Neural representation of action grammar structure in primate frontal cortex

Lucas Y. Tian[1,4,5]*, Daniel J. Hanuska[1,5], Kedar Garzón Gupta[1,6], Yue Liu[2,7], Xiao-Jing Wang[2], Joshua B. Tenenbaum[3,4], Winrich A. Freiwald[1,4,5]*

Abstract

Humans and other animals can solve new problems, even on the first attempt. This capacity to generate novel problem-solving behavior has been hypothesized to depend on brain mechanisms for recombining units of knowledge using systems of procedural rules, or *grammars*. Yet, whether and how the brain represents and implements grammars remains unclear, in large part due to the lack of neural recordings (typically performed in animal models) during grammar-based problem solving (typically studied in humans). Here, we address that gap, and in turn identify an underlying neural basis of grammatical behavior. We designed a visual–motor construction task in which macaque monkeys trace complex, often novel, geometric figures by generating a sequence of strokes, each a distinct shape, like a circle, dash, or chevron (action symbols[1]). Critically, sequencing was guided by a learned hierarchical action grammar $A^nB^mC^k$, meaning "repeat shape A, then repeat shape B, then repeat shape C", where the number of repetitions (n, m, and k) varied across problems. Behavior was internally generated (no cues guiding stroke order), problem-directed (to construct a specific image), and exhibited zero-shot generalization (successful on the first trial for novel problems, including harder ones)--key hallmarks of grammar use. To identify underlying neural substrates, we recorded from multiple frontal cortical areas previously implicated in rule use or action sequencing. Activity during drawing encoded key structural properties of the action grammar: (i) *shape index* (A, B, C), (ii) *abstract role* (e.g., the role *A* independent of the specific shape), and (iii) *ordinal position* within a shape repeat. Moreover, all three properties were strongest in a single region, the pre-supplementary motor area (preSMA). We suggest the possibility that this conjunction of representations is a signature of an implementation of an iterative algorithm resembling a "for-loop" program: internally tracking how many repetitions remain for each shape index and switching to the next when complete. Our study establishes a paradigm for studying the neural basis of grammar-

[1] Laboratory of Neural Systems, The Rockefeller University, USA

[2] Center for Neural Science, New York University, USA

[3] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

[4] Center for Brains, Minds and Machines, MIT & Rockefeller University, USA

[5] The Price Family Center for the Social Brain, Rockefeller University, USA

[6] Zuckerman Institute, Columbia University, USA

[7] College of Engineering & Computer Science, Florida Atlantic University, USA

* Co-corresponding authors: ltian@rockefeller.edu (L.Y.T.) and wfreiwald@rockefeller.edu (W.A.F.)

guided novel behavior and identifies in preSMA a dynamic representation of grammatical structure supporting such behavior.

Acknowledgements

We thank A. Rouse, M. Eldridge, and M Schieber for assistance with surgeries; X. Ma, T. Wu, D. Dolnik, S. Sharma, A. Urquieta, V. Calligy, Y. Tazi and other members of the Freiwald, Wang, and Tenenbaum labs for project feedback; K. Kay and T. Nigam for manuscript feedback; V. Sherman and A. Gonzalez for technical assistance; and L. Ying for administrative assistance. This work was supported by the National Institutes of Health, through the National Eye Institute (R01EY021594 to W.A.F.), the National Institute Of Mental Health (F32MH125573 to L.Y.T.), and the National Institute Of Neurological Disorders And Stroke (K99NS131585 to L.Y.T.), as well as the Simons Foundation's Collaboration on the Global Brain (876120SPI and AN-NC-GB-Pilot Extension-00002596-01 to X.-J.W., J.B.T., and W.A.F., and NC-GB-CULM-00003138 to X.-J.W.), the Center for Brains, Minds & Machines of the National Science Foundation (STC award CCF-1231216 to W.A.F. and J.B.T.), the Office of Naval Research (N00014-20-1-2292, Vannevar Bush Faculty Fellowship, to W.A.F., N00014-23-1-2040 to X.-J.W., and MURI N00014-21-1-2801 to J.B.T.), and the Air Force Office of Scientific Research (FA9550-22-1-0387 to J.B.T.).

Introduction

A hallmark of intelligence is the ability to solve novel problems by constructing new solutions that recombine known elements in novel ways. This capacity for *compositional generalization*[2–4] is often suggested to depend on internal systems of procedural rules[2,4–10], or *grammars*[2,5,6,8–10], that specify how discrete units of knowledge can be combined into structured composites. Because a finite set of rules can generate a vast, potentially infinite, set of possible solutions, grammars provide a principled account of zero-shot generalization, success on a novel problem on the first attempt. Grammatical structure has been characterized in a wide range of behaviors[6,8,9,11–21], including the structured reasoning that occurs in math and logic, and other seemingly less structured behaviors such as tool use, object use, musicianship, handwriting, drawing, speech, physical reasoning, and social reasoning. For example, when asked to "draw an animal that doesn't exist," children readily produce imaginary animals, such as a dog-like creature with six legs and three camel humps[15], consistent with a system of rules resembling "repeat body part X" or "attach part X to part Y".

Despite this behavioral evidence for grammars, whether and how they are implemented in the brain remains unclear. A key challenge is the lack of a model system that combines grammar-guided behavior, including tests of generalization, with neural recordings. To establish such a model system, we focused on three key behavioral hallmarks of grammatical ability. Behavior should be: (1) *internally generated*, in that action order is not directly cued by external stimuli; (2) *problem-directed*, organized around solving an immediate problem rather than reflecting a habitual or stimulus–response sequence; and (3) *zero-shot generalizing*, succeeding on novel and potentially harder problems on the first attempt, thereby ruling out memorized action sequences.

To date, prior studies combining neuronal recordings and animal behavior have not jointly demonstrated all three properties, although multiple lines of study have provided important insight. First, studies of "artificial grammar learning"[9,22,23] have found that many species can discriminate or learn grammatically structured regularities in stimulus sequences; however, these studies have typically relied on perceptual tasks, leaving unclear whether subjects can *generate behavior* consistent with the grammar. Second, recordings during motor sequencing have revealed diverse neural correlates, but have typically used explicit ordering cues[24,25], memorized sequences[26–30], or minimally constrained task settings[31–33], rather than assessing the ability to generalize rules. Third, other studies have assessed a variety of sequencing rules, but lacking neural recordings[34–37] or not involving zero-shot generalization[38–45], including to harder problems[46]. A fourth set of studies has demonstrated grammar-like sequencing with zero-shot generalization, but lack neural recordings. These include mirror grammars in macaques, including generalization to harder problems[47], and center-embedded grammars in macaques[48] and corvids[49]. Fifth, ethological studies have described grammatical structure in behaviors such as object manipulation[13],

foraging[50], and tool use[51] but these settings raise difficulties for experimental control, testing for novel behavior, and recording neural activity. In sum, although neuronal correlates of latent sequential structure have been described in a variety of paradigms[27–30,32,38–43,45,46,52], we currently lack one jointly probing grammatically structured behavior and neuronal activity. As a result, there remains much to be discovered about the underlying neural bases of grammatically structured behavior.

Here we address this gap by developing a drawing-like visual-motor construction task involving grammatically structured action sequencing (**Fig. 1a-d**) and then performing large-scale neural recordings. Macaques are presented with images of complex geometric figures which they must trace. Critically, they must do so by producing sequences of strokes (action symbols corresponding to learned shapes[1]) to construct the image, and in a manner consistent with the hierarchical action grammar $A^nB^mC^k$ (**Fig. 1a**), meaning "repeat shape A, then repeat shape B, then repeat shape C", where *A*, *B*, and *C* are abstract indices to which arbitrary shapes are associated by learning, and where the number of repetitions of each shape (n, m, and k) vary across problems. Monkeys successfully drew new images in a manner consistent with the underlying grammar, revealing a capacity for zero-shot generalization of grammatically structured action sequences.

To study the neural basis of this ability, we performed simultaneous recordings across multiple frontal cortical areas previously implicated in rule use or action sequencing, but never all directly compared in the same task. Doing so, we found a dominant role for pre-supplementary motor area (preSMA) in supporting this grammatical ability; in particular, preSMA activity dynamically encodes the latent structure of the $A^nB^mC^k$ grammar during drawing.

Results

Curriculum-based training of grammatically structured visual-motor construction

We trained monkeys to construct images by tracing a sequence of strokes, one for each shape component. On each trial, correct behavior was to produce a sequence consistent with the $A^nB^mC^k$ grammar, with specific shapes associated with each shape index (A, B, C) learned by the subject and stable throughout a multi-day experimental session (**Fig. 1a-d;** touchscreen setup in **Fig. S1;** Methods). Each shape was drawn with its associated action symbol—shape-associated stroke types that the subject had previously learned[1].

To train macaques, we developed a method involving a "curriculum" of problems (where each problem is a unique image with certain task parameters), increasing in difficulty along two dimensions: decreasing stroke-by-stroke guidance (**Fig. 1e**) and increasing image complexity (**Fig. 1f**), as depicted for an example experiment involving the grammar A^nB^m (*A* associated with lines and *B* associated with circles) (**Fig. 1b,c, e-h**). Stroke-by-stroke guidance involved one-by-one visual cueing of which component to draw (**Fig. 1e**). Over trials, we reduced this guidance until it was eliminated, at which point stroke choices were internally guided (**Fig. 1e**). Image complexity, defined as the number of total image components, started out low (two components), and increased over training (up to four in this experiment, with maximum two lines and two circles; **Fig. 1f**). Additionally, to motivate learning of generalizable grammars, a large variety of images were presented (see below).

Monkeys successfully completed the training curriculum. We quantified performance as *repeat length*, or the number of components drawn in repetition for a given shape ("performance", **Fig. 1g,** bottom), and compared this value to the *number of image components* for that shape (**Fig. 1g,** bottom), with perfect performance indicated by equality of *repeat length* to *number of components*. Over the two-day experiment, performance improved over the first day, and by the end of the second day was close to perfect even with no guidance (**Fig. 1h, i**). This was true even though problems were variable (six examples in **Fig. 1f**; quantification in **Fig. 1g**, top) and included a steady stream of novel images (**Fig. 1g**, middle).

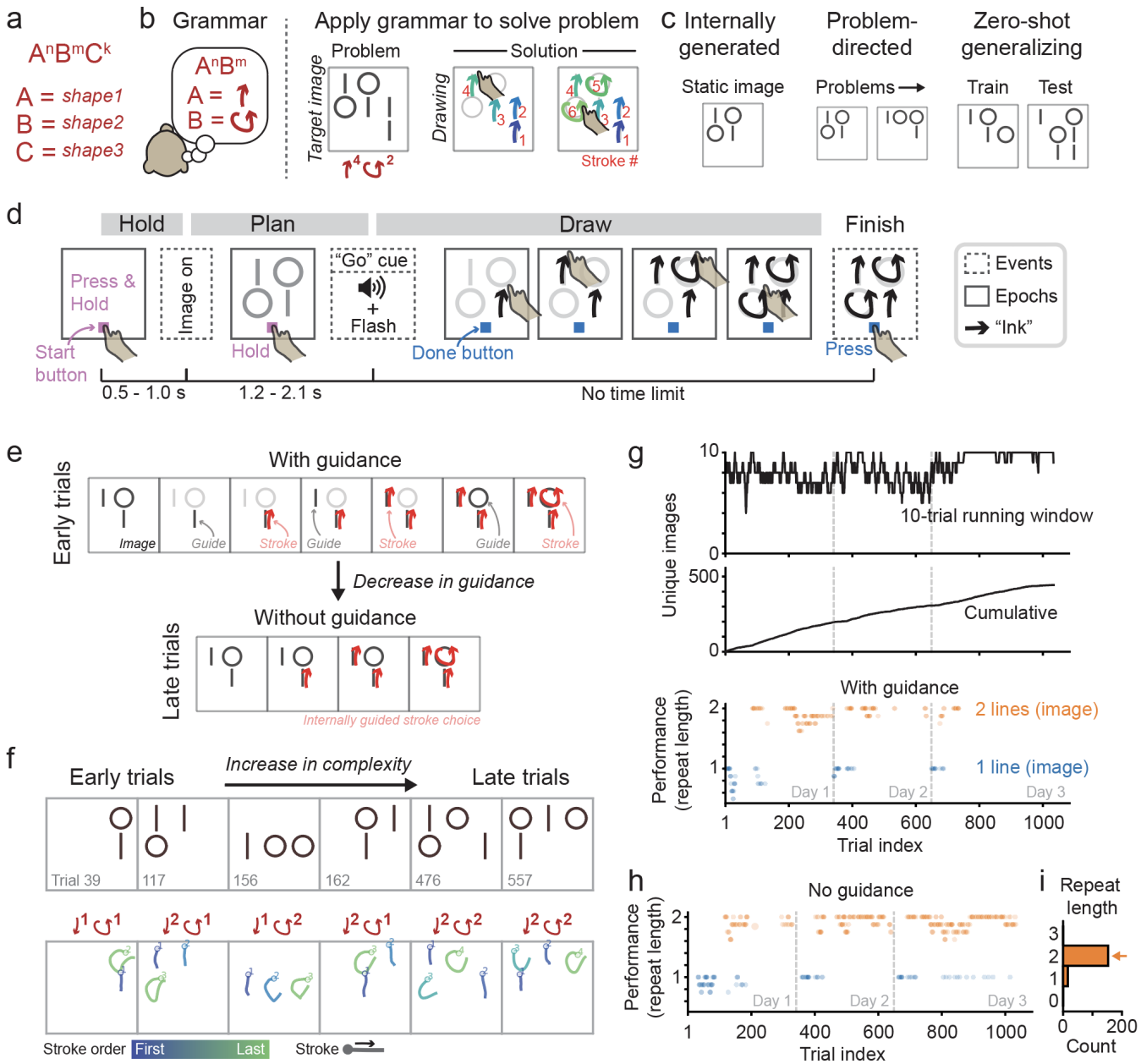

Figure 1. Curriculum-based training of grammatically structured visual-motor construction.

