



HAL
open science

The measurement, evolution, and neural representation of action grammars of human behavior

Dietrich Stout, Thierry Chaminade, Jan Apel, Ali Shafti, A. Aldo Faisal

► **To cite this version:**

Dietrich Stout, Thierry Chaminade, Jan Apel, Ali Shafti, A. Aldo Faisal. The measurement, evolution, and neural representation of action grammars of human behavior. *Scientific Reports*, 2021, 11, pp.13720. 10.1038/s41598-021-92992-5 . hal-03278117

HAL Id: hal-03278117

<https://hal.science/hal-03278117v1>

Submitted on 5 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

The measurement, evolution, and neural representation of action grammars of human behavior

Dietrich Stout^{1✉}, Thierry Chaminade², Jan Apel³, Ali Shafti⁴ & A. Aldo Faisal^{4,5,6,7✉}

Human behaviors from toolmaking to language are thought to rely on a uniquely evolved capacity for hierarchical action sequencing. Testing this idea will require objective, generalizable methods for measuring the structural complexity of real-world behavior. Here we present a data-driven approach for extracting action grammars from basic ethograms, exemplified with respect to the evolutionarily relevant behavior of stone toolmaking. We analyzed sequences from the experimental replication of ~2.5 Mya Oldowan vs. ~0.5 Mya Acheulean tools, finding that, while using the same “alphabet” of elementary actions, Acheulean sequences are quantifiably more complex and Oldowan grammars are a subset of Acheulean grammars. We illustrate the utility of our complexity measures by re-analyzing data from an fMRI study of stone toolmaking to identify brain responses to structural complexity. Beyond specific implications regarding the co-evolution of language and technology, this exercise illustrates the general applicability of our method to investigate naturalistic human behavior and cognition.

For more than 60 years, the serial ordering of behaviour has been a core topic for the cognitive and behavioural sciences^{1,2}. Enhanced capacities for complex action sequencing support distinctive human behaviors such as language³, imitation⁴, and tool use^{5,6}, and are fundamental to the flexibility that is a hallmark of human intelligence^{7,8}. It has been suggested that this implies a unitary neurocognitive foundation for human behavioural uniqueness across domains^{1,5,6}, but this remains controversial⁹. Although theory¹ and modelling¹⁰ suggest computational similarities across behaviours ranging from tool-use and foraging to language learning, empirical investigation outside domains with established notational systems (language, music, mathematics) has been limited by a lack of objective, generalizable methods for describing, quantifying, and comparing the sequential structure of diverse, real-world behaviours.

Research on motor sequence^{11–13}, implicit^{14,15}, and statistical learning¹⁶ has provided evidence of underlying neural and cognitive mechanisms, but has generally been limited to highly artificial tasks such as executing invariant key-press sequences or recognizing simple artificial grammars of known structure. This limits our ability to generalize findings to understand the learning and execution of the real-world skills^{17,18} of interest to fields ranging from sport science¹⁹ and surgery²⁰, to human behavioral ecology²¹, and comparative psychology²².

In human origins research specifically, investigation of long-standing hypotheses about the evolutionary relationships between tool making, language, and cognition have been hampered by the lack of an objective metric for the behavioural complexity of different ancient human technologies^{23–26}. Here we adopt a data-driven computational approach to this challenge by using grammatical pattern recognition algorithms to measure the structural complexity of behavioral sequences from modern tool-making replication experiments—effectively extracting action grammars for critical survival skills from the human evolutionary past. This allows us to isolate and compare the structural complexity of “noisy” natural behaviors that simultaneously vary across a wide range of other perceptual, motor, and kinematic dimensions, including identification of specific brain responses to this complexity.

We conducted 17 tool-making replication experiments and coded the behavior sequences that were generated (Fig. 1A,B). This sample includes 5 sequences for which upper limb movements and manual joint angles were recorded as part of a previous study²⁷, and 6 for which the tools and waste produced were analyzed and compared

¹Department of Anthropology, Emory University, Atlanta, GA, USA. ²Institut de Neurosciences de La Timone, Aix Marseille Université, Marseille, France. ³Department of Archaeology, Stockholm University, Stockholm, Sweden. ⁴Department of Bioengineering, Imperial College London, London, UK. ⁵Department of Computing, Imperial College London, London, UK. ⁶Integrative Biology, MRC London Institute of Medical Sciences, London, UK. ⁷Behaviour Analytics Lab, Data Science Institute, London, UK. ✉email: dwestout@emory.edu; a.faisal@imperial.ac.uk

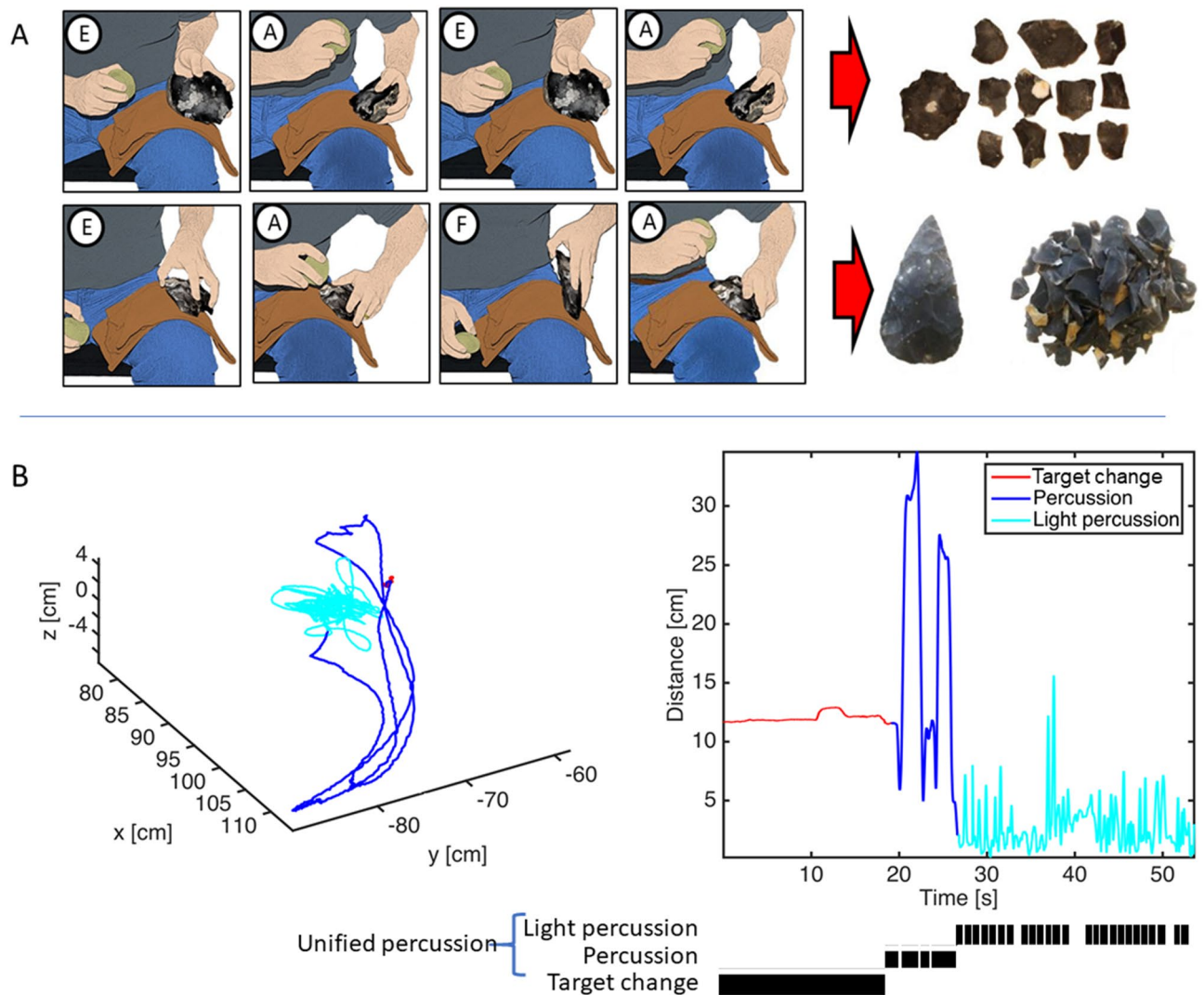


Figure 1. Symbolic Modeling of tool-making action sequences. (A) Oldowan (top) and Acheulean (bottom) action sequences were coded using 7 event codes (circled letters, see Materials & Methods). Products illustrated to the right, individual depicted is the first author. (B, Left) Spatial trajectory of the hammer stone during toolmaking. The X–Y plane is aligned with the plane of the percussion strike hand-arm movements. The core is centred at approximately 80 cm, –70 cm, 0 cm. (B, Right) Distance of the hammerstone from the centre of the core. Target change movements are coded in red. Percussion strike trajectories are in dark blue, light percussion movements (here for platform preparation) are in light blue. Below, the black bars indicate the corresponding timing and duration of the ethogram.

