**Title:** The Functional Brain Networks that Underlie Early Stone Age Tool Manufacture

**Authors:** Shelby S. Putt*[1,2], Sobanawartiny Wijeakumar[3], Robert G. Franciscus[4], John P. Spencer*[3]

[1]The Stone Age Institute, 1392 West Dittemore Road, Gosport, IN 47433, USA

[2]Center for Research into the Anthropological Foundations of Technology, Indiana University, Bloomington, IN, USA.

[3]School of Psychology, University of East Anglia, Lawrence Stenhouse Building 0.09, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK

[4]Department of Anthropology, University of Iowa, 114 Macbride Hall, Iowa City, IA 52242-1322, USA

*Corresponding authors: ssputt@indiana.edu, J.Spencer@uea.ac.uk

**Text:**

After 800,000 years of making simple Oldowan tools, early humans began manufacturing Acheulian handaxes around 1.75 million years ago (Ma). This advance is thought to reflect an evolutionary change in hominin cognition and language abilities. We used a neuroarchaeology approach to explore this hypothesis, recording brain activity via functional near-infrared spectroscopy (fNIRS) as modern human participants learned to make Oldowan and Acheulian stone tools in either a verbal or nonverbal training context. Here we show that Acheulian tool production requires the integration of visual, auditory, and sensorimotor information in the middle and superior temporal cortex, the guidance of visual working memory (VWM) representations in the ventral precentral gyrus (PrG), and higher-order action planning via the supplementary motor area (SMA), activating a brain network that is also involved in modern piano playing. The right analogue to Broca's area–which has linked tool manufacture and language in prior work[1,2]–was only engaged with verbal training. Acheulian toolmaking, therefore, may have more evolutionary ties to playing Mozart than quoting Shakespeare.

The human brain has increased in absolute and relative size in the last 2-3 million years, particularly the prefrontal and temporal cortices[3]. This increase in brain size undoubtedly coincided with the evolution of the distinctive features of modern human cognition[4]. Understanding the link between brain evolution and cognition remains a key scientific challenge because it is impossible to observe the functional brain activity of extinct human species to know how their brains operated. An innovative approach to this challenge is offered by the field of neuroarchaeology. Here, the idea is to use neuroscience methods and theories to investigate the evolution of the brain and cognition by capitalizing on the remnants of past material culture from the archaeological record[5].

Around 1.75 Ma, there was a revolutionary innovation in stone tool technology. Early *Homo* began to incorporate the bifacial, shaped core tools (handaxes and cleavers) of the Acheulian industry into their pre-existing repertoire of simple Oldowan flake and pebble tools (Fig. 1d). The addition of these multipurpose tools to the hominin toolkit allowed for the exploitation of a wider variety of energy-dense food items through functions such as butchery, woodworking, and digging[6,7,8]. The differences in technological complexity between these two tool types could be indicative of a shift in cognition and language abilities from a more ape-like to a human-like state. What changes in brain and cognition might have led to this advance?

Evidence suggests that the earliest members of the genus *Homo* display a trend toward a human-like organization of the brain[9]. Given this, the functional brain activity of modern humans as they reproduce the stone toolmaking behaviours of extinct hominins can shed light on the functional brain activity of these first hominin toolmakers. This is, of course, an inexact science. For instance, we cannot know to what extent the cognitive operations of modern humans resemble those of early humans, nor can we pinpoint the effect of modern culture and formal education on the cognitive operations of modern humans during toolmaking tasks. We can assume, however, that extinct hominins producing the same tool types and using the same operational sequence as modern humans likely possessed at least the minimum cognitive operations that modern humans use to complete the task.[10] Thus, the functional brain activity of modern humans can tentatively be used to infer the functional brain activity of earlier human species. Here, we examine functional brain activity as modern humans learned to make Early Stone Age (Oldowan and Acheulian) tools to shed light on the brain networks and cognitive skills that were needed to complete these tasks (Fig. 1a-c).

[Insert Fig. 1]

Acheulian stone tool manufacture is hypothesized to require more cognitive control and working memory than Oldowan tool manufacture[11]. This is because shaping a stone into a handaxe and maintaining a sharp edge along the entire piece (see Fig. 1d) requires the toolmaker to proceed through a series of complex action sequences that have an ambiguous goal hierarchy[12,13]. Nevertheless, activation of working memory neural circuits has been purportedly absent during replicative Oldowan and Acheulian tool production experiments[1,14] (but see Supplementary Fig. 1 and Supplementary Discussion). This may reflect the challenges of using neuroimaging techniques like functional magnetic resonance imaging (fMRI) to capture real-time brain activity during the act of making stone tools. For example, two fMRI studies attempted to simulate tool production by having participants observe videos of the toolmaking process[11,15], rather than actually knapping. These studies might have underestimated the role of working memory circuits because participants did not have to hold complex action sequences actively in mind during the imaging task.

Researchers have also hypothesized that a special co-evolutionary relationship exists between toolmaking and language because the earliest stage of stone toolmaking skill transmission appears to improve with verbal instructions[16]. Also, studies using positron emission tomography (PET), fMRI, and functional transcranial Doppler ultrasonography revealed that both behaviours activate overlapping brain regions and present similar cerebral blood flow lateralization signatures[1,14,15,17]. This suggests that language may have piggy-backed on the motor and hierarchical processing functions sub-served by the ventral precentral and inferior frontal gyrus (IFG), brain areas critically involved in Acheulian tool manufacture[2]. Because the learning context was not carefully controlled in these neuroimaging studies, however, this overlap could be the product of learning to knap by receiving verbal instructions from an interactive teacher. It is possible, for instance, that participants in these studies relied upon internal speech, recalled verbally delivered instructions, or enlisted specific language-based behavioural strategies because they learned with language instruction[18]. This may not mimic the learning context that existed during the Pleistocene when hominins likely did not possess modern language or the cognitive elements required for interactive teaching.

In the present study, we tested these hypotheses, examining the brain networks that underlie Early Stone Age toolmaking by firstly, using image-based fNIRS, a cutting-edge neuroimaging technique that measures changes in oxygenated and deoxygenated haemoglobin (oxy-Hb, deoxy-Hb) in the cortex. This approach produces reconstructed images of localized functional brain activity that can be directly compared to fMRI results[19,20]. Because fNIRS is less influenced by motion artefacts than fMRI, it was possible to use fNIRS to measure real-time, localized cortical activity as people made Oldowan and Acheulian tools. We predicted that fNIRS would detect a relative increase in the activation of brain areas involved in cognitive control and working memory during Acheulian tool production when contrasted with Oldowan tool production.

Secondly, we carefully controlled the learning context. We taught 31 participants to make both types of tools across seven learning sessions (Fig. 1e). During individual training sessions, fifteen of the participants learned to knap stone via *verbal* instruction by watching videos of a skilled knapper's actions as he demonstrated and explained how to knap (his face was not visible); sixteen of the participants learned to knap via *nonverbal* instruction using the same instructional videos, but with the sound turned off. Brain activity was measured while participants completed a motor baseline task that involved striking two rocks together without attempting to make flakes, as well as during an Oldowan task and an Acheulian task. We predicted that the two learning groups would show different neural activation patterns, with selectively greater activation in language-specific brain areas, including the right IFG, in the verbal instruction condition.

A two-way analysis of variance with Task (Oldowan, Acheulian) and Group (verbal, nonverbal) as factors (see Supplementary Discussion) replicated the Acheulian-biased activation in the left ventral PrG from previous PET research[1] (Fig. 2a). This area forms part of the VWM network[19] (see overlap between dark green and red in Fig. 2a). Working memory is not a uniquely human feature, but modern humans have been argued to possess an "enhanced working memory" that did not evolve until the late Pleistocene – much later than the onset of Acheulian tool manufacture[21]. Our findings suggest that even stone tool industries as ancient as the early Acheulian required working memory.

The analysis also revealed novel areas of activation associated with Acheulian toolmaking, including middle and superior temporal areas (Fig. 2b-c), as well as the SMA (Fig. 2d). The temporal areas are involved in complex sound processing, auditory short-term memory, and the integration of visual, auditory, and sensorimotor information in relation to tool use[22,23,24]. The SMA forms the cognitive control centre of a medial premotor system whose function is to plan complex action sequences, especially those requiring bimanual coordination[25]. The superior temporal gyrus (STG), middle temporal gyrus (MTG), and SMA are connected via white fibre tracts that coalesce at the insular cortex[26], which plays a critical role in guiding behaviour via attentional modulation[27]. While blood oxygenation concentrations in the insular cortex are too deep to record with fNIRS, this area has been implicated in stone tool production in previous work[15].

Acheulian toolmaking depends on the execution of a skilled striking platform setup to plan the direction, shape, and size of a series of flakes that will effectively thin and shape the piece[28,29]. The activation of bilateral temporal areas during the Acheulian task may signify that participants were holding in mind the varying sounds of impact to judge whether a platform was successfully prepared for removal of a flake. The ability to plan and execute a flexible sequence of actions to make a handaxe could be accomplished by integrating the working memory component of left ventral PrG with the complex motor planning of SMA and the auditory feedback and multimodal processing of STG and MTG via the insular circuit. Interestingly, this cognitive network is nearly identical to one that is active when trained pianists play the piano[30], consistent with our proposal that this network is critical for audiomotor integration. The relatively weak Oldowan activation in this network is also informative. In the Oldowan task, each strike is an independent event that attempts to create a flake with a sharp edge; there is little need to actively hold in mind a long chain of actions to meet the overarching goal of the task.

The ANOVA revealed four clusters where the instruction context had an effect on cortical activation during the toolmaking tasks (Supplementary Table 2). Post-hoc tests identified two areas where the Acheulian task significantly varied by group. A large cluster that includes the right temporal pole and *pars orbitalis* was activated in the nonverbal group and suppressed in the verbal group (Mann-Whitney $U = 55.0$; $p = 0.009$; Fig. 2e). The right temporal pole is a multimodal association cortex involved with semantic processing[31] and has strong connections to *pars orbitalis* and the insula[32]. The right orbital portion of the prefrontal cortex is known to be involved in decision-making and reward-related feedback[33]. This may indicate that the nonverbal group relied more extensively on auditory and visuospatial feedback while planning actions related to handaxe production. Post-experiment interviews support this claim. Only participants in the nonverbal group emphasized sound and tactile sensation as important to their thought process while knapping. Their descriptions also mentioned visuo-spatial imagery more often than descriptions produced by the verbal group.

The second cluster, *pars triangularis* of the right IFG, had significantly higher activation in the verbal group than the nonverbal group during the Acheulian task (Mann-Whitney $U$ = 198.0; $p$ = 0.001; Fig. 2f). This right hemisphere analogue to Broca's area participates in language functions, such as syntactic and sentence processing, especially in relation to context[34], as well as some non-language functions, such as response inhibition[35]. This suggests that participants who received verbal instruction may have engaged in inner speech during the Acheulian task, which is supported by post-experiment interviews (Supplementary Fig. 2). Critically, this cluster overlapped with the IFG cluster from previous work that led to the conclusion that language may have co-opted the neural circuits involved in toolmaking[1] (see yellow region in Fig. 2f). If language evolved by co-opting the motor areas of the brain that were used first for Early Stone Age tool manufacture, then we should observe activation of the right IFG in both groups as a result of the complex knapping task. Because this area shows elevated activation only among the verbal group participants, this suggests that language instruction in the modern learning context is responsible for right IFG activation in this and previous studies. Caution is urged, therefore, when interpreting results of neuroarchaeological studies that do not control for spoken language in the learning context.