(a) The general form of action grammars in our study. See main text.

(b) Schematic of task paradigm to test ability to learn and generalize an action grammar towards solving new drawing-like visual-motor construct problems. The task is to produce a sequence of tracing strokes to copy the target image, while applying the trained grammar, which for this problem (image) requires repeating "line" four times and then repeating "circle" twice.

(c) Three behavioral hallmarks of grammatically structured action sequences.

(d) Trial structure, showing three discrete events (dashed boxes) and sustained epochs (solid boxes). "Hold": press and hold a "start" button, enforcing a consistent posture. "Plan": subjects see the image but must maintain hold on "start" button. "Draw": subjects produce strokes. Black arrows represent "ink" (visible to the subject) marking touched locations. Subjects can report completion any time by pressing "done" button. They receive juice reward from producing a sequence consistent with the trained grammar. The strokes are learned action symbols, each a shape-associated stroke type the subject had previously learned[1], which must also (loosely) match the form of each shape component (Methods). "Buttons" always refers to touchscreen buttons.

(e) The task curriculum involved decreasing guidance over trials. Guidance involved cuing the correct stroke one-by-one (main text). It was decreased by gradually lowering the difference in contrast between the next shape and other shapes, until this difference reached zero, thus removing guidance ("Without guidance").

(f) The task curriculum involved increasing image complexity over trials. Image complexity was defined as the number of total shape components. In this experiment (panel b), complexity increased from two components to four (maximum two circles, two lines).

(g) Performance over the three days of training for trials with guidance (bottom), also plotting the number of unique images cumulatively (middle) and in a 10-trial running window (top). Performance was quantified as the number of repeated lines (shape index *A*), plotted separately for images with one (blue) and two (orange) lines.

(h) Performance for trials without guidance for the same experiment as (g).

(i) Summary of performance for two-line images on the last day of training (day 3). Orange arrow, expected length for correct performance.

Zero-shot generalization to harder problems

To test monkeys' strategies, we assessed what sequences they produce when evaluated on "test" problems not seen during training and designed to distinguish different candidate strategies. Test problems were interleaved with training problems after curriculum completion, and used images that were more complex (maximum six components, up to five circles and five lines, for the experiment in **Fig. 1, 2**) than experienced during training (maximum four components, up to two circles and two lines). To further increase novelty, test images included components at new screen locations (**Fig. 2a**).

Generalization of the grammar A^nB^m (**Fig. 2a**) predicts that monkeys will draw test problems using the correct repeat lengths, exceeding the maximum in training (**Fig. 2a,** right, "test"). An alternative possibility is a failure mode that resorts to either sequences that are random or based on a default directional bias, predicting poor generalization (**Fig. 2b**). Another alternative possibility is that subjects produce the specific sequences A^1B^1, A^1B^2, A^2B^1, and A^2B^2, which are sufficient to solve the training problems, but not the test problems (**Fig. 2c**). For example, an image with three *A*s and three *Bs* has no solution in that set of sequences; thus the subject may solve it by chaining A^2B^2 and A^1B^1 into a sequence AABBAB, which predicts low repeat length scores (**Fig. 2c**, right).

Performance on test problems was consistent with the grammar A^nB^m. This is evident in example drawings (**Fig. 2d**) and quantification of first-trial performance (**Fig. 2f**). Critically, subjects produced longer repeats than expected under the alternative hypotheses (**Fig. 2b,c**). Generalization was (i) *internally generated*, as images had no dynamic cues to guide action order; (ii) *problem-directed*, as each image represented a specific problem to solve in that moment; and (iii) *zero-shot generalizing*, as it occurred on the first attempt for novel, including harder, images—hallmarks of grammatical behavior (**Fig. 1c**).

We found that monkeys could learn new assignments of shapes to the indices (*A*, *B*) in different experimental sessions, such that the same test image was drawn differently (in different sessions) depending on the currently trained grammar (compare **Fig. 2d** to **Fig. 2e**; and **Fig 2f**, monkey D).

We also found that fine-grained behavioral structure was consistent with the $A^nB^mC^k$ grammar. We capitalized on the possibility that the timing of actions within sequences can reflect the underlying strategy—in particular, longer duration between actions in a sequence (i.e., gaps) may reflect cognitive processes including planning[53], decision making[54], and surprise due to violation of expectations, including of grammatical structure[55]. First, we tested for hierarchical structure, in the sense that each repetition of a given shape forms a high-level unit, such that AAABBC consists of three units (AAA, BB, C) (**Fig. S2a**). Consistent with this, gaps between shapes (e.g., A to B) were longer than gaps within shapes (e.g., A to A) (**Fig. S2b**). Second, we tested for ordinal structure of shape indices, or the understanding that A comes before B, which comes before C (**Fig. S2c**). Consistent with this, gaps where an expected shape index was skipped were slowed (e.g., A to C in AAACC is slower than A to B in AAABC) (**Fig. S2d**), possibly reflecting violation of an expectation of a B after A. Third, we tested for evidence that subjects consider the entire repeat length of a shape before initializing it, possibly reflecting planning or visual search for components of that shape[53,56] (**Fig. S2e**). Consistent with this, we found that gaps before repeat of a shape were slower if that upcoming repeat was longer (e.g., A to B is slower for ABBB vs. ABBC) (**Fig. S2f-h**). Thus, the duration of inter-stroke gaps reflects grammatical structure.

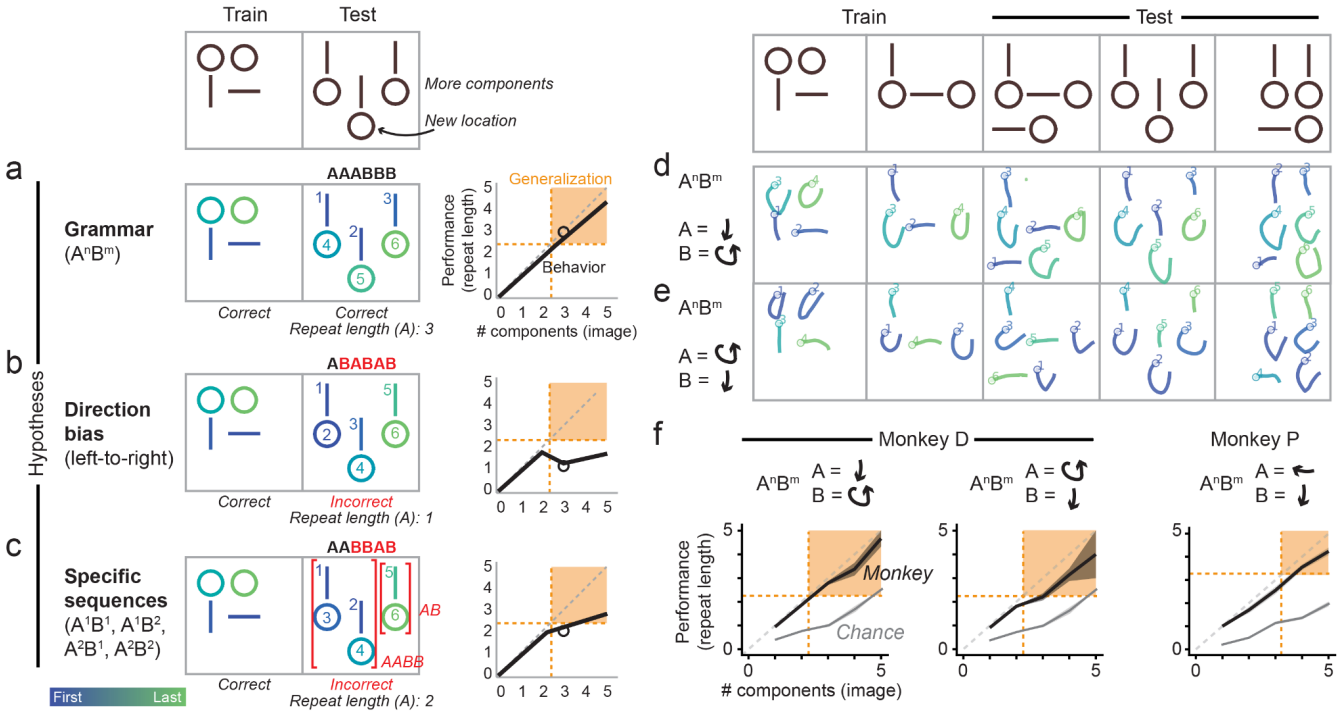

Figure 2. Zero-shot generalization to harder problems.

(a) Expected behavior from generalizing the A^nB^m grammar, producing the sequence AAABBB, resulting in a performance score (i.e., repeat length of shape *A*) of 3. Right: Expected performance vs. number of components (of *A*) in the image. Generalization should produce a function (black) with slope close to 1 (gray dashed). Open circle, the example test problem.

(b) An alternative to grammar generalization is failure that resorts to a directional bias (here, draw left to right), producing the sequence ABABAB for this example problem, resulting in a repeat length of 1 (i.e., only the first A). Right: Expected performance will produce repeats shorter than expected from generalizing A^nB^m.

(c) An alternative to grammar generalization is reuse of specific learned sequences (A^1B^1, A^1B^2, A^2B^1, and A^2B^2). For this test problem, the learned sequences AABB and AB are chained to produce AABBAB, which leads to a repeat of length 2. Right: expected performance.

(d) Monkey behavior. Shown are the first trials for three example test problems after training on A^nB^m with line (including both vertical and horizontal) assigned to *A* and circle to *B*. Note that for test problems, reward was provided for image completion regardless of whether their behavior was consistent with the grammar.

(e) Behavior for the same monkey in (e) after training on a new assignment between shapes and indices, with circle now assigned to *A* and lines to *B*. Drawings depict the first trials attempting the test problems after training.

(f) Summary of performance for monkey D for two experiments with different shape assignments (left, middle) and for monkey P (right). “Chance” is performance expected from random sequencing.

Hypothesized brain areas and neural activity correlates

What brain areas are involved? As candidates, we recorded from multiple frontal cortical areas each previously implicated in rule-based behavior or action sequencing. Prominent rule-associated areas (**Fig. 3a**, “rule use”) included dorsolateral and ventrolateral prefrontal cortex (dlPFC and vlPFC)[38,57–60], dorsal and ventral premotor cortex (PMd and PMv)[40,61–63], and frontal pole (FP)[64]. Prominent sequencing-associated areas (**Fig. 3a**, “sequencing”) included vlPFC and dlPFC[27,38,39,45,65–67], PMd[40,68], supplementary motor area (SMA)[26,69,70], and preSMA[26,28]. We also considered PMv as a candidate, because it is involved in this drawing task, in encoding action symbols[1]. We also recorded from primary motor cortex (M1), an area linked to motor control and not strongly to rules or sequencing[24,39,71], and which thus acts as a “negative control”. We recorded from these areas using chronically implanted multi-electrode arrays (sixteen 32-channel arrays, **Fig. 3b**, **Fig. S3**).

Recording from these areas together allowed us to, for the first time, directly compare them in the same task. To test involvement in grammatical sequencing, we assessed encoding of three structural properties associated with the $A^nB^mC^k$ grammar: *shape index*, *abstract role*, and *repetition ordinal position* (schematized and explained in **Fig. 3c-e**). The areas encoding these structural properties could in principle be those previously implicated in rules (**Fig. 3a**, “Rule use”) or those implicated in sequencing (**Fig. 3a**, “Motor sequencing”). Alternatively, none of the recorded areas may be involved (**Fig. 3a**, “None”), a single area may be preferentially involved (**Fig. 3a**, “Preferential”), or all areas may be equally involved (**Fig. 3a**, “Distributed”).

