on cortical motor representations of sets of actions in two different representational spaces. Previous work from our lab has shown that sets of actions can be learned in both visual and motor reference frames, and that each of these types of sets are learned differently [46]. Here we identified the cortical networks that represent sequence sets in visual and motor modalities to determine whether modality-specific training impacts these representations. We used RSA to isolate the representational pat-

terns associated with cue and movement sets. Next, we trained human subjects for ten days on either cue or motor sequences by remapping the cue-response pairs on each day [46] to determine whether there was an advantage in acquisition for sequences represented as cues or represented as actions. Hypothesis 2.1 Different cortical networks will encode sequences in different modalities (i.e., sensory sequences and motor sequences are learned along orthogonal channels). Hypothesis 2.2 Modality specific training will increase the representational distances for different sequences set *in those networks that encode for that sequence modality*. Aim 2 is addressed in chapter 3.

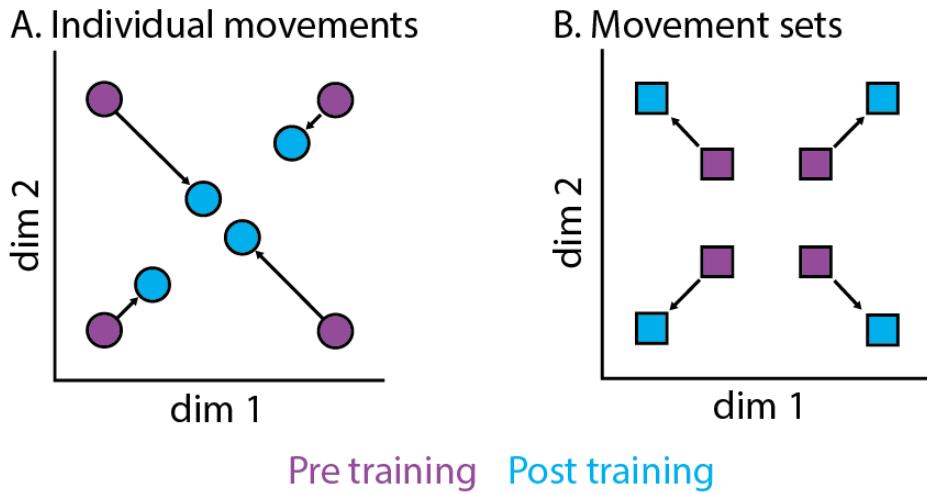


Figure 4: Visualization of the impact of binding on representational distances for both individual actions and action sets. A. Hypothesized distance changes for individual movements in Aim 1. Each circle represents the location of a pattern of activity across voxels. The distances between frequently paired movements should decrease with training more so than the distances between infrequently paired fingers. This would result in a decrease in the average distance H . B. Hypothesized distance changes for sets (visual or motoric) in Aim 2. Each square represents the location of a pattern of activity across voxels. After training, as you refine the learned sets, each set becomes more distinguishable. This results in a greater distance between sets after training. For clarity, only two dimensions are shown, in reality the points occupy a higher dimensional space.

2.0 BINDING AND INDIVIDUAL ACTIONS¹

2.1 INTRODUCTION

Being able to combine simple movements into coordinated sets of actions is critical to many everyday skills, such as typing on the computer or driving a manual transmission car [2]. Over the course of evolution the brain has solved this sequencing problem multiple times, resulting in many interacting algorithms that facilitate the consolidation of complex skills (for review see [78]). One of these algorithms is the process of set building, also known as chunking or binding [38]. Binding serial actions into sets improves computational efficiency during the production of complex actions by representing multiple movements under a single selection command [36]. To illustrate this process, consider the graphical model presented in Fig. 5A-B. On each trial, the manual response to a visual cue occurs through a hierarchical system of perception, selection (e.g., key), and motor planning (e.g., finger movement), that are all represented as latent states with their own independent sources of noise. In this example, the serial order of cues across trials follows a deterministic sequential order. Prior to training, each response is planned independently of the preceding trial. Once the order of cues is learned, the brain can consolidate the selection process so that a set of motor plans is represented under a single selection state. This selection state is triggered by the presentation of the first stimulus in the series, after which subsequent motor commands are cued by the internal state, rather than by the visual cues. This results in faster production of responses to items within a set, as well as a correlation in responses within bound sets due to their shared upstream command (Fig. 5C; [45, 44, 46]).

Many forms of non-sequential motor learning rely on the reorganization of movement

¹This chapter is reproduced from previously published work [1].

representations in motor networks [79]. Therefore, it is possible that action binding during sequence learning also alters internal motor representations of individual movements; however, this effect has been largely unexplored. Recent advances in representational analysis now allow precise quantification of the relationship between the cortical activity patterns for single finger movements using fMRI [73]. Using this approach, previous work has shown that the structure of individual fingers in primary motor cortex is organized in a way that is consistent with their co-articulation during natural hand movements [49]. Furthermore, artificial manipulations of pairwise finger correlations, by physically yoking two fingers together, alters the distance between finger representations in primary somatosensory cortex [77]. This suggests that elementary sensorimotor representations may be plastic and subject to changes over time and that multivariate pattern analyses on fMRI data are sensitive enough to detect these changes.

If individual actions are bound under a common motor command, then the internal representations of those actions, at some level of the motor hierarchy, should change over time in these areas that binding occurs. The naïve version of this model is that if two movements are executed repeatedly in a close temporal sequence, then the activation of one finger movement may already pre-activate the following movement. In the extreme, this model makes the prediction that two fingers that are regularly paired together in everyday actions will become enslaved together over time, thereby reducing behavioral flexibility [2]. It is therefore more likely that the process of binding alters the representation of contextually cued actions in upstream regions linked to more abstract response selection [47], such as the dorsal premotor cortex (PMd) or motor regions along the intraparietal sulcus. Wherever this binding process happens, the multivariate activity pattern for the two bound movements should become more similar in that region (Fig. 5D).

Here we tested the plasticity of individual action representations using a combination of behavioral analysis and event-related fMRI. Binding was measured behaviorally by looking at the degree of correlation between successive behavioral responses after training on a unimanual 32-item sequence. We also measured the population-level representations of visually-cued single finger movements in the cortex both before and after five weeks of training on the complex sequence. The simple plasticity hypothesis states that binding of serial

actions after consolidation of a motor sequence should make the neural representations for those actions more similar, thereby decreasing the representational distance between them.

2.2 RESULTS

Participants executed sequences in a serial reaction time task [4] on a laptop keyboard, in which each finger press was cued by a unique fractal image (Fig. 6A). Subjects learned the mapping of cue to finger press prior to the first day of training through visual feedback (red flash for incorrect responses). After a key press, the response time, measured as the elapsed time between cue onset and key press, was recorded and the next cue was presented following a 250 ms interval. Each day, participants were tested on trial blocks of random sequences (blocks 1,2,6) or trial blocks composed of a specific 32-element sequence (blocks 3,4,5,7). An additional control group received the same amount of training as the trained group - but here all blocks consisted of random sequences.

2.2.1 Behavior

To assess how training impacted performance, we compared the evolution of response times and accuracy across days for the Trained and Control groups. Fig. 6B illustrates all trial-wise responses during a single day for a subject in the Trained group. While responses during random trial blocks (black dots) remained relatively constant, the response times during sequence trial blocks (green dots) get steadily faster with training. The last two trial blocks were used to probe learning across time. On average both the Control (dashed line, Fig. 2C) and Trained subjects (dashed line, Fig. 6D) exhibited a general improvement in response speeds during the final random trial block (block 6). This general across-session speeding of responses during a trial block with random sequences likely reflects the improved learning of the cue-response mapping across days. During the final sequence block (block 7), however, sequence-specific responses in the Trained group also decreased rapidly across training days. Repeated measures ANOVA indicated a significant block x time effect: $F(368,23) = 15.366$,

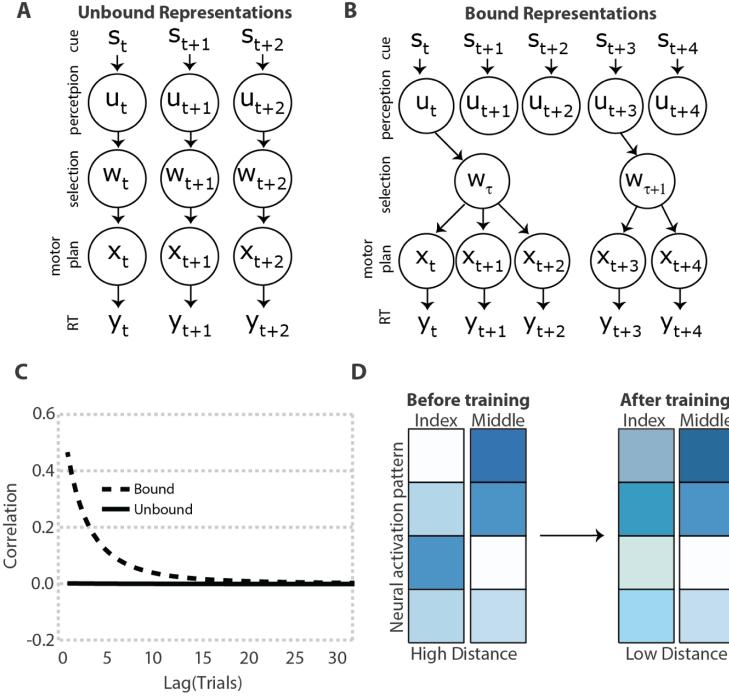


Figure 5: The process of response binding A. One each trial, (t), a visual stimulus (s) triggers an appropriate finger response (y), in this case reflecting a response time (RT). In the case of unbound actions, the visual perception (u), selection (w), and motor planning (x) processes are all represented as latent states that operate independently across trials. B. With training, the intermediary process of selection binds multiple motor plans together as a set. Each set of actions, τ , is triggered by the visual stimulus of the first item in the set. Subsequent actions are then internally triggered, rather than relying on external visual cues. This example shows two bound sets, a three item set followed by a two item set. C. The autocorrelation function of response times for bound actions (cyan) should exhibit a significant correlation across trials, while unbound actions (black) should not exhibit a temporal autocorrelation. D. A schematic of four hypothetical voxels in cortical sensory motor networks during the execution of either the index or middling finger, with darker colors reflecting stronger movement-evoked responses. Before training, each finger representation is associated with a unique neural activation pattern. After training, the representations of bound finger movements share more activation and the neural activation patterns are more similar.

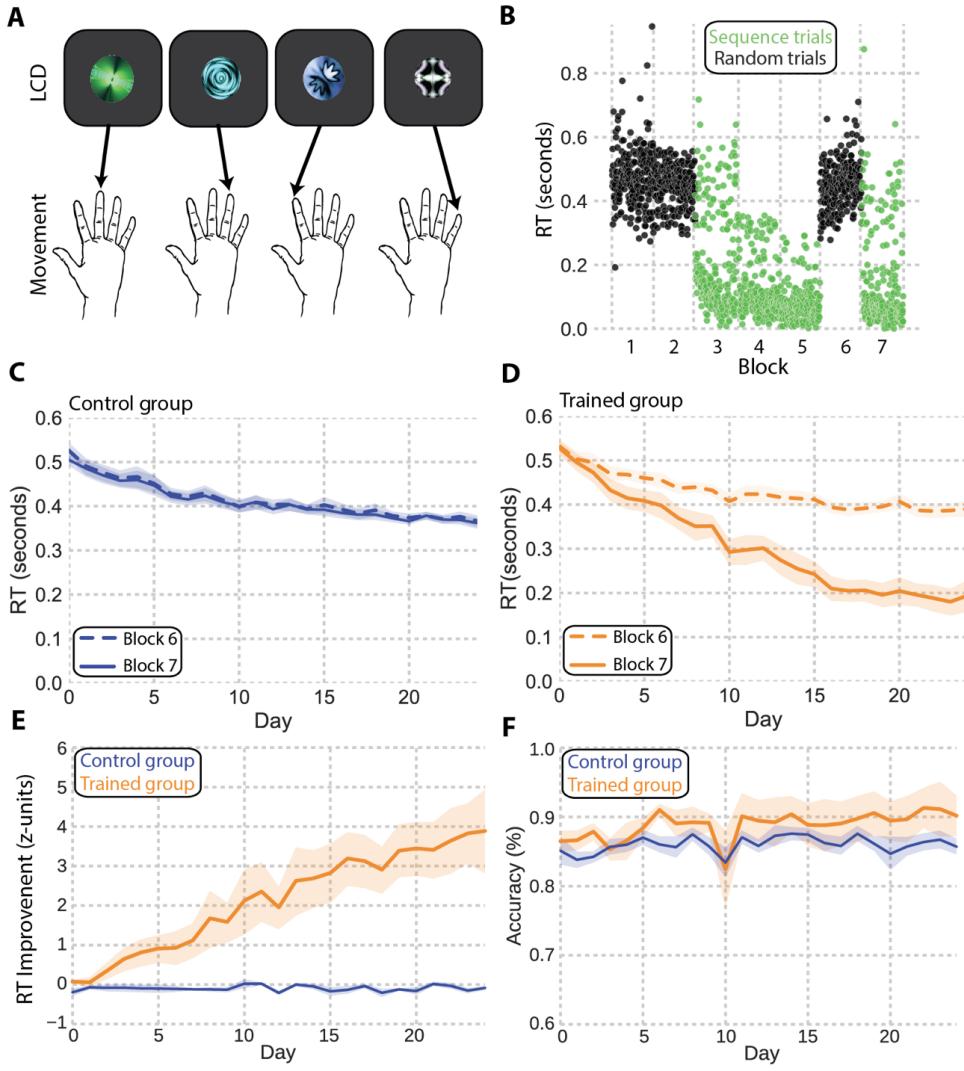


Figure 6: Task design and behavioral performance. A. Participants practiced a serial reaction time task in which each finger movement was prompted by a unique cue. B. Representative reaction time plot from Day 12. Each dot represents the response time on one trial. C. Reaction times for the Control group for random trials on blocks 6 and 7. D. Reaction times for the Trained group for the random trials (block 6) and sequence trials (block 7). E. Mean z-scored reaction times as a function of day for the Control group (blue) and Trained group (peach). F. Mean accuracy (correct trials/total trials) in the final trial block, as a function of day, for the Control group (blue) and Trained (peach) group. Shaded regions in panels C-F show standard error.

$p = 7.93 \times 10^{-41}$, with average response times dropping just below 200ms at the end of training (solid line, Fig. 6D). As expected, this effect was not observed in the Control group, $F(368,23) = 0.77$, $p = 0.76$, where the final trial block did not contain an embedded sequence (solid line, Fig. 6C). In order to capture sequence-specific changes in response speed, we normalized the mean response time for the final trial block (sequence in Trained group, random in Control group) by the mean and variance of response times during trial block 6 (random in both groups; see Methods). This analysis depicts a steady improvement in sequence specific response times across the 5 weeks for the Trained group, with sequence block responses approximately 4 standard deviations faster than the random trial blocks at the end of training (Fig. 6E). Repeated measures ANOVA indicated a significant group by time effect, $F(368,23) = 12.79$, $p = 1.67 \times 10^{-34}$. Unlike response speed, average accuracy during the final trial block gradually rose at a steady rate for both groups, saturating at around 90% for the Trained group and 85% for the Control group, with no significant between group differences $F(368,23) = 0.36$, $p = 0.99$.

There are several ways that responses could get faster during the sequence blocks [78]. The binding hypothesis (Fig. 5B), however, makes the specific prediction that serially successive actions that are bound under a shared motor plan should exhibit a correlation in their responses over time, as a consequence of arising from a common, high-level motor plan (Fig. 5C). For an index of binding, we used the autocorrelation of RTs during the last trial block for both groups [45]. Fig. 7 shows the autocorrelation functions for early (Day 1), middle (Day 12), and late (Day 24) stages of practice for the Trained (Fig. 7A) and Control (Fig. 7) groups separately. Participants in the Trained group showed no evidence of an autocorrelation in their RTs at Day 1. However, by the middle of training, a pronounced autocorrelation of temporally adjacent responses emerged that remained steady by the last day of practice. As expected, the Control group did not show any autocorrelation structure in RTs at any point during training, indicative of a lack of binding across responses.

We next examined the RT correlations of responses to each item in the 32-item sequence on each cycle of the sequence production. Before practice, this 32x32 correlation matrix does not show much structure, with all items approximately equally uncorrelated (Fig. 7C). After training, a clear structure in the correlations emerged, with local clusters of correlated

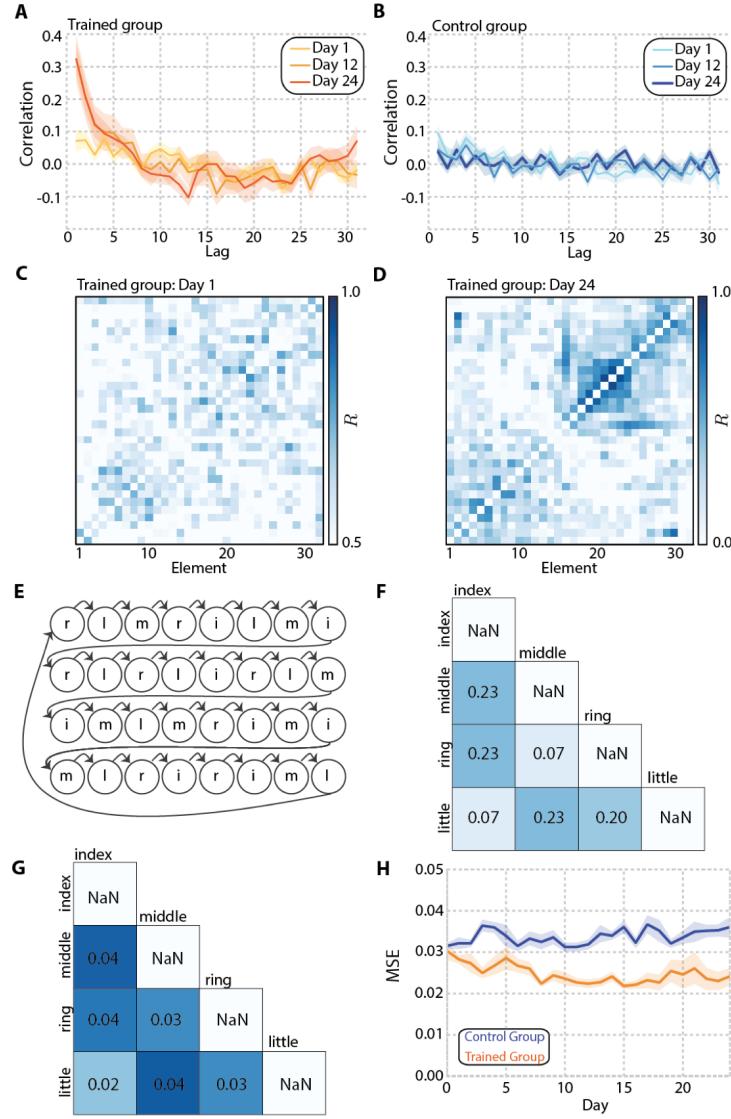


Figure 7: Binding in behavioral responses. A,B Mean autocorrelation function for lags 1-31 during early (day 1), middle (day 12) and late training (day 24) for the Control group (A) and Trained group (B). Shaded regions show standard error of the mean. C,D. Average correlation between each element in the sequence during the final trial block for the Trained group, during Day 1 (C) and Day 24 (D). E. The 32 element sequence showing frequency of each finger transition (i-index, m-middle, r-ring, l-little) F. Pairwise frequencies between each finger G. Average observed correlations between fingers at the end of training collapsed across subjects. H. The MSE between the pairwise frequencies (panel F) and observed correlation matrix computed separately for each subject. Smaller numbers indicate increased similarity to the expected pairwise frequencies (panel F). Shaded regions show standard error.

responses found along the diagonal of the matrix (Fig. 7D). If these clusters of correlated responses in the sequence reflect the inter-finger transition probabilities (Fig. 7E), then the pairing frequency of individual fingers should determine the degree of similarity between finger responses. The similarity between the observed correlations and expected correlations based on the pairwise frequencies (Fig. 7F) was computed using the mean squared error (MSE). The mean observed correlation matrix across all subjects on the final day of training is shown in Fig. 3G. There was increased similarity between the observed and expected correlations across days (Fig. 7H) in the Trained group, but the structure in the Control group remained unchanged, resulting in a significant group by time interaction, $F(368,23) = 1.90$, $p = 0.0079$.