[Insert Figs. 2-3]

Unique cortical areas recruited during the Oldowan task include the hand representation portions of the primary sensorimotor cortex in both hemispheres (Fig. 3a-b). This suggests the involvement of a lateral premotor system, which is dependent on external visual input to recognize and assign significance to external objects[25]. This is unsurprising, as the only goal of the Oldowan task is to visually identify ideal platforms and remove flakes until the core is exhausted. An evaluation of the video footage captured during the experiment and participant responses during an exit interview reveal that the absence of activation in these hand areas during the Acheulian task might have resulted from participants using the leg rather than the hand as a support for the core. Participants also took their time to evaluate progress more often during the Acheulian task than during the Oldowan task, which could have resulted in less activation in the hand motor areas.

The Oldowan task also appears to come under increased cognitive control when it has been learned in the absence of verbal instruction (Fig. 3c). For example, it is only in the nonverbal group that the left MFG, or frontal eye field, is activated (Mann-Whitney $U = 33.0$; $p < 0.001$). This area–also activated in the study by Stout et al.[1] (see yellow cluster in Fig. 3c)–forms part of the dorsal visual attention network[36]. The recruitment of this network only in the nonverbal condition suggests that learning to produce simple flakes without language requires increased attention to visuospatial demands. When learned verbally, Oldowan tool production elicits activity in the left dorsal PrG (Fig. 3d), an area that also is activated when passively reading action words related to the arm[37].

Considered together, our findings suggest that Oldowan tool manufacture relies on the coordination of visual attention and motor control to successfully remove simple flakes. It would not be surprising to find that a homologous cognitive network is active in wild chimpanzees when they skilfully crack nuts with stone tools[38], or even in capuchin monkeys when they strike two stones together, which can sometimes lead to unintentional flakes similar to those made by early hominins[39]. In sum, results of this experiment point to cognitive abilities that were more ape-like than human-like among hominin toolmakers prior to 1.8 Ma.

Acheulian tool manufacture, in addition, requires the integration of higher-order motor planning, working memory, and auditory feedback mechanisms to attend to information from multiple modalities as the toolmaker coordinates the different goals required by this more complex task. We propose that, like the processing of an auditory speech stream, Acheulian knapping requires the knapper to discriminate between knapping sounds and to assign meaning to those sounds based on how they relate to the hierarchy of goals involved in making a handaxe (e.g., how does this strike and its associated sound get me closer to setting up an ideal platform to remove a flake that will be long and thin enough to remove this nearby convexity; how does this strike and its associated sound relate to the overall shape of the handaxe that I am trying to achieve). Thus, the knapping of Acheulian tools may have played a role in fine-tuning this function in the STG, perhaps facilitating the evolution of neural connections involved in speech perception. Interestingly, the Acheulian technocomplex coincides in timing with the evolution of a derived middle ear anatomy in *Homo* that was more attuned to human speech frequencies[40,41]. Together, fossil and neuroarchaeological evidence now reveal that a major shift in hominin auditory processing occurred after *Homo* diverged from *Australopithecus* and *Paranthropus* and before the appearance of *H. heidelbergensis.*

The adoption of the Acheulian toolkit by early *Homo* also coincides in time with a more unpredictable environment, an increase in brain and body size, and a more diverse diet that relied upon tool-assisted hunting and foraging of large game animals and tough, fibrous plant products[42]. As reliable food items became scarcer in this unpredictable environment, those individuals who were capable of holding in mind multiple modes of information to guide and coordinate their motor behaviours likely experienced higher reproductive success because of their enhanced ability to produce complex tools. We speculate that this ability allowed these individuals and their offspring greater access to a diverse set of food resources.

Our findings do not neatly overlap with prior claims of a technological origin for language. There is more support for a working memory hypothesis, as VWM plays an active role in the network identified here that today allows modern humans to perform such behaviours as skilfully playing a musical instrument. Our data suggest that this cognitive network was probably necessary for early *Homo* to make Acheulian handaxes and might also have been important for other learned, complex behaviours. Additionally, a larger working memory capacity may have led to more complex imitative abilities, as Arbib has suggested[43]. We propose that selection for this integrated, multimodal network around 1.8 Ma in response to an unpredictable environment marked a turning point in the evolution of the hominin brain, leading to the expansion of prefrontal and temporal cortices[3], a more complex cognitive toolkit, and the evolution of a new species of *Homo.*

## Methods

*Experimental Design, Participants, and Procedure*

An *a priori* power analysis was performed for sample size determination based on data from a pilot study, comparing verbal with nonverbal instruction. Beta values from fifteen 20-s intervals of knapping were extracted from a channel that overlies anterior Broca's area. The effect size (Cohen's $d$ = 1.13) was considered to be large using Cohen's criteria[44]. With an alpha = 0.05 and power = 0.80, the projected sample size needed with this effect size is approximately 14 subjects per group[45].

Participants were recruited for the study via posted flyers that advertised for individuals interested in learning to make stone tools. Any persons interested in participating in the study received an online questionnaire that determined their eligibility to participate. They were screened for knapping experience, handedness, neurological, psychiatric, and physical handicaps, and drug use. Only individuals with no prior experience making stone tools were asked to participate. Because of evidence for abnormal language lateralization in left-handed and ambidextrous individuals[46], the Benton Neuropsychology Clinic Handedness test was administered during the screening process to determine the laterality quotient of potential subjects[47]. Only subjects who fell within the range of +75 - +100 points, or extreme right-handedness, were included in the experiment.

After positively demonstrating right-hand dominance and consenting to participate, subjects were asked about their psychiatric and neurologic history. Persons who had experienced traumatic brain injury (including stroke, anoxia and hypoxia, brain tumour, infections of the brain, etc.), loss of consciousness, a history of seizures, or severe learning disability were not included. Individuals with serious psychiatric disorders, such as autism, were excluded from the study. Additionally, the Drug Abuse Screen Test (DAST-10) was included to quantify the degree of drug abuse problems of potential subjects[48]. Individuals with a recent history of drug abuse show impairments in cognitive tasks[49]. Only persons who received a score of 2 or lower were permitted to participate. The study was approved by the IRB and Human Subjects Office at the University of Iowa (IRB ID #: 201304789), and all subjects signed an informed consent document prior to participating.

Participants were divided into two groups based on their performance during a manual dexterity test so that dexterity levels were equally distributed across groups. One group received verbal instruction while learning how to knap stone ($N$ = 15; 8 females, 7 males), and the other group received nonverbal instruction only ($N$ = 16; 8 females, 8 males). Manual dexterity was measured using the Minnesota Manual Dexterity Test (MMDT). This test assesses the manual dexterity required to place sixty round pegs with the dominant hand in specific places on a board[50]. While it is often used by physical and occupational therapists to determine baseline progress data from an injured patient, the MMDT has also proven to be a reliable and valid method for obtaining measures of manual dexterity in healthy adults[50,51]. For the final sample of included participants, the nonverbal group averaged 182.4 ± 17.5 s to place all sixty pegs in the holes on the board in three iterations, while the verbal group averaged 182.7 ± 16.9 s. There was no significant difference in dexterity between the two groups based on this assignment ($t = 0.06$, $p = 0.95$). Males, who averaged 181.4 ± 14.2 s, and females, who averaged 183.6 ± 19.5 s, also did not significantly differ from each other in their dexterity scores ($t = -0.34$, $p = 0.74$).

After screening and group assignment, participants attended their first practice session. One participant dropped out of the study halfway through this first session. Four additional participants were withdrawn after their first neuroimaging session because they had dark or thick hair that interfered with our ability to obtain high-quality NIRS signals. Finally, two subjects withdrew from the study before the final neuroimaging session for personal reasons. The final sample had 31 participants (nonverbal = 16, verbal = 15; 16 females, 15 males; age 24.0 ± 8.1 years [mean ± SD]) who completed the entirety of the experiment.

The participants individually attended seven 60-min knapping practice sessions, during which they learned how to knap stone tools by watching instructional videos. We chose video instruction rather than in-person instruction to ensure that every subject received the exact same instructions at the same rate and also to control for interactive teaching, as there is currently not enough evidence to confirm that early *Homo* was capable of interactive teaching. The videos featured an expert knapper with over 12 years of experience. His face was not visible in the frame, though his hands, lap, and torso were visible. This prevented the nonverbal group from picking up on any verbal cues that were communicated by the face. Both groups watched the same instruction videos; however, the nonverbal group watched a silent version. Each practice session proceeded in the following order: 1) a 10-min instruction video; 2) 20 min of practice; 3) the same 10-min instruction video; and 4) 20 min more to practice. Subjects were not able to manipulate the video in any way, for example, by pausing it. All the debitage created while knapping fell on a large tarpaulin mat. After the participants completed a core or core tool and were ready to move on to another rock, the core/core tool and its corresponding debitage were collected, bagged, and labelled with the rock number and other pertinent information for further analysis.

Each practice session introduced a new goal for them to meet, or reviewed and refined skills already introduced. The skills and tool types learned during practice sessions 1 and 2 were comparable to the skills and tool types of Oldowan simple tool production. This is a quick and expedient method of obtaining a sharp flake to use as a tool[52]. They learned how to recognize ideal striking angles on the raw material and tried to create flakes. They continued to practice making expedient flakes during the second practice session. The second video taught them how to recognize the best raw material for flaking. Subjects learned which materials fracture easily by trial and error. This also was communicated verbally to the verbal group. Practice sessions 3-7 introduced and reviewed skills involved in the production of the early Acheulian technocomplex, which involves a more efficient removal of flakes and the intentional shaping of a large cutting tool[53]. The third practice session video featured alternate flaking around a square edge as the main goal for this session, which is an important skill for making bifaces. The instruction video for the fourth practice session introduced core bifaces, and the instructor in the video demonstrated biface manufacture at a very slow rate. In the fifth practice session, the video began to focus more on primary thinning of a piece to remove large convexities. The sixth instruction video presented information on how to shape and refine a biface by trimming. Finally, the subjects were presented with an instruction video during the seventh practice session that focused on the entire process of bifacial reduction so that they could continue to practice the skills they learned from prior sessions.

For all practice and neuroimaging sessions, subjects were required to wear safety goggles, leather work gloves, and lap pads. They were also given the choice to wear a facemask to block out small particles of airborne silicates.

In addition to the training sessions, participants attended three 90-min neuroimaging sessions after the first, fourth, and seventh training sessions, during which they were video recorded and brain activity was observed using the TechEn CW6 system. They sat in a small room surrounded by black curtains. The experiment program was designed with EPrime software. The presentation of stimuli was synchronized with the CW6 system. Set-up involved measuring the participant's head to ensure the proper cap size, and measuring 10-20 landmarks to ensure proper cap placement on the head. Hair was cleared at each optode site. The 10-20 landmarks and positions of the sources and detectors on the head were then digitized.