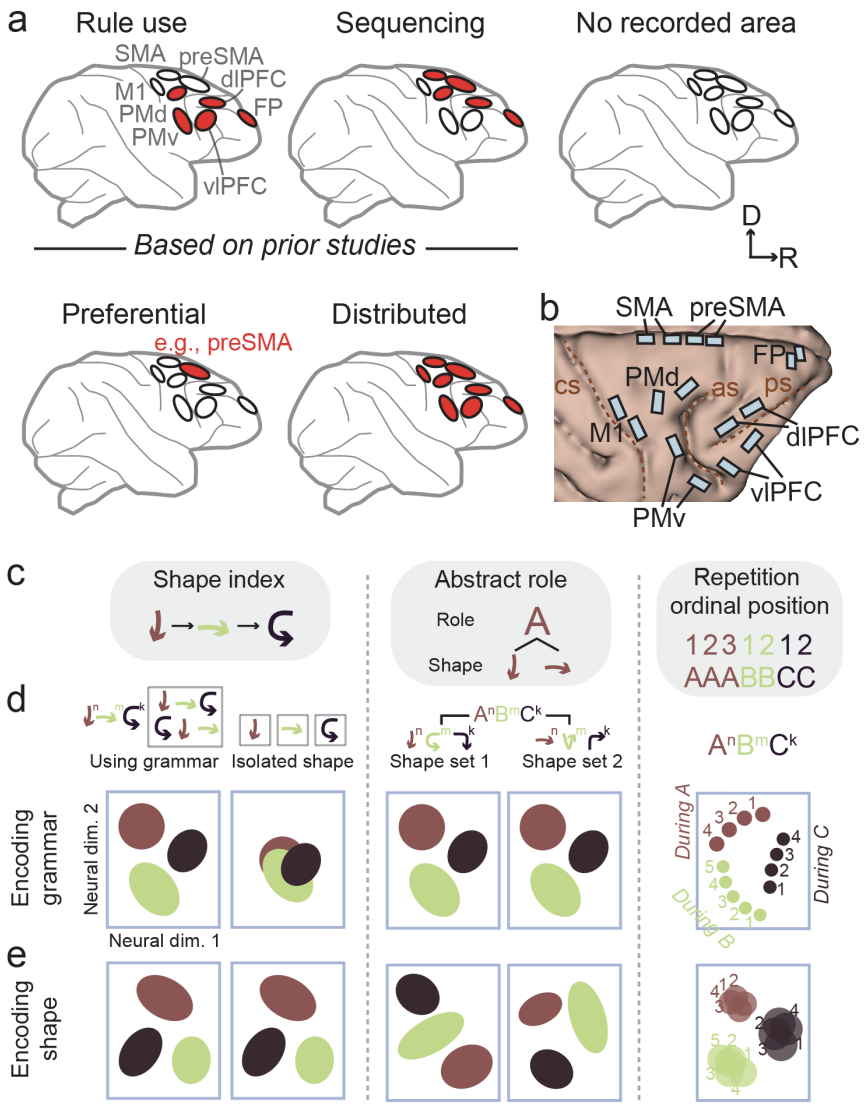

Figure 3. Hypothesized brain areas and neural activity correlates.

(a) Five hypotheses for sets of areas encoding grammatical structure, shown on the lateral surface of frontal cortex: (1) Areas previously implicated in use of abstract rules ("rule use"); (2) Areas previously implicated in action sequencing ("sequencing"); (3) None of the areas we are recording from ("no recorded area"); (4) A single area exhibiting stronger encoding of grammatical structure compared to other areas ("Preferential"); (5) Many, possibly all, areas exhibiting similarly strong encoding of grammatical structure ("Distributed"). Note that preSMA and SMA are depicted here on the lateral surface but are actually on the medial wall. R, rostral. D, dorsal.

(b) Recording sites, targeted to right frontal cortex, contralateral to the drawing hand, showing floating microelectrode arrays (32-channel, Microprobes) to scale on 3D rendering of brain (monkey P). SMA and preSMA electrodes were angled to target the medial wall (**Fig. S3**); cs, central sulcus; as, arcuate sulcus; ps, principal sulcus.

(c) Three latent properties of action sequences consistent with a representation of grammatical structure, tested for in each recorded area. See main text.

(d) Predicted neural activity encoding grammatical structure, depicted as population activity projected onto a two-dimensional subspace during drawing of index *A* (maroon ellipse), index *B* (green) and index *C* (black). "Shape index" encoding predicts separation of activity for each shape when used in a grammar ("Using grammar") that diminishes when the same shapes are used in a non-grammatical task drawing single shapes ("Isolated

shape"). "Abstract role" encoding predicts similar shape index encoding across two shape sets which assign different shapes to the same abstract roles (A, B, C). "Repetition ordinal position" predicts separation of activity during repetition of a given shape, possibly encoding ordinal position within the repeat (1, 2, 3, …).

(e) Predictions if activity encodes the shape or shape-associated action symbol rather than grammatical structure.

Encoding of *shape index*

Behavioral success (**Fig. 2**; and inter-stroke gap durations in **Fig. S2a-d**) implies an internal representation of the indices *A*, *B*, and *C* (**Fig. 3c**, "shape index"). Activity consistent with such a representation should exhibit distinct states when drawing *A*, *B*, or *C* ("shape encoding", **Fig. 3d**, "Using grammar"), and this effect should diminish for tasks involving drawing of a single shape component (**Fig. 3d**, "Isolated shape"), a control task with the visual and motor properties associated with each shape, but lacking grammatical structure.

Out of all the recorded areas, preSMA was the only one to strongly exhibit both shape encoding and its reduction in single-stroke tasks (**Fig. 4**). For the example preSMA unit (**Fig. 4a**) stroke-aligned activity strongly differs by the shape of that stroke (left panel), and this difference is greatly reduced when drawing these same shapes as isolated strokes (right panel). As a contrast, a unit in PMv (**Fig. 4b**) encodes the shape but does so similarly whether the stroke is in a grammatical sequence (left panel) or isolated (right panel), consistent with encoding shape-specific action symbols[1].

These encoding properties were evident at the population level. Two-dimensional linear projections of preSMA population activity exhibited shape-encoding activity states during grammatical sequencing (**Fig. 4c**, "Using grammar"), which diminished for single-stroke tasks (**Fig. 4c**, "Isolated shape"). In contrast, PMv activity encoded shapes in both tasks (**Fig. 4d**).

To quantify encoding of shape index, we devised a "shape encoding" metric, which reflects activity dissimilarity between sets of trials representing different shapes. It is the average pairwise Euclidean distance between all across-shape trials, subtracting the average within-shape Euclidean distance, normalized between 0 (no difference) and 1 (a ceiling, defined as the 98th percentile of pairwise Euclidean distances). Shape encoding was computed controlling for multiple potentially confounding behavioral factors, including the stroke's spatial location and the direction of movement preceding the stroke (Methods).

In preSMA, shape encoding was high when using the grammar (**Fig. 4e**, x-axis), and low when drawing single strokes (**Fig. 4e**, y-axis). In contrast to preSMA, other areas had relatively high shape encoding in both conditions (premotor and motor areas, **Fig. 4e**), or low encoding in both conditions (prefrontal and frontal polar areas, **Fig. 4e**). Thus, encoding of shape index is associated with activity in preSMA.

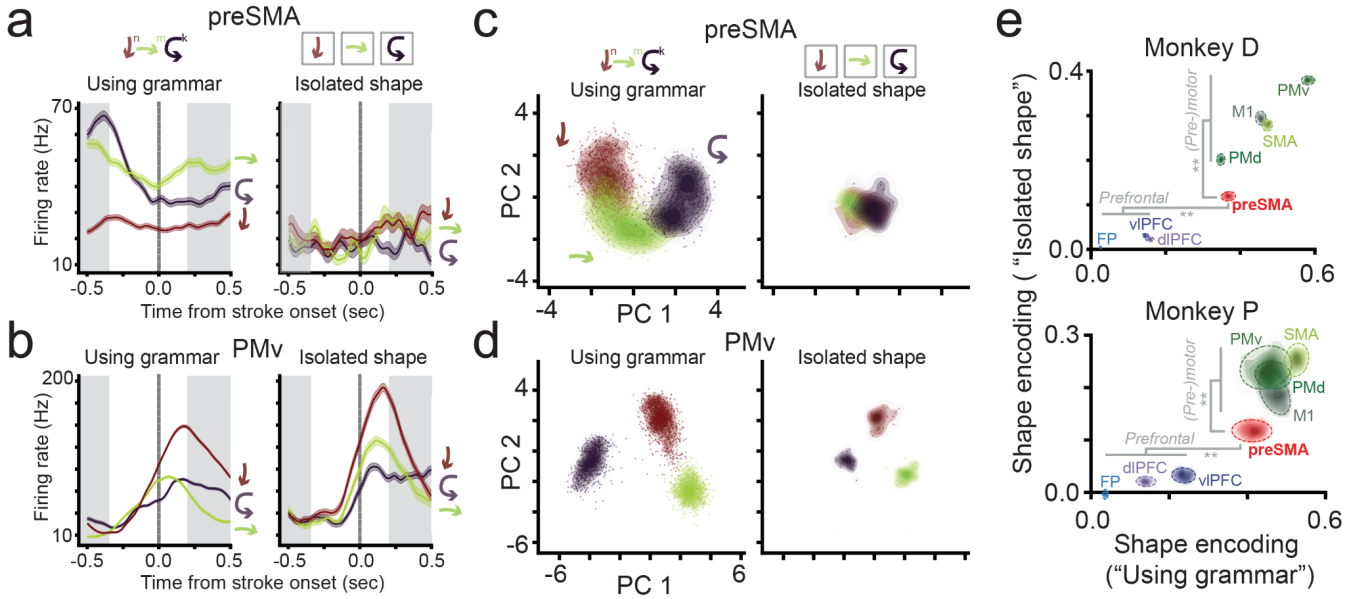

Figure 4. Encoding of *shape index*.

(a) An example unit in preSMA, showing mean activity for each shape during either repetition in the context of a grammar ("Using grammar") or as a single stroke ("Isolated shape"). White background indicates the time window aligned to stroke onset used for subsequent analysis (-0.35 to 0.2 sec).

(b) An example unit in PMv, recorded concurrently and plotted analogously to (a).

(c) Population activity for preSMA, projected to two dimensions using PCA, plotting for each shape a kernel density estimate (KDE; filled contours) and individual stroke datapoints, in the time window -0.35 to 0.2 sec relative to stroke onset.

(d) Population activity for PMv, plotted analogously to (c).

(e) Summary of pairwise shape encoding (see main text and Methods) while using the grammar (x-axis) and while drawing isolated strokes (y-axis). **, $p < 0.001$ for preSMA vs. each motor and premotor area in shape encoding (isolated shape); **, $p < 0.001$ for preSMA vs. each prefrontal area in shape encoding (using grammar). Statistical tests were performed using a linear model that controlled for behavioral variables that may influence shape encoding, and remain significant after Bonferroni correction for multiple comparisons (Methods).

Encoding of *abstract role*

A grammar's usefulness lies in its capacity to support generalization. One form of generalization occurs when a grammar defines abstract "roles" that can be bound to different concrete objects or actions, as in dynamic role–filler binding[72,73] and algebraic sequences[5,9,10] (**Fig. 3c**, "abstract role"). To test whether activity reflects abstract roles—or instead, reflects concrete shapes—we trained macaques on tasks involving the same $A^nB^mC^k$ grammar but using one of two different shape sets on randomly interleaved trials. This predicts that activity encoding shape index will be similar between the two shape sets (**Fig. 3d**), whereas encoding of concrete shape predicts they will be different (**Fig. 3e**).

We found evidence for encoding of abstract role, and it was strongest in preSMA. The two example preSMA units (**Fig. 5a, b**) encode shape index (colored red, green, black) similarly for shape set 1 (left panel) and 2 (right panel). In contrast, the two example PMv units (**Fig. 5c, d**) encode specific shapes. These effects are evident in population activity for preSMA (**Fig. 5e, f**) and PMv (**Fig. 5g, h**).

Consistent with encoding abstract role, preSMA exhibited relatively high shape encoding within each shape set (high on the x-axis in **Fig. 5i**) and low shape encoding across shape sets and within the same index (low on the y-axis in **Fig. 5i**). In contrast, for the other areas these two values were more correlated, such that shape encoding in motor and premotor areas was high in general (towards the top-right of **Fig. 5i**) and in the prefrontal areas was low in general (towards the bottom-left of **Fig. 5i**). Thus encoding of abstract role was strongest in preSMA.