2.2.2 Movement and cue set representations

In order to directly measure multivariate cortical representations of the individual cued movements, we used a rapid-event-related fMRI design consisting of presentations of each cued finger press followed by a period of fixation (Fig. 8A). A regions of interest (ROI) analysis was performed on the cortical motor network including primary motor cortex, M1; primary somatosensory cortex, S1; dorsal premotor cortex, PMd; ventral premotor cortex, PMv; supplementary motor area, SMA; and the superior parietal lobule, SPL. These regions were anatomically localized using Brodmann areas extracted from Freesurfer (see Methods). These regions are shown on the group average surface (Fig. 8C). In each of the cortical motor ROIs, we quantified the activity pattern related to each cued finger movement and then calculated a cross-validated Mahalanobis distance (crossnobis) between the activity patterns for each cued finger pair. If two cued fingers generate the same cortical activity patterns, then the corresponding distance between them will be 0. However, if two finger movements consistently generate dissimilar finger patterns, then the corresponding distance will be positive (Fig. 8B). Cross-validation across independent scanning sessions allows us to test the value of the distance estimates directly against zero (see [73, 72, 80]). The distances between every possible pair of fingers is summarized in a representational dissimilarity matrix for each ROI (Fig. 8D).

The representational structure we observed in primary motor and primary somatosensory cortex qualitatively matches previous reports [49], such that the index finger is furthest from the little finger, while the middle and ring fingers are close together. This pattern of representational distances is also similar to what is observed in the other cortical motor regions, although the overall between effector distances are smaller in these premotor regions (Fig. 8D). To confirm that each region has reliably different representations for the fingers, we computed the average cross-validated pairwise distance between all finger movements (Fig. 8B, see eq. 2.1). Average H scores greater than 0 indicate above-chance encoding [73]. Fig. 8E shows the mean H distribution computed from the surface-based searchlight (see eq. 2.2) across all voxels as kernel density estimates averaged across subjects (one distribution per subject for each ROI). In a control region, primary auditory cortex (A1), distances are symmetrically distributed about zero, indicating that one would not be able to reliably decode the cued-finger movement from this region. In sensorimotor regions, the distances were positively skewed, indicative of cued-finger movement encoding. In order to estimate the reliability of this encoding across subjects, we extracted the median distance value from each distribution for each subject and ROI. A one-sample t-test on those median values (one median per subject), after adjusting for multiple comparisons using a Bonferroni correction, found significant separation of cued finger representations (i.e., positive average distances) in the cortical sensorimotor areas, but not the A1 control region (Table 1). Thus, consistent with previous studies [49], the patterns of activity in the motor network can reliably discriminate individual effectors.

2.2.3 Stability of representational geometries

To determine whether the emergence of binding in the behavioral responses coincides with alterations of these representational distances of individual cued actions, we measured how average distances changed for each cortical motor ROI before and after training. More specifically, if binding results in the representations of frequently paired actions becoming more similar (Fig. 5D), then distances between frequently paired movements would decrease after practice only in the Trained group. When looking at all pairwise distances (Fig. 9A)

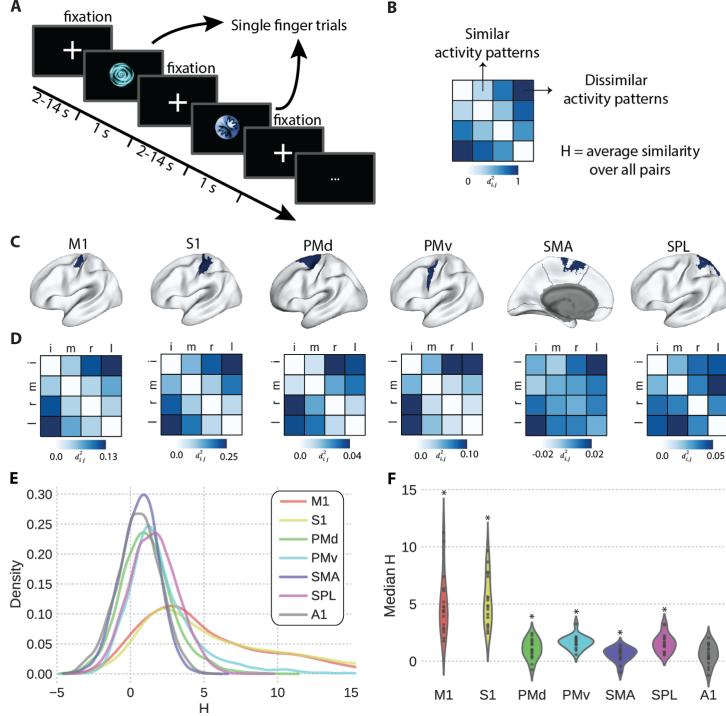


Figure 8: Multivariate activity patterns during cued finger movements. A. fMRI task schematic. Participants executed single finger movements on the button glove following a variable period of fixation. The cue-finger mapping was identical to that used during the training. B. Example of a representational dissimilarity matrix showing similar finger patterns that result in small distances and dissimilar finger patterns that result in large distances. The average crossnobis similarity (H) was used as a test statistic for assessing decoding in each ROI and for assessing representational plasticity. C. Regions of interest masks overlaid in blue on the group average surface. D. Average representational dissimilarity matrices for each region. Each colored square within the RDM indicates the distance between those two fingers (i=index, m=middle, r=ring, l=little) E. Average kernel density estimates of H across all subjects and all voxels in primary motor cortex (M1), primary somatosensory cortex (S1), premotor dorsal cortex (PMd), premotor ventral cortex (PMv), superior parietal lobule (SPL), supplementary motor area (SMA), and primary auditory cortex (A1). F. Violin plots show the distributions of median H values across subjects. Black circles inside plots show individual data. Asterisks indicate significance at $\alpha = 0.05$ after correcting for multiple comparisons (Bonferroni).

Table 1: Column “mean” indicates the average median H value in each region. T-statistics computed from one-sided t-test.* indicates significance after correcting for multiple comparisons.

Region	Mean	t(17)	p-value	95% CI
M1	4.92	7.91	* 4.23×10^{-7}	3.61, 6.23
S1	5.23	10.13	* 1.28×10^{-8}	4.14, 6.3
PMd	1.07	5.30	* 5.83×10^{-5}	0.64, 1.49
PMv	1.66	12.06	* 9.21×10^{-10}	1.37, 1.95
SMA	0.57	4.08	* 7.75×10^{-4}	0.28, 0.87
SPL	1.57	8.44	* 1.74×10^{-4}	1.18, 1.97
A1	0.57	2.78	0.013	0.14, 1.01

we were unable to find a reliable influence of sequence training on the average pattern distances in any cortical motor region (Fig. 9B). In most areas, the distances decreased only marginally for both Trained and Control groups together, but the finger patterns remained largely separable, with patterns exhibiting a high degree of stability. Across all regions, we failed to detect a reliable interaction between group and time that would be indicative of a training effect in representational distances (all $p > 0.26$, full statistics reported in Table 1). In order to evaluate the evidence in support of the null hypothesis that the interaction is not present, we conducted a JZS Bayes Factor (BF) ANOVA with uniform prior across all models [81] and found evidence supporting the null model that training does not influence distances. The BF’s ranged from 0.33-0.36 (Table 2), which can be considered positive evidence in support of the Null hypothesis [82].

Of course, changes in overall representational distances may not be sensitive enough to pick up changes in the representational distances of only a few finger pairs. The simple plasticity model we proposed in the introduction predicts that the greatest plasticity should be observed in the finger pairs most often executed together in the sequence (Fig. 4). If

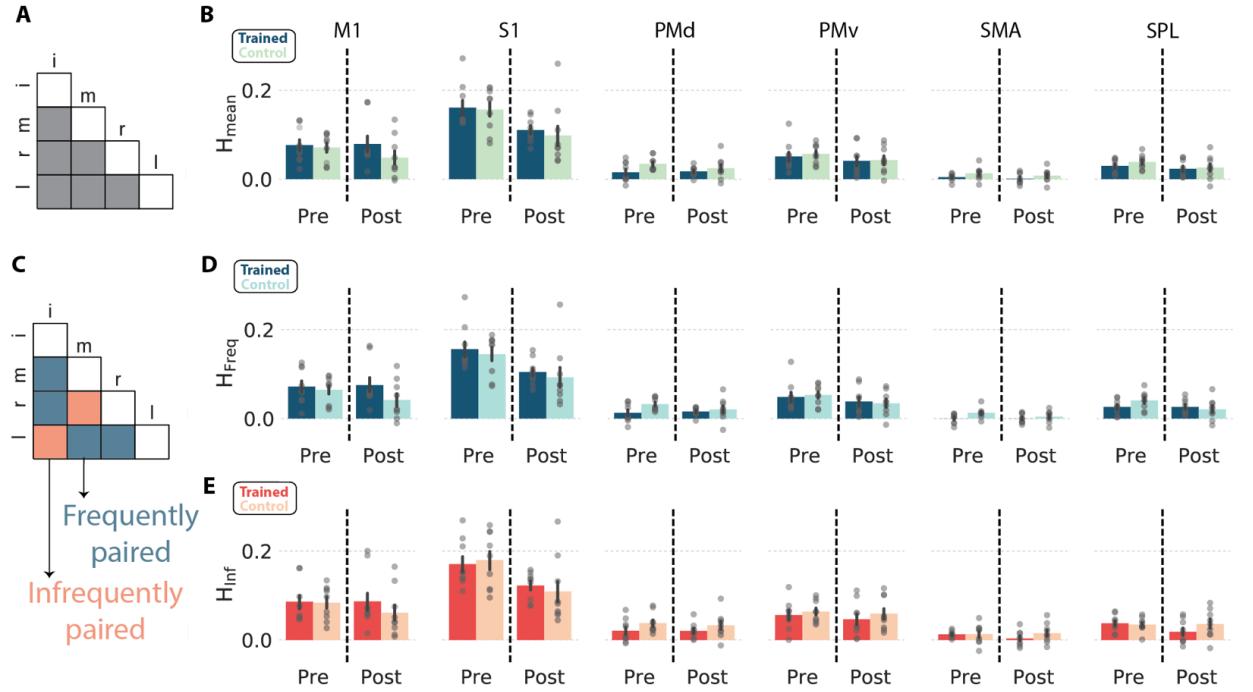


Figure 9: Stable representational distances after training. A. Pairwise finger distances included in overall distance analysis. B. Bar plots show mean ROI H values in the pre- and post-training scans separately for each group. Error bars show standard error. Gray circles are individual data points. C. Finger pair frequencies were asymmetrically distributed in the trained sequence (see Fig. 7F). Some finger pairs, e.g. index and little were infrequently paired, whereas other finger pairs e.g. index and middle were frequently paired. D,E. Bar plots show mean H for frequent pairs B (panel D) and infrequent pairs (panel E) in the Pre and Post scans separately for each group. Error bars show standard error. Gray circles are individual data points. No comparison was found to be statistically significant at $\alpha = 0.05$.

Table 2: F-statistics and p-values for testing significance of interaction effect (group x time) from repeated measures ANOVA for mean distances. Inclusion Bayes Factor (BF) is the ratio of the posterior over the prior probability of the model including the interaction term. BF<1 provide evidence that interaction effect is not present.

Region	F(1,16)	p-value	Inclusion BF
M1	1.35	0.26	0.348
S1	0.086	0.77	0.349
PMd	0.74	0.40	0.349
PMv	0.066	0.80	0.367
SMA	.042	0.84	0.333
SPL	0.34	0.57	0.35

the distances decreased for the more frequently paired effectors but increased for the less frequently paired effectors, this may result in a net change for the overall average distance near 0. To explore this possibility, we re-analyzed the distance changes by looking at the frequently and infrequently occurring finger pairs in the sequence structure itself (Fig. 9D,E). Based on the pairing frequencies, we identified four frequent finger pairs and two infrequent pairs (Fig. 9C). However, much like the overall distance patterns, we were unable to resolve focal changes in representational distances in either of the most frequently (Fig. 9D) or infrequently (Fig. 9E) paired effectors. Across all regions, two way repeated measures ANOVA indicated no significant group-by-time interaction for either frequently paired (all $p > 0.26$, full statistics provided in Table 3) or infrequently paired fingers (all $p > 0.13$, full statistics provided in Table 4). The Bayesian ANOVA revealed anecdotal evidence in favor of the null hypothesis for both the frequently (BFs: 0.68-0.89, Table 3) and infrequently (BFs: 0.53-0.60, Table 4) paired fingers.

Table 3: F statistics and p-values for testing significance of interaction effect (group x time) from repeated measures ANOVA for frequently paired fingers. Inclusion Bayes Factor (BF) is the ratio of the posterior over the prior probability of the model including the interaction term. BF<1 provide evidence that interaction effect is not present.

Region	F(1,16)	p-value	Inclusion BF
M1	1.36	0.26	0.87
S1	0.00065	0.98	0.86
PMd	1.27	0.28	0.89
PMv	0.27	0.61	0.68
SMA	1.05	0.32	0.83
SPL	2.86	0.11	0.84

Table 4: F statistics and p-values for testing significance of interaction effect (group x time) from repeated measures ANOVA for the infrequently paired fingers. Inclusion Bayes Factor (BF) is the ratio of the posterior over the prior probability of the model including the interaction term. BF<1 provide evidence that interaction effect is not present.

Region	F(1,16)	p-value	Inclusion BF
M1	1.16	0.30	0.54
S1	0.69	0.42	0.60
PMd	0.07	0.80	0.54
PMv	0.083	0.78	0.53
SMA	1.05	0.32	0.56
SPL	2.53	0.13	0.54

2.3 DISCUSSION

Here we examined whether the binding of serial actions during long-term sequence learning alters the cortical representations of individual cue-response pairings. We found that, during sequence production, temporally adjacent responses develop a high degree of correlation in their response speeds. This is consistent with participants binding multiple responses together under a unified command so as to reduce computational complexity (see also: [45, 36, 46]). Using a multivariate pattern analysis approach, based on the cross-validated Mahalanobis estimator, we also replicated previous studies showing that cortical motor areas reliably distinguish between activation patterns of individually cued finger responses [49]. We were, however, unable to find evidence for learning-related changes in this representational structure of cued finger responses in any of the cortical regions tested. Taken together, these findings show that the process of binding actions into chunked sets during long-term skill learning does not impact the representations of individual cued actions, suggesting that binding relies on changing more complex levels of representation beyond individual movements.

At first glance, the absence of plasticity in population level representations of individual actions that we observed appears to be incompatible with previous reports of plasticity in sensorimotor cortex. Kolansinki and colleagues [77] found that the representational distances of individual fingers shifted in S1 after physically yoking two fingers together for a period of 24 hours. In their study, the sensory representations of the two yoked fingers remained spatially and temporally identical, but the unyoked fingers altered their distances, suggesting a possible compensatory effect in the sensory representations themselves. In contrast, our task relied on training associations between temporally independent movements in a specific context. It is possible that, had we trained on chord-like movements, where multiple fingers are simultaneously engaged [83], for a longer period of time, we might have observed similar changes in cortical sensorimotor representations, a hypothesis that is left open to future studies.

Alternatively, there is a strong rationale for why single effector representations would remain stable in cortical sensorimotor networks, particularly motor execution areas like M1,

after long-term sequence learning. First, binding responses at the execution level may be a maladaptive strategy for maintaining a flexible movement repertoire [2]. For example, if index finger movements were consistently bound with middle finger movements because a single daily task required them to work together in sequential fashion, then they might exhibit a prepotent response in inappropriate contexts. In order to maximize flexibility, it would be beneficial for the movements to be bound at a more abstract motor planning stage, upstream from execution processes [84].

Of course, it is possible that there is plasticity in the representations of individual sensorimotor effectors during long-term sequence learning, but limitations in our experimental design may preclude identifying those changes. First, while the duration of training we used was longer than many sequence learning experiments in humans, five weeks may still not be enough time to lead to measurable representational changes in primary motor cortex. This concern is tempered by the fact that we were able to show strong evidence of action binding in the behavioral responses. Second, we could not look at finger representations in the striatum, where there is some evidence for binding [56, 85] as the voxel sizes in this study were too large to examine those the more fine grained subcortical representations. Future studies at a higher MRI field strength (e.g., 7T) may afford a better spatial resolution for picking up plasticity of sensorimotor representation in the striatum. Despite these limitations, our experiment clearly shows that five weeks of training on a complex unimanual sequence task does not alter the sensorimotor representations of individual effectors despite clear evidence of binding in the motoric actions. This suggests that execution level representations remain stable during learning and that proficiency is likely controlled by a higher level within the motor hierarchy.

2.4 MATERIALS AND METHODS

2.4.1 Participants

Eighteen right-handed participants (6 female, mean age: 26 years) were recruited locally from Carnegie Mellon University (CMU) and the University of Pittsburgh. Two authors (PB and TV) were included in the sample. All participants provided informed consent and were financially compensated for their time. All experimental protocols were approved by the Institutional review board at CMU.