with actual Paleolithic artifacts from the Middle Pleistocene site of Boxgrove, UK²⁸. Building on this and other previous research^{29–32}, we focused our current study on archaeologically documented tool-making methods from the early and late Lower Paleolithic, a period that witnessed a nearly threefold increase in hominin brain size. This allows us to empirically address the over 100 years of theorizing linking increasingly complex tool-making with brain evolution and language origins^{6,33–35}. The early (Oldowan, ca. 2.5 Mya) technology modeled here comprised the production of simple, sharp-edged stone flakes by striking one stone with another. The late (Late Acheulean, ca. 0.5 Mya) technology comprised the production of refined, teardrop-shaped “handaxes” through intentional shaping. We defined a shared action alphabet, consisting of 7 event types encompassing the elementary body movements and object transformations present in every sequence of both technologies (Fig. 2A), and applied two established sequence learning algorithms to the coded event sequences: Hidden Markov Models (HMM) and k-Sequitur.

Results

Hidden Markov modeling. HMM detects probabilistic regularities (hidden states) across sequences and can capture the structure of arbitrarily complex sequences given sufficient numbers of hidden states. The optimal number of hidden states provides a measure of structural complexity. We fitted HMMs to coded event sequences and computed the Bayesian Information Criterion (BIC) across different numbers of hidden states as a measure

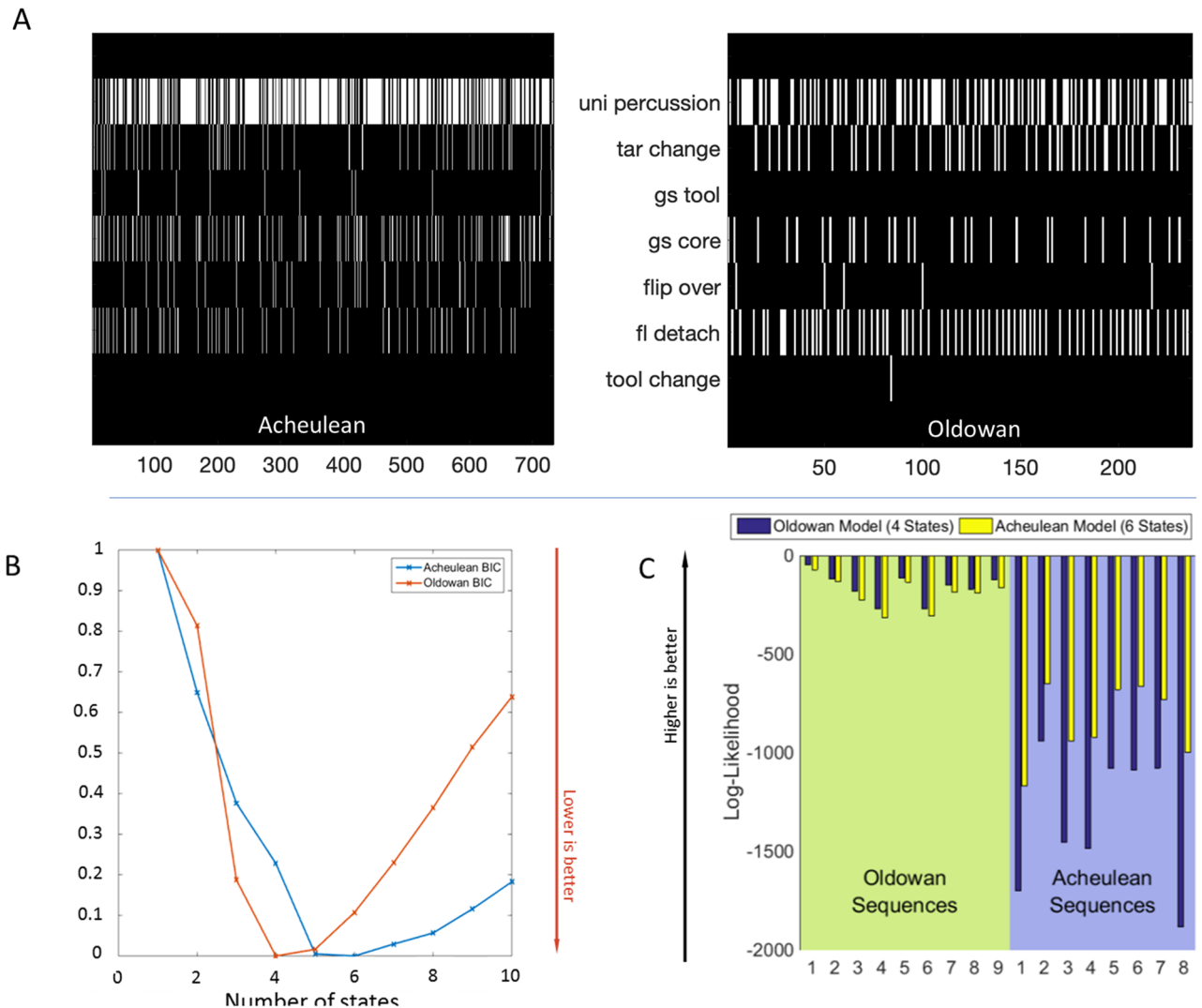


Figure 2. (A) Raster plot of two typical Acheulean (left) and Oldowan (right) tool making sequences. Each row in the raster plot represents the observation of one of the 7 actions. The symbolic representation has been abstracted from explicit timing information (cf. Fig. 1.B) to a time-scale invariant sequence representation that counts sequence position. (B) Bayesian Information Criterion values (less is better) across models with increasing numbers of hidden states. Red, Oldowan, Minimum = 4; Blue, Acheulean, Minimum = 6. (C) Log-Likelihood values indicating model fit (higher is better) across sequences. Fit for Oldowan sequences is better overall; Acheulean model fit to Oldowan data is better than Oldowan model fit to Acheulean data.

of model fit. BIC reached its minimum (less is better) at 4 hidden states for Oldowan and 6 for Acheulean data (Fig. 2B), indicating a 50% increase in complexity. These two models perfectly categorized the sequences (likelihood greater for correct model, Fig. 2C). The fit was better for both models on the simpler Oldowan sequences. The close fit of the Acheulean model to Oldowan data (but not vice versa) indicates that the former captures most of the structure of the latter, and that Oldowan sequences may be considered a subset of Acheulean sequences.

We therefore used the Acheulean HMM to test for further structure. We obtained the most likely hidden state sequences for the Oldowan and Acheulean data and then fitted a second, 2-state HMM onto these higher-order sequences (Fig. 3). We found that Acheulean sequences oscillate between two superordinate states-of-states (SoS) whereas Oldowan sequences remain in one). Thus, Acheulean sequences display an additional level of structure not expressed by Oldowan sequences.

Next, we fit the 6-state Acheulean HMM to Oldowan and Acheulean data and observed the probability of actions per hidden state as well as transitions between hidden states. Our Oldowan data are characterized by the repetition of one simple action “chunk” consisting of a relatively invariant sequence of states (3- > 4- > 2; Fig. 3A) that essentially corresponds to the removal of an individual flake and is entirely captured by SoS1 (Fig. 3B,C). Acheulean sequences are more variable (Fig. 3D), reflecting the addition within some flake removal chunks of a sub-operation archaeologists refer to as striking platform preparation. This involves repeated low-amplitude (see “Methods” section and Fig. 1B) chipping of striking surfaces to alter their sharpness, bevel, and placement relative to the midline. This operation is captured at the next level by SoS2 (Fig. 3E). SoS2 is less frequent in the early stages of our sequences (Fig. 3F) which is consistent with the presence of an initial “roughing out” stage in