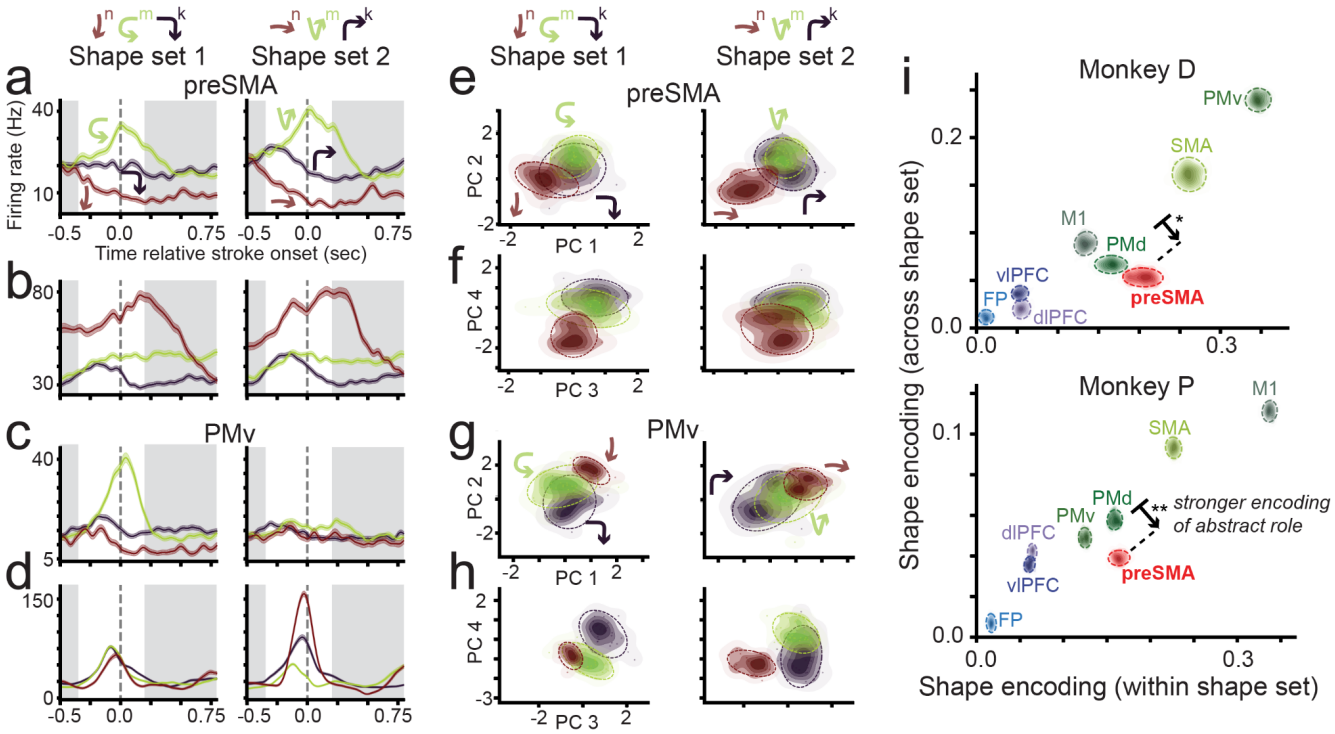

Figure 5. Encoding of *abstract role.*

(a) An example unit in preSMA, showing mean activity for each shape during execution of the $A^nB^mC^k$ grammar in a session switching trial-by-trial between two shape sets (left and right) that each assign a different shape to indices *A* (maroon), *B* (green), and *C* (black). Across trials there are six shapes, three each in sets 1 and 2. White background indicates the time window used for subsequent analysis (-0.35 to 0.2 sec).

(b) Another example unit in preSMA, recorded concurrently and plotted analogously to (a).

(c) An example unit in PMv, recorded concurrently and plotted analogously to (a).

(d) Another example unit in PMv, recorded concurrently and plotted analogously to (a).

(e) Population activity for preSMA, projected to PCA components 1 and 2, plotting a KDE and individual datapoints for each shape across two shape sets, using the time window of -0.35 to 0.2 sec relative to stroke onset.

(f) The same data as in (e) but projected to PCA components 3 and 4.

(g) Population activity for PMv, recorded concurrently and plotted analogously to (e).

(h) The same data as in (g) but projected to PCA components 3 and 4.

(i) Summary of pairwise shape encoding, either computed across shapes within each shape set and averaged across sets (x-axis) or across shape sets within each index and averaged across indices (y-axis). Activity towards the bottom-right of the plot indicates stronger encoding of abstract role. Statistics comparing brain areas using the ratio of shape encoding (across shape sets) to shape encoding (within shape set); monkey D: *, $p < 0.01$ for preSMA vs. PMd; $p < 0.005$ for preSMA vs. each other area (except FP, vlPFC, and dlPFC, which had low shape encoding and thus no well-defined ratio); monkey P: **, $p < 0.005$ for preSMA vs. each other area (except FP, which had low shape encoding and thus no well-defined ratio).

Encoding of *repetition ordinal position*

Do subjects internally represent progression, or ordinal position, within a shape repeat (**Fig. 3d**) and not just its shape index (**Fig. 3e**)? Consistent with the former, preSMA activity differed based on ordinal position, as in the example unit (**Fig. 6a**, **b**). In contrast, activity of the example PMv unit (**Fig. 6c**) did not strongly differ by ordinal position, and this was true across shape indices and repeat lengths (**Fig. 6d**). Population activity also differed strongly by ordinal position in preSMA (**Fig. 6e**) and not in PMv (**Fig. 6f**). We quantified ordinal encoding as the mean pairwise neural distance between all ordinal positions, computed separately for each shape, and found that it was strongest in preSMA (**Fig. 6g, h**).

Ordinal encoding could reflect position relative to repeat start ("number of items completed") or relative to repeat end ("number of items remaining"). To distinguish these possibilities we compared repeats of different length (**Fig. 6i**, top), which allows dissociating ordinal position relative to start (termed ord_{FWD}) from ordinal position relative to end (ord_{REV}). For example (**Fig. 6i**, bottom), if activity during production of the second stroke in a length-3 sequence (X^3) is similar to activity for the second stroke in a length-4 sequence (X^4), measured as low pairwise neural distance, this indicates encoding of ord_{FWD}. In contrast, if activity is more similar to the third stroke of X^4, this indicates encoding of ord_{REV}. Taking this approach, we found that activity encoded ord_{FWD} and ord_{REV} to different extents depending on ordinal position. Activity at the start of a repeat (i.e., the first and second strokes) more strongly encoded ord_{FWD} (**Fig. 6j**, top-left), while activity at the end encoded ord_{REV} (**Fig. 6j,** bottom-right). The change from encoding ord_{FWD} to ord_{REV} occurred around when three strokes remained in the repeat (**Fig. 6k**).

Activity more strongly encoding ord_{REV} (relative to ord_{FWD}) implies an internal representation of the number of remaining repetitions, consistent with maintenance of a (at least partial) plan for the upcoming repetitions. Having a plan would minimize the need for further decision making before each stroke, predicting faster gap durations between strokes. Consistent with this, sequences with stronger encoding of ord_{REV} also had shorter average gap duration (**Fig. 6l**).

Thus, preSMA encodes ordinal position within shape repeats, changing from encoding ord_{FWD} at sequence start to ord_{REV} at sequence end. Comparison of this neural property to speed of action indicates that stronger encoding of ord_{REV} may be a neural signature of a plan for the upcoming actions within a repeat.

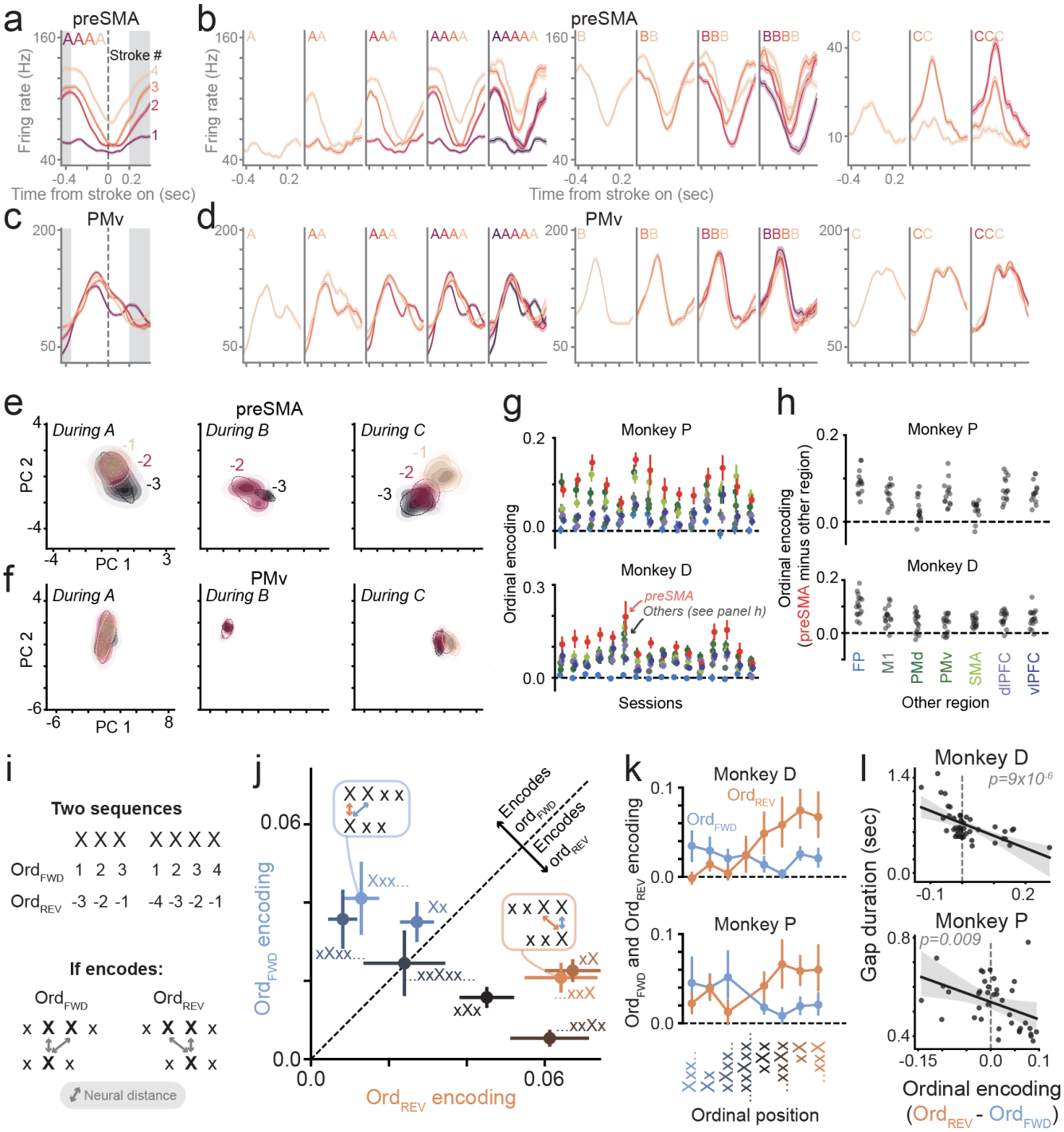

Figure 6. Encoding of *repetition ordinal position.*

(a) An example unit in preSMA, showing mean activity for each ordinal position (numbered, counting up from repeat onset) during repeat of *A* in length-4 *A* sequences using the $A^nB^mC^k$ grammar. White background indicates the time window used for subsequent analysis (-0.35 to 0.2 sec).

(b) Example unit activity in preSMA. Each subpanel is a distinct dataset during a specific length repeat of a specific shape. Mean activity is shown for each ordinal position. The same unit from (a) is plotted for *A* and *B* repeats and a different unit for *C* repeats.

(c) An example unit in PMv, recorded concurrently and plotted analogously to (a).

(d) Example unit activity in PMv recorded concurrently and plotted analogously to (b). This is the same unit from (c).

(e) Population activity for preSMA colored by ordinal position during repeats of either *A* (left), *B* (middle), or *C* (right), plotted in the same projection, using PCA. Shown are only data for strokes drawn at a single specific screen location, in order to control for movement-related confounds associated with location.

(f) Population activity for PMv, recorded concurrently and plotted analogously to (e).

(g) Summary of ordinal encoding, one mean value for each brain area (colors) in each session. Colors represent individual brain areas using the same mapping as (h).

(h) Summary of ordinal encoding, showing the difference between mean encoding in preSMA vs. each other area; n = 14 sessions (monkey P), 16 (D). Statistics: $p < 0.001$, for each preSMA vs. other area comparison, for both monkeys P and D.

(i) Schematic (top) of dissociation between ordinal position from sequence start (ord_{FWD}) and ordinal position from sequence end (ord_{REV}) by comparing sequences of different lengths (length 3 vs 4). Bottom: Two predictions for neural activity differences between these two sequences depending on whether activity encodes ord_{FWD} or ord_{REV}.

(j) Summary of pairwise mean +/- SEM ord_{FWD} encoding (y-axis) vs. ord_{REV} encoding (x-axis), computed separately for strokes differing by ordinal position (each point), with position indicated by color, ranging from blue (start of sequence) to orange (end of sequence). Ord_{REV} encoding is the mean pairwise neural distance across different ord_{REV} positions (keeping ord_{FWD} the same), while ord_{FWD} encoding is the neural distance across different ord_{FWD} positions (keeping ord_{REV} the same). Insets illustrate the procedure: each ordinal position for a given length sequence (e.g., the first X in XXX) was paired with a different length repeat for the same shape (e.g., XXXX) to contribute two values, one for ord_{FWD} encoding (blue) and the other ord_{REV} encoding (orange). This combines data for monkeys D and P.

(k) Summary of mean ord_{FWD} encoding (blue) and ord_{REV} encoding (orange) as a function of ordinal position.

(l) Relationship between gap duration (mean within sequence) and difference in ordinal encoding (ord_{REV} - ord_{FWD}), plotting results from ordinary least squares regression fit (95% bootstrapped CI). Each datapoint represents a single sequence (e.g., AAA), with gap duration averaged over gaps and ordinal encoding averaged over ordinal positions.