Each imaging session consisted of 1) a motor baseline task made up of 9 40-s blocks of activity segregated by 20-s rest periods to observe activation of motor-related brain areas while striking rocks together without the added element of actual knapping; 2) an Oldowan toolmaking task that was segregated into five 1-min blocks of activity with 15-s resting periods in between each block; and 3) an Acheulian toolmaking task segregated into fifteen 1-min blocks, separated by 15-s rest periods. The order of the tasks was not randomized during each imaging session nor was the length of resting periods; thus, there is some possibility that habituation effects impacted our results. These limitations should be addressed in future studies.

To eliminate the possibility of linguistic contamination, the experiment was designed so that all instructions were given via silent video with timing of events indicated by different tones, and subjects were instructed to not talk during the experiment. Subjects were told at the beginning of the experiment to perform the same activity that they viewed in the instruction videos, which preceded each new task or event. Instructions also included training on the meanings of different tones they would hear throughout the session that would signal whether to stop or start an action. Only data from the final neuroimaging session are included here because this was the first point when more than 90% of the surveyed participants were able to identify the different goals of the Oldowan and Acheulian tasks.

At each practice and neuroimaging session, subjects were presented with three or four local, granitic rocks of varying sizes that were naturally rounded for use as hammer stones. A goal of the training was to introduce the subjects to different qualities, shapes, and types of rock to fracture so that they would learn to select the blank of highest quality and the most workable edges from the three choices that they were always provided. Thus, a variety of unheated cherts from the Midwestern United States, Texas, and California were obtained from collectors in Missouri and Texas, though most of the material was Burlington chert, a fine- to medium-grained stone that is easy to flake[54]. Prior to being made available for the subjects to knap, each stone was assigned a unique, identifying label, weighed on a digital scale, and assigned a measurement of volume by the water displacement method. Spalls and cobbles ranged between 69.6 and 3000.0 g in mass (mean = 676.8 g) and had a volume between 20 and 1200 $cm^3$ (mean = 284.3 $cm^3$). Generally, smaller pre-made spalls of chert with edges of very acute angles were provided in the first two practice sessions. By the third and fourth practice sessions, the participants could choose from medium-sized spalls without cortex that had edges with more difficult angles, as well as rounded cobbles with cortex but with one or more flakes already removed to help them get started. A mix of small- to medium-sized spalls and cobbles were available to choose from for the Oldowan task during the neuroimaging sessions. Larger, more challenging pieces, many with square edges, were provided for the fifth, sixth, and seventh practice sessions and the Acheulian task during the neuroimaging sessions.

*Behavioural Data Acquisition and Processing*

A key issue when comparing different groups in neuroimaging studies that measure changes over learning is that participants might learn at different rates depending upon their group assignment. To examine this possibility, digital callipers were used to take measurements on cores and flake debris from both knapping tasks during the final neuroimaging session to determine whether one of the learning groups produced stone tools with greater skill than the other group (see Supplementary Discussion). All core and debitage pieces were collected after the completion of each finished core during the neuroimaging session. Any debitage that passed through a ¼" screen was discarded. The remaining pieces were labelled and measured. Each piece was weighed to the nearest tenth of a gram and allocated to a metric size category continuum as defined by the smallest of a series of nested squares on centimetre graph paper into which the piece would completely fit (i.e., 1 $cm^2$, 2 $cm^2$, 3 $cm^2$.., etc.). The maximum thickness was recorded for each piece. All non-core debitage was coded as a flake (either complete, proximal, or distal) or nonflake debitage shatter[55]. Any flakes with an intact striking platform underwent measurements for the maximum platform width and thickness.

These measurements were applied to a total of 5,757 debitage pieces that correspond to 72 cores, which were reduced by 30 of the participants in the study[*]. Relative knapping skill as determined by the debitage was measured using the following variables. The first set of variables

[*] Debitage output from the final neuroimaging session for one participant was not available for analysis.

measured correspond to flake and platform shape. Platform shape, determined by the ratio of maximum platform width to platform thickness, is a common method used to measure knapping skill[18,28,56], as platform shape contributes to the size and shape of the overall flake. The ratio of flake size to flake mass was also included to determine flake shape differences between the groups[18,56]. A larger ratio in both cases signifies a flake that is both relatively thin and elongated, which demonstrates the knapper's ability to remove desired flake tools in the case of the Oldowan task and long, thinning flakes for shaping the core tool in the case of the Acheulian task. We calculated the relative platform area ([platform width*platform thickness]/flake size) with the expectation that knappers of a higher skill level would produce smaller, thinner platforms relative to the size of the rest of the flake[28].

The second set of variables measured correspond to the efficient use of raw material, as inefficient use of raw material is indicative of low skill level[57]. We examined the proportion of intended flakes to unintended shatter fragments, both on low quality and high quality material[18,56], with the expectation that the assemblages of relatively more skilled knappers would include a higher percentage of flakes than the assemblages of less skilled knappers, demonstrating better control of the material. We also examined the proportion of whole flakes to flake fragments. Previous experimental research demonstrated that the assemblages of skilled knappers included more flake fragments than the assemblages of less skilled knappers, perhaps a result of skilled knappers striking the core at a higher velocity[56]. A clear sign of knapping skill in the case of the Oldowan task is the level of reduction of the cobble into usable flakes[56]. We measured this by determining the proportion of the original cobble's mass into flake, shatter, and unexploited core mass, with the expectation that the more skilled knappers would have a larger percentage of flake mass and a smaller percentage of unexploited core mass. Finally, we examined the relative number of missed strikes on cores and debitage (total number of missed strikes/original cobble mass), which can be observed as incipient cones of percussion, micro-flake scars or battered edges, and hammer stone marks[18]. While it is impossible to get an exact count of missed strikes by looking at the lithics alone, if one group were to have a higher number of missed strikes than the other, this would be indicative of less skill, signifying less manual control.

Forty-nine core tools (attempted bifaces) from the Acheulian task were analysed. Along with the measures described above, core tools were determined to be bifaces by the presence of two opposing faces and at least one bifacial edge. A bifacial edge is defined as any sharp edge that has been created by removing flakes near the same location that run across opposite planes of the stone. This would require the knapper to strike off one flake and then flip the piece over and use the newly created angle to remove a second flake, a technique known as alternate flaking. The proportion of successful bifaces was determined by dividing each group's total number of successful bifaces by the group's total number of attempted bifaces. The maximum breadth and thickness of each successful biface were recorded with digital callipers. The ratio of biface breadth to thickness is informative about the level of biface refinement, such that a refined handaxe should have a larger breadth relative to thickness, which would present as a larger ratio[28].

At the conclusion of the experiment, participants were asked questions related to their experience in the experiment, and their answers were recorded. Specifically, they were asked what they thought the goals were for the knapping tasks and whether or not they believed they achieved these goals. They were asked to explain how the two knapping tasks differed from each other, at what point in the experiment they understood there were differences between the two knapping tasks, and whether or not they used different strategies to achieve the different goals of the two tasks. They explained what they were generally thinking about while knapping, whether or not these thoughts included language. Finally, they were asked for their opinion on whether language would be beneficial for learning to knap. Some of their answers have been summarized in Supplementary Fig. 2.

*Designing the fNIRS Cap to Record from Target Regions of Interest (ROIs)*

Prior to the study, we identified a set of ROIs reported in three stone knapping studies that involve either PET or fMRI[1,14,15]. To further investigate the supposed involvement of the ventrolateral prefrontal cortex (vlPFC) during the transition to bifacial flaking, we also included coordinates from Table 2 in Badre and Wagner[58], which averages the coordinates for the vlPFC reported in six other studies. Similarly, to test for the involvement of the dorsolateral prefrontal cortex (dlPFC) during Early Stone Age tool manufacture, coordinates for dlPFC activation were compiled from Pessoa and colleagues[59,60].

Next, we used methods described in Wijeakumar et al.[19] to design a custom optode geometry to record from these ROIs. This involved digitizing candidate source and detector locations on an EasyCAP (Brain Products GmbH, Germany) using a Polhemus Patriot™ Motion Tracking System (Colchester, VT) and projecting these positions onto an adult atlas available in AtlasViewer GUI in the HOMER2 software package (www.nmr.mgh.harvard.edu/PMI/resources/Homer2)[61]. Final adjustments to the optode geometry were made after performing Monte Carlo simulations to create a sensitivity distribution for each source-detector pair (i.e., the sensitivity of each source-detector pair to detecting changes in absorption of NIR light) and visually inspecting whether these sensitivity volumes overlapped with the target ROIs. The end result was an optode geometry that recorded from all ROIs, including regions along the central sulcus, lateral prefrontal, superior temporal, and inferior parietal cortex.

*Image Acquisition and Processing*

fNIRS data were acquired at 25 Hz with a TechEn CW6 system with wavelengths of 690 nm and 830 nm. Light was delivered to a customized cap via fibre optic cables. The probe geometry had 12 sources and 24 detectors, creating 36 channels with a source-detector separation of 3 cm and two short source-detector channels with a separation of 1 cm (see Fig. 1c for optode coverage). HOMER2 software was employed to demean and convert the data into optical density (OD) units. A targeted principal component analysis (tPCA) was applied to data from the three tasks mentioned above to eliminate noise and motion artifacts[62]. We used a general linear model (GLM) to obtain beta values (β) for oxy-Hb and deoxy-Hb measures in every channel for all conditions in every task for each subject. Signals from the short source-detector channels were regressed from the rest of the channels to account for effects from superficial layers of the head.

The image reconstruction process is summarized briefly here (see Wijeakumar et al. for a more extensive explanation of this process[19,20]). Head 10-20 landmarks from the session that had the best symmetry were chosen as the reference for each subject. The landmarks from the other two sessions were transformed (linear) to fit this reference set of landmarks. The transformation matrices were applied to the corresponding source and detector positions. AtlasViewerGUI (available within HOMER2) was used to project the points onto an adult atlas using a relaxation algorithm. The projected geometry was used to run Monte Carlo simulations based upon a GPU-dependent Monte Carlo algorithm[63] for each session and subject. This resulted in sensitivity profiles (100 million photons) for each channel of the probe geometry for each session and subject. Head volumes and sensitivity profiles of channels were converted to NIFTII images. Subject-specific head volumes were skull-stripped and transformed to the head volume in the native atlas space using an affine transform (BRAINSFit in Slicer 3D). The transformation matrix obtained was applied to the sensitivity profiles to move them to the transformed head volume space (BRAINSResample in Slicer3D). Sensitivity profiles for all channels were thresholded to include voxels with an OD of greater than 0.0001[19]. These profiles were summed to create a session and subject-specific mask, and then these masks were summed across all sessions and subjects. Those shared voxels were used to create an intersection mask across participants.