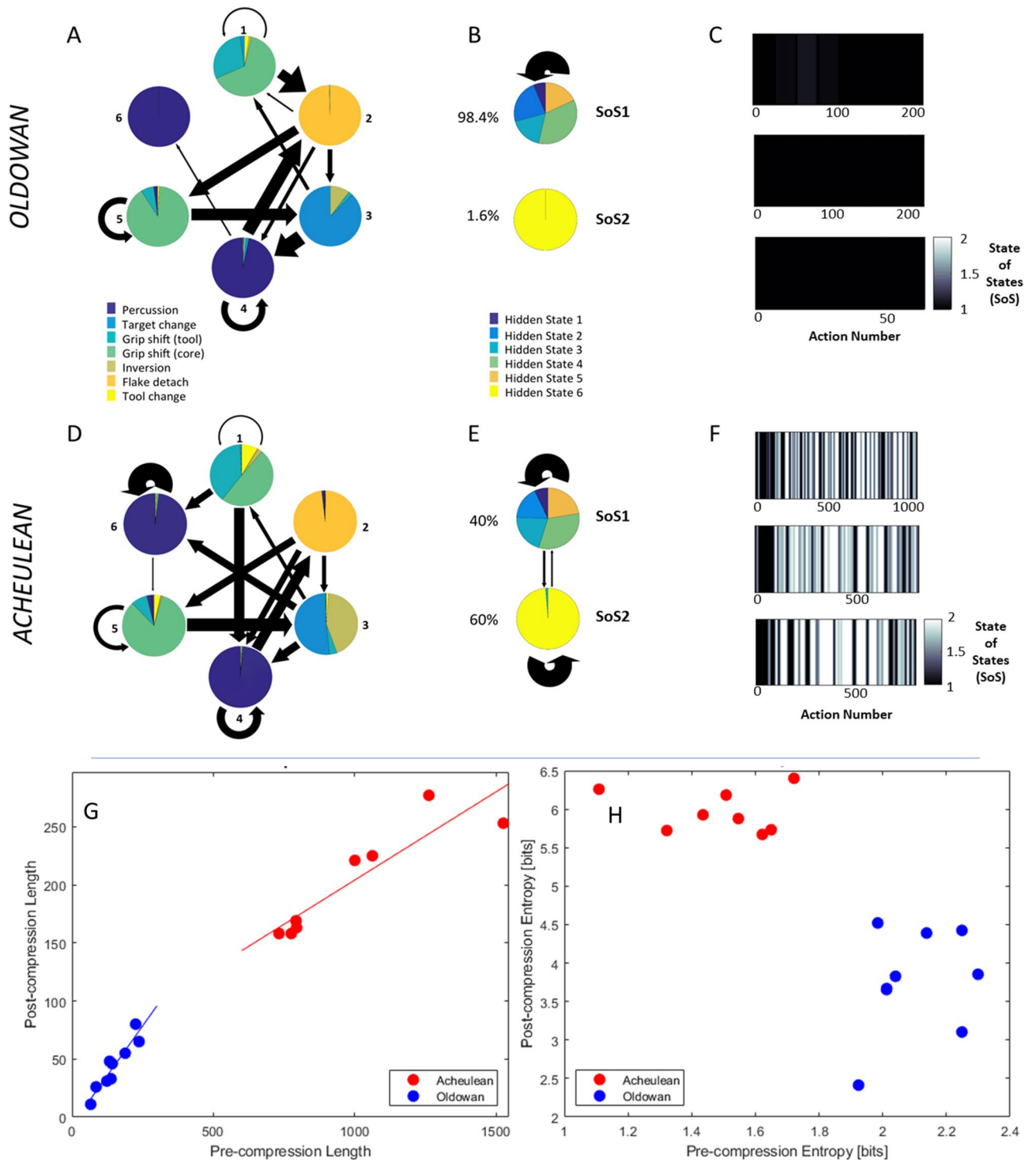


Figure 3. Results of grammatical analysis: HMM analysis (A–F) Empirical transition (arrows) and emission (pie charts) matrices of 6-state HMM fitted to all Oldowan (A) and Acheulean (D) sequences. Arrow thickness indicates transition probabilities between states (values < 5% not displayed). Pie chart area indicates probability of an action being performed in that state. Oldowan Hidden State 6 accounts for less than 2% of all data points. In the middle are similar illustrations of the superordinate “States-of-States” for Oldowan (B) and Acheulean (E) data. At right are examples of the running average State of States for Oldowan (C) and Acheulean (F) time-series. Black: everything in SoS 1; white: everything in SoS 2. Sequitur analysis (G,H): Effect of Sequitur compression on Acheulean (red) and Oldowan (blue) sequences, show clear differences between Oldowan and Acheulean sequences as measured by pre- and post-compression length (G). Note, how the slope of pre/post compression straight line fits is clearly different between Oldowan and Acheulean. Similarly, the information content within the uncompressed and compressed sequences as measured is clearly different (H).

handaxe manufacture prior to more refined shaping²³. Introspection by experienced tool-makers^{32,36} has previously suggested that platform preparation increases the complexity to Paleolithic action organization, but it has not previously been possible to test this intuition objectively or to quantify the magnitude of increase in a generalizable way. Our HMM method thus captures meaningful (i.e. goal directed) regularities in stone tool-making in a data-driven way that: 1) derives structure rather than imposing it, 2) respects the real variability underlying ideal characterizations, 3) enables objective quantification of grammatical complexity, and 4) is readily adaptable to the study of other sequential behaviors.

The Chomsky hierarchy in Formal Language Theory describes a series of increasingly powerful and inclusive computational systems, or grammars, differentiated by their memory resources^{1,37}. A simple Markov chain is a memoryless probabilistic system equating to a regular (finite-state) grammar that does not permit long distance dependencies. HMMs are dynamic Bayesian networks that asymptotically approximate supra-regular context-free grammars (with unbounded memory) through the progressive addition of hidden states. The increase we observed in the optimal number of hidden states from Oldowan to Acheulean thus provides a measure of increased grammatical complexity and memory requirements without positing infinite capacity. This modeling approach is consonant with the view that finite-state, probabilistic, and parallel computational models are cognitively and neurobiologically realistic^{7,37-39}. Others, however, contend that human cognition is in fact characterized by constitutively hierarchal processing using supra-regular resources, that humans have a tendency to employ such context-free solutions even when they are not actually necessary¹, and that such tree-based algorithms are more efficient for learning⁴⁰.

Context-free grammar fitting. We therefore pursued a second approach by fitting context-free grammars (CFGs) to the tool-making sequences. Formal Language Theory employs terminal symbols (in our case 7 event types) and non-terminal symbols (re-write rules expandable to terminal and/or non-terminal symbols) to generate strings. Ultimately, we have to solve the minimal grammar problem, i.e., finding the unknown grammatical rulebook that provides the simplest explanation given the observed sequence data. Whereas regular grammars and HMMs are driven by local relationships between symbols, CFGs capture nested dependencies of theoretically infinite length and depth. The standard algorithm to extract deterministic CFGs, Sequitur⁴¹, creates a new rule as soon as a terminal symbol pair is observed twice in a sequence and repeats this pair-wise aggregation, adding new levels of superordinate rules until the complete sequence is described. This makes Sequitur powerful but liable to detect a high number of spurious (occurring < 3 times) rules in the variable sequences generated by real human behavior. We therefore advance a simple variant of Sequitur, k-Sequitur, which requires a pair to occur k-times before generating a rule. Increasing k makes the grammar discovery process less sensitive to infrequent pairs and less prone to creating rules from noise. Nevertheless, both sequitur (2-sequitur) and higher-k versions produce the same qualitative results (see Supplementary Fig. S1):

In agreement with our HMM results, deterministic CFG extraction found Oldowan grammars to be a less complex sub-set of Acheulean grammars. Rule inference from combined Oldowan and Acheulean samples identified multiple rules that occur only in Acheulean sequences (Supplementary Fig. S2) and showed that the frequency of Acheulean-only rules increases at higher levels (0 at level 2, 1 at Level 3, 2 at Level 4, 5 at Level 5). No Oldowan-only rules were identified, even when rule inference was restricted to the Oldowan data set. CFG extraction achieved substantial compression of both Oldowan and Acheulean sequences (Fig. 3G), however the rate (inverse slope) of Acheulean compression was more than twice as great (7.69 vs. 2.94). This indicates that Acheulean sequences have more structure for rule-based compression, in an approximate 2:1 ratio paralleling our HMM finding of two Acheulean SoS vs. one Oldowan. Each post-compression Acheulean element (rule or terminal symbol) contains more information (measured as Shannon entropy: Fig. 3H), yet Acheulean grammars still require more non-terminal symbols (rules) to achieve a comparable fit to the data. These compression results are robust over increasing k values (k = 4, k = 8; see Supplementary Fig. S1). In addition to our bit-based representation of absolute entropy change as a result of compression, we also considered normalized entropy to ensure comparability. For this, each entropy value is divided by the theoretical maximum entropy, relating to a case of equiprobable states (see Supplementary Fig. S3). CFGs can parse regular strings, so fitting CFGs to our sequences in this way does not imply that supra-regular resources are required. It does show that the greater complexity and depth of Acheulean sequences is robust even assuming such resources.