Discussion

Trained to solve visual-motor construction problems requiring action sequencing with grammatical structure of the form $A^nB^mC^k$ (**Fig. 1**), macaques exhibited zero-shot generalization to solve novel, harder problems (**Fig. 2**). Comparing multiple motor, premotor, and prefrontal areas implicated in rule use and sequencing (**Fig. 3**), we found that preSMA exhibited the strongest encoding of each of three properties of $A^nB^mC^k$ structure: shape index (**Fig. 4**), abstract role (**Fig. 5**), and repetition ordinal position (**Fig. 6**). During drawing, preSMA population activity exhibits dynamics in which it occupies different activity states for different shape indices (**Fig. 4, 5**), and within each shape, activity encodes behavioral progression within its repeat (**Fig. 6**), initially encoding progression relative to the onset of the repeat, and then, with around three strokes left, changing to encoding the number of strokes remaining in the repeat (**Fig. 6j, k**). Overall, this study establishes a macaque paradigm for studying the neuronal basis of grammatically structured problem-solving behavior and, using this paradigm, provides insight into that basis by finding a dominant role for preSMA in encoding latent structure in this action grammar.

Insight into internal procedural rules underlying grammatical ability

Behavioral success demonstrates that the subjects understood the grammatical structure "draw all A, then draw all B, then draw all C." However, this does not reveal what underlying step-by-step procedural rules implement the behavior consistent with this grammar. In general, it is difficult to infer internal algorithms from behavior alone[74,75]. By establishing a macaque model, we were able to perform neural recordings that allow more direct observation of internal algorithms.

The grammatical structure represented in preSMA activity reveals aspects of internal procedures concerning both the sequencing of shape indices (A, B, C) and the sequencing of multiple repetitions within each shape index.

At the level of shape indices, multiple lines of evidence indicate that subjects understand shape indices and their ordinal structure (A then B then C). First, subjects correctly sequence indices even on test trials (**Fig. 2**). Second, preSMA activity encodes the identity of the currently repeated index (**Figs. 4, 5**). Third, behavioral timing shows longer inter-stroke gaps at transitions from one index to the next (e.g., A to B) compared to transitions within an index (e.g., A to A) (**Fig. S2a, b**), with even longer gaps when an expected intermediate index is omitted (e.g., A to C without B; **Fig. S2c, d**).

Notably, the finding of activity in preSMA encoding abstract roles (**Fig. 5**) implies that indices are represented at a level of abstraction that generalizes beyond specific shapes. Such an abstraction is thought to characterize some kinds of abstract grammars, supporting generalization across surface motor and sensory features—as in dynamic role–filler binding[72,73] and algebraic sequences[5,9,10]. Finding neural

evidence for abstract role encoding in non-human primates informs ongoing debates on the extent to which non-human animals represent such structure[5,6,8,9,17,76].

Within each shape index, what procedures govern the ability to draw all instances? One possibility is a procedure that relies on the persistence of the target image: “search for an instance of A, draw it, then search for another A, and continue until no unmatched instances remain.” This procedure resembles a while-loop in computer programming (“while A exists, draw A”). An alternative possibility is based on internally representing the target number of repetitions and comparing it to an internal representation of the number of completed actions: “count how many instances of A exist, then draw A that number of times while tracking progress.” This strategy resembles a for-loop program.

Although not definitive, our neural results provide evidence in favor of the latter “for loop” interpretation. We found that preSMA activity encodes the ordinal position of each action within a repeat (**Fig. 6**). Such internal tracking of ordinal position is unnecessary for a while-loop, but is a core requirement of a for-loop procedure. Additional behavioral evidence supports this interpretation: inter-stroke gaps at the start of a repeat are longer when the upcoming repeat contains more items (**Fig. S2e-h**), consistent with counting items or planning the forthcoming sequence.

Thus, a major insight from our approach combining grammatically structured behavior with simultaneous recordings is that it allows direct observation of neural activity reflecting otherwise latent procedures that would be difficult to infer from behavior alone. Recorded activity indicates hierarchically organized procedures; at the higher level, a procedure resembling “draw all A, then all B, then all C”; and at the lower level, evidence suggests a for-loop–like procedure for producing repetitions of each shape. Although further experiments will be needed to definitively characterize them, the present findings provide an initial window into the internal algorithms, and associated neural substrates, underlying grammatical action sequencing.

Reinterpretation of latent structure in action sequencing

Our findings relate to a growing body of work that finds evidence that the geometry of neural population activity in frontal cortex encodes the latent structure of action sequences, including in rodents[42,46] and primates[38,65]. Such activity has been proposed to represent abstract schemas or cognitive maps[42,46,65], internal models of behavioral, task-related or environmental structure which generalize across related tasks and may support planning and inference[42,46,77], working memory[38,65] and learning[42], among other functions.

Our results are compatible with aspects of these interpretations, but also suggest an additional possibility. By directly linking neural activity to a generative capacity in behavior, our results raise the possibility that these prior findings reflect, at least in part, implementation of rule-based procedures (a possibility also discussed in ref[38]). In this view, such activity not only represents a space of latent states in the world, but also states contributing to internal program-like algorithms[4,6,7,17,78,79], mirroring the distinction proposed between relational cognitive maps[80,81] and symbolic "language of thought" procedures[8,17,18,82,83].

This interpretation aligns with classic neuropsychological observations that frontal lobe damage selectively impairs the ability to internally organize complex actions in open-ended problems, despite preserved basic cognitive, motor and perceptual functions[84].

A central role for preSMA in grammatical sequencing

Prior studies of have identified various neural correlates of sequential structure in multiple brain areas, but have typically recorded from just a single, or a few, areas (see Introduction and **Fig. 3**). A major advance of our study is thus recording from many candidate areas together, allowing comparison of their relative contributions, and doing so while systematically assessing grammatically structured behavior. We found that encoding of grammatical structure was surprisingly weak in lateral prefrontal cortex (dlPFC and vlPFC), a region previously implicated in sequencing[27,38,39,65,67,88] and use of abstract rules[38,57,58,60]. Instead, grammatical structure was preferentially encoded in preSMA.

The representations we report in preSMA are consistent with the diverse neural correlates that have been identified in that region[28,52,71]. Prominent functions attributed to preSMA include involvement in performing voluntary and cognitively demanding actions[85], suppressing automatic actions for voluntary ones[86], representing temporal structure of action[87], learning new sequences[26], and planning and executing sequences[28]. Our results suggest that some of these prior neural observations may reflect involvement of preSMA in internal rule-based procedures for generating action sequences.

A frontal cortical system supporting compositional generalization

Together with our prior work identifying discrete action symbols in the current task[1], we have identified neural substrates of two key components of a compositional system: the elemental action units (action symbols represented in ventral premotor cortex, or PMv), and the procedures that recombine them (action grammars involving preSMA). PMv and preSMA are directly interconnected[89], and they are part of the putative macaque homolog of the multiple-demand network[90], a set of areas coactivated during abstract reasoning. The latter observation raises the possibility that preSMA and PMv may also contribute to problem solving in abstract non-motor tasks. Future studies may clarify how preSMA, PMv, and

connected areas implement algorithmic procedures in grammar-based recombination of action symbols, thus supporting compositional generalization.

References

1. Tian, L. Y. *et al.* Neural representation of action symbols in primate frontal cortex. *bioRxiv* 2025.03.03.641276 (2025) doi:10.1101/2025.03.03.641276.
2. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
3. Lake, B. M. & Baroni, M. Human-like systematic generalization through a meta-learning neural network. *Nature* **623**, 115–121 (2023).
4. Frankland, S. M. & Greene, J. D. Concepts and Compositionality: In Search of the Brain's Language of Thought. *Annu. Rev. Psychol.* **71**, 273–303 (2020).
5. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* **88**, 2–19 (2015).
6. Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S. & Sablé-Meyer, M. Symbols and mental programs: a hypothesis about human singularity. *Trends Cogn. Sci.* **26**, 751–766 (2022).
7. Ellis, K. *et al.* DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **381**, 20220050 (2023).
8. Quilty-Dunn, J., Porot, N. & Mandelbaum, E. The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behav. Brain Sci.* **46**, e261 (2023).
9. Ferrigno, S. Sequencing, Artificial Grammar, and Recursion in Primates. in *Primate Cognitive Studies* (eds Schwartz, B. L. & Beran, M. J.) 260–290 (Cambridge University Press, Cambridge, 2022). doi:10.1017/9781108955836.011.
10. Marcus, G. F. *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. (MIT Press, 2003).
11. Dehaene, S., Sablé-Meyer, M. & Ciccione, L. Origins of numbers: a shared language-of-thought for arithmetic and geometry? *Trends Cogn. Sci.* **29**, 526–540 (2025).
12. Tian, L., Ellis, K., Kryven, M. & Tenenbaum, J. Learning abstract structure for drawing by efficient motor program induction. in *Advances in Neural Information Processing Systems* vol. 33 2686–2697 (Curran Associates, Inc., 2020).
13. Hayashi, M. Perspectives on object manipulation and action grammar for percussive actions in primates. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140350 (2015).
14. Lake, B. M., Salakhutdinov, R. & Tenenbaum, J. B. Human-level concept learning through probabilistic program induction. *Science* **350**, 1332–1338 (2015).
15. Karmiloff-Smith, A. Constraints on representational change: Evidence from children's drawing. *Cognition* **34**, 57–83 (1990).
16. Zuidema, W. & De Boer, B. The evolution of combinatorial structure in language. *Curr. Opin. Behav. Sci.* **21**, 138–144 (2018).
17. Kazanina, N. & Poeppel, D. The neural ingredients for a language of thought are available. *Trends Cogn. Sci.* **27**, 996–1007 (2023).
18. Gallistel, C. R. Prelinguistic Thought. *Lang. Learn. Dev.* **7**, 253–262 (2011).
19. Cheney, D. L. & Seyfarth, R. M. *Baboon Metaphysics: The Evolution of a Social Mind*. (University of Chicago Press, Chicago, IL, 2008).
20. Herman, L. M., Uyeyama, R. K. & Pack, A. A. Bottlenose dolphins understand relationships between concepts. *Behav. Brain Sci.* **31**, 139–140 (2008).
21. Stout, D., Chaminade, T., Apel, J., Shafti, A. & Faisal, A. A. The measurement, evolution, and neural representation of action grammars of human behavior. *Sci. Rep.* **11**, 13720 (2021).
22. Wang, L., Uhrig, L., Jarraya, B. & Dehaene, S. Representation of Numerical and Sequential Patterns in Macaque and Human Brains. *Curr. Biol.* **25**, 1966–1974 (2015).

23. Wilson, B. *et al.* Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat. Commun.* **6**, 8901 (2015).
24. Zimnik, A. J. & Churchland, M. M. Independent generation of sequence elements by motor cortex. *Nat. Neurosci.* **24**, 412–424 (2021).
25. Mizes, K. G. C., Lindsey, J., Escola, G. S. & Ölveczky, B. P. Dissociating the contributions of sensorimotor striatum to automatic and visually guided motor sequences. *Nat. Neurosci.* **26**, 1791–1804 (2023).
26. Nakamura, K., Sakai, K. & Hikosaka, O. Effects of Local Inactivation of Monkey Medial Frontal Cortex in Learning of Sequential Procedures. *J. Neurophysiol.* **82**, 1063–1068 (1999).
27. Shima, K., Isoda, M., Mushiake, H. & Tanji, J. Categorization of behavioural sequences in the prefrontal cortex. *Nature* **445**, 315–318 (2007).
28. Tanji, J. Sequential Organization of Multiple Movements: Involvement of Cortical Motor Areas. *Annu. Rev. Neurosci.* **24**, 631–651 (2001).
29. Averbeck, B. B., Chafee, M. V., Crowe, D. A. & Georgopoulos, A. P. Neural activity in prefrontal cortex during copying geometrical shapes: I. Single cells encode shape, sequence, and metric parameters. *Exp. Brain Res.* **150**, 127–141 (2003).
30. Averbeck, B. B., Sohn, J.-W. & Lee, D. Activity in prefrontal cortex during dynamic selection of action sequences. *Nat. Neurosci.* **9**, 276–282 (2006).
31. Markowitz, J. E. *et al.* The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell* **174**, 44-58.e17 (2018).
32. Cohen, Y. *et al.* Hidden neural states underlie canary song syntax. *Nature* **582**, 539–544 (2020).
33. Chiang, F.-K., Wallis, J. D. & Rich, E. L. Cognitive strategies shift information from single neurons to populations in prefrontal cortex. *Neuron* **110**, 709-721.e4 (2022).
34. Terrace, H. S. The simultaneous chain: a new approach to serial learning. *Trends Cogn. Sci.* **9**, 202–210 (2005).
35. Avdagic, E., Jensen, G., Altschul, D. & Terrace, H. S. Rapid cognitive flexibility of rhesus macaques performing psychophysical task-switching. *Anim. Cogn.* **17**, 619–631 (2014).
36. Brannon, E. M. & Terrace, H. S. Ordering of the Numerosities 1 to 9 by Monkeys. *Science* **282**, 746–749 (1998).
37. Liao, D. A., Brecht, K. F., Veit, L. & Nieder, A. Crows “count” the number of self-generated vocalizations. *Science* **384**, 874–877 (2024).
38. Tian, Z. *et al.* Mental programming of spatial sequences in working memory in the macaque frontal cortex. *Science* **385**, eadp6091 (2024).
39. Mushiake, H., Saito, N., Sakamoto, K., Itoyama, Y. & Tanji, J. Activity in the Lateral Prefrontal Cortex Reflects Multiple Steps of Future Events in Action Plans. *Neuron* **50**, 631–641 (2006).
40. Ohbayashi, M., Ohki, K. & Miyashita, Y. Conversion of Working Memory to Motor Sequence in the Monkey Premotor Cortex. *Science* **301**, 233–236 (2003).
41. Berdyyeva, T. K. & Olson, C. R. Rank Signals in Four Areas of Macaque Frontal Cortex During Selection of Actions and Objects in Serial Order. *J. Neurophysiol.* **104**, 141–159 (2010).
42. Manakov, M. *et al.* Cognitive Graphs of Latent Structure in Rostral Anterior Cingulate Cortex. 2025.04.18.649507 Preprint at https://doi.org/10.1101/2025.04.18.649507 (2025).
43. Sawamura, H., Shima, K. & Tanji, J. Numerical representation for action in the parietal cortex of the monkey. *Nature* **415**, 918–922 (2002).
44. Geddes, C. E., Li, H. & Jin, X. Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. *Cell* **174**, 32-43.e15 (2018).
45. Ninokura, Y., Mushiake, H. & Tanji, J. Integration of Temporal Order and Object Information in the Monkey Lateral Prefrontal Cortex. *J. Neurophysiol.* **91**, 555–560 (2004).