The beta coefficients obtained for each channel, condition (within each task), and subject for oxy-Hb and deoxy-Hb were combined with the forward model results obtained from the Monte Carlo simulations to create voxel-based changes in oxy-Hb and deoxy-Hb concentration using image-reconstruction methods described by Wijeakumar et al[20]. Briefly, the image reconstruction problem can be formulated as the following generic equation:

$$Y = L\,.\,X \qquad (1)$$

where,

$$Y = \begin{bmatrix} \beta_{dOD}^{\lambda 1} \\ \beta_{dOD}^{\lambda 2} \end{bmatrix}$$

$$L = \begin{bmatrix} \varepsilon_{oxy-Hb}^{\lambda 1} \,.\, F^{\lambda 1} & \varepsilon_{deoxy-Hb}^{\lambda 1} \,.\, F^{\lambda 1} \\ \varepsilon_{oxy-Hb}^{\lambda 2} \,.\, F^{\lambda 2} & \varepsilon_{deoxy-Hb}^{\lambda 2} \,.\, F^{\lambda 2} \end{bmatrix}$$

$$X = \begin{bmatrix} \Delta oxy - Hb_{vox} \\ \Delta deoxy - Hb_{vox} \end{bmatrix}$$

Inverting $L$ to solve for $X$ results in an ill-conditioned and under-determined solution that might be subject to rounding errors. An alternative is to use Tikhonov regularization[64]. In this case, the above 'system' can be replaced by a regularized 'system'. The solution is given by the Gauss-Markov equation,

$$X = (L^T L + \lambda . I)^{-1} L^T . Y \qquad (2)$$

where $\lambda$ is a regularization parameter that determines the amount of regularization and $I$ is the identity operator.

The solution to (2) can be found by minimizing the cost function[65],

$$cost \min X = |L.X - Y|^2 + \lambda . |X - X_o|^2 \qquad (3)$$

where the size of the regularized solution is measured by the norm $\lambda . |X - X_0|^2$. $X_0$ is an *a priori* estimate of $X$, which is set to zero when no priori information is available. Here $X$ is determined for each chromophore and condition separately. Once Equation (3) is solved, there is now a voxel-wise estimate of the concentration data. Thus, the best estimate of the channel-wise concentration data for each condition (from the GLM) has been combined with information from the photon migration results to create an estimate of the voxel-wise concentration data for each chromophore, for each condition, and for each subject.

The resultant beta maps were intersected with the Intersection mask to restrict analyses to the voxels that were common to all sessions and subjects. Consequently, beta maps were obtained for each condition (within each task) and subject for oxy-Hb and deoxy-Hb concentration levels.

*Statistical Analysis*

The hemodynamic responses of the verbal and nonverbal groups and the Oldowan and Acheulian tasks were compared using two-way ANOVA tests for both the oxy-Hb and deoxy-Hb signals, conducted with the 3dMVM function in AFNI (Analysis of Functional Images)[66]. Resultant functional images of main effects and interactions were corrected for family-wise errors using the 3dClustSim function (corrected at alpha = 0.05, corresponding to a cluster size threshold of > 27 voxels). We analysed the highest-order effect in each spatially unique cluster; thus, main effect areas that overlapped with areas where an interaction occurred between Group and Task were interpreted based on the interaction effect.

Using the coordinates for the centre of mass of activation for each effect, we extracted the β values in these areas for the Oldowan and Acheulian tasks and the verbal and nonverbal groups. In cases of a significant interaction, the averaged β values of Task and Group were compared using the Wilcoxon signed-rank test and Mann-Whitney test, respectively. We also compared β values from the knapping conditions to the motor baseline conditions using the Wilcoxon signed-rank test to identify significant clusters that were unique to stone knapping and not simply general motor regions. Only those significant clusters where post-hoc tests determined knapping activation to be significantly higher than motor baseline activation were included in the final results discussed in the main text. Because the motor baseline task did not control for auditory stimulation while clicking rocks together, temporal cortex clusters were included in the final results, even if the signal in these regions was not significantly higher than the motor baseline signal.

To test for differences in knapping skill between the verbal and nonverbal groups, Kolmogorov-Smirnov, Mann-Whitney U, and Student's two-sided T-tests were employed for each variable related to the debitage and bifaces, with results considered significant at $p < 0.05$.

*Data Availability*

The datasets generated during the current study are available from the corresponding authors upon reasonable request.

**References:**


1. 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 T Roy Soc B* **363**, 1939-1949 (2008)
2. Stout, D., Chaminade, T. Stone tools, language and the brain in human evolution. *Phil T Roy Society B* **367**, 75-87 (2012)
3. Schoenemann, P. T. Evolution of the size and functional areas of the human brain. *Annu Rev Anthropol* **35**, 379-406 (2006)
4. Sherwood, C. C., Subiaul, F., Zawadzki, T. W. A natural history of the human mind: Tracing evolutionary changes in brain and cognition. *J Anat* **212**, 426-454 (2008)
5. Stout, D., Hecht, E. Neuroarchaeology. In *Human Paleoneurology* (Springer, New York, 2015)

6. Binneman, J., Beaumont, P. Use-wear analysis of two Acheulean handaxes from Wonderwerk Cave, Northern Cape. *South Afr Field Archaeol* **1**, 92-97 (1992)
7. Domínguez-Rodrigo, M., Serrallonga, J., Juan-Tresserras, J., Alcala, L., Luque, L. Woodworking activities by early humans: a plant residue analysis on Acheulian stone tools from Peninj (Tanzania). *J Hum Evol* **40**, 289-299 (2001)
8. Solodenko, N., *et al*. Fat residue and use-wear found on Acheulian biface and scraper associated with butchered elephant remains at the site of Revadim, Israel. *PLoS One* **10**, e0118572 (2015)
9. Balzeau, A., Holloway, R. L., Grimaud-Hervé, D. Variations and asymmetries in regional brain surface in the genus *Homo*. *J Hum Evol* **62**, 696-706 (2012)
10. Wynn, T. The intelligence of Later Acheulean hominids. *Man* **14**, 371-391 (1979)
11. Stout, D., Hecht, E., Nada, K., Bradley, B., Chaminade, T. Cognitive demands of Lower Paleolithic toolmaking. *PLoS ONE* **10**, e0121804 (2015)
12. Putt, S. S. "Human Brain Activity during Stone Tool Production: Tracing the Evolution of Cognition and Language," thesis, University of Iowa, Iowa City (2016)
13. Mahaney, R. A. Exploring the complexity and structure of Acheulean stoneknapping in relation to natural language. *Paleoanthropol* **2014**, 586-606 (2014).
14. Stout, D., Chaminade, T. The evolutionary neuroscience of tool making. *Neuropsychol* **45**, 1091-1100 (2007)
15. Stout, D., Passingham, R., Frith, C., Apel, J., Chaminade, T. Technology, expertise and social cognition in human evolution. *Eur J Neurosci* **33**, 1328-1338 (2011)
16. Morgan, T. J. H., *et al.,* Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* **6**(6029), 1-8 (2015)

17. Uomini, N. T., Meyer, G. F. Shared brain lateralization patterns in language and Acheulean stone tool production: A functional transcranial Doppler ultrasound study. *PLoS ONE* **8**, e72693.

18. Putt, S. S., Woods, A. D., Franciscus, R. G. The role of verbal interaction during experimental bifacial stone tool manufacture. *Lithic Technol* **39**, 96-112 (2014)

19. Wijeakumar, S., Spencer, J. P., Bohache, K., Boas, D. A., Magnotta, V. A. Validating a new methodology for optical probe design and image registration in fNIRS studies. *NeuroImage* **106**, 86-100 (2015)

20. Wijeakumar, S., Huppert, T., Magnotta, V. A., Buss, A. T., Spencer, J. P. Validating an image-based fNIRS approach with fMRI and a working memory task. *NeuroImage* **147**, 204-218 (2016).

21. Wynn, T., Coolidge, F. L. The implications of the working memory model for the evolution of modern cognition. *Int J Evol Biol* **2011**, 1-12 (2011)

22. Vouloumanos, A., Kiehl, K. A., Werker, J. F., Liddle, P. F. Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *J Cognitive Neurosci* **13**, 994-1005 (2001)

23. Lewis, W., Cortical networks related to human use of tools. *Neuroscientist* **12**, 211-231 (2006)

24. Leff, A. P., *et al*., The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain* **132**, 3401-3410 (2009)

25. Goldberg, G. Supplementary motor area structure and function: Review and hypotheses. *Behav Brain Sci* **8**, 567-616 (1985)

26. Augustine, J. R. Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res Rev* **22**, 229-244 (1996)

27. Menon, V., Uddin, L. Q. Saliency, switching, attention and control: A network model of insula function. *Brain Struct Funct* **214**, 655-667 (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. Gowlett, J. A. J. The elements of design form in Acheulian bifaces: Modes, modalities, rules and language. In *Lucy to Language: The Benchmark Papers* (Oxford University Press, Oxford, 2014)

30. Bangert, M., *et al*. Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage* **30**, 917-926 (2006)

31. Wong, C., Gallate, J. The function of the anterior temporal lobe: A review of the empirical evidence. *Brain Res* **1449**, 94-116 (2012)

32. Pascual, B., *et al.* Large-scale brain networks of the human left temporal pole: A functional connectivity MRI study. *Cereb Cortex* **25**, 680-702 (2015)

33. Rogers, R. D., *et al.* Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *J Neurosci* **19**, 9029-9038 (1999)

34. Vigneau, M. *et al*. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage* **54**, 577-593 (2011)

35. Levy, B. J., Wagner, A. D. Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Ann NY Acad Sci* **1224**, 40-62 (2011)

36. Corbetta, M., Schulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev* **3**, 201-215 (2002)

37. Hauk, O., Johnsrude, I., Pulvermüller, F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* **41**, 301-307 (2004)

38. Carvalho, S., Cunha, E., Sousa, C., Matsuzawa, T. Chaînes opératoires and resource-exploitation strategies in chimpanzee (*Pan troglodytes*) nut cracking. *J Hum Evol* **55**, 148-163 (2008)

39. Proffitt, T. et al. Wild monkeys flake stone tools. *Nature* **539**, 85-88 (2016)

40. Martínez, I., *et al.* Communicative capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. *Quatern Int* **295**, 94-101 (2013)

41. Quam, R.M., *et al.* Early hominin auditory ossicles from South Africa. *PNAS* **110**, 8847-8851 (2013)

42. Antón, S. C., Potts, R., Aiello, L. C. Evolution of early *Homo*: An integrated biological perspective. *Science* **345**, 1-13 (2014)

43. Arbib, M. A. From mirror neurons to complex imitation in the evolution of language and tool use. *Annu Rev Anthropol* **40**, 257-273 (2011)

**Correspondence:**

Correspondence and requests for materials should be addressed to Shelby S. Putt (ssputt@indiana.edu) or John P. Spencer (J.Spencer@uea.ac.uk).

**Acknowledgments:**

SSP acknowledges support from the Wenner-Gren Foundation (#8968), Leakey Foundation, Sigma Xi, the Scientific Research Society, and the University of Iowa. SSP held an American Fellowship from AAUW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Alexander Woods for his time and knapping expertise, the Lithic Casting Lab, Mera Hertel, and the Stone Age Institute for their contributions to Fig. 1, and Danielle Jones, Chloe Daniel, Elizabeth DeForest, Andriana Vega, Nicholas Fox, Anna Wells, Emily Hoeper, Emma Dellopolous, Madison Adams, Sean Allchin, and Graham Brua for their assistance in the lab.