2.4.2 Serial reaction time task

Participants were trained for 25 nonconsecutive days on a variant of the serial reaction time task [4]. All experimental procedures were performed on a laptop running Ubuntu 14.04. At the beginning of each training session, participants were instructed to place their right hand over the “h” (index), “j” (middle), “k” (ring), and “l” (pinky) key. Each trial consisted of a presentation of one of four unique fractal cues appearing on a black background. Each cue was uniquely mapped to one of four keys on the keyboard (Fig. 2A). The trial ended either when the participant executed a response or once the maximum response window expired (see below), depending on which event happened first. After a trial termination, the next cue was presented after a 250 ms inter-trial interval. Each trial block consisted of 256 trials and was followed by a rest period where the mean response time (RT) and accuracy for that block was provided to the participant. On each training day, participants completed 1792 trials, separated into 7 trial blocks. RT was calculated as the delay between stimulus presentation and a key press. Stimulus presentation and recording was controlled with custom written software in Python using the open source Psychopy package [86]. The software used for training is available on GitHub (CoAxLab, n.d.). Prior to the first session, subjects were assigned to either a Trained group ($n=9$) or a Control group ($n=9$). For participants in the Trained group, trial blocks were separated into two types: blocks of pseudo randomly ordered cues (Random; blocks 1,2,6) or blocks of deterministically ordered cues following an embedded 32-element sequence (Sequence; blocks 3,4,5,7). Fig. 2B shows

the blockwise structure for a single subject in the Trained group. Trials during the Random blocks were constrained such that repeated presentations of the same cue were excluded. This was done so that Random trial blocks would appear more similar to the Sequence trial blocks. The 32 element sequence presented on Sequence blocks consisted of the following key presses: 3-4-2-3-1-4-2-1-3-4-3-4-1-3-4-2-1-2-4-2-3-1-2-1-2-4-3-1-3-1-2-4 using the mapping (1-index finger, 2-middle finger, 3-ring finger, 4-little finger). Each Sequence block began in a random position of the sequence. For the first 2 blocks, the response threshold for each trial was set to 1000 ms. To encourage faster responses, the response window of blocks 3-5 was adaptively controlled such that the response window on one trial block was the mean plus one standard deviation of the RTs from the previous trial block. If that value fell below 200 ms or if the accuracy on the preceding block was less than 75%, the threshold was reset to 1000 ms. The threshold was removed for the final probe blocks (6 and 7) so that participants could move as quickly as they chose. For the Control group, the procedure was nearly identical to the Trained group, with the exception that all 7 blocks consisted of pseudorandomly ordered trials, i.e. there was no exposure to Sequence blocks.

2.4.3 Analysis of training data

Python code and source data to generate all figures is publicly available at [GitHub](#). All behavioral analysis during training focused on responses during the last two trial blocks (probe blocks) when no adaptive response window was applied: Random and Sequence conditions for the Trained group, Random and Random conditions for the Control group. Differences in response time (RT) and accuracy (percent correct responses) were measured as the difference in the means between the last two blocks, normalized by the standard deviation of values in trial block 6, i.e., z-scored difference in performance [45]. In the Trained group this reflects the sequence specific change in performance on each day. Since 3 subjects completed 24/25 days of training, average group visualizations are presented for day 24 so as to evaluate the same state of learning for all subjects. Binding was measured by computing the autocorrelation of the series of RTs within each probe trial block. The first 32 trials were excluded to remove the exponential decay as it distorts the autocorrelation

analysis [45]. The linear trend was then removed by regression and the residuals were used to calculate the autocorrelation function for lags 1 through 31, following the same procedure as described in [45, 46]. Since the autocorrelation function measures general associations across all sequential lags, it is not sensitive to specific associations between individual elements, and therefore cannot be used to measure binding between specific finger pairs. Therefore, we conducted a secondary analysis on the same data but examined pairwise correlations between each distinct element (1-32) in the sequence. Average correlations, ordered by sequence element, are shown in Fig. 7C-D. Binding between successive elements is reflected by increases in correlations before compared to after training. To measure how much the correlation between finger responses matches the statistical structure of the trained sequence, we collapsed the elementwise correlation matrices by finger identity (index, middle, ring, pinky), forming 4x4 observed correlation matrices. To measure the similarity of the observed binding structure to the expected binding structure, we computed the mean squared error between the finger pairing frequencies of the sequence and observed correlations. This gives a normalized similarity measure for how well the pattern of correlations in the behavioral responses matches the pairwise similarities of the trained sequence.

2.4.4 Imaging acquisition

Before and after training, all participants were scanned at the Scientific and Brain Research Center at Carnegie Mellon University on a Siemens Verio 3T magnet fitted with a 32-channel head coil. High-resolution T1-weighted anatomical images were collected for visualization and surface reconstruction (MPRAGE, 1 mm isotropic, 176 slices). A fieldmap with dual echo-time images (TR: 746 ms, TE1: 5.00 ms, TE2: 7.46 ms, 66 slices, 2 mm isotropic) was acquired to correct for fieldmap inhomogeneities. For the functional imaging sessions, we acquired 241 T2* weighted echo-planar imaging volumes (2 mm isotropic, TR: 2000ms, TE: 30.3 ms, MB factor: 3, 66 slices, A >> P, FoV: 192 mm, interleaved ascending order, flip angle: 79 deg, matrix size: 96x96x66, slice thickness: 2.00 mm). For the finger mapping task, we collected a total of 6 runs resulting in 1446 volumes. Functional images were oriented so as to maximize coverage of the entire cortex and cerebellum.

2.4.5 Neuroimaging tasks

We collected a set of finger mapping runs to estimate the activation patterns evoked by performing each distinct cue-response pair in isolation (i.e. not embedded within a sequence). The same stimuli from the behavioral experiments were projected on an MR-compatible LCD screen mounted at the rear of the scanner. Participants could see this screen through a mirror mounted on the head coil. Responses were recorded on a five-key MR compatible response glove (PST Inc.) placed under the right hand. Each effector (e.g., individual cue-response pairing) was presented in isolation on each trial with no structured order between trials. Thus, the paradigm only measured responses to individual cued movements, not the sequence itself. Each trial type was repeated 12 times per run totaling 72 trials per session. Subjects were instructed to press the cued key several times following stimulus presentation until the cue disappeared from the screen (1 second). Between runs, subjects were given the option to take several minutes of rest.

2.4.6 Imaging analysis

Functional imaging data were analyzed using [SPM8](#) and custom Matlab and Python functions. Raw functional EPI images were realigned to the first volume. No slice time correction was applied. These realigned images were then corrected for field distortions using the field maps. All analyses were performed in native functional space. Structural T1 images were used to reconstruct the pial and white surfaces using Freesurfer [87]. All custom code is publicly available (CoAxLab, n.d.). Raw imaging data is publicly available for download at <https://openneuro.org/datasets/ds001233>.

All analyses of task-related responses were performed using a region of interest (ROI) approach. Anatomical ROIs were defined separately for each subject, using the surface based Brodmann areas extracted from Freesurfer [88] following similar conventions as described in [68]. The hand voxels of the primary motor cortex (M1) were defined as the surface nodes with the highest probability of belonging to Brodmann area (BA) 4, 1 cm above and below the hand knob [89]. S1 was defined as the nodes in BA1 BA2, BA3a, or BA3b, 1 cm above and below the hand knob. Premotor cortex was defined as the nodes belonging to BA6 medial

(PMv) or lateral (PMD) to the medial frontal gyrus. Supplementary motor area (SMA) was defined as the voxels in BA6 along the medial wall. The Freesurfer atlas was used to define the superior parietal gyrus, as it is not defined by a unique Brodmann area. As a control ROI, we extracted the voxels belonging to primary auditory cortex as this region would not be expected to exhibit any significant decoding of the visually-cued finger patterns. Each surface based ROI was projected back into native functional space.

Analysis for effector representations was performed using representational similarity analysis (RSA) using the crossnobis estimator [70, 71, 72]. A GLM with regressors for each effector was fit for each mapping run, along with the six head motion regressors ($x, y, z, \text{pitch}, \text{yaw}, \text{roll}$). Omissions and incorrect key presses were regressed out of the model. Raw time series were orthogonalized by eigenvector decomposition and projected into the principal component space to minimize model bias in the decoding. To estimate the differences between finger patterns, we used a cross-validated Mahalanobis distance between prewhitened regression coefficients for each effector [50, 72]. The cross-validated Mahalanobis distance has the advantage over other distance measures in that it is unbiased, since noise is orthogonalized across runs, resulting in an expected distance of 0 if a voxel or region does not reliably distinguish two finger patterns [49]. The estimated distance (\hat{d}) between the patterns (u) of two fingers (i, j) was averaged across every pair (m, l) of runs (M) resulting in (6 choose 2) = 15 folds using the following equation:

$$\hat{d}_{i,j}^2 = \frac{1}{M} \sum_{m,l:m \neq l}^M (u_i^m - u_j^m)^T (u_i^l - u_j^l) \quad (2.1)$$

The pairwise distances between each of the fingers are summarized in a representational dissimilarity matrix (Fig. 8B). To test for encoding and plasticity within each voxel or ROI, we extracted the average distance between each pair of fingers pattern ($K = 4$) using the following equation:

$$H = \sum_{k=1}^K \frac{d_{i,j}^2}{K(K-1)} \quad (2.2)$$

To examine the extent of finger representations across all of cortex, we conducted a surface-based searchlight [90], assigning every surface node an H value based on the local

($p=160$) patterns surrounding an approximately 10 mm radius. Values for the number of voxels (p) and radius were chosen based on previous studies [50]. This searchlight approach enabled us to examine the entire H distribution across all voxels in each of the ROIs to confirm that each region reliably discriminated individual effectors. Due to the observed positive skew, we extracted the median H for all regions across all subjects and conducted a one sample t-test against 0, in order to establish whether a region reliably decoded the single finger movement representations. For tests of plasticity, changes in representational distances were compared using the patterns across the top 150 voxels from each ROI, rank-ordered by average distance, similar to the number of voxels used in previous studies [48], because representational geometries are highly sensitive to the number of voxels that make up a pattern [90]. We computed H separately for the pre and post training sessions and each ROI. A repeated measures ANOVA was used to examine the influence of training on distances in each ROI. Bayesian repeated measures ANOVA with a JZS prior over all models was used to determine the inclusion Bayes Factor to measure the extent to which the data supported inclusion of the interaction effect [81]. The guidelines in [82] were used to interpret the weight of the evidence in support of the null hypothesis.

3.0 BINDING AND SETS OF ACTIONS

3.1 INTRODUCTION

A sequence, such as a piano sonata, can be represented in multiple reference frames, including spatially (e.g., extrinsic, visual space) and motorically (e.g., intrinsic, joint space) [5, 91, 92, 93]. For example, a sequence of notes in a melody can be represented visually as positions of the keys on the keyboard, or they can be represented as individual finger movements. Indeed, converging empirical evidence suggests that natural sequence skill learning relies on learning serial orders in multiple modalities [94]. This is because representing sequences in each modality offers different advantages. For example, learning a sequence as a series of goals, e.g., the letters “f-t-w”, confers advantages when the same goal needs to extend across different environments, e.g., a computer keyboard versus a keypad on a phone. On the other hand, learning a sequence as a series of actions, e.g., outstretch the arm and sweep the hand in a clockwise manner, allows for maximizing execution efficiency even when the end goal may be different, e.g., waxing a car or blocking a karate blow. Thus sequential skill learning is not a unitary process, but rather a learning process that can occur along multiple distinct channels.

While the idea that sequence learning can occur in multiple modalities is not new [5], it remains unclear to what degree binding is associated with reorganization of the underlying movement representations, as opposed to the goal representations. There is extensive behavioral evidence that sets of movements are bound with practice to reduce the computational load during execution, but the reference frame this occurs in [36, 46] is unknown. Both binding movements and binding goals could lead to the observed performance improvements. However, these two mechanisms entail different plasticity in the underlying

population representations.

Here we mapped the networks that encode sequences of movements and sequences of visual cues using a remapping paradigm that allows for independently presenting serial orders of visual cues and serial orders of finger movements [46]. Representational similarity analysis (RSA) [70] was used to resolve where each set type was encoded. This approach enabled probing for modality specific plasticity of the underlying population representations after human participants reached high levels of proficiency and demonstrated robust behavioral evidence of binding. We predicted that the distances between elements of the bound set would converge resulting in increased dissimilarity between sets (see Fig. 4). This increased dissimilarity would alter the representational geometries by increasing the average between set distance in a modality specific way.

3.2 RESULTS

3.2.1 Dissociating visual and motor set encoding

We first isolated how cue and movement sets are differentially encoded within cortical networks. Participants executed 4-element finger movement sequences on a button glove with their right hand while laying in an MRI scanner. Each scan lasted 90 minutes. Participants were scanned on 2 consecutive days to avoid fatigue. The task schematic is shown in Fig. 10. Between each of eight independent scanning runs, participants learned a novel mapping of visual cue to movement that was rehearsed until threshold performance (95% accuracy) over the previous 20 trials was achieved (after a minimum of 100 trials), which required approximately 2-4 minutes. The resulting cue and movement combinations associated with movement and cue sets over each of the 8 runs is shown in Table 3.2.1. Without remapping between individual runs, the movements and the cues are colinear, which means that the associated BOLD patterns of activity could carry information related to the movement, the visual cue, or some mixture of the two (as the representations did in the analysis of the individual movements in ch. 2). With remapping between runs, the patterns of activity

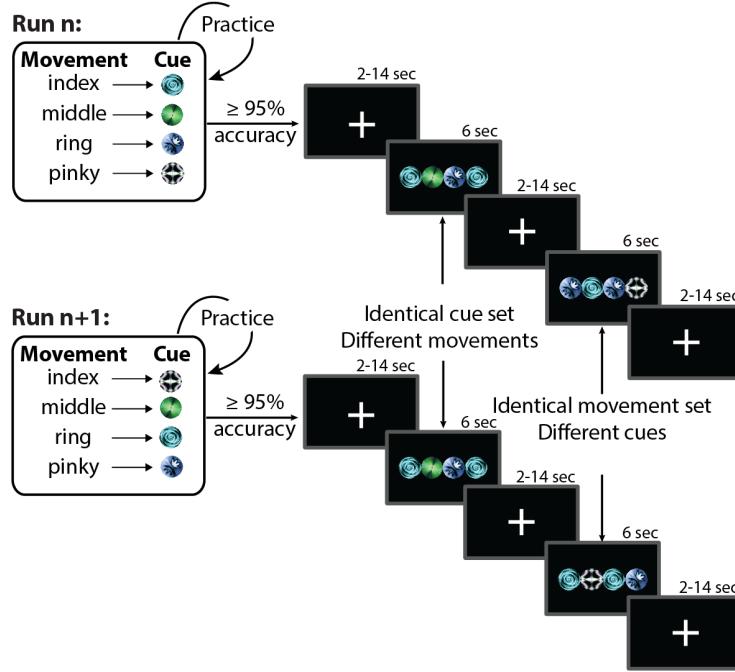


Figure 10: fMRI task used to dissociate movement and cue set representations. Prior to each imaging acquisition run, participants learned a mapping of finger movement to cue by practicing the pairings until 95% accuracy over the preceding 20 trials (after a minimum of 100 trials) was achieved. Following this practice, participants executed the various 4 element sets during the scan. During each trial, 4 of the stimuli were presented simultaneously on the screen. A correct response resulted in the stimulus disappearing. An incorrect response resulted in that cue flashing red. After all 4 movements were completed correctly, or when the time on that trial exceeded 6 seconds, the fixation cross reappeared. Each set was executed once during a trial. Responses were recorded on a button glove. Variable periods of fixation between each set presentation were drawn from an exponential with a mean of 6 seconds. For each of the runs, the mapping of movement to cue was randomly assigned, which required relearning a novel mapping. This broke the colinearity between the movement and cue, such that the executed sets across runs were composed of either identical movements with different cues or identical cues with different movements. This permitted assessing each of the resulting networks in isolation. The movement and cue set mappings for every run is provided in Table 3.2.1.

associated with the movements and the cues are dissociable, due to the fact that, across runs, movements sets are associated with distinct cues and cue sets are associated with distinct movements. This allows for dissociating the patterns of activity related to either cue or movement sequences. These patterns were analyzed with two approaches. The first approach consisted of a whole brain cortical searchlight to identify any differences in encoding and the second approach consisted of a more targeted region-of-interest based analysis to more precisely characterize the representational geometries within anatomically defined regions of interest spanning motor and visual areas.

3.2.2 Movement and cue sets are distinctly represented

In order to resolve the overall discriminability of each sequence type throughout the cortex, we conducted a surface-based cortical searchlight [90]. The searchlight indicates the crossnobiis distance at each voxel between each of the sets for either the cue or response type (Fig. 11). Positive H values indicate that the patterns of activity encompassing that voxel were reliably dissimilar across runs, and therefore can be used as a test for encoding.

Encoding of movement sets encompassed a broad set of areas in the left hemisphere contralateral to the executing hand. The greatest dissimilarity between movements sets was observed in a cluster spanning primary motor (M1) and primary somatosensory (S1) cortex around the hand representation within the central sulcus (Fig. 11A). Within that same cluster, encoding of movement sets extended anteriorly along the precentral gyrus into premotor cortex (PMD). The next largest cluster showing movement set encoding was observed within the superior parietal lobule (SPL). Several smaller clusters showing weaker encoding were observed within the motor cingulate, the posterior parietal cortex, and insular cortex.

In contrast to movement set encoding, cue set encoding was partially bilateral, specifically around the occipital pole (OCPOLE) in the left and right hemisphere (Fig. 11B). In the left hemisphere, encoding of the cue sets extended laterally from the occipital pole into the inferior frontal sulcus (ITS), at the approximate location of the lateral occipital complex (LOC) but extending more anterior. In addition, we observed two cue encoding regions

Table 5: Movement cue mapping associated with scanning runs 1-4 (day 1) and scanning runs 5-8 (Day 2). This ordering is shown for a single subject. The orderings were randomized across subjects, such that all participants experience the same remapped set but in different orders. Key: Index: 2, Middle: 3, Ring: 4, Little: 5. Movement set: M_S , Cue set: C_S Movement: Mvt.