Our CFG results reveals that the greater complexity of Acheulean sequences is due to long strings of repeated percussions, the removal of which eliminates Oldowan/Acheulean differences in compression rate (Supplementary Fig. S4). These strings comprise the same repeated, low-amplitude chipping of striking platforms (Fig. 1B) extracted as SoS2 in our HMM analysis and corresponding to the tool-making operation known as platform preparation²⁸. HMM and CFG methods thus converge, not only to quantify the greater complexity of Acheulean sequences, but also to extract a key technological element of the instrumental structure of Acheulean tool making that largely accounts for this difference.

Discussion

The results show that our grammar extraction methods are able to discover the instrumental structure of behavior directly from the structure of action sequences coded using a minimalistic and objective ethogram, without requiring subjective functional or intentional interpretations by the observer. These methods are easily generalizable to other behaviors and, in the specific case of Paleolithic tool making, provide new means to investigate the archaeological record of technological change. By using a single elementary action alphabet, we can consider variation within as well as between putative behavior categories in strictly equivalent terms, treating variation as a source of information rather than noise⁴². This provides a method for studying the structure and neurocognitive

foundations of complex and variable real-world behaviors in a way that complements ongoing research into structured sequence learning that uses artificial experimental tasks.

To date, motor sequence learning research has generally employed simple sequences of basic motor actions (e.g. finger movements, key-presses) to address questions about the time-course and neural mechanisms of learning (e.g. ^{12,13,43}). Serial Reaction Time Tasks¹⁵ add a sensory element by making responses contingent on a (usually visuospatial) cue. This enables study of more complex sequences, such as "high-order" (n-gram, $n > 1$) Markov chains, to address additional questions including the contributions and neural foundations of explicit vs. implicit learning⁴⁴. This has also been a focus in the statistical learning literature, which has classically employed artificial grammar learning^{14,16} paradigms to investigate discrimination (indicated by button press or looking time) of valid vs. violation sequences. Among other things, this work has demonstrated the relevance of statistical learning to human language comprehension^{16,45,46} and explored the language-relevant capacities of nonhumans^{40,47,48}.

Despite the quality and quantity of research on motor and statistical learning, however, important questions remain to be addressed. For example, it is controversial whether structured sequencing is supported by a single, domain-general mechanism^{49–51} or by parallel computations in multiple, modality- or task-specific systems^{9,52}. Similarly, it remains unclear to what extent mechanisms for sequence perception overlap with those involved in sequence production⁵³. Such unresolved questions hinder attempts to determine if observed species differences in sequence learning reflect general cognitive constraints or the particular sensory, motor, and motivational features of different experimental paradigms^{40,54}. We suggest that new approaches to the study of naturalistic behaviors, such as the action grammar extraction methods presented here, may help to address these issues.

Well-controlled artificial experimental paradigms will continue to be an essential tool for progress, but run the risk of producing results that are not generalizable to real-world behavior^{17,18,55}. For example, motor sequence and statistical learning experiments to date have not addressed the complex actions (e.g. bimanual, multi-joint, transitive), reciprocal stimulus–response contingencies (action influences as well as being constrained by context), and multisensory (tactile, kinesthetic, auditory, visual) cue integration that are characteristic of real-world skills. These omissions will be problematic if sequence processing mechanisms are holistic and emergent rather than easily decomposable into modular components^{55,56}. In the statistical learning literature, the paradoxical observation of similar behavior⁵⁴ and partial neuroanatomical overlap⁵⁷ for auditory vs. visual sequence learning combined with a lack of transfer, interference, or within-subject correlation across modalities⁵² suggests this may indeed be the case. Our method for extracting grammars from real-world behaviors addresses this issue by allowing us to quantify structural complexity while maintaining the irreducible complexity of the natural behavior. This creates the prospect of directly investigating the computational demands and neurocognitive substrates of complex sequential behaviors actually exhibited "in the wild" by both humans and non-humans.

In our chosen example of stone tool making, it is not possible to control lower-level kinematic, spatial, and visual features without altering the higher-level action structure that emerges from them and thus compromising the basis for analogy with real, archaeologically documented behaviors⁵⁸. Grammar extraction allowed us to identify the abstract structure emerging from this low-level variability and to relate it back to particular goal-directed actions observable in the archaeological record. Specifically, we found variation in structural complexity to be driven by implementation of a technical operation known as platform preparation, which is important for establishing control over the size, shape, and location of detached flakes, and thus over the form of the finished piece⁵⁹. The earliest currently documented instances of platform preparation date to approximately 500,000 years ago^{28,60} and may be related to the rapid encephalization that also occurred during the late Middle Pleistocene.

Structurally, platform preparation comprises a series of core rotations and repeated percussions embedded within a basic flake removal sequence^{26,36}. Behaviorally, it requires adaptation of kinematic details (Fig. 1B) to different proximate objectives that are themselves defined by attention to subtle material properties of the core^{61,62} that determine fracture patterns. Neurophysiological experiments with artificial paradigms implementing such multiple, context-dependent stimulus–response rules reveal rostral-caudal gradients of abstraction in frontal cortex generally⁶³, and in bilateral inferior frontal gyrus specifically for the selection of action chunks⁵⁰. However, such studies employ discrete sets of simple, predefined perceptual stimuli and motor responses that do not require the multi-level extraction of units (segmentation) as well as relations between units (parsing) from a complex and continuous action stream as is characteristic of natural behaviors like stone tool making or language comprehension. Such issues of perceptual-motor complexity and temporal scale are increasingly relevant to debates about the neurocognitive mechanisms of structured sequence processing^{64–66}.

To explore the potential application of our approach to neurophysiological research, we used HMM and CFG grammar extraction to measure the complexity of action sequences in Oldowan and Acheulean video stimuli from a published fMRI study of tool-making action observation³¹. To generate a continuous complexity measure from HMM, we used the difference in likelihood (measured by Akaike Information Criterion, see methods) between more (6 state) and less (4 state) complex models fit to the stimulus sequences. For CFG, we simply used the compression ratio. Results (Fig. 4) indicate that the two very different HMM and CFG metrics capture partially overlapping stimulus processing demands in the brain.

The conjunction of the two covariates reveals regions of left parietal operculum and rIFG (*pars opercularis*) that are specifically responsive to stimulus complexity irrespective of measurement method. This is consistent with prior evidence of inferior frontoparietal functional activation^{31,32,67–69} and structural remodeling⁷⁰ in response to stone tool making, which have previously been interpreted using informal, qualitative estimations of technological complexity and reverse inference from published studies of other tasks. Parameterization of complexity through grammar extraction now allows direct identification of activity driven by sequence structure and more specific localization to cognitive control^{71,72} and multisensory integration⁷³ regions of inferior frontal and parietal cortex. Together with structural and comparative evidence, this provides functional, archaeological, and neuroanatomical grounding for longstanding hypotheses of tool-language co-evolution^{5,6,33–35,74}. Interestingly, the HMM covariate also picks up more widespread response across occipitotemporal and parietal portions of

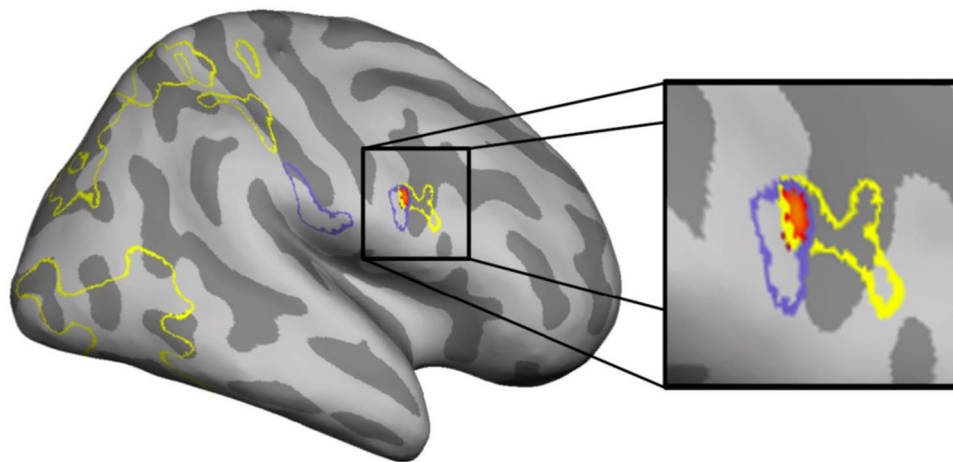


Figure 4. Covariance of BOLD response with tool-making stimulus structural complexity. Outlines corresponding to areas yielding significant correlation with the CFG (purple) and HMM (yellow) covariates describing action sequence complexity. Heatmap clusters represent the minimum of these two correlations where they overlap voxel-wise in the right pars opercularis of the inferior frontal gyrus (right).

a “dorsal attention”⁷⁵ or “multimodal integration”⁷³ network. This suggests that the probabilistic, context-based description of structure provided by HMM may be a better fit to brain activity than the rule-based compression employed by CFG and raises the possibility that action grammar extraction could be used to test explicit hypotheses about neurocognitive mechanisms in the same way that has been done for language comprehension^{76,77}.