46. El-Gaby, M. *et al.* A cellular basis for mapping behavioural structure. *Nature* 1–10 (2024) doi:10.1038/s41586-024-08145-x.
47. Jiang, X. *et al.* Production of Supra-regular Spatial Sequences by Macaque Monkeys. *Curr. Biol.* **28**, 1851-1859.e4 (2018).
48. Ferrigno, S., Cheyette, S. J., Piantadosi, S. T. & Cantlon, J. F. Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Sci. Adv.* **6**, eaaz1002 (2020).
49. Liao, D. A., Brecht, K. F., Johnston, M. & Nieder, A. Recursive sequence generation in crows. *Sci. Adv.* **8**, eabq3356 (2022).
50. Byrne, R. W. & Russon, A. E. Learning by imitation: A hierarchical approach. *Behav. Brain Sci.* **21**, 667–684 (1998).
51. Howard-Spink, E. *et al.* Nonadjacent dependencies and sequential structure of chimpanzee action during a natural tool-use task. *PeerJ* **12**, e18484 (2024).
52. Conen, K. E. & Desrochers, T. M. The Neural Basis of Behavioral Sequences in Cortical and Subcortical Circuits. in *Oxford Research Encyclopedia of Neuroscience* (Oxford University Press, 2022). doi:10.1093/acrefore/9780190264086.013.421.
53. Sternberg, S., Monsell, S., Knoll, R. L. & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. in *Information processing in motor control and learning* 117–152 (Elsevier, 1978).
54. Wong, K.-F. & Wang, X.-J. A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *J. Neurosci.* **26**, 1314–1328 (2006).
55. Malassis, R., Dehaene, S. & Fagot, J. Baboons (Papio papio) Process a Context-Free but Not a Context-Sensitive Grammar. *Sci. Rep.* **10**, 7381 (2020).
56. van Opheusden, B. *et al.* Expertise increases planning depth in human gameplay. *Nature* **618**, 1000–1005 (2023).
57. Freedman, D. J., Riesenhuber, M., Poggio, T. & Miller, E. K. Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science* **291**, 312–316 (2001).
58. Wallis, J. D., Anderson, K. C. & Miller, E. K. Single neurons in prefrontal cortex encode abstract rules. *Nature* **411**, 953–956 (2001).
59. Mansouri, F. A., Freedman, D. J. & Buckley, M. J. Emergence of abstract rules in the primate brain. *Nat. Rev. Neurosci.* **21**, 595–610 (2020).
60. Bernardi, S. *et al.* The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* **183**, 954-967.e21 (2020).
61. Romo, R., Hernández, A. & Zainos, A. Neuronal Correlates of a Perceptual Decision in Ventral Premotor Cortex. *Neuron* **41**, 165–173 (2004).
62. Okuyama, S., Kuki, T. & Mushiake, H. Recruitment of the premotor cortex during arithmetic operations by the monkey. *Sci. Rep.* **14**, (2024).
63. Wise, S. P. & Murray, E. A. Arbitrary associations between antecedents and actions. *Trends Neurosci.* **23**, 271–276 (2000).
64. Hogeveen, J. *et al.* What Does the Frontopolar Cortex Contribute to Goal-Directed Cognition and Action? *J. Neurosci.* **42**, 8508–8513 (2022).
65. Xie, Y. *et al.* Geometry of sequence working memory in macaque prefrontal cortex. *Science* **375**, 632–639 (2022).
66. Petrides, M. Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *J. Neurosci.* **15**, 359–375 (1995).
67. Fujii, N. Representation of Action Sequence Boundaries by Macaque Prefrontal Cortical Neurons. *Science* **301**, 1246–1249 (2003).

68. Ohbayashi, M., Picard, N. & Strick, P. L. Inactivation of the Dorsal Premotor Area Disrupts Internally Generated, But Not Visually Guided, Sequential Movements. *J. Neurosci.* **36**, 1971–1976 (2016).
69. Russo, A. A. *et al.* Neural Trajectories in the Supplementary Motor Area and Motor Cortex Exhibit Distinct Geometries, Compatible with Different Classes of Computation. *Neuron* **107**, 745-758.e6 (2020).
70. Shima, K. & Tanji, J. Neuronal Activity in the Supplementary and Presupplementary Motor Areas for Temporal Organization of Multiple Movements. *J. Neurophysiol.* **84**, 2148–2160 (2000).
71. Tanji & SHima. Role for supplementary motor area cells in planning several movements ahead. (1994).
72. Fodor, J. A. & Pylyshyn, Z. W. Connectionism and cognitive architecture: A critical analysis. *Cognition* **28**, 3–71 (1988).
73. Smolensky, P. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif. Intell.* **46**, 159–216 (1990).
74. Lakretz, Y. & Dehaene, S. Recursive Processing of Nested Structures in Monkeys? Two alternative accounts. Preprint at https://doi.org/10.31234/osf.io/k8vws (2021).
75. Bursley, J. K. Evidence for Recursive Operations in Human Cognition. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37365938 (2020).
76. Penn, D. C., Holyoak, K. J. & Povinelli, D. J. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* **31**, 109–130 (2008).
77. Jensen, K. T. *et al.* A mechanistic theory of planning in prefrontal cortex. 2025.09.23.677709 Preprint at https://doi.org/10.1101/2025.09.23.677709 (2025).
78. Zylberberg, A., Dehaene, S., Roelfsema, P. R. & Sigman, M. The human Turing machine: a neural framework for mental programs. *Trends Cogn. Sci.* S136466131100088X (2011) doi:10.1016/j.tics.2011.05.007.
79. Piantadosi, S. T. *et al.* Why concepts are (probably) vectors. *Trends Cogn. Sci.* **28**, 844–856 (2024).
80. Camp, E. THINKING WITH MAPS*. *Philos. Perspect.* **21**, 145–182 (2007).
81. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
82. *The Routledge Handbook of Philosophy of Animal Minds*. (Routledge, Taylor & Francis Group, London New York, 2019).
83. Beck, J. Do Nonhuman Animals have a Language of Thought? in *The Routledge Handbook of Philosophy of Animal Minds* (Routledge, 2017).
84. SHALLICE, T. & BURGESS, P. W. DEFICITS IN STRATEGY APPLICATION FOLLOWING FRONTAL LOBE DAMAGE IN MAN. *Brain* **114**, 727–741 (1991).
85. Nachev, P., Kennard, C. & Husain, M. Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* **9**, 856–869 (2008).
86. Isoda, M. & Hikosaka, O. Switching from automatic to controlled action by monkey medial frontal cortex. *Nat. Neurosci.* **10**, 240–248 (2007).
87. Mendoza, G., Méndez, J. C., Pérez, O., Prado, L. & Merchant, H. Neural basis for categorical boundaries in the primate pre-SMA during relative categorization of time intervals. *Nat. Commun.* **9**, 1098 (2018).
88. Ninokura, Y., Mushiake, H. & Tanji, J. Representation of the Temporal Order of Visual Objects in the Primate Lateral Prefrontal Cortex. *J. Neurophysiol.* **89**, 2868–2873 (2003).
89. Barbas, H. & Pandya, D. N. Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J. Comp. Neurol.* **256**, 211–228 (1987).

90. Mitchell, D. J. *et al.* A Putative Multiple-Demand System in the Macaque Brain. *J. Neurosci.* **36**, 8574–8585 (2016).
91. Neupane, S., Fiete, I. & Jazayeri, M. Mental navigation in the primate entorhinal cortex. *Nature* **630**, 704–711 (2024).
92. Tafazoli, S. *et al.* Building compositional tasks with shared neural subspaces. *Nature* 1–9 (2025) doi:10.1038/s41586-025-09805-2.
93. Mollazadeh, M. *et al.* Spatiotemporal Variation of Multiple Neurophysiological Signals in the Primary Motor Cortex during Dexterous Reach-to-Grasp Movements. *J. Neurosci.* **31**, 15531–15543 (2011).
94. Saleem, K. S. & Logothetis, N. K. *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates*. (Academic Press, 2012).
95. He, S., Dum, R. & Strick, P. Topographic organization of corticospinal projections from the frontal lobe: motor areas on the lateral surface of the hemisphere. *J. Neurosci.* **13**, 952–980 (1993).
96. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
97. Berger, M., Agha, N. S. & Gail, A. Wireless recording from unrestrained monkeys reveals motor goal encoding beyond immediate reach in frontoparietal cortex. *eLife* **9**, e51322 (2020).
98. Lu, M.-T., Preston, J. B. & Strick, P. L. Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *J. Comp. Neurol.* **341**, 375–392 (1994).
99. Gentilucci, M. *et al.* Functional organization of inferior area 6 in the macaque monkey. (1988).
100. Schaffelhofer, S. & Scherberger, H. Object vision to hand action in macaque parietal, premotor, and motor cortices. *eLife* **5**, e15278 (2016).
101. Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* **3**, 131–141 (1996).
102. Díaz, H. *et al.* Contextual neural dynamics during time perception in the primate ventral premotor cortex. *Proc. Natl. Acad. Sci.* **122**, e2420356122 (2025).
103. He, S., Dum, R. & Strick, P. Topographic organization of corticospinal projections from the frontal lobe: motor areas on the medial surface of the hemisphere. *J. Neurosci.* **15**, 3284–3306 (1995).
104. Nakamura, K., Sakai, K. & Hikosaka, O. Effects of Local Inactivation of Monkey Medial Frontal Cortex in Learning of Sequential Procedures. *J. Neurophysiol.* **82**, 1063–1068 (1999).
105. Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the Quantity of Visual Items in the Primate Prefrontal Cortex. *Science* **297**, 1708–1711 (2002).
106. Romo, R., Brody, C. D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
107. Wilson, F. A. W., Scalaidhe, S. P. Ó. & Goldman-Rakic, P. S. Dissociation of Object and Spatial Processing Domains in Primate Prefrontal Cortex. *Science* **260**, 1955–1958 (1993).
108. Tsujimoto, S., Genovesio, A. & Wise, S. P. Evaluating self-generated decisions in frontal pole cortex of monkeys. *Nat. Neurosci.* **13**, 120–126 (2010).
109. Miyamoto, K., Setsuie, R., Osada, T. & Miyashita, Y. Reversible Silencing of the Frontopolar Cortex Selectively Impairs Metacognitive Judgment on Non-experience in Primates. *Neuron* **97**, 980-989.e6 (2018).
110. Mansouri, F. A., Koechlin, E., Rosa, M. G. P. & Buckley, M. J. Managing competing goals — a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* **18**, 645–657 (2017).
111. Hwang, J., Mitz, A. R. & Murray, E. A. NIMH MonkeyLogic: Behavioral control and data acquisition in MATLAB. *J. Neurosci. Methods* **323**, 13–21 (2019).
112. Pachitariu, M., Sridhar, S., Pennington, J. & Stringer, C. Spike sorting with Kilosort4. *Nat. Methods* **21**, 914–921 (2024).

113. Testard, C. *et al.* Neural signatures of natural behaviour in socializing macaques. *Nature* **628**, 381–390 (2024).
114. Liu, Y., Brincat, S. L., Miller, E. K. & Hasselmo, M. E. A Geometric Characterization of Population Coding in the Prefrontal Cortex and Hippocampus during a Paired-Associate Learning Task. *J. Cogn. Neurosci.* **32**, 1455–1465 (2020).
115. Willett, F. R. *et al.* Hand Knob Area of Premotor Cortex Represents the Whole Body in a Compositional Way. *Cell* **181**, 396-409.e26 (2020).