	Day 1	Set	Mvt.	Cues	Day 2	Set	Mvt.	Cues
Run1	M_{S1}	2,3,5,2	C,D,B,C		Run5	M_{S1}	2,3,5,2	C,A,D,C
	M_{S2}	4,5,3,5	A,B,D,B			M_{S2}	4,5,3,5	B,D,A,D
	M_{S3}	3,2,4,3	D,C,A,D			M_{S3}	3,2,4,3	A,C,B,A
	M_{S4}	5,4,2,4	B,A,C,A			M_{S4}	5,4,2,4	D,B,C,B
	C_{S1}	4,5,3,4	A,B,D,A			C_{S1}	3,4,5,3	A,B,D,A
	C_{S2}	2,3,5,3	C,D,B,D			C_{S2}	2,5,4,5	C,D,B,D
	C_{S3}	5,4,2,5	B,A,C,B			C_{S3}	4,3,2,4	B,A,C,B
	C_{S4}	3,2,4,2	D,C,A,C			C_{S4}	5,2,3,2	D,C,A,C
Run2	M_{S1}	2,3,5,2	B,D,A,B		Run6	M_{S1}	2,3,5,2	B,D,C,B
	M_{S2}	4,5,3,5	C,A,D,A			M_{S2}	4,5,3,5	A,C,D,C
	M_{S3}	3,2,4,3	D,B,C,D			M_{S3}	3,2,4,3	D,B,A,D
	M_{S4}	5,4,2,4	A,C,B,C			M_{S4}	5,4,2,4	C,A,B,A
	C_{S1}	5,2,3,5	A,B,D,A			C_{S1}	2,5,3,2	A,B,D,A
	C_{S2}	4,3,2,3	C,D,B,D			C_{S2}	4,3,5,3	C,D,B,D
	C_{S3}	2,5,4,2	B,A,C,B			C_{S3}	5,2,4,5	B,A,C,B
	C_{S4}	3,4,5,4	D,C,A,C			C_{S4}	3,4,2,4	D,C,A,C
Run3	M_{S1}	2,3,5,2	C,B,D,C		Run7	M_{S1}	2,3,5,2	C,B,D,C
	M_{S2}	4,5,3,5	A,D,B,D			M_{S2}	4,5,3,5	A,D,B,D
	M_{S3}	3,2,4,3	B,C,A,B			M_{S3}	3,2,4,3	B,C,A,B
	M_{S4}	5,4,2,4	D,A,C,A			M_{S4}	5,4,2,4	D,A,C,A
	C_{S1}	4,3,5,4	A,B,D,A			C_{S1}	3,2,5,3	A,B,D,A
	C_{S2}	2,5,3,5	C,D,B,D			C_{S2}	4,5,2,5	C,D,B,D
	C_{S3}	3,4,2,3	B,A,C,B			C_{S3}	2,3,4,2	B,A,C,B
	C_{S4}	5,2,4,2	D,C,A,C			C_{S4}	5,4,3,4	D,C,A,C
Run4	M_{S1}	2,3,5,2	B,D,C,B		Run8	M_{S1}	2,3,5,2	D,C,A,D
	M_{S2}	4,5,3,5	A,C,D,C			M_{S2}	4,5,3,5	B,A,C,A
	M_{S3}	3,2,4,3	D,B,A,D			M_{S3}	3,2,4,3	C,D,B,C
	M_{S4}	5,4,2,4	C,A,B,A			M_{S4}	5,4,2,4	A,B,D,B
	C_{S1}	4,2,3,4	A,B,D,A			C_{S1}	5,4,2,5	A,B,D,A
	C_{S2}	5,3,2,3	C,D,B,D			C_{S2}	3,2,4,2	C,D,B,D
	C_{S3}	2,4,5,2	B,A,C,B			C_{S3}	4,5,3,4	B,A,C,B
	C_{S4}	3,5,4,5	D,C,A,C			C_{S4}	2,3,5,3	D,C,A,C

along the superior frontal sulcus (SFS), anterior to PMD. Two other smaller clusters were observed, one within the angular gyrus, and another within the anterior cingulate.

Based on these observations, no region appeared to encode both sets irrespective of modality (i.e., movement and visual cue). To explore this further, we performed a complementary analysis examining integration of set encoding at the individual subject level. For every voxel, we counted how many subjects showed encoding ($H > 0$) of both movement and cue sets. Across the entire cortical surface, we observed only small clusters showing at maximum 60% of subjects encoding both sets at any voxel (Fig. 12). Together, these two analyses strongly support the idea that cue and movement set encoding does not overlap in their cortical representations.

3.2.3 Region of interest analysis

To study the representational structures associated with cue and movement sets, we defined eight anatomical regions of interest, largely independently of the searchlight results. Five anatomical regions, including M1, S1, PMD, SMA, and SPL were chosen based on a previous MVPA neuroimaging study that showed sequential encoding within these regions [48]. In addition to these regions, we added the OCPOL, the ITS, and the SFS, which broadly encompassed the areas showing cue set encoding in the searchlight 11. This approach did not exhaustively test every small cluster that showed encoding at the group level in the searchlight, due to the fact some of the smallest clusters were not well defined anatomically.

The pairwise distances were summarized in a representational dissimilarity matrix in each anatomically defined region of interest, for the movement and cue (Fig. 13). To test whether a region differentially encoded cue and movement sets, we conducted pairwise t-tests of the average cue and movement distances. Follow-up 1-sample, 1-sided t-tests compared the average distances to 0 to confirm significant encoding (i.e. $H > 0$).

3.2.4 Representational dissimilarity matrices of sets

Movement set encoding but not cue set encoding was observed in M1, S1, PMD, and SPL (Fig. 13A). Within M1 and S1, the distances systematically differed between pairs. For

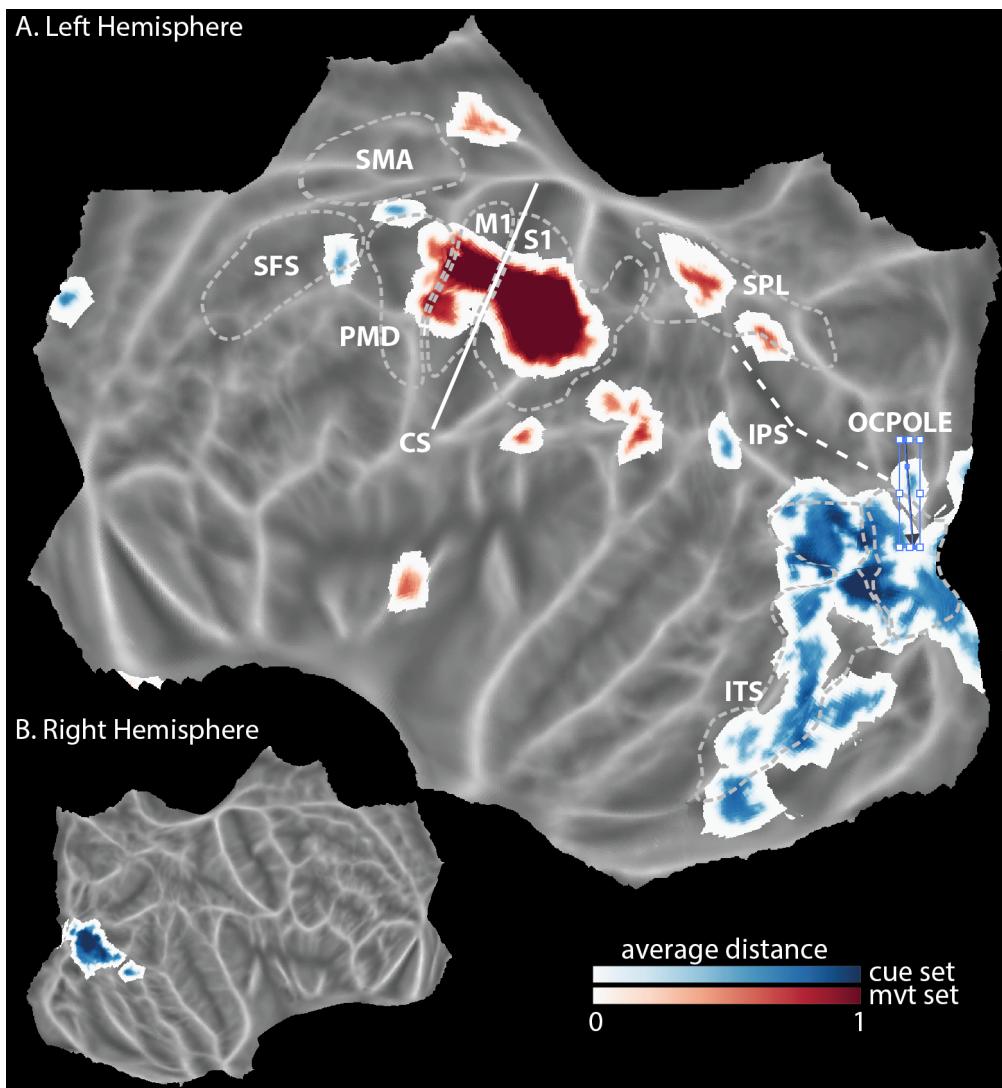


Figure 11: Whole brain surface searchlight showing distinction of movement set and cue set encoding. A. The flattened cortical surface from the left hemisphere showing the average discriminability across movement sets (red) and cue sets (blue). B. Same as A but right hemisphere. Significance was assessed using a t-test ($H > 0$) with FDR correction at 0.05 and cluster extent threshold of 30mm. Major sulci are indicated with arrows, inferior frontal sulcus: IFS, central sulcus: CS, superior temporal sulcus: STS, intraparietal sulcus: SPS. Anatomically defined regions of interest from a single subject and projected to the group surface are shown with dotted lines. Occipital pole: OCPOLE, inferior temporal sulcus: ITS, superior parietal lobule: SPL, primary somatosensory cortex: S1, dorsal premotor cortex: PMD, supplementary motor area: SMA, superior frontal sulcus: SFS. Left hemisphere: LH, right hemisphere: RH.

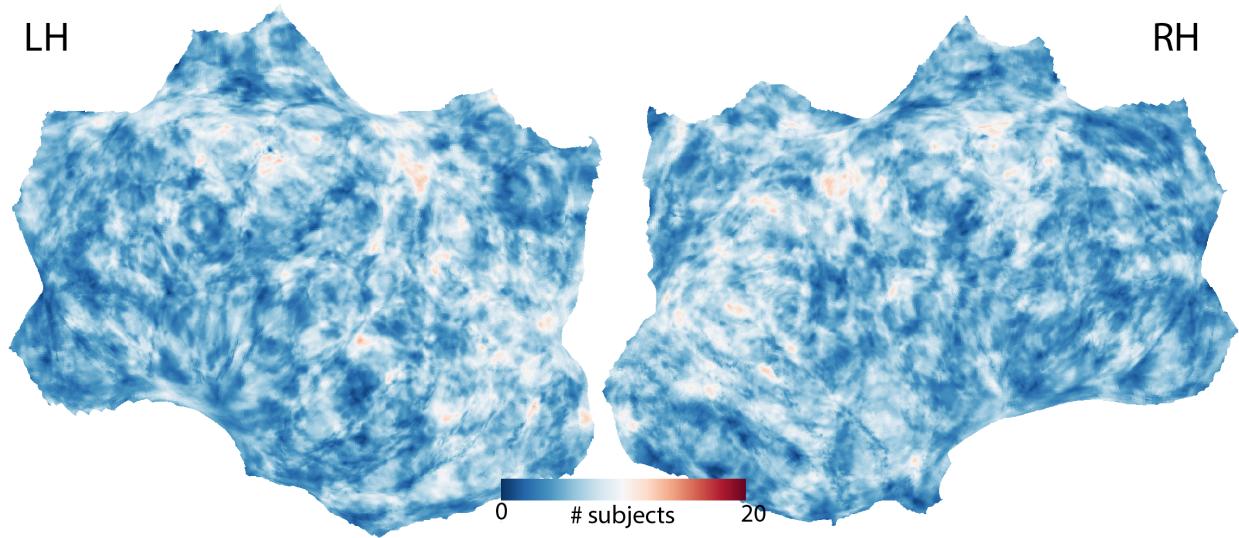


Figure 12: Flat maps showing encoding of showing the number of subjects which encoded both cue and movement sets across all of cortex computed using a surface searchlight. Left Hemisphere (LH), Right hemisphere (RH).

example, compared to the other pairwise distances, Set₁ and Set₃ and Set₂ and Set₄ exhibited higher similarity. This could potentially reflect the constituent movements of those specific sets. For example Set₁ begins with index→middle and Set₃ begins with middle→index, while sets Set₂ begins with ring→little and Set₄ begins with little→ring (see 3.2.1). That is, these sets begin with the same two fingers but in a different order. For other set pairs (e.g. Set₁ and Set₄) this symmetry is not present. The representational structure within PMD appeared qualitatively similar to what was observed in M1 and S1, but overall showed smaller distances between all pairs. The distances between sets within SPL appeared more similar.

Cue set encoding but not movement set encoding was observed along the ITS, and the OCPOLE by assessing the differential encoding of each set type as above (Fig. 13B). Encoding at the occipital pole is nearly uniformly distant, and may reflect a simple encoding of low level visual features of the stimulus, rather than anything due to set structure.

The SFS and SMA were not found to reliably encode either movement or cue sets. Lack of encoding within SMA was not surprising given the lack of any observed clusters within this region in the searchlight. Lack of encoding within the SFS, where there is a significant cluster, may be the result of its small size, since the representational distances were extracted across the entire region.

Overall, the representational structures computed from the anatomically defined regions of interest largely confirmed what was observed in the whole brain searchlights, with movement set encoding and cue set encoding being entirely independent.

3.2.5 Distinguishing first item effects

One potential confound concerning the previous observations is the fact that, for a given run, each set started with a unique cue and a unique finger (Table 3.2.1). Therefore, it is not clear if the consistency of the first item is driving the observed encoding or the consistency of the entire set. This distinction is crucial because if an area is truly encoding sequence structure, the discriminability of patterns of activity across sets must be due to the consistency of the entire set and not simply the first item. The logic of this argument arises from recently

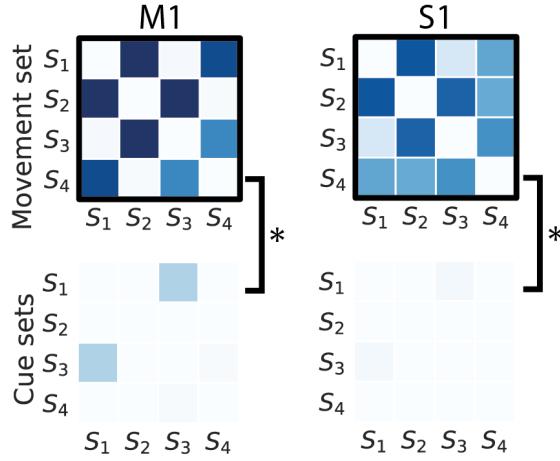
published work that has suggested that sequence-related activity in primary motor cortex can be fully explained by a linear weighting of the component movements, especially the first finger, suggesting that M1 does not encode a genuine sequence representation [95]. Therefore, while we observe encoding of the movement sets in M1 (Figs. 11A, 13A), this could be due to the individual response patterns being driven by a strong weighting on the first finger. This study was not designed to specifically address this question, but it is possible to use the other modalities sets as a control analysis to test whether the consistency of the entire set matters or only the consistency of the first item in the set.

Therefore we investigated the RDMs associated with the first items (either first fingers, or first cues). To assess first finger RDMs, for each run, we re-categorized the cue sets according to what finger each cue set began with. This was possible because each cue set began with a unique cue, and thus was initiated by a unique finger. Similarly, to assess the first cue RDMs, for each run, we re-categorized the movement sets according to what cue each set began with. Thus, across runs, the cue sets were used as a control for the movement set's first finger consistency effect, and the movement sets were used as a control for the cue set's first cue effect. If the consistency of the set did not matter, and the encoding results were being driven by first item effects, then we would expect to see reliable structure in the first item RDMs, and significant ($H > 0$) distances. However, unlike for the sets of movements, the first item RDMs did not show any reliable structure related to either the first finger or the first cue, displaying overall insignificant distances between all pairs in every ROI and for each set type (Fig. 14). This result is consistent with the idea that the observed discriminability between sets (Fig. 13) is driven by the consistency of the entire set, and not just a single item.

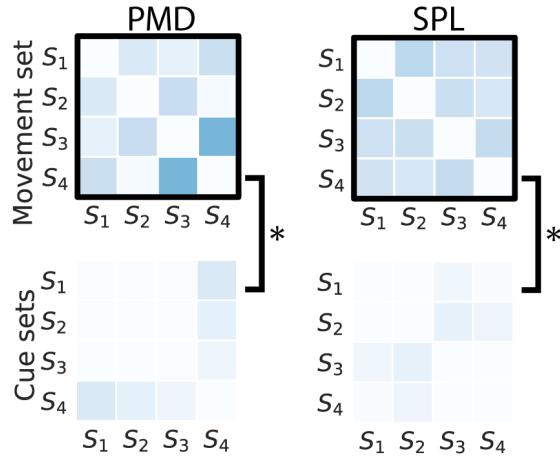
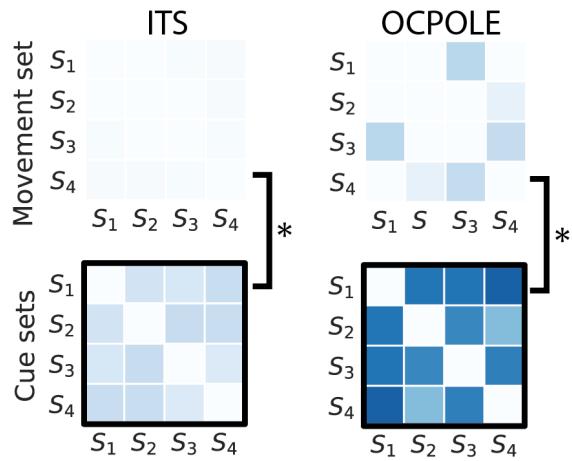
3.2.6 Cue and movement sequence learning

Having established that independent networks encode sequences in different modalities, we next asked whether sequence training selectively increased the representational distances between sets in a specific modality. To answer this question, we split the same participants into two groups and trained the first group on a sequence composed of the cue sets and

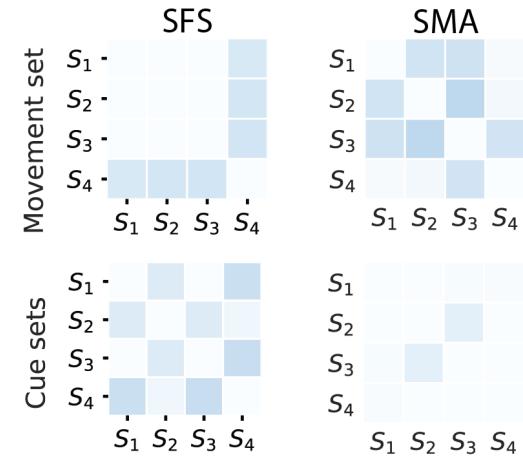
A. Movement set encoding



B. Cue set encoding



C. No encoding



0.000 $d_{i,j}^2$ 0.025

Significant difference between set types
paired t-test,
 $p<0.0062$, (Bonferroni corrected)

Significant encoding of set type
1 sample t-test, $H>0$,
 $p<0.0031$ (Bonferroni corrected)

Figure 13: Representational structures of movement sets and cue sets in anatomically defined regions of interest. A. Representational dissimilarity matrices of both movement and cue sets within regions that showed movement set encoding but not cue set encoding. B. Representational dissimilarity matrices of both movement and cue sets within regions that showed cue set encoding but not movement set encoding C. Representational dissimilarity matrices of both movement and cue sets in regions that did not show encoding of either type. Abbreviations as in Fig. 11.