**Author Contributions:**

S.S.P. and R.G.F. developed the concept for the study. S.S.P. and J.P.S. designed the study. S.S.P. performed the experiment and carried out the analyses, under the supervision of S.W. and J.P.S. S.S.P. and J.P.S. wrote the manuscript, with contributions from S.W. and R.G.F.

**Competing Interests Statement:**

The authors declare no competing interests.

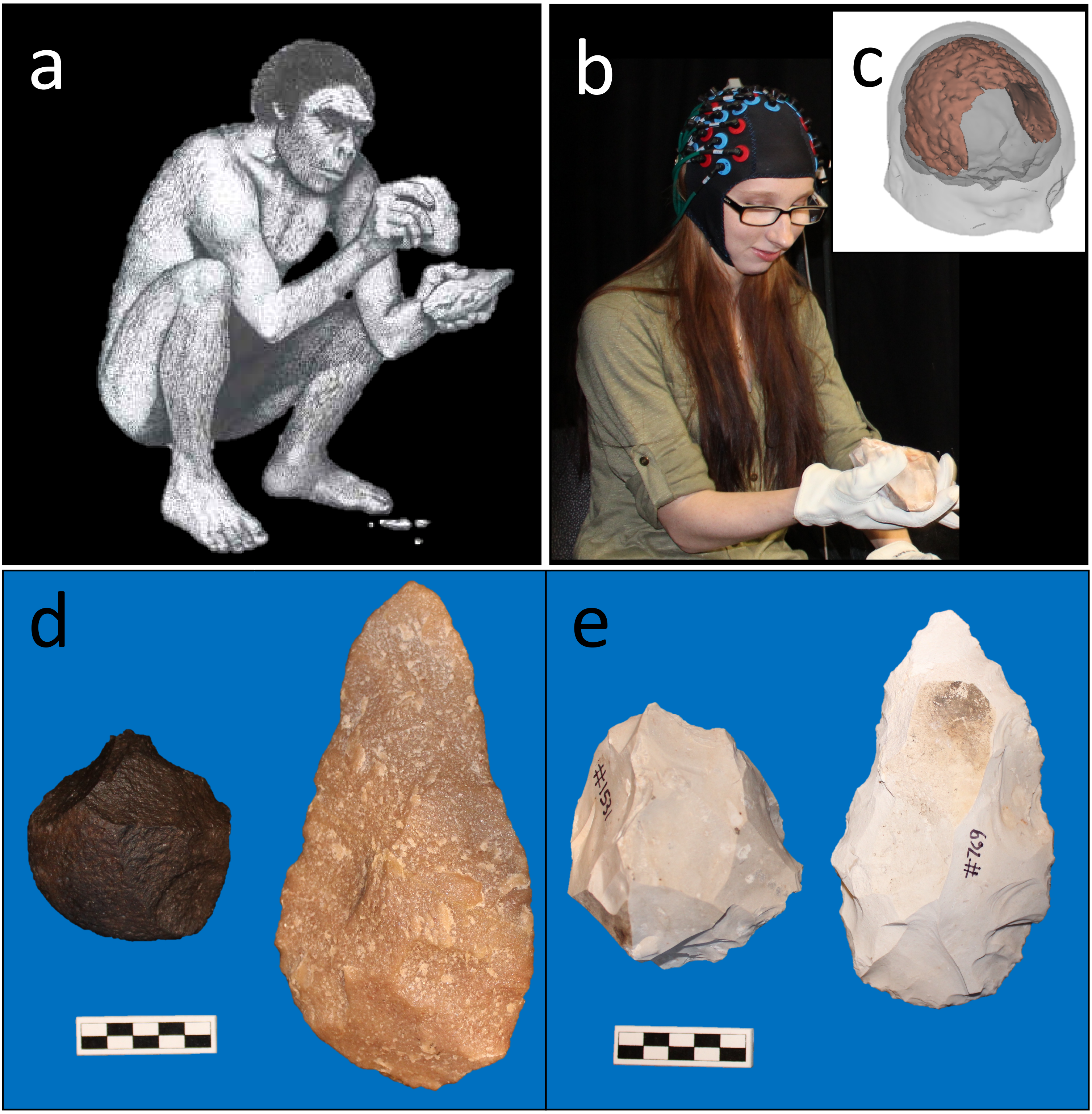
**Figure Legends**

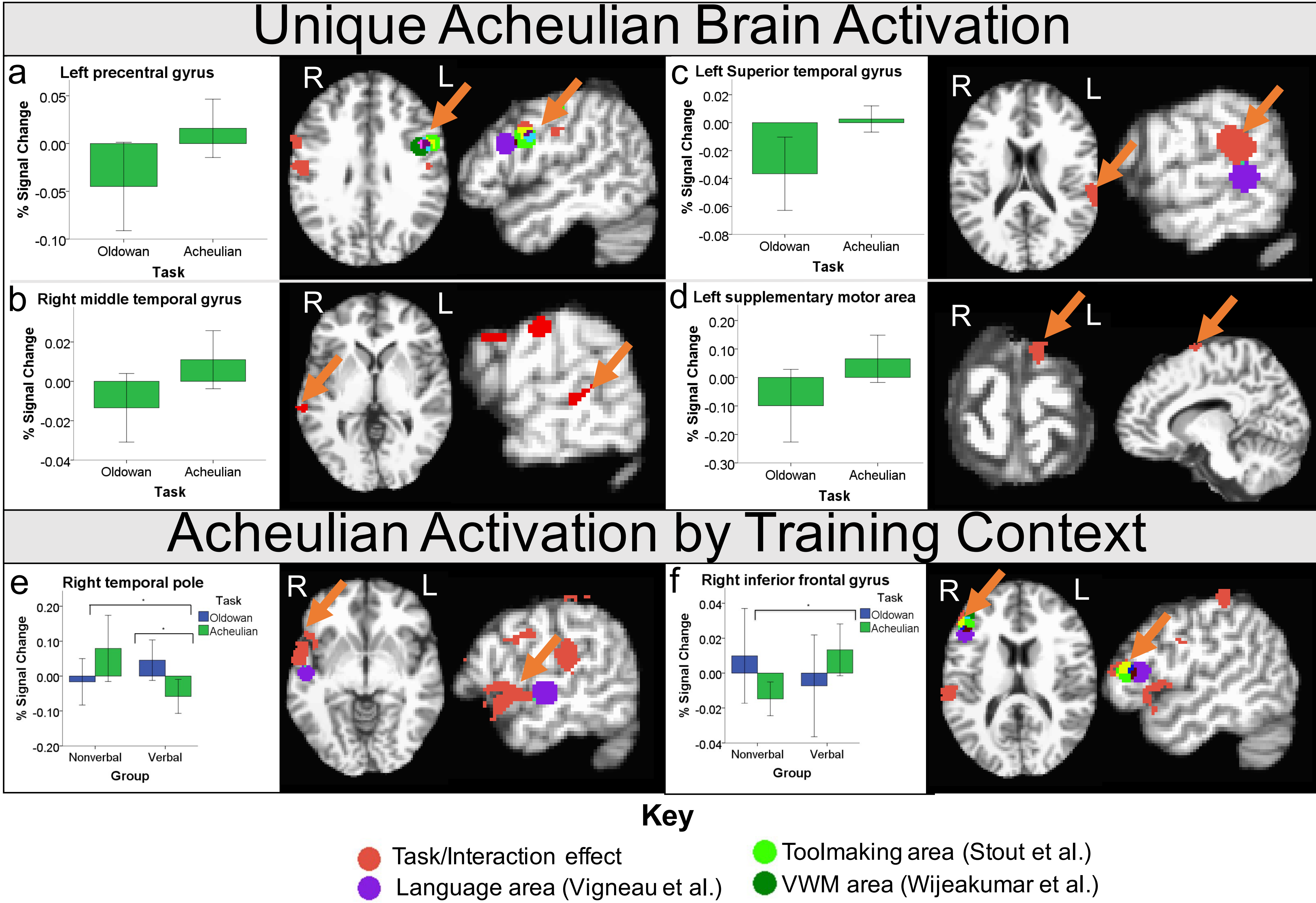
**Fig. 1.** The lithic reduction processes of early *Homo* (a) were replicated by 31 modern human subjects while we used functional near-infrared spectroscopy (b) to record regional brain activity from portions of the frontal, parietal, and temporal cortices of the brain (c). Both Oldowan (left) and Acheulian (right) tools from the archaeological record (d) were reproduced by the participants in the study (e).

**Fig. 2.** Acheulian Activation and the Effect of Training Context. Spheres (8mm in diameter) show overlap between the current results (red; $N = 31$, $F = 4.19$, $p < 0.05$) and previous findings

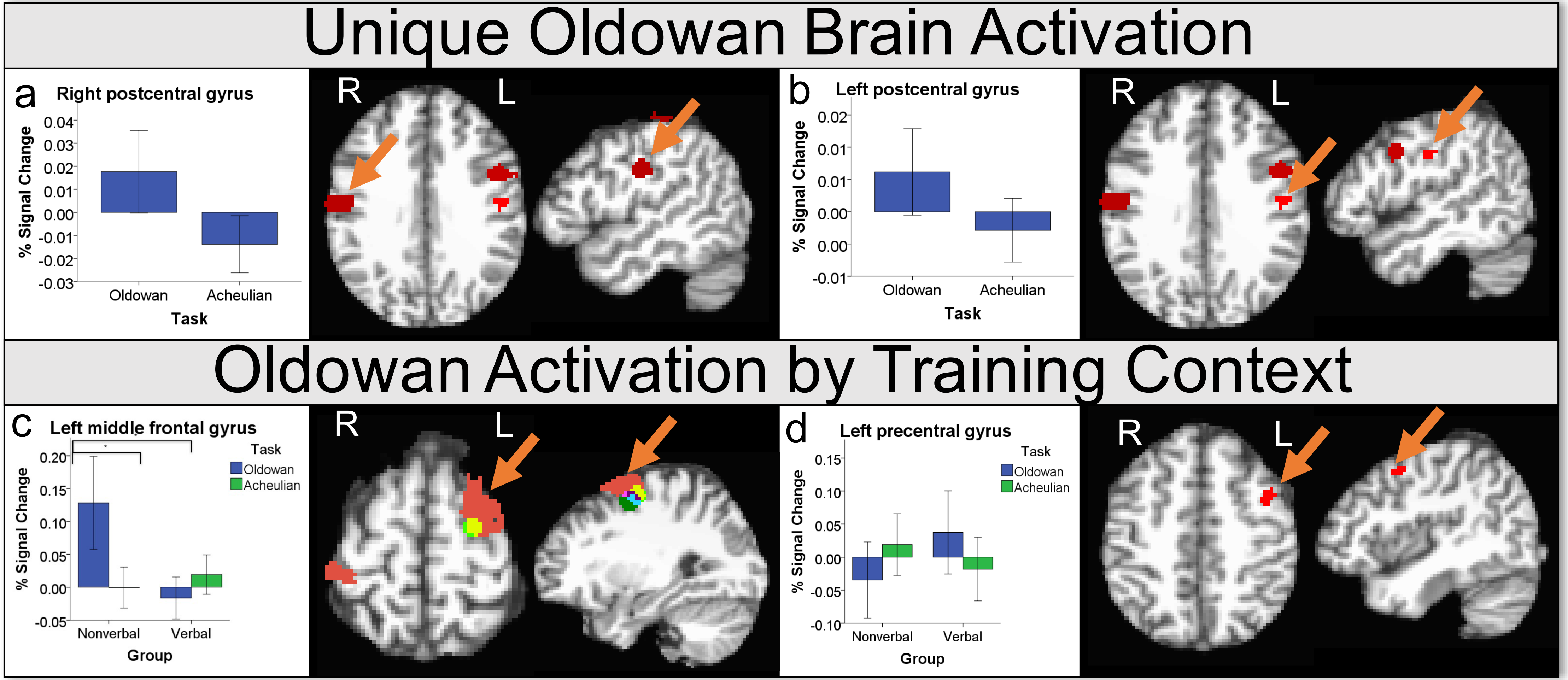
from a neuroarchaeological study[1] (light green), a language meta-analysis[34] (purple), and a VWM meta-analysis[19] (dark green). All other colours reflect overlap among these clusters (see text for additional detail). Error bars represent 95% confidence intervals. % Signal Change is in μM units. Starred brackets in (e) and (f) indicate significant differences revealed via post-hoc tests where $p < 0.05$ (see Methods).