Methods

Subjects and surgical procedures

Data were acquired from two adult male macaques (*Macaca mulatta*, average weights 17 kg (S1) and 10 kg (S2), average ages 9 years (S1) and 7 years (S2)). This sample size was chosen to match the standard for neural recording studies of behaving monkeys[91,92]. All animal procedures complied with the NIH Guide for Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of the Rockefeller University (protocol 24066-H).

After undergoing initial task training in their home cages, subjects underwent two surgeries, the first to implant an acrylic head implant with a headpost, and the second to implant electrode arrays. Both surgeries followed standard protocol, including for anesthetic, aseptic, and postoperative treatment. In the first surgery, a custom-designed MR-compatible Ultem headpost was implanted, surrounded by a bone cement cranial implant, or “headcap” (Metabond, Parkell and Palacos, Heraeus), which was secured to the skull using MR-compatible ceramic screws (Rogue Research). After a six month interval, to allow bone to grow around the screws and for the subject to acclimate to performing the task during head fixation via the headpost, we performed a second surgery to implant 16 floating microelectrode arrays (32-channel FMA, Microprobes for Life Science), following standard procedures[93]. Briefly, after performing a craniotomy and durotomy over the target area, arrays were inserted one by one stereotactically, held at the end of a stereotaxic arm with a vacuum suction attachment (Microprobes). Using vacuum suction allowed us to release the arrays, after insertion, with minimal mechanical perturbation by turning off the suction. After all arrays had been implanted, the dura mater was loosely sutured and covered with DuraGen (Integra LifeSciences). The craniotomy was closed with bone cement.

We used standard density arrays (1.8 mm x 4 mm) for all areas, except SMA and preSMA, for which we used four high density arrays (1.6 mm x 2.95 mm). Four additional electrodes on each array served as reference and ground. Two arrays were targeted to each of multiple areas of frontal cortex, with locations identified stereotactically, and planned using brain surface reconstructions derived from anatomical MRI scans (3D Slicer 5.6.2). Locations were selected based on their published functional and anatomical properties (see below), anatomical sulcal landmarks, and a standard macaque brain atlas[94].

814 During surgery, locations were further adjusted based on cortical landmarks, and to avoid visible blood
815 vessels. Arrays were implanted in the right hemisphere (contralateral to the arm used for drawing).

816 Array locations are depicted in **Fig. 3** and **Fig. S3**, confirmed with intraoperative photographs. For
817 M1, we targeted hand and arm representations (F1), directly medial to the bend of the central sulcus
818 (which corresponds roughly to the intersection of the central sulcus and the arcuate spur if the latter were
819 extended caudally), based on retrograde labeling from spinal cord and microstimulation of M1[95] and M1
820 recordings[96]. For PMd, we placed both arrays lateral to the precentral dimple, with one (more caudal)
821 array directly medial to the arcuate spur (the arm representation[95–97], F2), and the other more rostral
822 (straddling F2 and F7). For PMv, we targeted areas caudal to the inferior arm of the arcuate sulcus (F5),
823 which are associated with hand movements based on retrograde labeling from spinal cord[95] and M1[98],
824 microstimulation[99] and functional studies[61,100,101], as well as with decision making[61,102]. These areas
825 contain neurons interconnected with PFC[98]. For SMA (F3) and preSMA (F6), we targeted the medial wall
826 of the hemisphere, with the boundary between SMA and preSMA defined as the anterior-posterior
827 location of the genu of the arcuate sulcus, consistent with prior studies finding significant differences
828 across this boundary in anatomical connectivity (e.g., direct spinal projections in SMA but not preSMA[103])
829 and function[28,104]. SMA arrays were largely in the arm representation[103]. For dlPFC, we targeted the
830 region immediately dorsal to the principal sulcus (46d), following prior studies of action sequencing[27,38,66]
831 and other cognitive functions[59]. For vlPFC, we targeted the inferior convexity ventral to the principal
832 sulcus, with one (more rostral) array directly ventral to the principal sulcus (46v) and the other rostral to
833 the inferior arm of the arcuate sulcus (45A/B), based on evidence for encoding of abstract concepts in
834 regions broadly spanning these two locations[57,105,106], including a possibly heightened role (compared to
835 dlPFC) in encoding abstract concepts in a manner invariant to temporal or spatial parameters[45,106,107]. For
836 FP, we targeted a rostral location similar to prior recording and imaging studies (one array fully in area
837 10, the other straddling 9 and 10)[108,109], including areas associated with executive functions[110]. In general,
838 array locations targeted the cortical convexity immediately next to sulci, instead of within the banks, in
839 order to allow shorter insertion depths that minimize the risk of missing the target. The exceptions were
840 SMA and preSMA in the medial wall, for which this was not possible. To avoid damaging the superior

841 sagittal sinus, we positioned the arrays laterally (2 mm from midline) and slanted the electrodes medially
842 (**Fig.** S3).
843 The lengths of each electrode were custom designed to target half-way through the gray matter,
844 and to vary substantially across the array, to maximize sampling of the cortical depth. Electrode lengths
845 spanned (in mm) 1.5 - 3.5 (M1), 1.5 - 3.1 (PMd, PMv), 2.8 - 5.8 (SMA, preSMA), 1.5 - 2.5 (dlPFC, vlPFC),
846 and 1.5 - 2.6 (FP) for subject 1, and 1.7 - 3.75 (M1), 1.5 - 3.3 (PMv), 1.5 - 3.1 (PMd), 2.65 - 5.95 (SMA,
847 preSMA), 1.75 - 3.15 (dlPFC), 1.35 - 3.2 (vlPFC), and 1.6 - 2.9 (FP) for subject 2. Reference electrodes
848 were longer (6 mm) to anchor the arrays. 28 electrodes were Pt/Ir (0.5 MΩ) and 4 were Ir (10 kΩ). Array
849 connectors (Omnetics A79022) were housed in custom-made Ultem pedestals (Crist), which were
850 secured with bone cement onto the cranial implant. Four pedestals were used per subject, holding 5, 5,
851 4, and 2 connectors each.
852
853 **Behavioral task**
854 Task overview
855 The general task structure and experimental setup are described in a prior publication[1]. In brief, subjects
856 were seated comfortably in the dark with head restrained by headpost fixation. They faced a touchscreen
857 (Elo 1590L 15" E334335, PCAP, 768 x 1024 pixels, refresh rate 60 Hz, with matte screen protector to
858 reduce finger friction) that presented images and was drawn on. The touchscreen location was optimized
859 to allow each subject to easily draw at all relevant locations on the screen (23 to 26 cm away; diagram in
860 **Fig. S1**). Both subjects decided on their own over the course of learning to perform the task with the left
861 hand. The chairs were designed to minimize movements of the torso and legs (by using a loosely
862 restricting "belly plate") and the non-drawing arm (by resting on the belly plate and having movement
863 restricted to within the chair). Gravity-delivered reward (water-juice mixture) was controlled by the
864 opening and closing of a solenoid pinch valve (Cole-Parmer, 1/8" ID). Subjects were water-regulated,
865 with careful monitoring that consumption met the minimum requirement per day (typically exceeding it),
866 and body weight was closely monitored to ensure good health. The task was controlled with custom-
867 written software, using the MonkeyLogic (2.2.45) behavioral control and data acquisition MATLAB

868 package[111]. (PC: Windows 10 Pro, Intel Core i7-4790K, 32GB RAM; DAQ: National Instruments PCIe-
869 6343). All stimuli (images of line figures defined as point sets, with points rendered large enough to
870 appear as continuous curves) were also generated with custom-written MATLAB (R2021a) code. Images
871 were presented in a "workspace" area on the screen (16.6 cm x 16.9 cm, corresponding to approximately
872 37° by 38° visual angle). Shape components in images were on average 4.0 cm (9°) (maximum of width
873 and height).
874
875 Task Training
876 The procedure for training subjects on the basic drawing task are described in a prior publication[1].
877
878 Trial structure
879 Trial structure is summarized in **Fig. 1d** (screen schematic in **Fig. S1**) and described in detail in a prior
880 publication[1]. Screen image changes (including image presentation and other trial events) were recorded
881 using photodiodes (Adafruit Light Sensor ALS-PT19) and sounds were recorded using an electret
882 microphone (Adafruit Maxim MAX4466, 20-20KHz). We performed eye tracking (ISCAN), but did not
883 enforce eye fixation.
884
885 **Behavioral data analysis**
886 Preprocessing of touchscreen data
887 Touchscreen data were represented as time series of (x, y) coordinates in units of pixels (conversion:
888 33.6 pixels/cm) and sampled at 60 Hz, which we upsampled to 500 Hz (performed in MonkeyLogic to
889 align all behavioral signals, including trial event markers and eye tracking), and low-pass filtered below
890 the highest frequency in the drawing movements (15 Hz). Strokes were segmented based on the time of
891 first touch (onset) and the time of last touch (offset) with 500 Hz resolution.
892
893 **Neural recordings**

894 Recordings were acquired using a Tucker-Davis Technologies (TDT) system, including headstage (Z-
895 Series 32 Channel Omnetics, LP32CH-Z), amplifier (PZ5M-256), processor (RZ2), and storage (RS4),
896 sampled at 25 kHz (local reference mode), controlled with TDT Synapse (v98) software run on a Windows
897 10 PC (Intel Core i7-3770, 32GB RAM), and saved to disk. Analog and digital task-related signals,
898 including behavioral events (photodiode, audio, and trial event markers) and eye tracking (ISCAN, 125
899 Hz), were synchronized to external triggers recorded by the neural data acquisition system.
900
901 **Neural data preprocessing**
902 Spike sorting
903 We extracted for later analysis both isolated single-unit (SU) and multi-unit (MU) spike clusters from the
904 stored broadband signal following procedures combining Kilosort[112] (v2.5) and custom-written software.
905 Spikes times were then converted to firing rate functions by smoothing with a 0.025 s Gaussian kernel
906 (0.01 s slide). All of these procedures are described in a prior publication[1].
907
908 **Neural data analyses**
909 Dimensionality reduction of population activity
910 We performed dimensionality reduction on the neural population activity using principal
911 components analysis (PCA), in general because high-dimensional noise can reduce the interpretability
912 of the Euclidean distance[113], using methods described in a prior publication[1].
913
914 Computing "neural distance"
915 To quantify the similarity of population activity between two sets of trials, such as trials for conditions A
916 and B, each a specific conjunctive value of task-relevant variables, we devised a “neural distance” metric,
917 as described in a prior publication[1]. In brief, inspired by the “normalized distance” in Liu et al[114], it is the
918 average pairwise Euclidean distance across conditions A and B, minus the average within-condition
919 distance. This subtraction ensures the useful property that this distance is unbiased, in that the expected
920 value of neural distance between two sets of trials sampled from the same distribution is zero (unlike the

921 mean Euclidean distance, which is biased upwards[115]). In addition, the resulting distance is normalized
922 by dividing by an upper-bound distance to normalize it between 0 and 1. Neural distance is defined as:

923 $$\underline{D^*_{AB} = D_{AB} - (D_{AA} + D_{BB})/2}$$

924 where the normalized Euclidean distance between sets of trials (indexed by $\underline{i}$ and $\underline{j}$) in conditions A and
925 B is:

926 $$D_{AB} = \left\langle \frac{1}{d_{\max}(t)} \left\langle \left\| \boldsymbol{x}_i(t) - \boldsymbol{x}_j(t) \right\| \right\rangle_{i \in A, j \in B} \right\rangle_{t \in \{\mathrm{t}_1, .., \mathrm{t}_\mathrm{n}\}}$$

927 Here, $\mathbf{x}_i(t)$ is the population activity vector at time t (in a window between times t_1 and t_n), and $d_{\max}$ is an
928 upper-bound (98th percentile) of the distances between all pairs of different trials combined across all
929 conditions.
930
931 <u>Computing a variable's encoding strength</u>
932 To compute how strongly a given variable is encoded in population activity (e.g., "shape encoding" in **Fig.**
933 **4e**), we computed the mean effect of that variable on population activity, in terms of neural distance, while
934 controlling for the other relevant variables. In general, shape encoding was the average pairwise neural
935 distance between conditions differing by shape, controlling for behavioral variables expected to affect
936 neural activity and thus possibly confound the analyses: screen location (discretized as a grid), screen
937 location of the previous stroke (effectively controlling for the movement direction during the gap before
938 the current stroke). Ordinal encoding (**Fig. 6g, h**) was the average pairwise neural distance between sets
939 of trials differing in ordinal position, computed separately for each (shape index, repeat length) condition,
940 and then averaged over those conditions. We controlled for the behavioral variables of location and
941 previous location. To compute $\mathrm{ord}_{\mathrm{FWD}}$ and $\mathrm{ord}_{\mathrm{REV}}$ encoding (**Fig. 6j-l**), we performed the following. For
942 each condition defined by (ordinal position, shape, repeat length) we computed one value for $\mathrm{ord}_{\mathrm{FWD}}$ and
943 one for $\mathrm{ord}_{\mathrm{REV}}$. $\mathrm{Ord}_{\mathrm{REV}}$ encoding is the mean pairwise neural distance across different $\mathrm{ord}_{\mathrm{REV}}$ positions
944 (keeping $\mathrm{ord}_{\mathrm{FWD}}$ the same), while $\mathrm{ord}_{\mathrm{FWD}}$ encoding is the neural distance across different $\mathrm{ord}_{\mathrm{FWD}}$ positions
945 (keeping $\mathrm{ord}_{\mathrm{REV}}$ the same), controlling for the behavioral variables of location and previous location. This
946 procedure is illustrated in the insets in **Fig. 6j**.