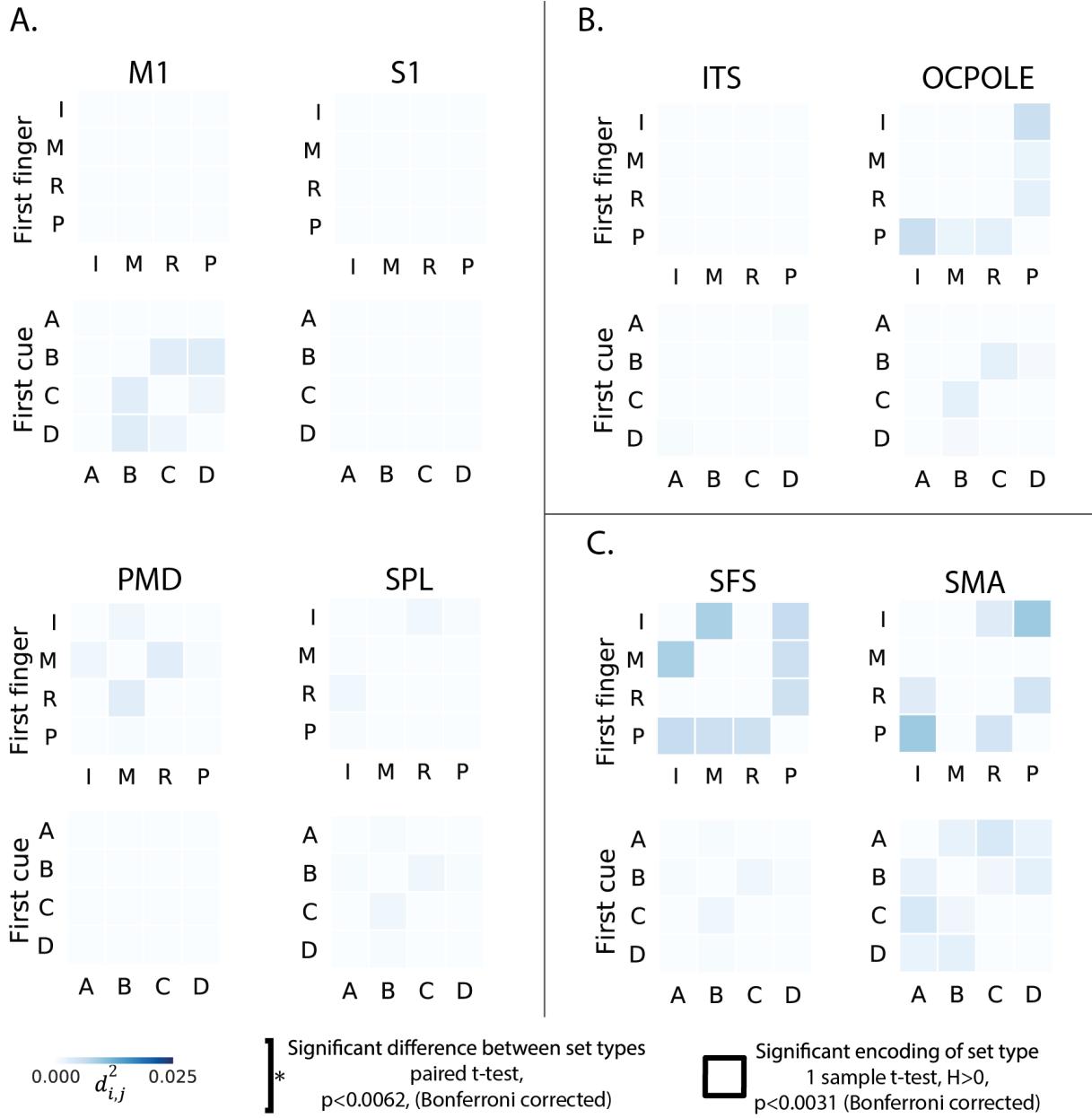


Figure 14: Representational structures of first items in anatomically defined regions of interest. A. Representational dissimilarity matrices of first finger and first cues within regions that showed movement set encoding but not cue set encoding in Fig. 13. B. Representational dissimilarity matrices of first finger and first cues within regions that showed cue set encoding but not movement set encoding in Fig. 13. C. Representational dissimilarity matrices of first finger and first cues within regions that did not show any encoding Fig. 13. Abbreviations as in Fig. 11.

the second group on a sequence composed of the movement sets. We modified the training paradigm described in ch.2 to focus on specific sets, because here we were interested in the representations of sets and not the representations of individual movements.

Similar to the task design used in ch.2, participants executed sequences in a serial reaction time task [4] on a laptop keyboard, in which each finger press was cued by a unique fractal image. Subjects learned a novel mapping of cue to finger press prior to each day, before testing, as in the between run remapping sessions of the imaging experiment (see Fig. 10). This remapping procedure was designed in order to maintain consistency across days of the serial ordering of either the movements, but not the cues, or the cues, but not the movements. Sequence-specific response time was measured as in ch.2 (i.e. by normalizing the mean response time for the final trial block by the mean and variance of response times during trial block 6). As in ch.2, training consisted of 7 blocks, each of which consisted of 256 trials. Blocks 1, 2, and 6 were composed of random trials and blocks 3, 4, 5, and 7 were composed of a specific 16-element sequences. Importantly, during the training blocks (1-5), there was a one second pause between every 4th element. This pause was used to encourage participants to selectively bind 4 element sets. This pause was omitted during the final probe blocks (6 and 7) to observe performance and set boundaries during more naturalistic sequence production. It is these final blocks that were used for analysis. If the subjects were binding sets in groups of 4 elements, as intended, then there should be lag 4 peaks during natural sequence production. Reaction times demonstrating the block and trial structure from a representative subject in the cue group is shown in Fig. 15. The cue group executed a 16 element sequence composed of the 4 cue sets that were previously tested during the imaging experiment, and the movement group executed a 16 element sequence composed of the 4 movement sets. Training lasted approximately 30 minutes each day for 10 days, after which the participants returned to be scanned under the same imaging protocol as before to assess the plasticity of the trained sets.

We first confirmed that both groups showed evidence of learning by comparing the evolution of response times and accuracy across days. Both groups successfully learned to perform the target sequence with high accuracy and speed. Accuracy for both groups exceeded 95% during the final day of training, and sequence specific response times were performed ap-

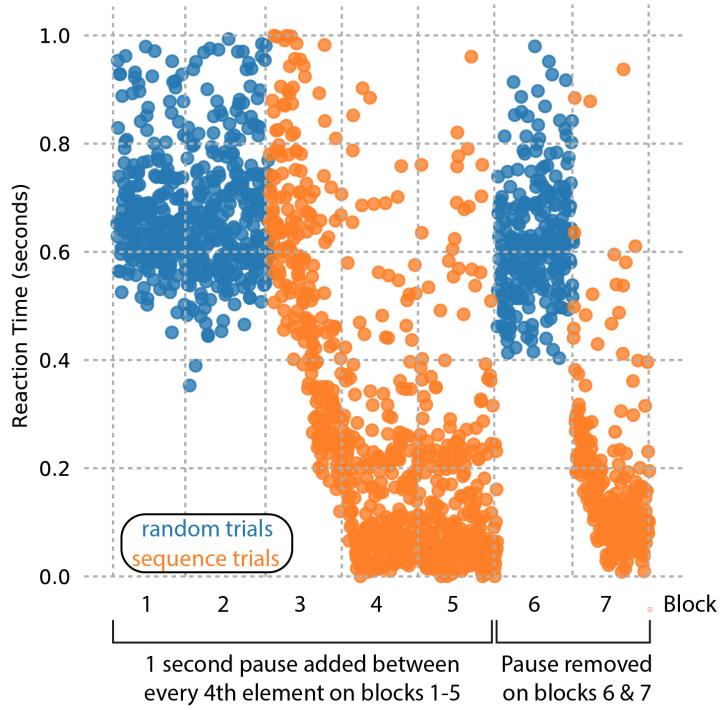


Figure 15: Representative reaction time plot from Day 5 from a participant in the cue set group. Each circle shows a single cue-response trial (1756 trials total). Response times for the sequence trials (orange circles) show increasing speed on block 3, indicating learning of the sequence, but relatively constant times for the random trials (blue circles). On blocks 1-5, there was a 1 second pause between every 4th element, which encouraged participants to group the sequence in sets of 4 elements. This pause was removed for blocks 6 (random trials) and 7 (sequence trials). This pause is excluded from the measurement of reaction time, which is computed from the onset of the cue (which occurs after the pause) and the onset of the key press.

proximately 3 standard deviations faster than the random trials by the end of training. Repeated measures ANOVA indicated a significant main effect of time for both accuracy, $F(9, 189) = 3.88, p = .00015$, and RT $F(9, 189) = 16.06, p < .00001$, indicating a significant increase in both accuracy and response time throughout the training period. However, there was no group x time interaction effect in terms of either accuracy ($F(9, 189) = 1.51, p = 0.14$), or RT ($F(9, 189) = 1.47, p = 0.15$). Together these results indicate that both types of sequences were performed with a high degree of accuracy and speed, with no significant group difference detected.

We next examined whether there was any emergence of binding in either group, using the same autocorrelation method as before (ch.2). Both groups showed non-zero correlations suggestive of binding (Fig. 17). In addition, there was a pronounced peak in the correlation function every 4th lag that was especially evident in the cue group and whose magnitude appeared to decrease with training. These peaks were expected due to the fact that participants were trained during the practice blocks with pauses every 4 trials. Therefore, the participants learned to execute the sequence in sets of 4 elements. When considering all lags, we did not detect a significant group x time interaction in the magnitude of correlation ($F(9, 189) = 0.741, p = 0.67$), suggesting that binding indices across the two groups did not differ. Followup tests on the average of the first three lags ($F(8, 189) = 0.966, p = 0.470$), and the 4th lag ($F(9, 189) = 0.517, p = 0.861$), did not result in any significant differences between groups (Fig. 18). Thus, there was no detectable difference in binding using these measures across 10 days of practice between groups that learned a sequence composed of cue sets, and a group that learned a sequence composed of movement sets.

3.2.7 Effects of training on the set representations

To determine whether binding in the behavioral responses coincides with altered representational distances between sets of actions, we measured how the average pairwise distances changed for each cortical ROI before and after training. We hypothesized that if the movements within a set are bound, then there should be an increase in the average distance between set pairs, due to the fact that the patterns generating a particular set became more

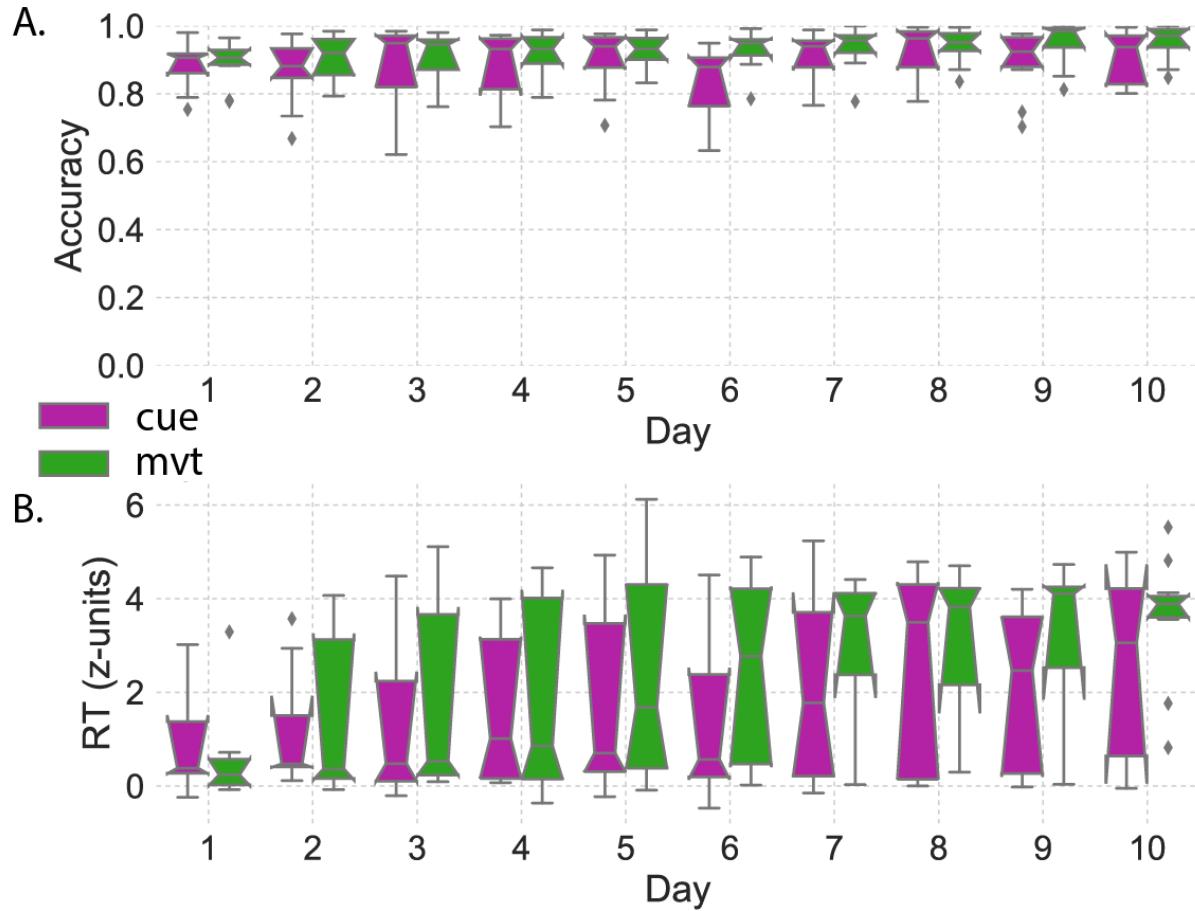


Figure 16: Both cue and motor sets were learned and executed with high speed and accuracy. Boxplots show quartiles (box), medians (horizontal lines), 95% CI of the median (notch) and outliers (diamonds) for accuracy (A) and z-scored reaction time (B) averaged across all 256 trials during the final sequence probe block, separated by group, across each of the 10 days of training.

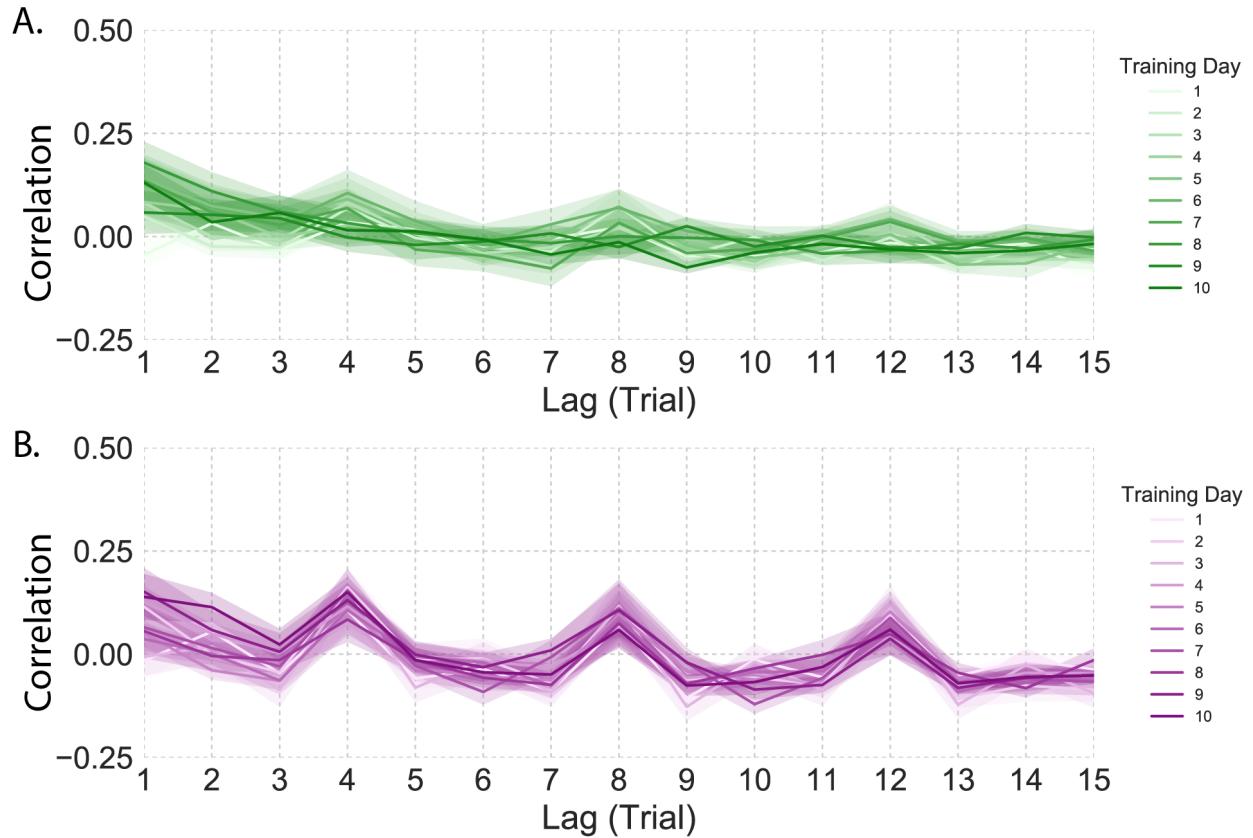


Figure 17: Both the group trained on the motor sets and the group trained on the cue sets showed emergence of binding. The correlation was computed by comparing the sequence of response times during the final probe block with lagged copies of itself. A. Movement group: correlation between lags 1-15 of the 16 element sequence shown for each day of training separated by training day. B. Same as A for the cue group. Peaks in the correlation functions at lag 4 are the result of training the sequence in 4 element sets (see methods: serial reaction time task). Error bars show standard error.

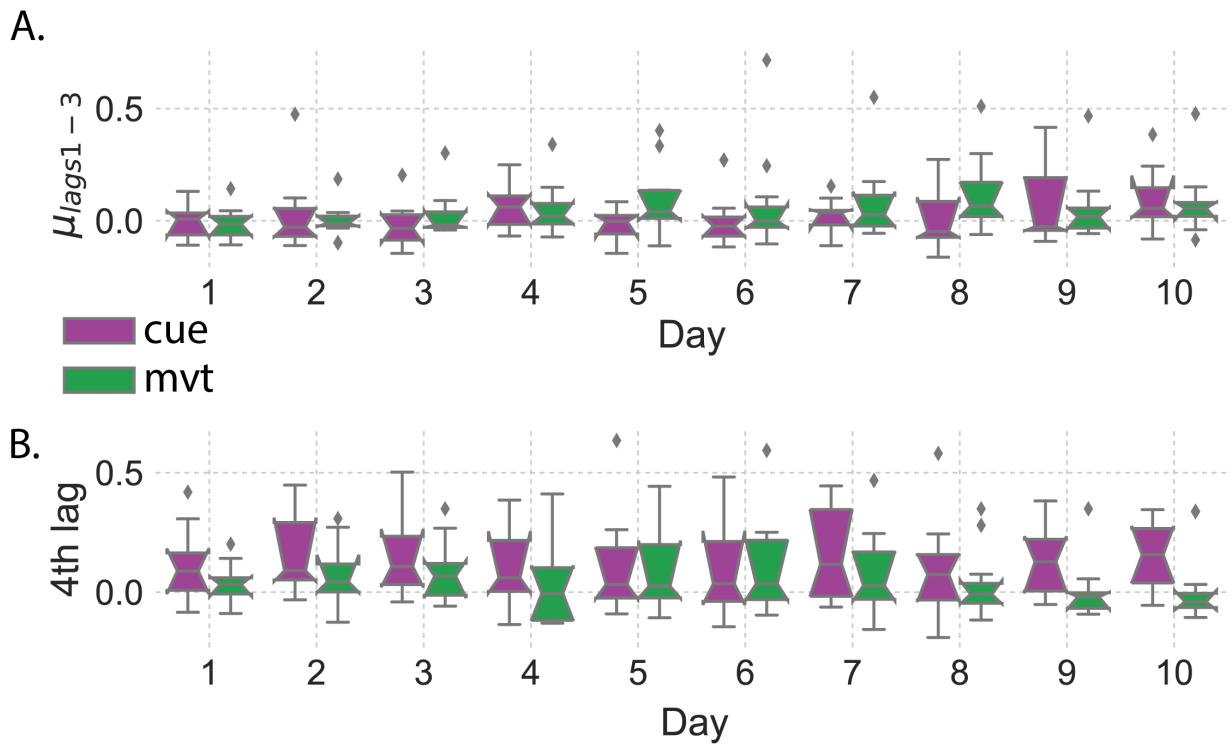


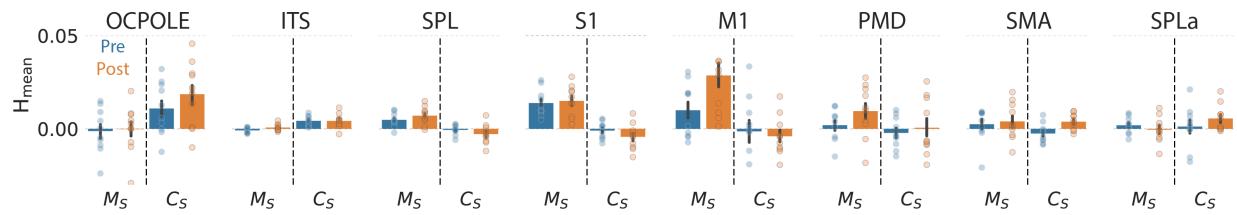
Figure 18: Binding measurements of the average of the first three lags and the 4th lag (17) did not significantly differ. A. The average of the correlation function over the first three lags plotted on each day for both groups separately. B. Same as A for the 4th lag (peaks in 17). Plotting conventions as in 16.

distinct from one another (Fig. 4B). In addition, this should occur in a modality specific manner, with increases in the cue set (C_S) distances for the cue group and an increase in the movement set (M_S) distances for the motor group. If correct, this would result in a change in the magnitude of the average between pair distance (H).