**Fig. 3.** Oldowan Activation and the Effect of Training Context. Spheres (8mm in diameter) show overlap between the current results (red; $N = 31$, $F = 4.19$, $p < 0.05$) and previous findings from a neuroarchaeological study[1] (light green) and a VWM meta-analysis[19] (dark green). All other colours reflect overlap among these clusters (see text for additional detail). Error bars represent 95% confidence intervals. % Signal Change is in μM units. Starred brackets in (c) indicate significant differences revealed via post-hoc tests where $p < 0.05$ (see Methods).


a
b
c
d
e
#151
#769



Unique Acheulian Brain Activation
a
Left precentral gyrus
% Signal Change
0.05
0.00
-0.05
-0.10
Oldowan
Acheulian
Task
R
L
c
Left Superior temporal gyrus
% Signal Change
0.02
0.00
-0.02
-0.04
-0.06
-0.08
Oldowan
Acheulian
Task
b
Right middle temporal gyrus
% Signal Change
0.02
0.00
-0.02
-0.04
Oldowan
Acheulian
Task
d
Left supplementary motor area
% Signal Change
0.20
0.10
0.00
-0.10
-0.20
-0.30
Oldowan
Acheulian
Task
Acheulian Activation by Training Context
e
Right temporal pole
% Signal Change
0.20
0.10
0.00
-0.10
-0.20
Nonverbal
Verbal
Group
Task
Oldowan
Acheulian
f
Right inferior frontal gyrus
% Signal Change
0.04
0.02
0.00
-0.02
-0.04
Nonverbal
Verbal
Group
Task
Oldowan
Acheulian
Key
Task/Interaction effect
Language area (Vigneau et al.)
Toolmaking area (Stout et al.)
VWM area (Wijeakumar et al.)



Unique Oldowan Brain Activation
a
Right postcentral gyrus
% Signal Change
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
Oldowan
Acheulian
Task
R
L
b
Left postcentral gyrus
% Signal Change
0.02
0.01
0.01
0.00
0.00
-0.01
Oldowan
Acheulian
Task
R
L
Oldowan Activation by Training Context
c
Left middle frontal gyrus
Task
Oldowan
Acheulian
% Signal Change
0.20
0.15
0.10
0.05
0.00
-0.05
Nonverbal
Verbal
Group
R
L
d
Left precentral gyrus
Task
Oldowan
Acheulian
% Signal Change
0.15
0.10
0.05
0.00
-0.05
-0.10
Nonverbal
Verbal
Group
R
L

# The Functional Brain Networks that Underlie Early Stone Age Tool Manufacture

*Supplementary Information*

Shelby S. Putt*[1,2], Sobanawartiny Wijeakumar[3], Robert G. Franciscus[4], John P. Spencer*[3]

[1]The Stone Age Institute.

[2]Center for Research into the Anthropological Foundations of Technology, Indiana University.

[3]School of Psychology, University of East Anglia.

[4]Department of Anthropology, University of Iowa.

*Corresponding authors: ssputt@indiana.edu, J.Spencer@uea.ac.uk

## Supplementary Discussion

*Behavioural Results*

Every participant was successful at removing flakes from a cobble by the final neuroimaging session. The statistical analysis of the debitage-related variables did not reveal a clear pattern of higher skill among one group over the other (Supplementary Table 1). Both groups produced a similar proportion of flakes to shatter on both low quality and high quality material, though the verbal group made significantly more whole flakes relative to flake fragments. The verbal group, on average, produced more flake mass than the nonverbal group, though this difference was not significant. Otherwise, both groups reduced a similar amount of shatter mass from the original cobble during the Oldowan task, leaving a similar amount of material unexploited on the core. The verbal group had fewer missed strikes than the nonverbal group, which could signify increased skill, but this difference was not significant. Some significant differences between the groups occurred among the measures of flake shape. The flakes produced by the nonverbal group had a shape that was significantly thinner and longer than those produced by the verbal group, which replicates the results of a previous study that looked at differences in knapping skill between verbally- and nonverbally-instructed novices in an interactive teaching environment[18]. The flakes made by the nonverbal group also had significantly smaller platforms relative to flake size than the verbal group, though platform shape on its own did not significantly differ between groups.

There is no evidence from the statistical analysis of the core tools that one group excelled over the other (Supplementary Table 1). Both groups had an almost identical proportion of successful bifaces (verbal = 0.652, nonverbal = 0.650). The verbal group's bifaces had a larger average breadth to thickness ratio than the nonverbal group, but this difference was not significant. These results imply that the two groups reached similar levels of skill, and any differences in localized neural activation reflect the type of tool constructed (Oldowan, Acheulian) and the training context (verbal, nonverbal).

**Supplementary Table 1.** Group differences in knapping skill using debitage and core variables

| | Nonverbal | | | Verbal | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | S.D. | N | Mean | S.D. | Statistic | Sig. |
| Platform Shape (Width/Thickness) | 1719 | 3.68 | 2.42 | 1609 | 3.50 | 2.07 | $1.22^{D}$ | 0.103 |
| Flake Shape (Size/Mass) | 3157 | 2.96 | 4.16 | 2711 | 2.34 | 3.85 | $4.78^{D}$ | <0.001* |
| Relative Platform Area | 1710 | 27.54 | 33.1 | 1604 | 33.41 | 36.78 | $2.81^{D}$ | <0.001* |
| Proportion of Flakes to Shatter | 35 | 0.89 | 0.10 | 36 | 0.89 | 0.11 | $612.00^{U}$ | 0.836 |
| Proportion of Flakes on Low Quality Material | 32 | 0.84 | 0.15 | 35 | 0.85 | 0.16 | $589.00^{U}$ | 0.715 |
| Proportion of Flakes on High Quality Material | 31 | 0.93 | 0.09 | 30 | 0.91 | 0.12 | $422.50^{U}$ | 0.535 |
| Proportion of Flakes to Flake Fragments | 35 | 0.57 | 0.18 | 36 | 0.66 | 0.17 | $2.35^{t}$ | .021* |
| Proportion of Flake Mass Removed | 35 | 0.48 | 0.20 | 36 | 0.56 | 0.22 | $771.00^{U}$ | 0.105 |
| Proportion of Shatter Mass Removed | 35 | 0.11 | 0.12 | 36 | 0.12 | 0.18 | $693.00^{U}$ | 0.469 |
| Proportion of Remaining Core Mass | 35 | 0.32 | 0.20 | 36 | 0.28 | 0.22 | $490.00^{U}$ | 0.107 |
| Relative Number of Missed Strikes | 35 | 0.15 | 0.22 | 36 | 0.12 | 0.10 | $671.00^{U}$ | 0.637 |
| Biface ratio (Breadth/Thickness) | 13 | 1.89 | 0.34 | 15 | 2.00 | 0.54 | $0.62^{t}$ | 0.538 |

*Significant at $p < 0.05$

$^{D}$A Kolmogorov-Smirnov test was applied to cases with non-normal distributions and unequal variances. Statistic here refers to a *D* statistic.

$^{U}$A Mann-Whitney U test was applied to cases with non-normal distributions and equal variances. Statistic here refers to a *U* statistic.

$^{t}$Student's t-test was applied to cases with normal distributions and equal variances. Statistic here refers to a *t* statistic.

*Preliminary Meta-Analysis*

Previous neuroarchaeological research suggests that stone knapping behaviours do not require working memory involvement but do overlap with language-processing areas[1,14,15,17]. This interpretation was based mainly on the lack of activation in the dlPFC, which is considered to be an important component of the working memory system. This claim may have been premature, however, as working memory is a distributed neural system with multiple integrated, cortical regions. Indeed, a recent ALE meta-analysis of neuroimaging studies focused on stone knapping reveals that working memory plays an essential role in stone knapping, especially during Acheulian tool replication[67]. By plotting the coordinates of eight significant clusters from a recent neuroarchaeological study[1] in the same space as the coordinates from a visual working memory (VWM) meta-analysis[19] and a language-processing meta-analysis that includes phonological, lexico-semantic, and sentence processing neuroimaging studies[34], we also found that stone knapping functional activation not only overlaps with language centres but also overlaps with the VWM network, a fact that has been overlooked in previous studies (Supplementary Fig. 1).

*Neuroimaging Results*

fNIRS is unique in that it simultaneously measures the changes in concentration of both oxy-Hb and deoxy-Hb. Here we present the results for both chromophores. Supplementary Table 2 shows the list of all significant ANOVA results related to the oxy-Hb signal. The results reported in the main text (highlighted in grey in Supplementary Table 2) reflect active clusters with the highest-order effect (an Interaction effect in the case of overlap between a Main Effect and an Interaction; a Main Effect otherwise) that were also significantly higher than the motor baseline task. Effects that were not significantly greater than motor baseline but lie within the temporal cortex were also included because we did not control for sound production in the motor baseline task. In total, we focused on six clusters that showed a significant effect of Task and four clusters where Oldowan and Acheulian toolmaking were modulated by the linguistic context of training. Note that all of the Group main effects were subsumed by an overlapping Task x Group interaction.

Supplementary Table 3 shows the list of all significant ANOVA results related to concentrations of deoxy-Hb. The ANOVA revealed multiple clusters showing a significant main effect of Group and Task, as well as significant Group x Task interactions. Ten of these significant clusters overlapped spatially with significant oxy-Hb clusters. In nine of these clusters, there was an inverse relationship between deoxy-Hb and oxy-Hb (highlighted in grey in Supplementary Table 3). Oxy-Hb and deoxy-Hb signals tend to be negatively correlated with each other[68]. Critically, six of these nine clusters overlapped with the oxy-Hb results reported in the main text, including the right temporal pole, left STG, right PoG, left MFG, and two areas in the left PrG (Supplementary Fig. 3). These deoxy-Hb results lend further support to the conclusions reached in the main text.

