Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications

Article *in* Journal of Human Evolution · July 2019

DOI: 10.1016/j.jhevol.2019.05.010

CITATIONS
0

READS
162

3 authors:

Justin Pargeter
New York University
54 PUBLICATIONS **484** CITATIONS
SEE PROFILE

Nada Khreisheh
Emory University
10 PUBLICATIONS **104** CITATIONS
SEE PROFILE

Dietrich Stout
Emory University
54 PUBLICATIONS **2,917** CITATIONS
SEE PROFILE

Journal of Human Evolution 133 (2019) 146–166

Contents lists available at ScienceDirect

Journal of Human Evolution

journal homepage: www.elsevier.com/locate/jhevol

Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications

Justin Pargeter [a, b, *], Nada Khreisheh [c], Dietrich Stout [a]

[a] Department of Anthropology, Emory University, Atlanta, GA, USA
[b] Rock Art Research Institute, School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa
[c] The Ancient Technology Centre, Cranborne, Dorset, UK

ARTICLE INFO

Article history:
Received 25 February 2019
Accepted 22 May 2019

Keywords:
Skill acquisition
Social transmission
Handaxes
Experimental archaeology
Executive function
Acheulean

ABSTRACT

Despite its theoretical importance, the process of stone tool-making skill acquisition remains understudied and poorly understood. The challenges and costs of skill learning constitute an oft-neglected factor in the evaluation of alternative adaptive strategies and a potential source of bias in cultural transmission. Similarly, theory and data indicate that the most salient neural and cognitive demands of stone tool-making should occur during learning rather than expert performance. Unfortunately, the behavioral complexity and extensive learning requirements that make stone knapping skill acquisition an interesting object of study are the very features that make it so challenging to investigate experimentally. Here we present results from a multidisciplinary study of Late Acheulean handaxe-making skill acquisition involving twenty-six naïve participants and up to 90 hours training over several months, accompanied by a battery of psychometric, behavioral, and neuroimaging assessments. In this initial report, we derive a robust quantitative skill metric for the experimental handaxes using machine learning algorithms, reconstruct a group-level learning curve, and explore sources of individual variation in learning outcomes. Results identify particular cognitive targets of selection on the efficiency or reliability of tool-making skill acquisition, quantify learning costs, highlight the likely importance of social support, motivation, persistence, and self-control in knapping skill acquisition, and illustrate methods for reliably reconstructing ancient learning processes from archaeological evidence.

1. Introduction

Stone artifacts provide some of the most prolific and highest-resolution evidence of Paleolithic behavior and its spatiotemporal variation. This evidence has been used to investigate evolving hominin behavioral ecology/adaptive strategies (e.g., Shea, 2017; Režek et al., 2018), cultural transmission (e.g., Lycett and Bae, 2010; Tennie et al., 2017; Stout et al., in press, motor skills (e.g. Williams-Hatala et al., 2018), and neurocognitive evolution (e.g., Stout et al., 2015; Wynn and Coolidge, 2016; Bruner et al., 2018). The actual process of learning to make stone tools, however, remains understudied despite its central relevance to each of these research directions. Thus, investments in skill learning are often neglected when evaluating the costs and benefits of alternative adaptive strategies and we know little about the learning challenges that may have biased cultural transmission and influenced the evolution and distribution of technological traits (e.g., Roux, 1990; Henrich, 2016). Theory and results from prior research on stone tool learning (e.g., Eren et al., 2011; Hecht et al., 2015; Stout et al., 2015) similarly lead us to expect that the most salient neural and cognitive demands of tool-making should occur during learning rather than expert performance.

For such reasons, skill acquisition has emerged as a central component of current theoretical approaches to human evolution, which emphasize powerful feedback relations between social learning, brain size, and life history strategies (e.g., Kaplan et al., 2000; Antón et al., 2014; Isler and Van Schaik, 2014; González-Forero and Gardner, 2018) leading to the emergence of the unique human adaptive complex of cooperation, sharing, and the intergenerational reproduction of complex subsistence skills (Hill et al., 2009). We have elsewhere described this constellation of factors as the human 'technological niche' (Stout and Khreisheh, 2015; Stout and Hecht, 2017). Given such clear reasons for interest, it is unfortunate that pragmatic challenges and methodological

* Corresponding author.
E-mail address: justin.pargeter@nyu.edu (J. Pargeter).

https://doi.org/10.1016/j.jhevol.2019.05.010
0047-2484/

limitations continue to make it difficult to study the acquisition of complex, real-world skills like stone tool-making. Despite valuable earlier investigations (Roux et al., 1995; Stout, 2002; Winton, 2005) and increasing recent attention (e.g., Nonaka et al., 2010; Eren et al., 2011; Rein et al., 2014; Stout et al., 2014, 2015; Schillinger et al., 2014, 2017; Putt et al., 2014, 2017; Hecht et al., 2015; Lombard, 2015; Morgan et al., 2015; Lombao et al., 2017) we are still a long way from really understanding the mechanisms underlying the reproduction of knapping skills, even in recent humans.