Fig. 19 summarizes the average distances across all the cortical regions of interest both before and after the 10 days of training. The distances between sets within a specific modality were computed in order to separate the potential effects of training on that modality. Across all regions, the average movement set separability and the average cue set separability was similar before and after training. We observed a slight increase in the average separation for some of the conditions, for example within the OCPOLE for the cue group between cue sets. However, the cue group also showed an increase in separability of the movement sets (which were untrained) in M1. When looking at all pairwise distances (Fig. 19) we were unable to find a reliable influence of sequence training on the average pattern distances in any of the regions of interest for either the cue (Fig. 19A) or the movement group (Fig. 19B).

Table 3.2.7 shows the results of both frequentist and Bayesian repeated measures ANOVA for the main effects of training (pre-post), modality (mvt-cue) and the interaction between training and modality (Int.) in each of the regions of interest. This testing was conducted on each training group separately. The first column shows the results of testing for the presence of a main effect of training, which would indicate a change in set separability. The 2nd column shows the main effects of movement versus cue separability, which indicates a difference in modality encoding within a particular region. Those tests recapitulated the results of the previous section (3.2.2). The third column indicates the results of testing for an interaction effect between training and modality, which would be an indicator of modality specific plasticity. None of the interaction effects were significant (after Bonferroni correction). Thus, there was no evidence that between set distances within these regions of interest changed with learning in either group. In the cue group, the Bayes Factors revealed mostly anecdotal to mixed evidence for the null, with the greatest amount of evidence in support of plasticity of the untrained modality within M1. For the movement group, across all regions, the Bayes Factors supported the null with anecdotal evidence in SPL, and moderate to strong evidence in all other regions.

A. Cue group



B. Movement group

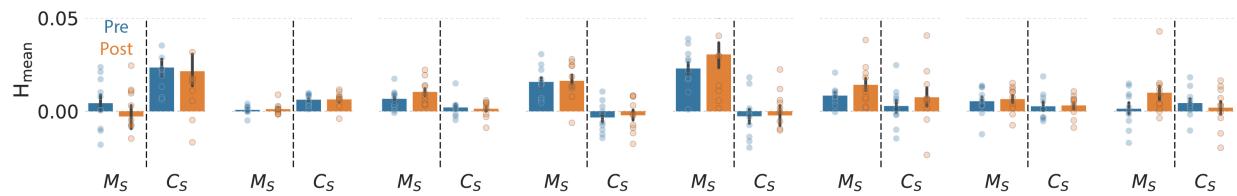


Figure 19: Average distances (H_{mean}) did not significantly change before and after training.

A. Average distances across all subjects within the cue group before and after training shown for each anatomically defined region of interest (Fig. 3.2). B. Same as A for the movement group. Error bars show standard error. Movement sequence: M_S , cue sequence: C_S . Circles are individual data points. ROI abbreviations as in Fig. 11.

Table 6: F statistics and p-values for main effects of differences in encoding of movement and cue sets (TrainedModality), differences in pre and post training encoding (prepost), and the interaction between set type and training (Int.), computed using repeated measures ANOVA. Log Bayes Factors (BF) indicate the ratio of the posterior over the prior probability of the model including that effect.

Cue group	prepost			TrainedModality			Int.		
	F(1,9)	p-value	logBF	F(1,9)	p-value	BF	F(1,9)	p-value	logBF
OCPOLE	1.289	0.286	-0.692	9.68	0.012	4.11	1.125	0.317	-0.374
ITS	0.607	0.456	-0.827	34.99	0.0002	8.702	0.852	0.38	-0.449
SPL	.00018	0.989	-0.85	23.068	0.001	9.144	3.797	0.083	0.061
S1	0.614	0.45	-1	32.005	0.0003	16.7	1.644	0.232	-0.45
M1	1.838	0.208	0.275	12.457	0.006	4.819	6.957	0.027	1.07
PMD	1.79	0.213	-0.494	4.867	0.55	-0.191	0.015	0.904	-1.07
SMA	2.222	0.17	-0.164	1.003	0.343	-0.812	1.909	0.2	-0.852
SFS	0.249	0.63	-1.233	1.102	0.321	-0.865	1.683	0.227	-1.227

Mvt group	prepost			TrainedModality			Int.		
	F(1,9)	p-value	logBF	F(1,9)	p-value	BF	F(1,9)	p-value	logBF
OCPOLE	0.484	0.504	-0.977	7.311	0.024	3.257	0.24	0.636	-0.845
ITS	0.057	0.817	-1.314	62.968	0.000025	10.394	0.045	0.837	-1.027
SPL	1.13	0.316	-0.704	17.653	0.002	4.334	0.829	0.386	-0.211
S1	0.065	0.805	-1.134	35.113	0.00022	12.85	0.023	0.883	-0.998
M1	0.76	0.406	-0.964	16.378	0.003	9.264	0.619	0.452	-0.654
PMD	1.797	0.213	-0.521	4.867	0.055	-0.209	0.015	0.904	-1.044
SMA	0.061	0.81	-1.392	3.816	0.083	-0.615	0.05	0.282	-1.825
SFS	1.797	0.213	-0.957	4.867	0.055	-1.041	0.015	0.904	-0.873

3.3 DISCUSSION

Here, using a novel paradigm that dissociates visual and motor responses, we identified two systems that represent visual and motoric sequence sets separately. Cue set encoding was observed along the ventral visual stream, while a dorsal stream and cortical motor areas appeared to encode motoric sets. Two separate groups that were trained for 10 days on either cue sets or movement sets exhibited largely similar behavioral patterns in terms of improvements in accuracy, increases in speed, and measurements of binding. Despite clear evidence of learning in both groups, the average representational distances between sets did not significantly differ following training. This suggests that our hypothesis regarding a reorganization of sets does not explain the binding mechanism.

Our results on movement set encoding are consistent with a long line research that has implied a role of M1, S1, PMd and SPL in sequence learning and execution (for reviews, see [96, 97]). In addition, we detected several smaller clusters showing movement set encoding within regions that have also been shown to be involved in hand and motor control including motor cingulate [98], posterior parietal cortex [99], and insular cortex [100]. At the same time, we did not detect any movement set encoding within SMA (Fig. 11A, Fig. 13A), which is an area that has previously been implicated in the execution of sequences [101] as well as the chunking of movements into sets. One possible explanation is that SMA only encodes set representations when the visual stimuli and the corresponding movement are consistent.

Cue set encoding was partially bilateral, specifically around the occipital pole. It is likely that encoding in primary visual areas is driven by low level visual features and not necessarily the sequence itself, since the cued sets were presented as a single image. We also observed cue set encoding extending ventrally from a region approximately situated within the lateral occipital complex, known to be tuned to visual statistics [14], but extending anteriorly as well. The two visual regions anterior to dorsal premotor cortex are situated at the approximate location of the frontal eye fields [102], which could be the result of stereotyped saccades for specific sets. We also observed encoding within the angular gyrus, and within the anterior cingulate, which could be related to the attentional demands of the task [103].

Perhaps most importantly, we did not observe any overlap in movement or cue set encod-

ing in either the searchlight analysis, nor in the region of interest analysis, suggesting that learning motor sets and cue sets may occur through primarily independent channels. Therefore, there is no single sequence learning area but multiple areas that represent sequences in different ways.

The null findings that we observed with respect to the stability of the average within-set distances suggest that our hypothesis – that reorganization of the movement representations is a mechanism underlying binding – is incorrect. If the absence of plasticity after modality-specific sequence learning in those set representations is accurate, then it suggests that the mechanisms underlying binding and sequential skill learning do not rely on altering or impacting initial representations that are established early during learning.

At the same time, it is possible that the specific type of RSA that we conducted is insensitive to detecting changes in cortical representations after sequences become bound into unified sets. The foundation of our approach is the cross-validated Mahalanobis (crossnobis) distance measure [70, 74, 75]. The crossnobis distance measure has many useful properties, such as a meaningful 0 point, as positive crossnobis distances indicate that two conditions are reliably different from one another [73]. We chose this measure because we did not have a strong prior on the representational structure of either visual or motor sets. Indeed, exploring the representational structures of the visual and motor sets was one of the goals of this study. Alternative approaches based on pattern component modeling are capable of revealing more detailed insights about the representational structure of a given region, for example whether a region encodes a linear combination of individual finger movements [95]. Our study design was suboptimal for this type of analysis, due to the fact that we did not test multiple sequences with the same combination of finger presses or cue presentations with different orderings. While each sequence began with a unique cue or finger across runs, some sequences included fingers that were not used in other sets. In addition, we did not have multiple sequences that began with the same cue or finger, but differed in the subsequent key presses/cues. However, we did find that the consistency of the entire set, and not the first item specifically, was necessary to obtain reliable encoding.

A final limitation is the fact that we did not examine the representational structures within subcortical areas, like the basal ganglia, that have been shown to be involved se-

quential skill learning, particularly the representation of set boundaries e.g: [56, 57, 85]. It is possible that the basal ganglia do encode both types of sequences (i.e. independent of modality). Our analysis was restricted to cortical areas because the signal to noise ratio for this particular task is high in cortex and the analysis method has been validated [49]. The basal ganglia are small and we lacked the appropriate spatial resolution with 3T fMRI to sample at a high enough spatial resolution to resolve representational patterns in this subcortical area. Ongoing investigations at higher resolution (7T) will hopefully overcome this limitation and extend the investigation of sequence representations into crucial subcortical nuclei that regulate motor skill consolidation.

3.4 MATERIALS AND METHODS

3.4.1 Participants

20 healthy adults (mean age: 26, 9 females) were recruited from the Pittsburgh population. All participants provided informed consent and were financially compensated for their time. Experimental protocols were approved by the Institutional review board at Carnegie Mellon University.

3.4.2 Scanner Task

Participants executed sets of four visually cued finger movements, which were recorded on a button glove while in the scanner. The cue to response mapping was randomized on each run, such that the collinearity between cue and response was minimized. Between each imaging session ($n=8$), participants learned a new mapping, and were required to reach 95% accuracy before the next scan would begin.

3.4.3 Serial reaction time task

Participants were trained for 10 non-consecutive days on a variant of the serial reaction time task [4]. All experimental procedures were performed on a laptop running Ubuntu 14.04. At the beginning of each training session, participants were instructed to place their right hand over the “h” (index), “j” (middle), “k” (ring), and “l” (pinky) key. On each day, each cue was uniquely mapped to one of four keys on the keyboard such that every day consisted of a novel mapping between cues and responses, analogous to the between run procedure used during the fMRI acquisition. Participants learned this mapping through several repeated cycles with the appropriate key to press visually presented on the screen. Participants then had to reach 95% accuracy on a set of trials during which the visual feedback was not presented. This ensured that participants were not attempting to perform the sequence task without knowing the mapping, which would have confounded the results.

Each day of training consisted of 1792 trials separated into 7 blocks. Each trial consisted of a presentation of one of four unique fractal cues appearing on a black background. The trial ended either when the participant executed a response or once the maximum response window expired, depending on which event happened first. After a trial termination, the next cue was presented after a 250 ms inter-trial interval. Each trial block consisted of 256 trials and was followed by a rest period where the mean response time (RT) and accuracy for that block was provided to the participant. RT was calculated as the delay between stimulus presentation and a key press. Stimulus presentation and recording was controlled with custom written software in Python using the open source Psychopy package [104, 86]. The software used for training is available on GitHub (CoAxLab, n.d.).

Prior to the first session, subjects were assigned to either a cue group ($n=10$) or a movement group ($n=10$). Blocks 1, 2, and 6 consisted of randomly ordered trials and blocks 3, 4, 5, and 7 consisted of deterministically ordered cues following an embedded 16-element sequence. This sequence was composed of either the 4 cue sets (cue group), each set consisting of 4 cues, or the 4 movement sets (movement group), again each consisting of 4 movements. Sequence blocks 3, 4, and 5 included a prolonged gap (1000ms) between every 4th element, in order to encourage specific binding of that group of movements or cues. This gap was

removed during the final block (block 7), such that participants could execute the entire sequence with their preferred timing. The final block was the sole block used to analyze the response times, accuracy, and binding.

3.4.4 Analysis of training data

Differences in response time (RT) and accuracy (percent correct responses) were measured as the difference in the means between the last two blocks, normalized by the standard deviation of values in trial block 6, i.e., z-scored difference in performance [45]. Accuracy was measured as the percentage correct trials during the final block. Binding was measured by computing the autocorrelation of the series of RTs within each probe trial block. The first 16 trials were excluded to remove the exponential decay, as it distorts the autocorrelation analysis [45]. A 5th order polynomial was fit to the remaining RTs and the residuals from that fit were used to calculate the autocorrelation function for lags 1 through 15, following the same procedure as described in [45, 46]. It was found that a 5th order polynomial produced a better fit than the linear fit and generated less trend in the residuals.

3.4.5 Imaging acquisition

Participants were scanned on two consecutive days at the Scientific and Brain Research Center at Carnegie Mellon University on a Siemens Verio 3T magnet fitted with a 32-channel head coil. On the first day, high-resolution T1-weighted anatomical images were collected for visualization and surface reconstruction (MPRAGE, 1 mm isotropic, 176 slices). Fieldmaps with dual echo-time images (TR: 746 ms, TE1: 5.00 ms, TE2: 7.46 ms, 66 slices, 2 mm isotropic) were acquired both days to correct for fieldmap inhomogeneities. For each of the functional imaging sessions, we acquired 241 T2* weighted echo-planar imaging volumes (2 mm isotropic, TR: 2000ms, TE: 30.3 ms, MB factor: 3, 66 slices, A >> P, FoV: 192 mm, interleaved ascending order, flip angle: 79 deg, matrix size: 96x96x66, slice thickness: 2.00 mm). Scans were oriented to incorporate all of the cortex and as much cerebellum as allowed. 4 runs were collected on each day resulting in 8 total runs (1928 volumes total).

3.4.6 Imaging data analysis

Functional imaging data were preprocessed using SPM8. Raw functional EPI images were realigned to the first volume and corrected for field map inhomogeneities using the voxel displacement map generated from the field maps, separately for each day. Statistical analyses were performed at the individual subject level in native functional space. No slice timing correction was applied. For visualization, a 2x2x2 FWHM kernel was applied on the group level distance maps (Fig. 11).

3.4.7 Representational Similarity Analysis

Analysis of movement and cue set encoding was performed using representational similarity analysis using the crossnobis estimator [70, 71, 72]. Three separate GLMs were used to evaluate set encoding, first finger effects and first cue effects. For the analysis of set encoding, a GLM with regressors for each cue (4 regressors) and movement (4 regressors) set was fit for each run, along with nuisance regressors for head motion (x,y,z, pitch, yaw, roll). Two additional GLMs with regressors for each finger movement (index, middle, ring, little) or each cue (A, B, C, D), along with the nuisance regressors were also fit to evaluate the strength of encoding sets relative to the first item effects, which are thought to account for some portion of the hemodynamic response in primary motor cortex [50]. Distances were estimated between the patterns of prewhitened beta coefficients (u_i) in each ROI or surrounding a particular voxel for the surface searchlight using the crossnobis estimator across runs (M). Distances were normalized by the number of voxels (P) in each ROI which varies across participants due to differences in brain size and curvature.

$$\hat{d}_{i,j}^2 = \frac{1}{MP} \sum_{m,l:m \neq l}^M (u_i^m - u_j^m)^T (u_i^l - u_j^l) \quad (3.1)$$

The representational dissimilarity matrices shown in Fig. 13 show these distances between each cue or motor set pair. For the searchlights shown in Fig. 11 and for tests of statistical significance, we computed the average distance (H) across all pairs (K) using the

following equation:

$$H = \sum_{k=1}^K \frac{d_{i,j}^2}{K(K-1)} \quad (3.2)$$

3.4.8 Whole brain searchlight

To examine the extent of modality specific set encoding across all of cortex, we performed a surface-based searchlight [90], assigning every surface node an H value based on the local p=160 patterns surrounding an approximately 10 mm radius. Individual encoding maps were smoothed with a 2x2x2mm FWHM Gaussian kernel, and projected to the average surface. Significance was assessed at each voxel using a one sample t-test on the distances ($H > 0$). Multiple comparisons correction was applied using FDR ($q < 0.05$). Clusters were thresholded at a 30mm extent.

3.4.9 Regions of interest

Regions of interest were selected similarly to previously published reports [78]. The hand voxels of the primary motor cortex (M1) were defined as the surface nodes with the highest probability of belonging to Brodmann area (BA) 4, 1 cm above and below the hand knob [89]. S1 was defined as the nodes in BA1 BA2, BA3a, or BA3b, 1 cm above and below the hand knob. Premotor dorsal cortex was defined as the nodes belonging to BA6 lateral to the medial frontal gyrus. Supplementary motor area (SMA) was defined as the voxels in BA6 along the medial wall. The Freesurfer atlas was used to define the superior parietal lobule, the occipital pole, the superior frontal sulcus, and the inferior frontal sulcus. For each region of interest, we extracted the patterns across all voxels within that roi. These patterns were compared using equation 3.1 and 3.2 in order to generate the average RDMs for each ROI and the average distances across all patterns to statistically compare cue and motor set encoding in each region.