**Supplementary Table 2.** Regions of significant activation (oxy-Hb) as determined by a two-way ANOVA between Group (verbal and nonverbal) and Task (Oldowan and Acheulian)[1].

| Localization | | Sig. Effect[2] | MNI Coordinates (mm) x | y | z | Volume ($mm^3$) | *M* Δoxy-Hb (μM) ± SEM |
|---|---|---|---|---|---|---|---|
| **Task main effect** | | | | | | | |
| Left | Superior temporal gyrus | A>O | -60.8 | -31.9 | 17.7 | 3600 | 6.18 ± 0.06 |
| Left | Precentral gyrus*[α] | O>A | -31.7 | -4.3 | 59.7 | 3584 | 6.26 ± 0.07 |
| Right | Postcentral gyrus[α] | A>O | 46 | -25.2 | 62 | 1688 | 5.07 ± 0.05 |
| Right | Postcentral gyrus* | O>A | 58.5 | -14.7 | 32.3 | 1624 | 6.55 ± 0.12 |
| Left | Precentral gyrus* | A>O | -50.2 | 5.8 | 33.5 | 1104 | 4.92 ± 0.05 |
| Right | Middle temporal gyrus | A>O | 67.7 | -33.6 | 2.8 | 536 | 4.39 ± 0.02 |
| Right | Precentral gyrus[α] | A>O | 61.9 | 7 | 28.7 | 432 | 5.81 ± 0.20 |
| Left | Supplementary motor area* | A>O | -9.9 | 1.4 | 75.7 | 352 | 4.73 ± 0.07 |
| Left | Postcentral gyrus* | O>A | -50.7 | -14.2 | 32.8 | 320 | 5.18 ± 0.11 |
| **Group main effect** | | | | | | | |
| Right | Rolandic operculum[α] | NV>V | 63.4 | -12.3 | 11.6 | 6904 | 7.03 ± 0.10 |
| Left | Inferior parietal lobule*[α] | NV>V | -55.2 | -31.4 | 38.9 | 6312 | 7.38 ± 0.08 |
| Left | Superior frontal gyrus*[α] | NV>V | -22.5 | -0.7 | 65.8 | 5688 | 6.48 ± 0.06 |
| Right | Postcentral gyrus[α] | NV>V | 36.3 | -33.1 | 71 | 328 | 5.30 ± 0.13 |
| **Group x Task interaction** | | | | | | | |
| Right | Temporal pole | V: O>A; A: NV>V | 57.3 | 9.6 | -5.8 | 4968 | 6.83 ± 0.08 |
| Left | Middle frontal gyrus* | NV: O>A; O: NV>V | -27.9 | -1.4 | 64.9 | 4928 | 8.40 ± 0.13 |
| Right | Supramarginal gyrus | O: NV>V | 63.7 | -26 | 19.9 | 4008 | 6.39 ± 0.07 |
| Left | Supramarginal gyrus | V: A>O; O: NV>V | -55.5 | -42.6 | 33 | 2456 | 5.13 ± 0.04 |
| Right | Postcentral gyrus | V: A>O; O: NV>V | 46.7 | -32 | 62.8 | 1864 | 6.88 ± 0.13 |
| Right | Postcentral gyrus | NV: A>O | 60.3 | -2 | 30.2 | 1192 | 5.78 ± 0.11 |
| Right | Inferior frontal gyrus* | A: V>NV | 51.4 | 37.2 | 13.5 | 776 | 4.79 ± 0.05 |
| Left | Precentral gyrus* | NS | -40.3 | 6.5 | 46.2 | 624 | 5.00 ± 0.06 |

[1]Grey highlighted areas reflect active clusters with the highest-order effect (an Interaction effect in the case of overlap between a Main Effect and an Interaction; a Main Effect otherwise) that were also significantly higher than the motor baseline task.

[2]A=Acheulian, O=Oldowan, V=Verbal, NV=Nonverbal, NS=Not significant

*Indicates cluster where knapping activation is significantly higher than motor baseline activation

[α]Main effect subsumed by an Interaction effect. Note that localization labels reflect the centre of mass of each cluster using MNI labelling conventions; thus, labels used for overlapping main effects and interactions might differ because the centres of mass differed.

**Supplementary Table 3.** Regions of significant activation (deoxy-Hb) as determined by a two-way ANOVA between Group (verbal and nonverbal) and Task (Oldowan and Acheulian).[1]

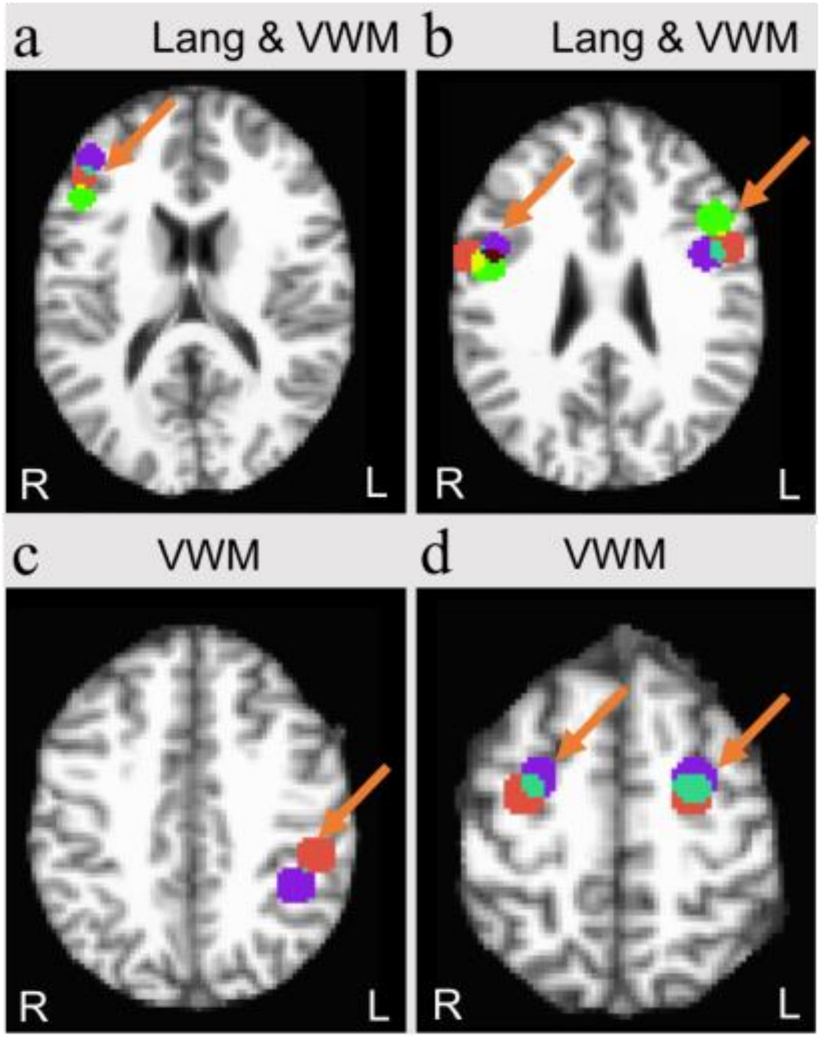
| | Localization | Sig. Effect[2] | MNI Coordinates (mm) x | y | z | Volume ($mm^3$) | *M* Δdeoxy-Hb (μM) ± SEM |
|---|---|---|---|---|---|---|---|
| | **Task main effect** | | | | | | |
| Right | Precentral gyrus[α] | O>A | 43.1 | -16.2 | 59.5 | 2728 | 5.24 ± 0.04 |
| Left | Precentral gyrus | O>A | -49.2 | -2.6 | 33.9 | 944 | 5.18 ± 0.08 |
| Right | Postcentral gyrus[α] | A>O | 58.9 | -12.7 | 29.6 | 920 | 5.27 ± 0.08 |
| Left | Superior temporal gyrus | O>A | -62.1 | -37.6 | 14.7 | 744 | 4.91 ± 0.06 |
| Right | Inferior frontal gyrus | A>O | 53.0 | 29.1 | 15.7 | 704 | 5.02 ± 0.08 |
| Left | Middle frontal gyrus | O>A | -39.3 | 30.0 | 41.0 | 600 | 4.94 ± 0.07 |
| Left | Inferior frontal gyrus | O>A | -53.4 | 15.7 | 34.5 | 400 | 4.57 ± 0.05 |
| Right | Superior parietal lobule[α] | A>O | 36.0 | -59.8 | 60.4 | 384 | 4.46 ± 0.04 |
| | **Group main effect** | | | | | | |
| Left | Superior temporal gyrus | NV>V | -59.6 | -17.4 | 10.7 | 1920 | 5.74 ± 0.07 |
| Right | Postcentral gyrus | V>NV | 55.9 | -24.0 | 45.8 | 1120 | 5.74 ± 0.11 |
| Left | Precentral gyrus[α] | NV>V | -41.9 | 0.7 | 50.0 | 648 | 4.83 ± 0.05 |
| Left | Middle frontal gyrus | V>NV | -46.4 | 28.2 | 34.0 | 640 | 5.00 ± 0.06 |
| | **Group x Task interaction** | | | | | | |
| Right | Precentral gyrus | NV: A>O; V: O>A; O: V>NV | 52.3 | -4.6 | 49.9 | 2704 | 6.25 ± 0.09 |
| Left | Precentral gyrus | O: NV>V | -40.4 | 7.9 | 46.9 | 2384 | 6.33 ± 0.09 |
| Left | Precentral gyrus | NS | -26.7 | -0.6 | 58.8 | 2016 | 5.38 ± 0.06 |
| Right | Superior parietal lobule | NS | 40.8 | -50.0 | 62.4 | 1120 | 5.52 ± 0.08 |
| Right | Superior temporal gyrus | NV: O>A; A: V>NV | 59.2 | -21.6 | 6.9 | 1000 | 4.80 ± 0.04 |
| Left | Superior temporal gyrus | V: A>O; A: V>NV | -64.4 | -9.6 | 0.4 | 928 | 4.66 ± 0.04 |
| Right | Postcentral gyrus | O: NV>V | 59.7 | -12.0 | 29.7 | 856 | 5.30 ± 0.09 |
| Right | Supramarginal gyrus | NS | 59.3 | -45.6 | 35.5 | 856 | 4.72 ± 0.04 |
| Right | Middle frontal gyrus | NV: O>A | 41.6 | 13.3 | 55.5 | 784 | 5.17 ± 0.09 |
| Left | Paracentral lobule | O: V>NV | -7.9 | -32.2 | 78.2 | 632 | 4.83 ± 0.06 |
| Right | Inferior frontal gyrus | O: NV>V | 60.0 | 17.5 | 3.2 | 504 | 5.27 ± 0.09 |
| Right | Postcentral gyrus | NS | 33.2 | -37.5 | 66.2 | 440 | 4.65 ± 0.05 |
| Right | Middle temporal gyrus | V: A>O; A: V>NV | 68.9 | -20.6 | -7.6 | 424 | 4.66 ± 0.04 |
| Right | Inferior frontal gyrus | NV: O>A | 63.3 | -12.6 | 29.3 | 224 | 4.65 ± 0.08 |

[1]Grey highlighted rows represent clusters that overlap and share an inverse relationship with significant oxy-Hb clusters (see Supplementary Table 1).