The remarkable expressive power of human language derives from an ability to recombine a relatively small set of discrete units into a vast array of meaningful structures^{3,29}. At the outset of the Cognitive Revolution, Lashley⁷ used the example of language to argue that *all* skilled behavior (and associated neural activity) is organized in this hierarchical fashion. This insight was subsequently applied to the specific case of stone tool making by Holloway⁶. Sixty-odd years later, however, we are still struggling with what Lashley identified as the “essential problem of serial order”: defining the “generalized schemata of action which determine the sequence of specific acts” that he termed “the syntax of the act.” Here we developed objective and generalizable methods for defining and quantifying these structures (action grammars) along with their neural correlates from raw behavioral data. While the detailed kinematics of hand actions to produce a tool vary from trial-to-trial considerably^{27,42}, we found an invariant hierarchical structure underpinning performance at the symbolic level of actions. Our analytical approach does not postulate the existence of such “action grammars” a priori, but instead identifies them from raw behavioral data using machine learning techniques, showing that even with the same alphabet of actions qualitatively more complex artefacts can be produced by using measurably more complex action grammars.

This usefulness for analysis is complementary to our recent work on synthesis of behavior, i.e., when learning the generation of complex behavior in robotics and machine learning. In these artificial intelligence domains, we have shown that an action grammar approach to structure the learning algorithm’s representations of behavior does help speed up learning of complex tasks, it confers the ability to transfer learned skills between related tasks (which may often differ in kinematics but less so in their action grammars) and boosts the human-interpretability of the how a robotic system may plan and execute a complex task^{78,79}.

Thus, our action grammar behavior analysis approach’s promises not only broad utility for behavioral and social sciences, but also the finding that our automatic identification of action grammars maps to distinct neural correlates offers the potential for novel quantitative approaches to study hierarchical structure of complex behavior with benefits across many applications from dexterous prosthetics⁸⁰, the training of complex skills in surgery⁸¹ or sports⁸² to human-like artificial intelligence.

Methods

Methods for this study have previously been made available as a preprint⁸³ and are reproduced here with minor clarifications here. De-identified data, descriptors, and code sufficient to produce results published here are available on the FigShare community archive <https://doi.org/10.6084/m9.figshare.14703687.v1>.

Tool replication. Tool making videos analyzed for this study were produced by video recording two research participants experimentally replicating Paleolithic tool forms. This research was approved by the University College London Research Ethics Committee [0603/001], all experiments were performed in accordance with relevant guidelines and regulations, participant identities were not revealed, and each participant provided written informed consent. Participants were both expert stone tool makers with decades of experience. These video-recorded replication experiments included 9 instances of Oldowan knapping and 8 instances of Acheulean knapping. In each experiment, a piece of flint was worked until either completely exhausted (Oldowan) or successfully shaped into a refined handaxe (Acheulean). Six of the handaxes produced in these experiments have previously

been described and compared to archaeological examples for the Middle Pleistocene site of Boxgrove²⁸. Kinematics from a different subset of the experiments (3 Oldowan, 2 Acheulean) have been published²⁷.

Experimental replication is a long-established research method in archaeology, especially with respect to flaked stone technology⁵⁸. Our tool-making experiments drew upon this background to model simple flake production (cf. “Oldowan”, “Mode 1”, “Mode C”⁸⁴, here termed “Oldowan”) and refined handaxe shaping (cf. “Later Acheulean handaxe”, “Mode 2”, “Mode E2”⁸⁴, here termed “Acheulean”). Previous experimentation has shown that a wide range of Oldowan forms may be replicated through hard-hammer free-hand percussion without intentional core shaping⁸⁵, whereas other techniques (e.g. bipolar, passive hammer) produce diagnostic traces that are less common in the archaeological record⁸⁶. Although there is some evidence of structured reduction strategies in the Oldowan (e.g. preference for unifacial vs. bifacial flaking^{87,88}) it is possible to produce most or all Oldowan forms through unstructured (cf. “mindless”⁸⁹ or “least effort”⁸⁵) flaking. We thus instructed our subjects to knap Oldowan experiments in an opportunistic fashion, following the definition of “simple debitage” provided by⁹⁰. For Acheulean experiments, subjects were instructed to produce “refined” Acheulean handaxes of the kind known from the site of Boxgrove (with which subjects were familiar). This included the use of soft hammers and simple platform preparation (faceting), both of which are attested in the Boxgrove archaeological collection²⁸. Experimental handaxes produced were comparable in refinement and debitage morphology to those from Boxgrove²⁸.

Paleolithic tool making occurred over a vast time period and many millions of square miles and encompasses substantial variation that could not be included in our experiments. The methods we did select are considered broadly representative of early and late Lower Paleolithic technology, and details of the production techniques employed match those documented in specific archaeological collections. We thus consider our training protocol to be both generally representative and specifically accurate in re-creating Paleolithic tool-making action sequences.

Event coding. We defined an action alphabet consisting of 7 event types encompassing the elementary body movements and object transformations present in every sequence of both technologies. Events were transcribed from video-recordings using Etholog 2.25⁹¹. Events were defined as follows:

- Percussion: Striking core with percussor (hammerstone or antler billet).
- Target Change: A change in the location of percussion on the core.
- Grip Shift Core: Repositioning of the hand grasping the core.
- Grip Shift Tool: Repositioning of the hand grasping the percussor.
- Inversion: Flipping over the core without otherwise reorienting.
- Flake detach: Removal of a flake (judged to be) > 20 mm.
- Tool Change: Exchange of one percussor for another.

This provides a minimalistic alphabet intentionally designed to limit the need for subjective interpretation and to avoid building prior hypotheses into the coding scheme. In particular, any attempt to infer the intention of the knapper (e.g., identifying a flake detachment as “preparatory” or “thinning”) was avoided. Much richer description of knapping actions in terms of technological function is both possible and informative^{25,92}, but was not in line with our aim to develop a data-driven and generalizable method. The coding scheme was developed through pilot work with the MRI stimulus videos (Table 1 in³¹) and the new replication experiments reported here to be complete (every action on the core or percussor is coded), exclusive (no action could have two codes), and unambiguous. During this pilot work, we removed and/or clarified the criteria of any codes that generated uncertainly or inconsistency during trial coding.

For example, we initially recorded an eighth event type, “Light Percussion” (cf. Fig. 1B), which was not subsequently employed in analysis. This event was defined as “Striking core with percussor using small amplitude arm movements not intended to detach flakes > 20 mm” and was omitted because: 1) it required interpretation, 2) it did not occur in Oldowan sequences, and 3) it might be ambiguous with the Percussion event. Thus, we treated all instances of “light percussion” simply as percussion. However, this gesture—typical of a technical operation known as “striking platform preparation”—was rediscovered by our HMM and Sequitur analyses based purely on sequential structure analysis, thus providing a validation of our iterative approach to developing a reliable and unambiguous ethogram and of our data-driven analytic approach. While the actual alphabet used here is specific to stone tool making, this approach to coding could be generalized to any sequential behavior.

Hidden Markov modeling. We fit Hidden Markov Models (HMM) to the action sequences using the Baum-Welch algorithm implemented in Kevin Murphy’s Bayes’ Net Toolbox. As the algorithm is very sensitive to the initial estimates of the transition and emission matrices, we fit each data set 1000 times for each number of states by randomly varying the initial condition and only picked the HMM with the highest log-likelihood. To compare HMMs with different number of hidden states with each other, we computed the Bayesian Information Criterion (BIC) which gives a measure of model fitness penalised by the number of free parameters in the model.