3.4.10 Statistical testing

We conducted both classical ANOVA and Bayesian ANOVA as both tests provide useful information, and Bayesian ANOVA is especially useful to interpret non-significant effects and the evidence for interaction and main effects across different models. Repeated measures ANOVA was conducted to compare the main effects of cue and response set encoding, training effects on the representations, and the interaction effect between set type and training. Bayesian repeated measures was conducted in R using JASP [81]. The reported inclusion log Bayes Factors are calculated using the ratio of the posterior over the prior probability of the model including the effect of interest.

3.4.10.1 Software Accessibility Code to reproduce all analyses is available on GitHub [105]. The software for reproducing the experimental tasks in the scanner was written in python using Psychopy [104]. The code for reproducing the figures is available as a jupyter notebook [105]. The raw imaging data was preprocessed using custom SPM8 scripts.

3.4.10.2 Data Availability All the data for recreating the figures is available on GitHub [105]. The raw imaging data will be made available at <https://openneuro.org/>.

4.0 SUMMARY AND CONCLUSIONS

4.1 FINAL SUMMARY

The ability to bind multiple movements into unified sets is a central mechanism underlying the ability to consolidate habitual sequential skills. In the previous two chapters, we examined different features associated with binding before, during, and after reaching expert level performance at a sequence production task. In two separate studies, we observed clear evidence that robust correlations between reaction times of temporally independent actions during sequence production emerges with training (see also [45, 46]). Importantly, these emergent correlational patterns between temporally independent actions appeared to operate on distinct time scales compared to the simple reductions in reaction time or improvements in accuracy that also accompany sequential skill learning [45]. The emergence of correlations in sequential movements is consistent with the movements falling under a new common motor command that initiates multiple movements together, as depicted in Fig. 2. In chapter 2, we showed that the correlational structure of bound movements appears to follow the first order transition probabilities between movements in the sequence. We also showed that the cortical representations of the individual movements were stable after prolonged training. In chapter 3, we showed that these correlations in sequential responses can be detected when sequences are represented in both visual and motoric modalities; however, unlike our previous work [46] we did not detect an advantage for training in the visual modality over the motoric modality. In the cortical networks linked to visually-cued movement production, we found a clear dissociation of networks that encode movement sets and visual cue sets. While this apparent independence of networks that encode sensory and motor sequences has been previously postulated [5], it has never directly tested within the

same subjects. This result suggests that there is no central sequence representation common across sequence modalities, but multiple representations of a sequence in different reference frames.

4.2 METHODOLOGICAL LIMITATIONS

While our behavioral findings were robust and clear, there are several possible reasons that could explain why we failed to detect any training-related changes of movement representations in cortical motor networks. One concern, for example, is the reliance on the crossnobis estimator [73]. While this estimator is unbiased [80], the variance on some of the analyses reported here appears to be high. This fact, coupled with variability in performance, suggests that the average distances may only appear stable when in fact a subtle change in representational relationships occurs. Tempering this criticism is the fact that one of the properties of the crossnobis estimator is a meaningful zero point. This means that the expected average distance of a region that does not encode a particular task condition is zero. In regions where we had a strong prior for encoding, such as M1 and PMd, the distances between conditions were positive. In regions where we had no strong prior for encoding, such as primary auditory cortex, the expected distances were zero. This provides confidence that we are reliably measuring encoding of action representations in these areas. In addition, the Bayes Factors provide some support that the effect we hypothesized is not present, rather than that the data were too noisy to detect it. Together, the signal-to-noise ratio seemed adequate to answer our main question by showing that the training paradigm did not alter the cortical representations.

A second methodological limitation for both neuroimaging analyses is the lack of power to observe what is likely a relatively modest effect size. However, previous studies of sensory representational plasticity provide a reasonable measure of the true effect size, suggesting we are reasonably powered [77]. While it is true that the number of samples in both studies was comparatively low for a typical univariate functional imaging study (at 10 participants per group), several design choices alleviate this concern. First, we collected a substantial amount

of data per subject. In study 1 (chapter 2), each subject was scanned for approximately 2 hours before training, and 2 hours after training, with 6 identical and independent imaging sessions per run. In study 2 (chapter 3), each subject was scanned for approximately 3 hours before and after training, each involving 8 independent imaging sessions per run. This relatively large volume of data per subject enabled us to obtain robust estimates of the population patterns of interest. Second, due to the fact that we had independent estimates of these patterns for each individual subject, we were also capable of treating each subject as their own experiment, similar to non-human primate studies that report estimates for a handful of animals. Individual subject data were analyzed with non-parametric permutation tests for both imaging studies, and the results of those tests did not alter our conclusions. Thus, while the number of subjects was modest, we do not believe that our results are simply the result of insufficient power.

A third methodological limitation is the focus on the cortical motor representations. Notably, we did not examine subcortical representational plasticity. This is a concern given the fact that a lot of evidence, in both electrophysiological [56, 57, 54, 55] and imaging [85, 106] studies, points to the striatum as being a location where motor sets may be built (for a review see [107]). The main reason that we did not examine the representations within the striatum is that we did not observe robust separation of the digits as we did throughout the cortical motor network. Therefore, the signal-to-noise ratio for these particular experiments and imaging protocol is insufficient to address questions about binding within the basal ganglia. Future studies in our lab at 7T MRI, with much higher spatial resolution, should be sufficient to resolve the topography within and address this question.

A final methodological limitation is the reliance on the average distance H as the sole indicator of representational plasticity. There are several alternative measures that could have been used. One possibility is that the geometries are stable, but that the overall representational extent increases. If a greater neuronal population is recruited for a given task, such as executing a set of finger movements, that would not necessarily be reflected in a change in the distance. Indeed, it has been shown that practice expands the cortical representations of trained sequences compared to untrained sequences [48]. We did not report on this type of analysis because it is not as clearly related to binding as to sequence learning, more broadly

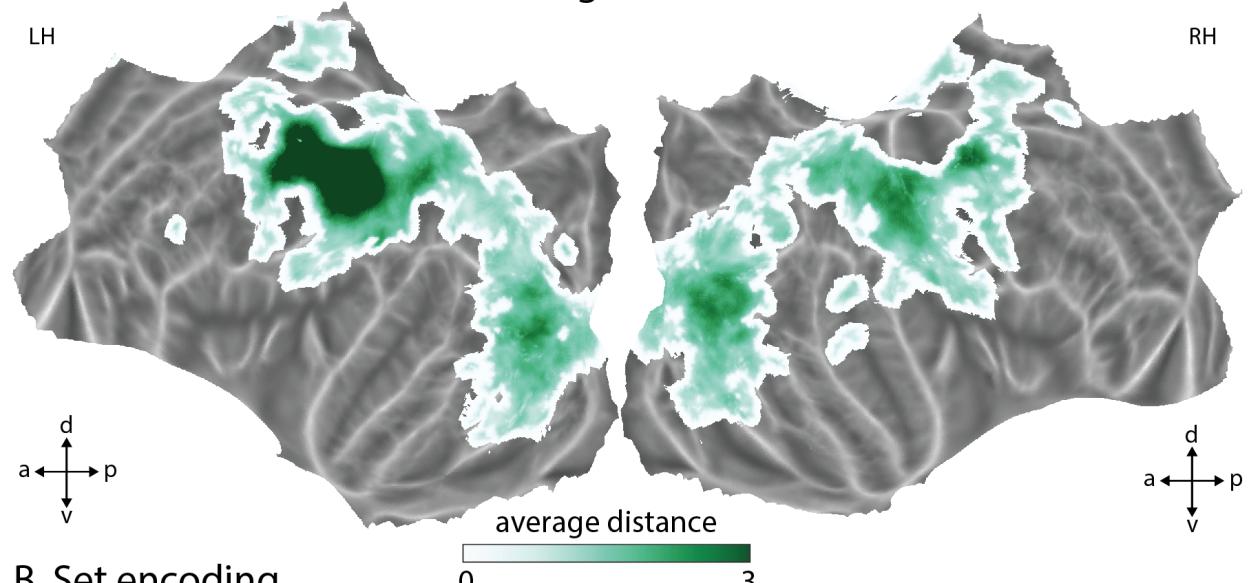
construed. Another alternative indicator of plasticity would be identifying changes in representational connectivity. For example, one could measure the RDM correlation throughout the motor network, to determine if higher order motor planning regions are more dynamic. This correlation is again somewhat tangential to the main question of a binding mechanism, and thus we did not address it here.

4.3 INSIGHTS

Fig. 20 summarizes the similarities between the cortical networks shown to encode individual cued-movements (chapter 2) and the cortical networks that encoded action sets represented in the two modalities (chapter 3). The set maps (panel B) show clusters that occupy what is almost a strict subset of the areas encoding individual cued-movements (panel A), for example, in the right hemisphere ipsilateral to the executing hand. This is consistent with the idea that learning sequential actions is a specialty function of visuomotor planning. This fact also raises the possibility that the set encoding is simply a weighted combination of individual movements. However, the covariance patterns associated with the action sets did not appear to be copies of the first finger or first cue effects (compare Fig. 13 & Fig. 14). Therefore, in order to obtain reliable encoding, the entire set had to be consistent, and not just the first item within that set. A recent paper that explored sequence representations in primary motor cortex found that the local patterns of activity could be sufficiently explained by a linear combination of the component finger movements [95]. In our data, the representational structure of the sets is suggestive of a genuine sequence representation even within primary motor cortex, but the study design lacks the appropriate detail to do such an analysis. Future analyses will examine the extent to which the representational structures of sets (Fig. 13 & Fig. 14) can be predicted by the single finger representations.

The behavioral findings reported in chapters 2 and 3 are not the first reports of an emergent autocorrelative structure in sequential movements [45, 46]; however, we have advanced the current understanding of this effect by showing that naturalistic binding follows the statistical structure of the trained sequence. In addition, we showed that training on explicit

A. Individual movement encoding



B. Set encoding

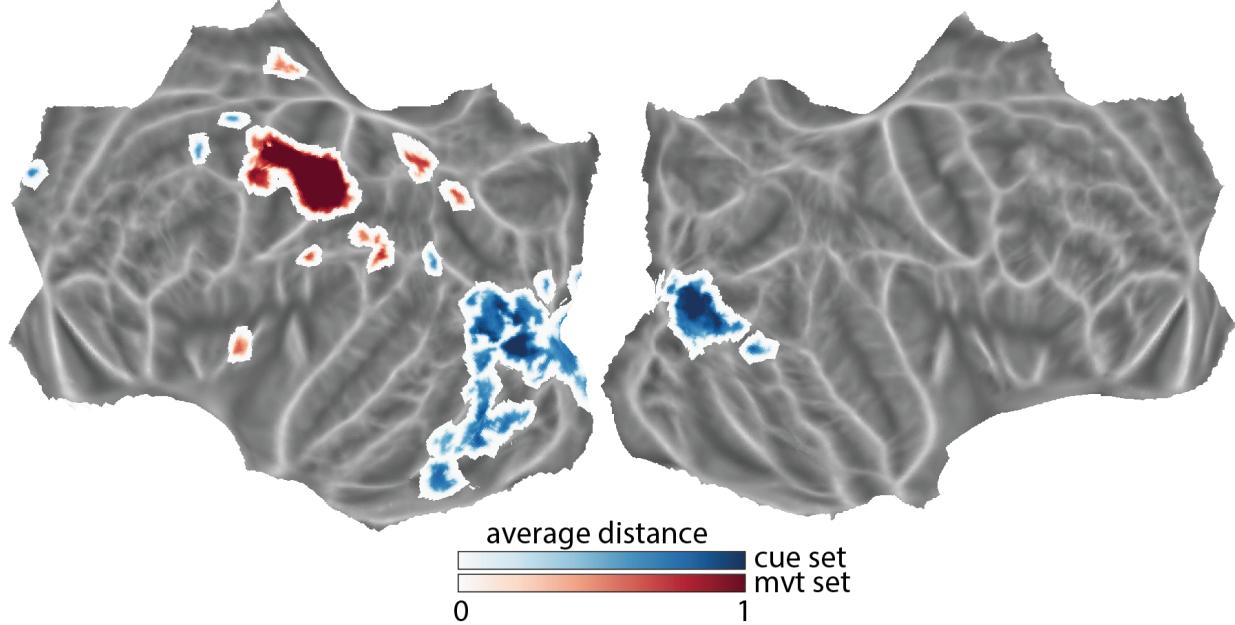


Figure 20: Difference in individual movement encoding and movement set encoding across the entire brain. A. Average distance between individual cue-response pairs (chapter 2). B. Average distance between cue (shown in blue) or movement (shown in red) sets (chapter 3). Distances were computed at each voxel using a surface based searchlight and projected onto a flat map of the average surface. Panel B shows the same data as Fig. 11. Note that the scales between panels are not identical.

set structures in either the visual or motor modalities does not have an observable effect in terms of degree of binding. The fact that we did detect an advantage in the visual group in [46] but not here may be accounted for by differences in training regimens between the two studies. Unlike in [46] where natural boundaries of subsets of the sequence were required to be learned in a similar way as done in chapter 2, in the experiment reported in chapter 3 we included pauses between every 4th movement in order to facilitate binding specific to those 4 element sets. Thus, by explicitly demarcating a substructure of the sequence, we may have removed an advantage for learning to bind items together that we previously observed for sequences represented in a visual modality.

The emergence of a correlation in response times across movements suggests the existence of a skill representation that generates a common motor command at some, likely intermediate, level of the motor hierarchy. Based on previous studies [45, 46], we hypothesized that binding would involve a reorganization of the neural population patterns that encode the individual actions. In particular, we hypothesized that binding would involve a shift in the representational geometries of either individual movements (reduced distances after training) or sequential sets (increased distances after training) (Fig. 4). A major part of this dissertation consisted of investigating the representational structures, in cortical motor networks, associated with both individual movements (chapter 2) and movement sets (chapter 3) both before and after extensive sequential training. As stated in the specific aims, we postulated that the population patterns of the elemental movements that compose bound sets would converge, and that this convergence would alter the representational geometry associated with those sets. Importantly, we presumed that this training-related change would happen in higher level motor planning areas and not low level execution areas like primary motor or primary somatosensory cortex. Such a reorganization would have been suggestive of the neural mechanism underlying binding, which is not currently understood. However, we did not detect any consistent changes in the representational geometries associated with either individual movements or movements sets. This fact suggests that binding may not be associated with a plasticity mechanism that alters the relative patterns of representation of movements.

One reason for this stability in the cortical representations of actions may simply be

the limited scope of training. The cortical representational geometries that we observed for individual movements within M1 and S1 (Fig. 8) were consistent with previous reports that have indicated the organizational structure of the hand is best explained by the covariance patterns associated with everyday use [49]. Since 30 minutes of training constitutes only a small proportion of the total hand use time throughout the day, such training does not significantly alter the natural statistics of hand use. In addition, noted in chapter 2, binding movements together at this level of the motor hierarchy could reduce behavioral flexibility [2]. Perhaps it is not that surprising that the representational structures within primary motor and somatosensory cortex are therefore stable. It is perhaps more surprising that the geometries were also stable in higher order planning regions, such as PMd and SMA, where the geometries are not as well described by natural statistics (Fig. 8). In any event, while the training time was limited, there was robust behavioral evidence of learning as well as the emergence of binding, which should have moved the representations if our hypotheses concerning the binding mechanisms were correct.

Even before training, the representational patterns associated with particular movement sets were highly dissimilar (Fig. 11). This result is consistent with previous research that has shown both trained and untrained sequences are encoded in both primary motor and higher order planning regions such as PMd, PMv and IPS [48]. Given that this pretraining discriminability requires consistent pairwise differences across imaging runs, one explanation as to why we did not observe set plasticity is that the representations are very quickly constructed and remain stable regardless of how training occurs.

While binding was clearly visible in the behavior of both experiments, the consequence of this plasticity does not appear to impact the representational geometries in cortical motor networks, suggesting that the mechanism underlying binding does not involve a reorganization of representational patterns of movements. This finding is somewhat puzzling in light of the substantial body of work that has shown representational geometries do change in other learning contexts, for example in category learning [17, 108] and somatosensory perception [77]. What could account for the fact that representational geometries are known to be plastic in other contexts and not here? One possible clue as to why these representations exhibit such high stability is recent research using brain computer interfaces to investigate

motor learning in non-human primates. Neural activation patterns occupy low dimensional subspaces, or intrinsic manifolds, within motor cortex, presumably resulting from the underlying circuitry. These intrinsic manifolds have been shown to constrain the degree of learning in point to point reaching tasks such that within manifold perturbations are much easier to learn than perturbations orthogonal to the manifold [109]. Furthermore, when learning does occur, the existing neural population patterns are stable, but the same population patterns become reassociated with different movements [110]. In the context of RSA, this result could mean that the average distance between all neural population patterns, which we used as a summary statistic, would remain constant. While a direct link between intrinsic manifolds and the representational geometries requires a thorough analysis of the voxel-by-voxel covariance matrices across different conditions, the absence of plasticity we reported would be consistent with the reassociation learning strategy. This would suggest not looking for changes in the average distance (H), but changes in which patterns are associated with what movement.

4.4 CONCLUSION

Together, the findings from these studies show that, after extensive practice of a motor sequence, the brain learns to bind temporally adjacent actions. This binding process naturally arises from learning the temporal order of associations between sequentially repeated actions, regardless of whether sequences are represented as sets of movements or sets of visual cues. Neurally, binding appears to have little impact on the macroscale cortical patterns of activity associated with either individual movements or isolated movement sets represented in either a motor or sensory modality. Despite robust impacts of skill learning on behavior, binding does not coincide with fundamental changes in the underlying relationships between sensorimotor representations.

BIBLIOGRAPHY

- [1] Patrick Beukema, Jörn Diedrichsen, and Timothy Verstynen. Binding during sequence learning does not alter cortical representations of individual actions. *bioRxiv*, page 255794, 2018.
- [2] Karl Spencer Lashley. *The problem of serial order in behavior*. Bobbs-Merrill, 1951.
- [3] Stanislas Dehaene, Florent Meyniel, Catherine Wacongne, Liping Wang, and Christophe Pallier. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1):2–19, 2015.
- [4] Mary Jo Nissen and Peter Bullemer. Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, 19(1):1–32, 1987.
- [5] Okihide Hikosaka, Kae Nakamura, Katsuyuki Sakai, and Hiroyuki Nakahara. Central mechanisms of motor skill learning. *Current opinion in neurobiology*, 12(2):217–222, 2002.
- [6] Tim Curran and Steven W Keele. Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(1):189, 1993.
- [7] Jenny R Saffran, Richard N Aslin, and Elissa L Newport. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928, 1996.
- [8] Pierre Perruchet and Sébastien Pacton. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, 10(5):233–238, 2006.
- [9] Elisabeth A Karuza, Elissa L Newport, Richard N Aslin, Sarah J Starling, Madalina E Tivarus, and Daphne Bavelier. The neural correlates of statistical learning in a word segmentation task: An fmri study. *Brain and language*, 127(1):46–54, 2013.
- [10] Travis Meyer, Suchitra Ramachandran, and Carl R Olson. Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *Journal of Neuroscience*, 34(28):9332–9337, 2014.