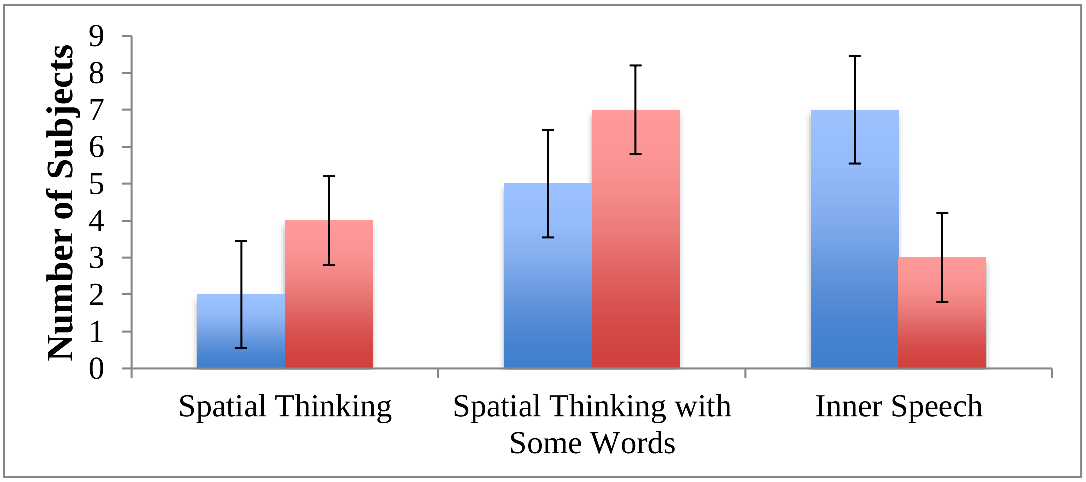
[2]A=Acheulian, O=Oldowan, V=Verbal, NV=Nonverbal, NS=Not significant

[α]Main effect subsumed by an Interaction effect. Note that localization labels reflect the centre of mass of each cluster using MNI labelling conventions; thus, labels used for overlapping main effects and interactions might differ because the centres of mass differed.

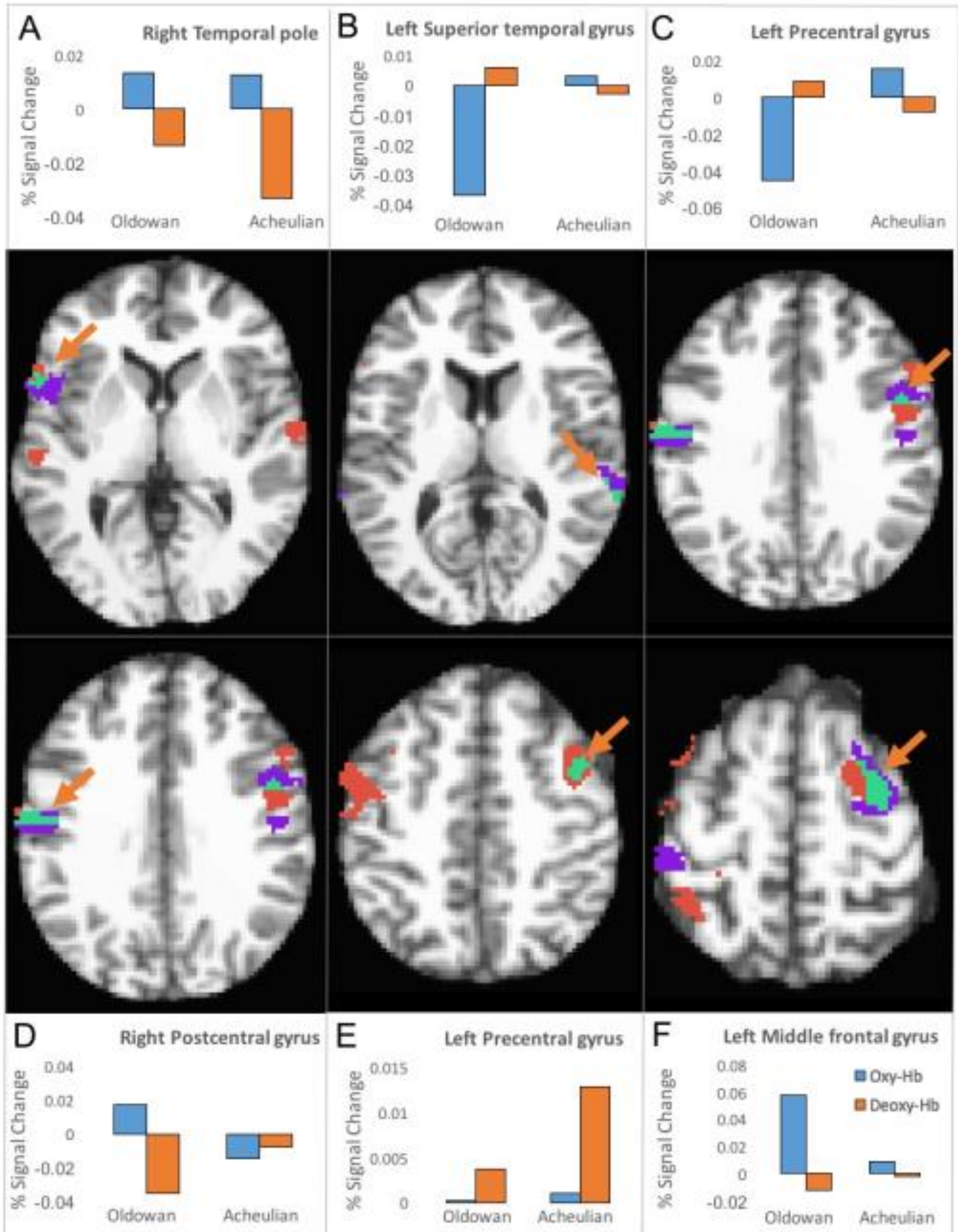
## Supplementary Figures




**Supplementary Figure 1.** Areas of functional overlap between a prior study of Early Stone Age knapping[1] (red), language-processing[34] (light green), and/or VWM[19] (purple), including (a) right IFG (*pars triangularis*), (b) bilateral ventral PrG, (c) left inferior parietal lobule, and (d) bilateral dorsal PrG. Overlap between spheres is represented by turquois, mauve, and yellow colours. This figure demonstrates that stone knapping overlaps with the VWM network to an even greater extent than it overlaps with language centres.




**Supplementary Figure 2.** Post-experiment interview subject responses by group (nonverbal = red; n = 14, verbal = blue; n = 14) to the question, "Did you think with language while knapping?" Subjects' responses were coded as one of three categories. A completely negative response to the question was coded as 'Spatial Thinking.' Responses that indicated minimal involvement of inner speech while thinking about the task were coded as 'Spatial Thinking with Some Words,' and participants who emphasized inner speech as their main mode of thinking or mentioned recalling entire phrases from the instruction videos were coded as 'Inner Speech.' Error bars represent standard error.




**Supplementary Figure 3.** Relationship between significant, overlapping oxy-Hb (purple) and deoxy-Hb (red) clusters ($N = 31$). Overlap between clusters is represented by turquoise. Bar plots compare relative oxy-Hb (blue) and deoxy-Hb (orange) concentrations across tasks. % Signal Change is in μM units.

**Supplementary References**


44. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Laurence Erlbaum Associates, Hillsdale, NJ, ed. 2, 1988)
45. Soper, D.S. A-priori sample size calculator for Student's T-tests [Software]. Available from http://www.danielsoper.com/statcalc (2014)
46. Szaflarski, J.P. *et al.* Language lateralization in left-handed and ambidextrous people: fMRI data. *Neurol* **59**, 238-244 (2002)
47. Oldfield, R.C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsy* **9**, 97-113 (1971)
48. Skinner, H.A. The drug abuse screening test. *Addict Behav* **7**, 363-371 (1982)
49. London, E.D., Ernst, M., Grant, S., Bonson, K., Weinstein, A. Orbitofrontal cortex and human drug abuse: Functional imaging. *Cereb Cortex* **10**, 334-342 (2000)
50. Yankosec, K.E., Howell, D. A narrative review of dexterity assessments. *J Hand Ther* **22**, 258-270 (2009)
51. Desrosiers, J., Rochette, A., Hebert, R., Bravo, G. The Minnesota Manual Dexterity Test: Reliability, validity and reference values studies with healthy elderly people. *Can J Occup Ther* **64**, 270–276 (1997)
52. Toth, N. The Oldowan reassessed: A close look at early stone artifacts. *J Archaeol Sci* **12**, 101-120 (1985)
53. Stout, D. Stone toolmaking and the evolution of human culture and cognition. *Phil T Roy Soc B* **366**, 1050-1059 (2011)
54. Hoard, R.J., Anglen, A.A. Lithic analysis. *Plains Anthropol* **48**, 36-50 (2003)
55. Andrefsky, W. *Lithics: Macroscopic Approaches to Analysis* (Cambridge University Press, Cambridge, ed. 2, 2005)
56. Toth, N., Schick, K., Semaw, S. A comparative study of the stone tool-making skills of *Pan, Australopithecus*, and *Homo sapiens*. In *The Oldowan: Case Studies into the Earliest Stone Age* (Stone Age Institute Press, Gosport, IN 2006)
57. Bamforth, D.B., Finlay, N. Introduction: Archaeological approaches to lithic production skill and craft learning. *J Archaeol Met Theor* **15**, 1-27 (2008)
58. Badre, D., Wagner, A. D. (2005). Frontal lobe mechanisms that resolve proactive interference. *Cereb Cortex* **15**, 2003–2012
59. Pessoa, L., Gutierrez, E., Bandettini, P. A., Ungerleider, L. G. Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron* **35**, 975-987 (2002)
60. Pessoa, L., Ungerleider, L. G. Neural correlates of change detection and change blindness in a working memory task. *Cereb Cortex* **14**, 511-520 (2004)
61. Huppert, T. J., Diamond, S. G., Franceschini, M. A., Boas, D. A. HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl Optics* **48**, D280-D298 (2009)
62. Yücel, M. A., Selb, J., Cooper, R. J., Boas, D.A. Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy. *J Innovat Opt Health Sci* **7**, 1350066 (2014)
63. Fang, Q., Boas, D. Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. *Opt Express* **17**, 20178-20190 (2009)
64. Tikhonov, A. Solution of incorrectly formulated problems and the regularization method. *Sov Mathematics-Doklady* **5**, 1035-1038 (1963)

65. Calvetti, D., Morigi, S., Reichel, L., Sgallari, F. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *J Comput Appl Math* **123**, 423-446 (2000)
66. Chen, G., Adleman, N. E., Saad, Z. S., Leibenluft, E., Cox, R. W. Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate general linear model. *NeuroImage* **99**, 571-588 (2014)
67. Mahaney, R. A. "Cognition and Planning in Paleolithic Technology: Studies in Experimental Archaeology," thesis, Indiana University, Bloomington (2015)
68. Cui, X., Bray, S., Reiss, A. L. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage* **49**, 3039-3046 (2010)