Statistically comparing brain regions in strength of variable encoding

In analyses that compare brain areas in their encoding strengths for particular variables (**Figs. 4e, 5i, 6g, h**), for a given brain-area pair of interest, we then fit a linear least-squares regression model to test for an effect of brain region on the variable's encoding strength y, controlling for trial-condition pair:

$$y = \beta_0 + \beta_r X_r + \sum_{j=1}^{N_c} \gamma_j Z_j + \epsilon$$

where X_r is 0 or 1 depending on brain region, and Z_j is an indicator variable for trial-pair condition, with γ_j as their coefficients, and ϵ is a noise term. Finally, we extracted the p-value for β_r (two-sided t-test), which represents the significance of the difference between these two regions in encoding strength for this variable.

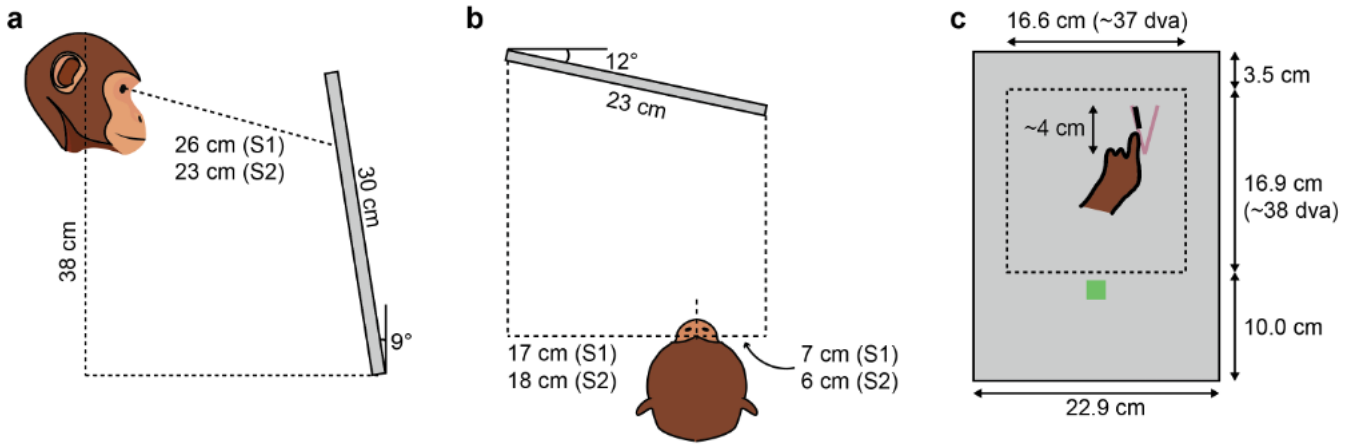

Figure S1. Behavioral task setup.

(a) Schematic of subject relative to touchscreen, profile view, during recording sessions. Subjects 1 and 2 were positioned at different distances to accommodate their individual anatomies and postures. The screen was slanted slightly to optimize the ability to see and reach to the same part of the screen (the workspace at the top of the screen; see panel c).

(b) Schematic of subject relative to screen, top view. Subjects were positioned to the right to accommodate reaching with the hand used for drawing (left).

(c) Schematic of screen during trial, with component locations and sizes to scale. The finger is tracing over the figure (purple-gray “V”), leaving a black trail of “ink” behind. The average size of shapes was 4.0 cm (maximum of width and height). The thickness and color of the figure and “ink” are to scale. The subject can press the “done” button (green square) at any time to report completion. Dashed line (not visible to the subject) indicates the workspace where images and drawings occurred.

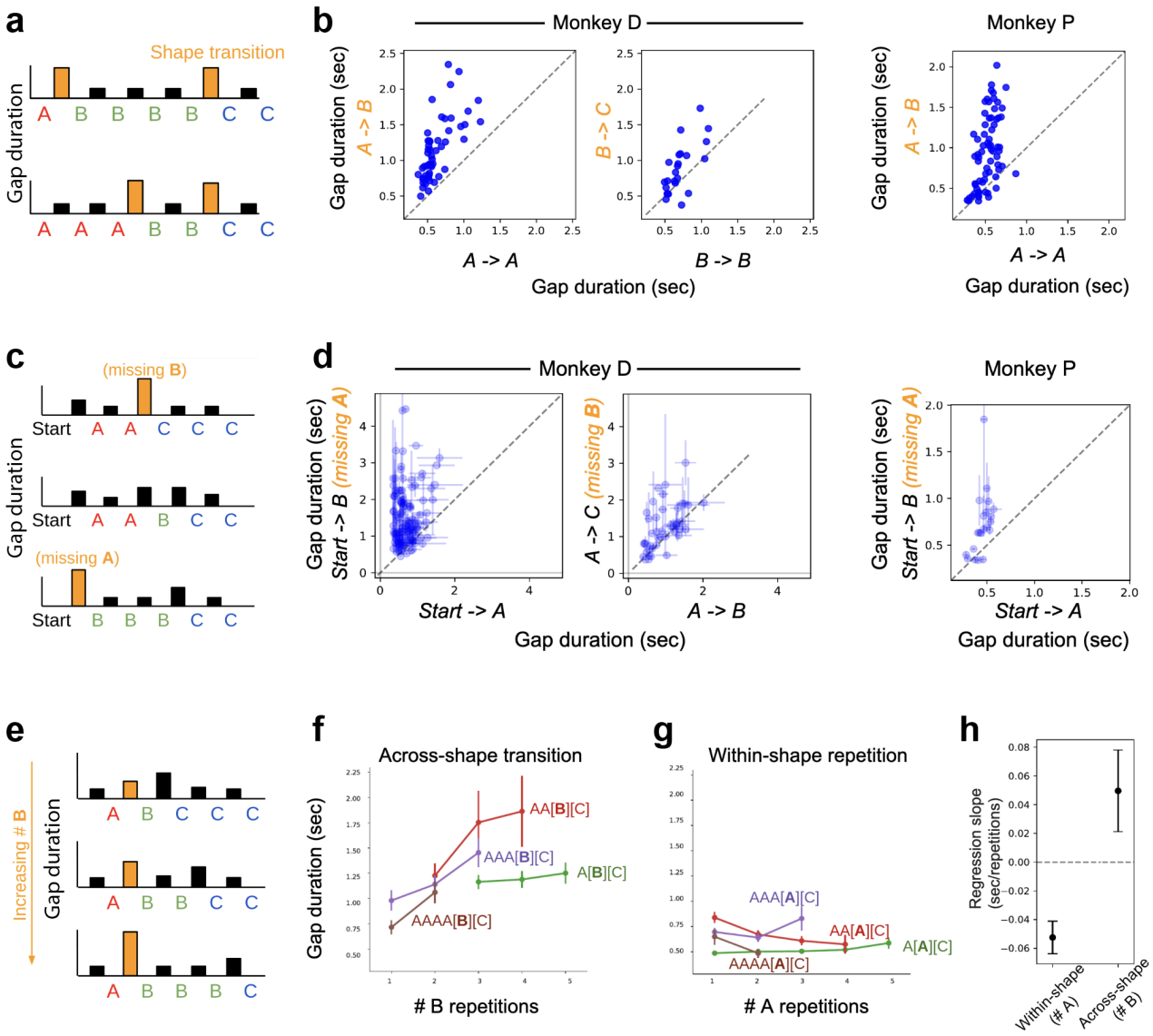

Figure S2. Gap timing reflects grammatical structure.

(a) Schematic of predicted gap timing consistent with hierarchical structure. If each shape repeat forms a high-level unit, then gaps should be slower between shapes ("shape transition") compared to within shapes.

(b) Mean gap duration for transitions across shapes (y-axis) vs. transitions within shapes (x-axis).

(c) Schematic of predicted gap timing consistent with understanding of ordinal structure of shapes. Transitions when an expected shape is skipped (e.g., B is skipped in AACCC) should be slow.

(d) Mean gap duration for transitions where an expected shape is missing (y-axis) vs. a matched condition when the expected shape is present (x-axis).

(e) Schematic of predicted gap timing consistent with consideration of the entire repeat of a shape before initializing it, with longer upcoming sequences involving longer gaps before initializing that sequence.

(f) Mean gap duration preceding a repeat of B as a function of repeat length of B (x-axis) controlling for different sequential contexts (colors) defined by the preceding sequence (varying numbers of *A*s). The sequence is defined by both context and # repetitions; e.g., if context is AA[B]C and repeat length is 3, then this is the sequence AABBBC.

(g) Mean gap duration preceding a repeat of A, plotted analogously to (f). Note the lack of increasing gap duration preceding increasing number of repetitions of A.

(h) Summary of effect of length of upcoming repeat on the duration of the gap preceding the repeat, plotting 95% CI estimated using linear mixed effects modeling (fixed effect of repeat length, random effect of context, performed separately for within-shape and across-shape transitions).

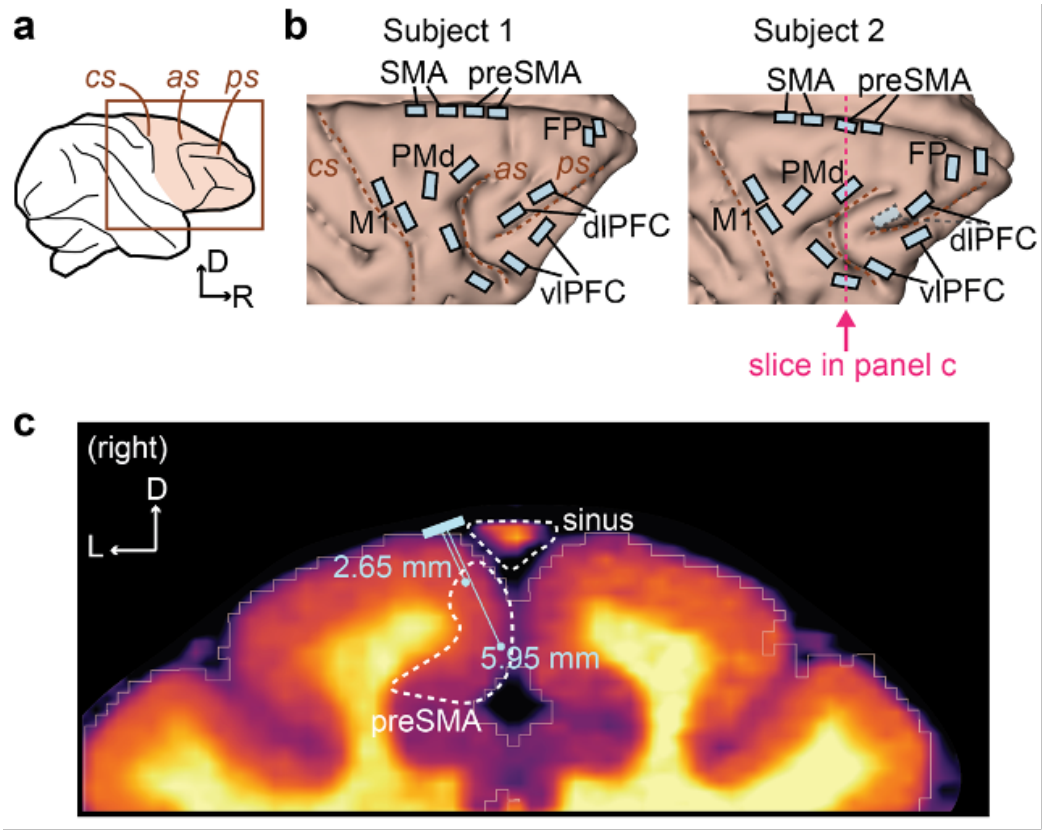

Figure S3. Multi-area neuronal recordings across frontal cortex and angled implantation of medial wall arrays.

(a) Recordings were targeted to right frontal cortex, contralateral to the drawing hand. cs, central sulcus; as, arcuate sulcus; ps, principal sulcus. R, rostral. D, dorsal.

(b) Arrays to scale on 3D rendering of brain. The caudal dlPFC array for subject 2 malfunctioned. SMA and preSMA electrodes were angled to target the medial wall (see panel c). Panel b (subject 1) is reproduced from Fig. 3.

(c) Coronal MRI section showing SMA and preSMA arrays angled to avoid the superior sagittal sinus and target the medial wall, depicted for a preSMA array. Arrays were implanted ~2 mm lateral to the midline, and angled medially. The array is depicted as a blue rectangle and two lines represent the shortest (2.65 mm) and longest (5.95 mm) electrodes. This location and angle is our best estimate based on the surgical plan, stereotactic coordinates, and intra-operative photographs. The other two SMA and one preSMA arrays were similar. D, dorsal. L, lateral.