From the 6 state Acheulean HMM we obtained the most likely state-sequences through the Oldowan and Acheulean action sequences by using the Viterbi algorithm. To investigate whether the obtained hidden state sequence, contained more structure, we fitted a second, 2-state HMM onto the state sequences. As previously described, 1000 runs were performed to obtain the best-fitting HMM. Using the Viterbi algorithm again gives rise to a hidden state sequence within the hidden state sequences, a hidden “States of States” (SoS) sequence.

Deterministic context-free grammar fitting. Any stochastic regular grammar can be represented by a uniquely corresponding HMM where outputs correspond to terminal symbols. Left regular stochastic grammars—because they are strictly equivalent to first order Hidden Markov Models—can only model phenomena with very short memory. Stochastic Context-Free Grammars represent a super-set of stochastic grammars which can feature long term memory and very hierarchical organization.

Sequitur⁴¹ is a recursive algorithm that infers a hierarchical structure in the form of a context-free grammar from a sequence of discrete symbols. We chose Sequitur because we are concerned here with the smallest grammar problem (simplest explanation according to Occam's Razor) which has been applied to musical scores, DNA sequences and are, due to their simple nature, very data efficient⁹³. Other methods are based on data compression approaches which are difficult to apply for short sequences (such as our ethogram data) due to these methods having fixed and data-size dependent information that need to be represented, but the fixed size components outweigh the benefits of the variable length representation^{94–96}.

The Sequitur algorithm constructs a grammar by substituting repeating symbol digrams in the given sequence with new rules and therefore produces a concise representation of the sequence. The algorithm works by scanning a sequence of terminal symbols and building a list of all the symbol pairs which it has read. Whenever a second occurrence of a pair is discovered, the two occurrences are replaced in the sequence by a non-terminal symbol, the list of symbol pairs is adjusted to match the new sequence, and scanning continues. If a pair's non-terminal symbol is used only in the just created symbol's definition, the used symbol is replaced by its definition and the symbol is removed from the defined nonterminal symbols. Once the scanning has been completed, the transformed sequence can be interpreted as the top-level rule in a grammar for the original sequence. The rule definitions for the non-terminal symbols which it contains can be found in the list of symbol pairs. Those rule definitions may themselves contain additional non-terminal symbols whose rule definitions can also be read from elsewhere in the list of symbol pairs.

For example:

Input sequence: *the little cat chases the mouse the little cat catches the mouse the big cat chases the little cat the little cat runs away from the big cat*

Compressed sequence: *r2 chases r3 r2 catches r3 r5 chases r2 r2 runs away from r5*

Grammar:

- *Root -> r2 chases r3 r2 catches r3 r5 chases r2 r2 runs away from r5*
- *r2 -> the little cat (used 4 times)*
- *r3 -> the mouse (used 2 times)*
- *r5 -> the big cat (used 2 times)*

We ran Sequitur on each sequences in both the Acheulean and the Oldowan data sets and enumerated all the rules found across both data sets. After inferring rules from the combined Acheulean and Oldowan data set, we found that some rules only occurred in Acheulean sequences (Supplementary Fig. S2).

The Sequitur algorithm reduces the length of the sequences by replacing terminal symbol strings with aggregating rule strings. This compresses the sequence by reducing its redundancy. Figure 3G shows that sequences in our Oldowan and Acheulean samples share common compressible structure within samples but are distinct across samples. This is indicated by the fact that their pre and post-compression lengths are linear and have distinct slopes. Linear regression fit for Acheulean data is $R^2 = 0.9852$ with slope = 0.13; for Oldowan data $R^2 = 0.9982$ and slope = 0.34. The inverse slope on this plot corresponds to the data compression rate through rule extraction.

Sequitur as a compression algorithm is loss-less, in that reverse applying the rules recovers the original sequence error free, and thus the same information is communicated by fewer symbols. This contrasts with the hidden states of the HMM that only capture probabilistically a higher order structure. A Sequitur compressed sequence must have more information per character and this gain in information density can be quantified using Shannon's entropy measure. Shannon's entropy is computed directly as the log probability of each symbol averaged over all symbols. A sequence with equally probable use of all symbols has the highest entropy, while a sequence using only a single symbol has an entropy of 0. Entropy thus measures how unpredictable a symbol is. We plotted the pre and post compression entropies in Fig. 3H. Pre-compression entropy of Acheulean sequences is considerably lower than that of Oldowan sequences due to the much higher frequency of percussion events. However, post-compression entropy is considerably higher for Acheulean sequences than Oldowan sequence. Thus, pre-compression Acheulean elements (rules + symbols) carry less information than Oldowan elements whereas after compression the reverse is true.

fMRI covariates. In order to generate covariates for fMRI analysis it was necessary to produce continuous measures of complexity for the 20 s video stimuli. For HMM, we first applied the method described above to each stimulus and then evaluated how well the stimulus was explained by the two respective (Acheulean 6 hidden state vs. Oldowan 4 hidden state) HMM models. Sequence length was both short and variable (stimuli were controlled for time rather than number of actions), so we employed the Akaike Information Criterion [AIC] which, unlike BIC, is not directly dependent on sample size in order to avoid confounding sequence length with model likelihood. Differences in AIC between models indicate the relative strength of evidence in their favor. Because our models differ in complexity, this difference provides a continuous measure of how complex (i.e. Acheulean-like vs. Oldowan-like) each short stimulus sequence is compared to models derived from our entire corpus. As a lower AIC indicates a more probable model, decreasing values for Acheulean – Oldowan AIC indicate increasing stimulus complexity and we predict a negative correlation with BOLD response measured by fMRI (cf. Figure 4).

For CFG, we applied the same deterministic grammar extraction approach discussed above. However, in our main analysis, sequitur was applied separately to each sequence. To generate a CFG covariate comparable to our HMM AIC metric, it was necessary to generate a global set of rules derived over the entire corpus to which individual stimuli could be compared. We thus fitted sequitur to the complete set of all sequences in one run. This provided us with a sequitur parse using compressed rules for the entire corpus. We then broke down the compressed rules and matched them to the individual stimulus sequences and computed the basic metrics (as for the long sequences) for these matched compressed sequences. The compression ratio for each stimulus provides a straightforward measure of complexity, we used post- over pre-compression values so that our CFG metric would parallel our HMM metric in matching decreasing values with increasing complexity and predicting negative correlation with BOLD.

fMRI analyses. Experimental paradigm and participants were presented ref.¹⁸. Briefly, 10 Naïve, 10 Trained and 5 Expert subjects observed 20-s videos of an expert demonstrator performing two tool-making methods of differing complexity and antiquity: the simple ‘Oldowan’ method documented by early tools 2.5 million years ago; and a more complex ‘Late Acheulean’ method used to produce refined tools 0.5 million years ago. In the present SPM analysis, the two categories of tool making were defined as two conditions, and complexity scores (HMM and CFG) were added as covariates describing each stimulus in two individual subject analyses.

The effect of these covariates combined across the two categories of stimuli were entered in two multisubject analyses across the 20 non-expert participants, thresholded at $p < 0.05$ FDR-corrected at the cluster level (Fig. 4). Experts were omitted due to a small sample size insufficient to properly assess confounding expertise and automaticity effects^{64,65}. To confirm the overlap in left parietal and right frontal cortices between the two analyses, a conjunction (“&”) was calculated between the T-maps describing the voxels yielding significant negative correlation with the two covariates. This analysis yielded two clusters, one in the left parietal operculum and one in the posterior part of the right inferior frontal gyrus corresponding to the pars opercularis according to the Anatomy toolbox⁹⁷.

Data availability

De-identified data, descriptors, and code sufficient to produce results published here are available on the FigShare community archive <https://doi.org/10.6084/m9.figshare.14703687.v1>.