- [11] Suchitra Ramachandran, Travis Meyer, and Carl R Olson. Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *Journal of neurophysiology*, 115(1):355–362, 2015.
- [12] Hajime Mushiake, Masahiko Inase, and Jun Tanji. Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements. *Journal of neurophysiology*, 66(3):705–718, 1991.
- [13] Nicholas B Turk-Browne, Brian J Scholl, Marcia K Johnson, and Marvin M Chun. Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, 30(33):11177–11187, 2010.
- [14] Nicholas B Turk-Browne, Brian J Scholl, Marvin M Chun, and Marcia K Johnson. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, 21(10):1934–1945, 2009.
- [15] Caterina Breitenstein, Andreas Jansen, Michael Deppe, Ann-Freya Foerster, Jens Sommer, Thomas Wolbers, and Stefan Knecht. Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage*, 25(3):958–968, 2005.
- [16] Erica R Appleman, Genevieve Albouy, Julien Doyon, Alice Cronin-Golomb, and Bradley R King. Sleep quality influences subsequent motor skill acquisition. *Behavioral neuroscience*, 130(3):290, 2016.
- [17] Anna C Schapiro, Lauren V Kustner, and Nicholas B Turk-Browne. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, 22(17):1622–1627, 2012.
- [18] Anna C Schapiro, Emma Gregory, Barbara Landau, Michael McCloskey, and Nicholas B Turk-Browne. The necessity of the medial temporal lobe for statistical learning. *Journal of cognitive neuroscience*, 26(8):1736–1747, 2014.
- [19] S-I Higuchi and Yasushi Miyashita. Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings of the National Academy of Sciences*, 93(2):739–743, 1996.
- [20] Norbert J Fortin, Kara L Agster, and Howard B Eichenbaum. Critical role of the hippocampus in memory for sequences of events. *Nature neuroscience*, 5(5):458, 2002.
- [21] Kara L Agster, Norbert J Fortin, and Howard Eichenbaum. The hippocampus and disambiguation of overlapping sequences. *Journal of Neuroscience*, 22(13):5760–5768, 2002.
- [22] Christopher J MacDonald, Kyle Q Lepage, Uri T Eden, and Howard Eichenbaum. Hippocampal “time cells” bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749, 2011.

- [23] Jessica K Wilson, Bengi Baran, Edward F Pace-Schott, Richard B Ivry, and Rebecca MC Spencer. Sleep modulates word-pair learning but not motor sequence learning in healthy older adults. *Neurobiology of aging*, 33(5):991–1000, 2012.
- [24] Rebecca MC Spencer, Michelle Sunm, and Richard B Ivry. Sleep-dependent consolidation of contextual learning. *Current Biology*, 16(10):1001–1005, 2006.
- [25] Matthew P Walker, Tiffany Brakefield, Alexandra Morgan, J Allan Hobson, and Robert Stickgold. Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, 35(1):205–211, 2002.
- [26] Aaron M Bornstein and Nathaniel D Daw. Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, 35(7):1011–1023, 2012.
- [27] Haline E Schendan, Meghan M Searl, Rebecca J Melrose, and Chantal E Stern. An fmri study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron*, 37(6):1013–1025, 2003.
- [28] Alexander J Barnett, Edward B O’Neil, Hilary C Watson, and Andy CH Lee. The human hippocampus is sensitive to the durations of events and intervals within a sequence. *Neuropsychologia*, 64:1–12, 2014.
- [29] Tim Curran. Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of cognitive neuroscience*, 9(4):522–533, 1997.
- [30] Daniel Tranel, Antonio R Damasio, Hanna Damasio, and Joan P Brandt. Sensorimotor skill learning in amnesia: additional evidence for the neural basis of nondeclarative memory. *Learning & Memory*, 1(3):165–179, 1994.
- [31] Juliane Döhring, Anne Stoldt, Karsten Witt, Robby Schönfeld, Günther Deuschl, Jan Born, and Thorsten Bartsch. Motor skill learning and offline-changes in tga patients with acute hippocampal ca1 lesions. *Cortex*, 89:156–168, 2017.
- [32] Yue Du and Jane E Clark. New insights into statistical learning and chunk learning in implicit sequence acquisition. *Psychonomic bulletin & review*, 24(4):1225–1233, 2017.
- [33] Christopher Summerfield and Floris P De Lange. Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, 15(11):745, 2014.
- [34] George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.
- [35] Oleg Solopchuk, Andrea Alamia, Etienne Olivier, and Alexandre Zénon. Chunking improves symbolic sequence processing and relies on working memory gating mechanisms. *Learning & Memory*, 23(3):108–112, 2016.

- [36] Pavan Ramkumar, Daniel E Acuna, Max Berniker, Scott T Grafton, Robert S Turner, and Konrad P Kording. Chunking as the result of an efficiency computation trade-off. *Nature communications*, 7:12176, 2016.
- [37] David A Rosenbaum, Sandra B Kenny, and Marcia A Derr. Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 9(1):86, 1983.
- [38] Willem B Verwey. Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3):544, 1996.
- [39] Willem B Verwey, Elger L Abrahamse, Marit FL Ruitenberg, Luis Jiménez, and Elian de Kleine. Motor skill learning in the middle-aged: limited development of motor chunks and explicit sequence knowledge. *Psychological research*, 75(5):406–422, 2011.
- [40] Willem B Verwey, Elger L Abrahamse, and Luis Jiménez. Segmentation of short keying sequences does not spontaneously transfer to other sequences. *Human movement science*, 28(3):348–361, 2009.
- [41] Steve W Kennerley, Katsuyuki Sakai, and MFS Rushworth. Organization of action sequences and the role of the pre-sma. *Journal of neurophysiology*, 91(2):978–993, 2004.
- [42] Willem B Verwey, Robin Lammens, and Jack van Honk. On the role of the sma in the discrete sequence production task: a tms study. *Neuropsychologia*, 40(8):1268–1276, 2002.
- [43] Willem B Verwey and Teun Eikelboom. Evidence for lasting sequence segmentation in the discrete sequence-production task. *Journal of motor behavior*, 35(2):171–181, 2003.
- [44] Daniel E Acuna, Nicholas F Wymbs, Chelsea A Reynolds, Nathalie Picard, Robert S Turner, Peter L Strick, Scott T Grafton, and Konrad P Kording. Multifaceted aspects of chunking enable robust algorithms. *Journal of neurophysiology*, 112(8):1849–1856, 2014.
- [45] Timothy Verstynen, Jeff Phillips, Emily Braun, Brett Workman, Christian Schunn, and Walter Schneider. Dynamic sensorimotor planning during long-term sequence learning: the role of variability, response chunking and planning errors. *PLoS One*, 7(10):e47336, 2012.
- [46] Brighid Lynch, Patrick Beukema, and Timothy Verstynen. Differentiating visual from response sequencing during long-term skill learning. *Journal of cognitive neuroscience*, 29(1):125–136, 2017.
- [47] Jörn Diedrichsen and Katja Kornysheva. Motor skill learning between selection and execution. *Trends in cognitive sciences*, 19(4):227–233, 2015.

- [48] Tobias Wiestler and Jörn Diedrichsen. Skill learning strengthens cortical representations of motor sequences. *eLife*, 2:e00801, jul 2013.
- [49] Naveed Ejaz, Masashi Hamada, and Jörn Diedrichsen. Hand use predicts the structure of representations in sensorimotor cortex. *Nature neuroscience*, 18(7):1034, 2015.
- [50] Atsushi Yokoi, Spencer A. Arbuckle, and Jörn Diedrichsen. The role of human primary motor cortex in the production of skilled finger sequences. *Journal of Neuroscience*, 38(6):1430–1442, 2018.
- [51] Machiko Ohbayashi, Nathalie Picard, and Peter L Strick. Inactivation of the dorsal premotor area disrupts internally generated, but not visually guided, sequential movements. *Journal of Neuroscience*, 36(6):1971–1976, 2016.
- [52] Michel Desmurget and Robert S Turner. Motor sequences and the basal ganglia: kinematics, not habits. *Journal of Neuroscience*, 30(22):7685–7690, 2010.
- [53] Pierre-Luc Tremblay, Marc-Andre Bedard, Dominic Langlois, Pierre J Blanchet, Martin Lemay, and Maxime Parent. Movement chunking during sequence learning is a dopamine-dependant process: a study conducted in parkinson’s disease. *Experimental brain research*, 205(3):375–385, 2010.
- [54] Kyle S Smith and Ann M Graybiel. A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, 79(2):361–374, 2013.
- [55] Theresa M Desrochers, Ken-ichi Amemori, and Ann M Graybiel. Habit learning by naive macaques is marked by response sharpening of striatal neurons representing the cost and outcome of acquired action sequences. *Neuron*, 87(4):853–868, 2015.
- [56] Xin Jin, Fatuel Tecuapetla, and Rui M Costa. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature neuroscience*, 17(3):423, 2014.
- [57] Xin Jin and Rui M Costa. Shaping action sequences in basal ganglia circuits. *Current opinion in neurobiology*, 33:188–196, 2015.
- [58] Xin Jin and Rui M Costa. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, 466(7305):457, 2010.
- [59] I Kermadi, Y Jurquet, M Arzi, and JP Joseph. Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Experimental Brain Research*, 94(2):352–356, 1993.
- [60] Ari E Kahn, Elisabeth A Karuza, Jean M Vettel, and Danielle S Bassett. Network constraints on learnability of probabilistic motor sequences. *arXiv preprint arXiv:1709.03000*, 2017.

- [61] Scott T Grafton, Eliot Hazeltine, and Richard Ivry. Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7(4):497–510, 1995.
- [62] VB Penhune and J Doyon. Cerebellum and m1 interaction during early learning of timed motor sequences. *Neuroimage*, 26(3):801–812, 2005.
- [63] Stephane Lehericy, Habib Benali, Pierre-Francois Van de Moortele, Melanie Pelegrini-Issac, Tobias Waechter, Kamil Ugurbil, and Julien Doyon. Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. *Proceedings of the National Academy of Sciences of the United States of America*, 102(35):12566–12571, 2005.
- [64] Russell A Poldrack, Fred W Sabb, Karin Foerde, Sabrina M Tom, Robert F Asarnow, Susan Y Bookheimer, and Barbara J Knowlton. The neural correlates of motor skill automaticity. *Journal of Neuroscience*, 25(22):5356–5364, 2005.
- [65] Jinhu Xiong, Liangsoo Ma, Binquan Wang, Shalini Narayana, Eugene P Duff, Gary F Egan, and Peter T Fox. Long-term motor training induced changes in regional cerebral blood flow in both task and resting states. *Neuroimage*, 45(1):75–82, 2009.
- [66] Ingo Meister, Timo Krings, Henrik Foltys, Babak Boroojerdi, Mareike Müller, Rudolf Töpper, and Armin Thron. Effects of long-term practice and task complexity in musicians and nonmusicians performing simple and complex motor tasks: Implications for cortical motor organization. *Human brain mapping*, 25(3):345–352, 2005.
- [67] Nicholas F Wymbs and Scott T Grafton. The human motor system supports sequence-specific representations over multiple training-dependent timescales. *Cerebral cortex*, 25(11):4213–4225, 2014.
- [68] Tobias Wiestler and Jörn Diedrichsen. Skill learning strengthens cortical representations of motor sequences. *Elife*, 2, 2013.
- [69] Nathalie Picard, Yoshiya Matsuzaka, and Peter L Strick. Extended practice of a motor skill is associated with reduced metabolic activity in m1. *Nature neuroscience*, 16(9):1340, 2013.
- [70] Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4, 2008.
- [71] Hamed Nili, Cai Wingfield, Alexander Walther, Li Su, William Marslen-Wilson, and Nikolaus Kriegeskorte. A toolbox for representational similarity analysis. *PLoS computational biology*, 10(4):e1003553, 2014.
- [72] Alexander Walther, Hamed Nili, Naveed Ejaz, Arjen Alink, Nikolaus Kriegeskorte, and Jörn Diedrichsen. Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage*, 137:188–200, 2016.

- [73] Jörn Diedrichsen and Nikolaus Kriegeskorte. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS computational biology*, 13(4):e1005508, 2017.
- [74] Nikolaus Kriegeskorte, Marieke Mur, Douglas A Ruff, Roozbeh Kiani, Jerzy Bodurka, Hossein Esteky, Keiji Tanaka, and Peter A Bandettini. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141, 2008.
- [75] Nikolaus Kriegeskorte and Rogier A Kievit. Representational geometry: integrating cognition, computation, and the brain. *Trends in cognitive sciences*, 17(8):401–412, 2013.
- [76] Spencer A Arbuckle, Atsushi Yokoi, J Andrew Pruszynski, and Jörn Diedrichsen. Stability of representational geometry across a wide range of fmri activity levels. *bioRxiv*, page 266585, 2018.
- [77] James Kolasinski, Tamar R Makin, John P Logan, Saad Jbabdi, Stuart Clare, Charlotte J Stagg, and Heidi Johansen-Berg. Perceptually relevant remapping of human somatotopy in 24 hours. *Elife*, 5, 2016.
- [78] Patrick Beukema, Jörn Diedrichsen, and Timothy Verstynen. Binding during sequence learning does not alter cortical representations of individual actions. *bioRxiv*, 2018.
- [79] Randoff J Nudo, Garrett W Milliken, W Merzenich Jenkins, and Michæl M Merzenich. Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *Journal of Neuroscience*, 16(2):785–807, 1996.
- [80] Jörn Diedrichsen, Serge Provost, and Hossein Zareamoghaddam. On the distribution of cross-validated mahalanobis distances. *arXiv preprint arXiv:1607.01371*, 2016.
- [81] JASP Team et al. Jasp (version 0.8.5). *Computer software*, 2017.
- [82] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.
- [83] Timothy Verstynen, Jorn Diedrichsen, Neil Albert, Paul Aparicio, and Richard B Ivry. Ipsilateral motor cortex activity during unimanual hand movements relates to task complexity. *Journal of neurophysiology*, 93(3):1209–1222, 2005.
- [84] Aaron L Wong, Martin A Lindquist, Adrian M Haith, and John W Krakauer. Explicit knowledge enhances motor vigor and performance: motivation versus practice in sequence tasks. *Journal of neurophysiology*, 114(1):219–232, 2015.
- [85] Nicholas F Wymbs, Danielle S Bassett, Peter J Mucha, Mason A Porter, and Scott T Grafton. Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans. *Neuron*, 74(5):936–946, 2012.

- [86] Jonathan W Peirce. Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1-2):8–13, 2007.
- [87] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.
- [88] Bruce Fischl, Niranjini Rajendran, Evelina Busa, Jean Augustinack, Oliver Hinds, BT Thomas Yeo, Hartmut Mohlberg, Katrin Amunts, and Karl Zilles. Cortical folding patterns and predicting cytoarchitecture. *Cerebral cortex*, 18(8):1973–1980, 2007.
- [89] TA Yousry, UD Schmid, H Alkadhi, D Schmidt, A Peraud, A Buettner, and P Winkler. Localization of the motor hand area to a knob on the precentral gyrus. a new landmark. *Brain: a journal of neurology*, 120(1):141–157, 1997.
- [90] Nikolaas N Oosterhof, Tobias Wiestler, Paul E Downing, and Jörn Diedrichsen. A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage*, 56(2):593–600, 2011.
- [91] Genevieve Albouy, Stuart Fogel, Bradley R King, Samuel Laventure, Habib Benali, Avi Karni, Julie Carrier, Edwin M Robertson, and Julien Doyon. Maintaining vs. enhancing motor sequence memories: respective roles of striatal and hippocampal systems. *Neuroimage*, 108:423–434, 2015.
- [92] K Witt, N Margraf, C Bieber, J Born, and G Deuschl. Sleep consolidates the effector-independent representation of a motor skill. *Neuroscience*, 171(1):227–234, 2010.
- [93] Daniel A Cohen, Alvaro Pascual-Leone, Daniel Z Press, and Edwin M Robertson. Off-line learning of motor skill memory: a double dissociation of goal and movement. *Proceedings of the National Academy of Sciences of the United States of America*, 102(50):18237–18241, 2005.
- [94] Patrick Beukema and Timothy Verstynen. Predicting and binding: interacting algorithms supporting the consolidation of sequential motor skills. *Current Opinion in Behavioral Sciences*, 20:98–103, 2018.
- [95] Atsushi Yokoi, Spencer A Arbuckle, and Jörn Diedrichsen. The role of human primary motor cortex in the production of skilled finger sequences. *Journal of Neuroscience*, 38(6):1430–1442, 2018.
- [96] Eran Dayan and Leonardo G Cohen. Neuroplasticity subserving motor skill learning. *Neuron*, 72(3):443–454, 2011.
- [97] Julien Doyon. Motor sequence learning and movement disorders. *Current opinion in neurology*, 21(4):478–483, 2008.
- [98] Richard P Dum and Peter L Strick. Cingulate motor areas. In *Neurobiology of cingulate cortex and limbic thalamus*, pages 415–441. Springer, 1993.

- [99] Jean-Alban Rathelot, Richard P Dum, and Peter L Strick. Posterior parietal cortex contains a command apparatus for hand movements. *Proceedings of the National Academy of Sciences*, 114(16):4255–4260, 2017.
- [100] Gereon R Fink, Richard SJ Frackowiak, Uwe Pietrzik, and Richard E Passingham. Multiple nonprimary motor areas in the human cortex. *Journal of neurophysiology*, 77(4):2164–2174, 1997.
- [101] Jun Tanji. Sequential organization of multiple movements: involvement of cortical motor areas. *Annual review of neuroscience*, 24(1):631–651, 2001.
- [102] Marine Vernet, Romain Quentin, Lorena Chanes, Andres Mitsumasu, and Antoni Valero-Cabré. Frontal eye field, where art thou? anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in integrative neuroscience*, 8:66, 2014.
- [103] Matthew Botvinick, Leigh E Nystrom, Kate Fissell, Cameron S Carter, and Jonathan D Cohen. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758):179, 1999.
- [104] Jonathan W Peirce. Generating stimuli for neuroscience using psychopy. *Frontiers in neuroinformatics*, 2:10, 2009.
- [105] Patrick Beukema. Sequence mapping code. <https://github.com/pbeukema/ModMap>, 2018.
- [106] Ovidiu Lungu, Oury Monchi, Geneviève Albouy, Thomas Jubault, Emanuelle Ballarin, Yves Burnod, and Julien Doyon. Striatal and hippocampal involvement in motor sequence chunking depends on the learning strategy. *PLoS One*, 9(8):e103885, 2014.
- [107] Ann M Graybiel and Scott T Grafton. The striatum: where skills and habits meet. *Cold Spring Harbor perspectives in biology*, 7(8):a021691, 2015.
- [108] Anna C Schapiro, Nicholas B Turk-Browne, Kenneth A Norman, and Matthew M Botvinick. Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, 26(1):3–8, 2016.
- [109] Patrick T Sadtler, Kristin M Quick, Matthew D Golub, Steven M Chase, Stephen I Ryu, Elizabeth C Tyler-Kabara, M Yu Byron, and Aaron P Batista. Neural constraints on learning. *Nature*, 512(7515):423, 2014.
- [110] Matthew D Golub, Patrick T Sadtler, Emily R Oby, Kristin M Quick, Stephen I Ryu, Elizabeth C Tyler-Kabara, Aaron P Batista, Steven M Chase, and M Yu Byron. Learning by neural reassociation. *Nature neuroscience*, page 1, 2018.