Received: 1 March 2021; Accepted: 18 June 2021

Published online: 02 July 2021

References

- Fitch, W. & Martins, M. D. Hierarchical processing in music, language, and action: Lashley revisited. *Ann. N. Y. Acad. Sci.* **1316**, 87–104 (2014).
- Lashley, K. in *Cerebral mechanisms in behavior* (ed L. A. Jeffress) 112–136 (John Wiley, 1951).
- Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: What is it, who has it and how did it evolve?. *Science* **298**, 1569–1579 (2002).
- Byrne, R. & Russon, A. E. Learning by imitation: A hierarchical approach. *Behav. Brain Sci.* **21**, 667–721 (1998).
- Greenfield, P. M. Language, tools, and brain: The development and evolution of hierarchically organized sequential behavior. *Behav. Brain Sci.* **14**, 531–595 (1991).
- Holloway, R. Culture: A human domain. *Curr. Anthropol.* **10**, 395–412 (1969).
- Botvinick, M. & Weinstein, A. Model-based hierarchical reinforcement learning and human action control. *Phil. Trans. R. Soc. B Biol. Sci.* <https://doi.org/10.1098/rstb.2013.0480> (2014).
- Duncan, J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010).
- Fedorenko, E., Duncan, J. & Kanwisher, N. Language-selective and domain-general regions lie side by side within Broca’s area. *Curr. Biol.* **22**, 2059–2062 (2012).
- Kolodny, O., Edelman, S. & Lotem, A. Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proc. R. Soc. B* **282**, 20150353 (2015).
- Jenkins, I., Brooks, D., Nixon, P., Frackowiak, R. & Passingham, R. Motor sequence learning: A study with positron emission tomography. *J. Neurosci.* **14**, 3775–3790 (1994).
- Bassett, D. S., Yang, M., Wymbs, N. F. & Grafton, S. T. Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* **18**, 744 (2015).
- Kawai, R. *et al.* Motor cortex is required for learning but not for executing a motor skill. *Neuron* **86**, 800–812 (2015).
- Reber, A. S. Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* **6**, 855–863 (1967).
- Nissen, M. J. & Bullemer, P. Attentional requirements of learning: Evidence from performance measures. *Cogn. Psychol.* **19**, 1–32 (1987).
- Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
- Matusz, P. J., Dikker, S., Huth, A. G. & Perrodin, C. (MIT Press, 2019).
- Ingram, J. N. & Wolpert, D. M. *Progress in Brain Research* Vol. 191 3–29 (Elsevier, 2011).
- Rein, R. & Memmert, D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *Springerplus* **5**, 1410. <https://doi.org/10.1186/s40064-016-3108-2> (2016).
- Forestier, G., Riffaud, L., Petitjean, F., Henaux, P.-L. & Jannin, P. Surgical skills: Can learning curves be computed from recordings of surgical activities?. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 629–636 (2018).
- Schniter, E., Gurven, M., Kaplan, H. S., Wilcox, N. T. & Hooper, P. L. Skill ontogeny among Tsimane forager-horticulturalists. *Am. J. Phys. Anthropol.* **158**, 3–18. <https://doi.org/10.1002/ajpa.22757> (2015).
- Byrne, R. W. *Evolving Insight: How It Is We Can Think about Why Things Happen*. (Oxford University Press, 2016).
- Mahaney, R. A. Exploring the complexity and structure of Acheulean stoneknapping in relation to natural language. *PaleoAnthropology* **2014**, 586–606 (2014).
- Perreault, C., Brantingham, P. J., Kuhn, S. L., Wurz, S. & Gao, X. Measuring the complexity of lithic technology. *Curr. Anthropol.* **54**, S397–S406 (2013).

25. Muller, A., Clarkson, C. & Shipton, C. Measuring behavioural and cognitive complexity in lithic technology throughout human evolution. *J. Anthropol. Archaeol.* **48**, 166–180 (2017).
26. Stout, D. Stone toolmaking and the evolution of human culture and cognition. *Phil. Trans. R. Soc. B Biol. Sci.* **366**, 1050–1059 (2011).
27. Faisal, A., Stout, D., Apel, J. & Bradley, B. The Manipulative Complexity of Lower Paleolithic Stone Toolmaking. *PLoS ONE* **5**, e13718 (2010).
28. Stout, D., Apel, J., Commander, J. & Roberts, M. Late Acheulean technology and cognition at Boxgrove, UK. *J. Archaeol. Sci.* **41**, 576–590 (2014).
29. Belić, J. J. & Faisal, A. A. Decoding of human hand actions to handle missing limbs in Neuroprosthetics. *Front. Comput. Neurosci.* **9** (2015).
30. Hecht, E. E. *et al.* Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Struct. Funct.* <https://doi.org/10.1007/s00429-014-0789-6> (2014).
31. Stout, D., Passingham, R., Frith, C., Apel, J. & Chaminade, T. Technology, expertise and social cognition in human evolution. *Eur. J. Neurosci.* **33**, 1328–1338. <https://doi.org/10.1111/j.1460-9568.2011.07619.x> (2011).
32. Stout, D., Toth, N., Schick, K. D. & Chaminade, T. Neural correlates of Early Stone Age tool-making: Technology, language and cognition in human evolution. *Philos. Trans. R. Soc. Lond. B* **363**, 1939–1949 (2008).
33. Ambrose, S. Paleolithic technology and human evolution. *Science* **291**, 1748–1753 (2001).
34. Engels, F. in *Philosophy of Technology* (eds R. C. Scharff & V. Dusek) 71–77 (Blackwell, 2003 [1873]).
35. Stout, D. & Hecht, E. E. Evolutionary neuroscience of cumulative culture. *Proc. Natl. Acad. Sci.* **114**, 7861–7868. <https://doi.org/10.1073/pnas.1620738114> (2017).
36. Moore, M. W. in *Stone tools and the evolution of human cognition* (eds April Nowell & Iain Davidson) 13–43 (University Press of Colorado, 2010).
37. Petersson, K.-M., Folia, V. & Hagoort, P. What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* **120**, 83–95 (2012).
38. Donoso, M., Collins, A. G. E. & Koechlin, E. Foundations of human reasoning in the prefrontal cortex. *Science* **344**, 1481–1486. <https://doi.org/10.1126/science.1252254> (2014).
39. Frank, S. L., Bod, R. & Christiansen, M. H. How hierarchical is language use?. *Proc. R. Soc. B Biol. Sci.* **279**, 4522–4531 (2012).
40. Jiang, X. *et al.* Production of supra-regular spatial sequences by macaque monkeys. *Curr. Biol.* **28**, 1851–1859. e1854 (2018).
41. Nevill-Manning, C. G. & Witten, I. H. Identifying hierarchical structure in sequences: A linear-time algorithm. *J. Artif. Intell. Res. (JAIR)* **7**, 67–82 (1997).
42. Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. *Nat. Rev. Neurosci.* **9**, 292–303 (2008).
43. Dayan, E. & Cohen, L. G. Neuroplasticity subserving motor skill learning. *Neuron* **72**, 443–454 (2011).
44. Robertson, E. M. The serial reaction time task: Implicit motor skill learning?. *J. Neurosci.* **27**, 10073–10075 (2007).
45. Christiansen, M. H. & Chater, N. *Creating Language: Integrating Evolution, Acquisition, and Processing*. (MIT Press, 2016).
46. Krishnan, S., Watkins, K. E. & Bishop, D. V. M. Neurobiological basis of language learning difficulties. *Trends Cogn. Sci.* **20**, 701–714. <https://doi.org/10.1016/j.tics.2016.06.012> (2016).
47. Fitch, W. T. & Hauser, M. D. Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380. <https://doi.org/10.1126/science.1089401> (2004).
48. Wilson, B. *et al.* Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat. Commun.* **6**, 8901 (2015).
49. Patel, A. D. Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681 (2003).
50. Koechlin, E. & Jubault, T. Broca's Area and the hierarchical organization of human behavior. *Neuron* **50**, 963–974 (2006).
51. Fadiga, L., Craighero, L. & D'Ausilio, A. Broca's area in language, action, and music. *Ann. N. Y. Acad. Sci.* **1169**, 448–458. <https://doi.org/10.1111/j.1749-6632.2009.04582.x> (2009).
52. Frost, R., Armstrong, B. C., Siegelman, N. & Christiansen, M. H. Domain generality versus modality specificity: The paradox of statistical learning. *Trends Cogn. Sci.* **19**, 117–125. <https://doi.org/10.1016/j.tics.2014.12.010> (2015).
53. Martins, M. J., Bianco, R., Sammler, D. & Villringer, A. Recursion in action: An fMRI study on the generation of new hierarchical levels in motor sequences. *Hum. Brain Mapp.* (2019).
54. Milne, A. E., Petkov, C. I. & Wilson, B. Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience* **389**, 104–117. <https://doi.org/10.1016/j.neuroscience.2017.06.059> (2018).
55. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: Correcting a reductionist bias. *Neuron* **93**, 480–490. <https://doi.org/10.1016/j.neuron.2016.12.041> (2017).
56. Bechtel, W. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. (Psychology Press, 2009).
57. Milne, A. E., Wilson, B. & Christiansen, M. Structured sequence learning across sensory modalities in humans and nonhuman primates. *Curr. Opin. Behav. Sci.* **21**, 39–48 (2018).
58. Eren, M. I. *et al.* Test, model, and method validation: The role of experimental stone artifact replication in hypothesis-driven archaeology. *Ethnoarchaeology* **8**, 103–136. <https://doi.org/10.1080/19442890.2016.1213972> (2016).
59. Schick, K. D. & Toth, N. *Making Silent Stones Speak: Human Evolution and the Dawn of Technology*. (Simon & Schuster, 1993).
60. Wilkins, J. & Chazan, M. Blade production ~ 500 thousand years ago at Kathu Pan 1, South Africa: Support for a multiple origins hypothesis for early Middle Pleistocene blade technologies. *J. Archaeol. Sci.* **39**, 1883–1900 (2012).
61. Magnani, M., Rezek, Z., Lin, S. C., Chan, A. & Dibble, H. L. Flake variation in relation to the application of force. *J. Archaeol. Sci.* **46**, 37–49 (2014).
62. Nonaka, T., Bril, B. & Rein, R. How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. *J. Hum. Evol.* **59**, 155–167. <https://doi.org/10.1016/j.jhevol.2010.04.006> (2010).
63. Badre, D. & D'Esposito, M. Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nat. Rev. Neurosci.* **10**, 659–669 (2009).
64. Thiessen, E. D. What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Phil. Trans. R. Soc. B Biol. Sci.* **372**, 20160056 (2017).
65. Hasson, U. The neurobiology of uncertainty: Implications for statistical learning. *Phil. Trans. R. Soc. B Biol. Sci.* **372**, 20160048. <https://doi.org/10.1098/rstb.2016.0048> (2017).
66. Saffran, J. R. & Kirkham, N. Z. Infant statistical learning. *Annu. Rev. Psychol.* **69**, 181–203 (2018).
67. Stout, D. & Chaminade, T. The evolutionary neuroscience of tool making. *Neuropsychologia* **45**, 1091–1100 (2007).
68. Putt, S. S., Wijekumar, S., Franciscus, R. G. & Spencer, J. P. The functional brain networks that underlie Early Stone Age tool manufacture. *Nat. Hum. Behav.* **1**, 0102 (2017).
69. Putt, S. S. J., Wijekumar, S. & Spencer, J. P. Prefrontal cortex activation supports the emergence of early stone age toolmaking skill. *Neuroimage* **199**, 57–69. <https://doi.org/10.1016/j.neuroimage.2019.05.056> (2019).
70. Hecht, E. E. *et al.* Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Struct. Funct.* **220**, 2315–2331 (2015).
71. Aron, A. R., Robbins, T. W. & Poldrack, R. A. Inhibition and the right inferior frontal cortex: One decade on. *Trends Cogn. Sci.* **18**, 177–185 (2014).
72. Levy, B. J. & Wagner, A. D. Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Ann. N. Y. Acad. Sci.* **1224**, 40–62. <https://doi.org/10.1111/j.1749-6632.2011.05958.x> (2011).

73. Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H. & Johnson, K. A. Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *J. Neurosci.* **32**, 10649. <https://doi.org/10.1523/JNEUROSCI.0759-12.2012> (2012).
74. Stout, D. & Chaminade, T. Stone tools, language and the brain in human evolution. *Phil. Trans. R. Soc. B Biol. Sci.* **367**, 75–87. <https://doi.org/10.1098/rstb.2011.0099> (2012).
75. Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
76. Brennan, J. Naturalistic sentence comprehension in the brain. *Lang. Linguist. Compass* **10**, 299–313. <https://doi.org/10.1111/lnc3.12198> (2016).
77. Nelson, M. J. *et al.* Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl. Acad. Sci.*, 201701590 (2017).
78. Lange, R. T., & Faisal, A. Semantic RL with Action Grammars: Data-Efficient Learning of Hierarchical Task Abstractions. NeurIPS WS Deep Reinforcement Learning (also arXiv preprint [arXiv:1907.12477](https://arxiv.org/abs/1907.12477)) (2019).
79. Beyret, B., Shafti, A., & Faisal, A. A. Dot-to-Dot: Explainable hierarchical reinforcement learning for robotic manipulation. IEEE IROS (also arXiv preprint [arXiv:1904.06703](https://arxiv.org/abs/1904.06703)) (2019).
80. Xiloyannis, M., Gavriel, C., Thomik, A. A. & Faisal, A. A. Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1785–1801 (2017).
81. Reznick, R. K. & MacRae, H. Teaching surgical skills—Changes in the wind. *N. Engl. J. Med.* **355**, 2664–2669. <https://doi.org/10.1056/NEJMra054785> (2006).
82. Haar, S., van Assel, C. M. & Faisal, A. A. Motor learning in real-world pool billiards. *Sci. Rep.* **10**(1), 1–13 (2020).
83. Stout, D., Chaminade, T., Thomik, A., Apel, J. & Faisal, A. A. Grammars of action in human behavior and evolution. *bioRxiv*. <https://doi.org/10.1101/281543> (2018).
84. Shea, J. J. Lithic modes A-I: A new framework for describing global-scale variation in stone tool technology illustrated with evidence from the east mediterranean levant. *J. Archaeol. Method Theory* **20**, 151–186 (2013).
85. Toth, N. The Oldowan reassessed: A close look at early stone artifacts. *J. Archaeol. Sci.* **12**, 101–120 (1985).
86. Harmand, S. *et al.* 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature* **521**, 310–315 (2015).
87. Stout, D., Semaw, S., Rogers, M. J. & Cauche, D. Technological variation in the earliest Oldowan from Gona, Afar, Ethiopia. *J. Hum. Evol.* **58**, 474–491. <https://doi.org/10.1016/j.jhevol.2010.02.005> (2010).
88. Stout, D., Rogers, M. J., Jaeggi, A. V. & Semaw, S. Archaeology and the origins of human cumulative culture: A case study from the earliest Oldowan at Gona, Ethiopia. *Curr. Anthropol.* <https://doi.org/10.17605/OSF.IO/UUVBVW> (2019).
89. Moore, M. W. The design space of stone flaking: Implications for cognitive evolution. *World Archaeol.* **43**, 702–715 (2011).
90. Inizan, M.-L., Reduron-Ballinger, M., Roche, H. & Tixier, J. *Technology and Terminology of Knapped Stone*. (C.R.E.P., 1999).
91. Ottoni, E. B. EthoLog 2.2: a tool for the transcription and timing of behavior observation sessions. *Behav. Res. Methods Instrum. Comput.* **32**, 446–449 (2000).
92. Roux, V. & David, E. in *Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behaviour* (eds Valentine Roux & Blandine Brill) 91–108 (McDonald Institute for Archaeological Research, 2005).
93. Nevill-Manning, C. G. & Witten, I. H. Identifying hierarchical structure in sequences: A linear-time algorithm. *J. Artif. Intell.* **7**, 67–82 (1997).
94. Kieffer, J. C. & Yang, E. H. Grammar-based codes: A new class of universal lossless source codes. *IEEE Trans. Inf. Theory* **46**(3), 737–754 (2000).
95. Yang, E. H. & Kieffer, J. C. Efficient universal lossless data compression algorithms based on a greedy sequential grammar transform—Part one: Without context models. *IEEE Trans. Inf. Theory* **46**(3), 755–777 (2000).
96. Kieffer, J. C. & Yang, E. H. Grammar based codes: A new class of universal lossless source codes. *IEEE Trans. Inf. Theory* **46**(3), 737–754 (2000).
97. Eickhoff, S. B. *et al.* Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *Neuroimage* **36**, 511–521 (2007).

Acknowledgements

This work was supported by the Commission of the European Communities Research Directorate-General Specific Targeted Project Number 029065, “Hand to Mouth: A framework for understanding the archaeological and fossil records of human cognitive evolution” and National Science Foundation (USA) Grant SMA-1328567 and a UKRI Turing AI Fellowship Grant EP/V025449/1 (AAF). The individual depicted in Fig. 1A is the first author (DS). We thank Andreas Thomik for his work developing and applying pattern recognition methods, Francois Belletti for software development support, and James Steele and Daniel Wolpert for support and encouragement that made this project possible at the start.

Author contributions

D.S. conceived the study and conducted the replication experiments. T.C. analyzed brain imaging data. A.A.F. developed the action grammars and behaviour analytics methods. AAF and AS analysed the data. JA contributed to ethogram development and coded videos. DS, AAF, and TC wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92992-5>.

Correspondence and requests for materials should be addressed to D.S. or A.A.F.

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

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021