The current study advances beyond previous efforts by combining a relatively large sample of naïve learners ($n = 17$) with a long-term (~25–90 hours) thoroughly documented training program including instructor notes and periodic skill assessments, trainee self-evaluations, video-recordings, and extensive data on the lithic products. This training program occurred as part of a larger study of tool-making skill and cognition that also included structural and functional magnetic resonance imaging (MRI), eye-tracking, linguistic performance measures, psychometric testing, a flake prediction task, and debitage analyses to be reported elsewhere. As in previous work (e.g., Faisal et al., 2010, Stout et al., 2008, 2011, 2015, 2018; Hecht et al., 2014), this larger study focuses on Late Acheulean handaxe production, seeking to further refine our understanding of its cognitive, behavioral, and evolutionary significance.

In this report, we focus on the fundamental problem of quantifying variation in knapping skill over the training period and across individuals. This is an obvious prerequisite for further investigation of learning demands and trajectories, the nature and causes of individual differences, and the effects of experimental manipulations. Pragmatically it provides a criterion for estimating the minimum training time needed to address particular questions during research design and for evaluating the success of training protocols during interpretation.

1.1. The experimental archaeology of knapping skill

Experimental archaeology aims to identify causal relations linking observable archaeological residues to past human behaviors. Since no experiment can be a perfect replication of the past, this requires balancing the frequently competing demands of relevance to actual past conditions (i.e., external validity) vs. experimental control (i.e., internal validity; Flenniken, 1984; Thomas, 1986; Lycett and Eren, 2013; Eren et al., 2016; Lin et al., 2018). As Eren et al. (2016) argued, appropriate trade-offs between external and internal validity are determined by specific research questions and a diversity of approaches is desirable.

For some questions, it may be most appropriate to use highly artificial experiments. For example, machine flaking (e.g., Magnani et al., 2014) allows precise control over core morphology and force application variables in order to reveal the basic fracture mechanical properties of stone that necessarily constrained the technological activities and products of past knappers (Lin et al., 2018). Similarly, studies of social transmission have examined handaxe shape and size copying error using experimental tasks, like drawing on an iPad (Kempe et al., 2012) or carving plasticine (Schillinger et al., 2014) or foam (Schillinger et al., 2015, 2016), that afford high degrees of control, relatively large sample sizes, and the mitigation of "safety and feasibility concerns" (Schillinger et al., 2014: 131). Use of such tractable tasks and materials allows the study of handaxes as a simplified 'model artifact' relevant to more general questions about cultural microevolutionary processes (Schillinger et al., 2016).

However, the more direct application of results from such approaches to the interpretation of the archaeological record may require strong assumptions about the (ir)relevance of artificial experimental manipulations to particular questions. For example, Schillinger et al. (2015) found that observing a demonstrator carve a foam block ('imitation' condition) yielded higher shape copying fidelity than observing products alone ('emulation' condition) and argued that this corroborates the likely importance of imitation in maintaining Acheulean shape homogeneity. This is reasonable, but it is also possible that the mechanics of actually knapping stone might be sufficiently constraining to eliminate this effect of learning condition, and/or that it would disappear over longer (i.e., >20 min) learning periods. Indeed, the plausibility of the former is supported by the fact that, even within the foam carving paradigm, tool selection (knife vs. peeler) also affects copy fidelity (Schillinger et al., 2016). Similarly, controlled fracture experiments have demonstrated causal relations between manipulated platform variables and resulting flake morphology, but these relationships leave large amounts of the variation encountered in actual archaeological assemblages unexplained (Archer et al., 2018). More accurate prediction of flake morphology is possible using holistic 3D morphometric approaches to platform variation, but this method has thus far left the actual features and technical behaviors driving these relations unresolved (Archer et al., 2018).

We would argue that, in the case of stone tool-making skill acquisition, our understanding of the actual importance of 'naturalistic' factors, such the physical properties of stone, or the amount, timing, and content of practice, is often too limited to identify factors that can be safely eliminated for the purposes of experimental tractability and/or internal validity (cf. Eren et al., 2011). At the same time, our ignorance regarding the actual learning practices and contexts that occurred in the past (e.g., presence/absence of active teaching and imitation; Tennie et al., 2016; Gärdenfors and Högberg, 2017; Stout et al., 2019) makes pursuit of high external validity through the approximation of past conditions equally problematic.

Previous studies of knapping skill acquisition processes (as opposed to cross-sectional comparisons of knappers at different skill levels; e.g. Nonaka et al., 2010) have generally navigated this quandary by prioritizing internal validity and experimental tractability in order to test hypotheses about past learning processes. For example, controlled studies of multiple learning conditions (e.g., observation vs. verbal or gestural instruction) and iterated learning transmission chains have been enabled by brief training periods (see Table 1 for stone tool-making experiments), video-recorded demonstrators (Putt et al., 2017; Cataldo et al., 2018), and the use of non-stone (Geribàs et al., 2010; Schillinger et al., 2016) or proxy lithic raw materials (Putt et al., 2014; Morgan et al., 2015). These studies have produced important results and methodological innovations, but implications for human evolution remain ambiguous. Whiten (2015), for example, discussed the difficulty of assessing the actual adequacy of different learning

Table 1
Subject training times from previous stone tool making experiments.

Training time	Reference
No training	Duke and Pargeter (2015); Geribàs et al. (2010)
5 minutes	Cataldo et al. (2018)
5–15 minutes	Lombao et al. (2017)
1 hour	Bril et al. (2010)
1.5 hours	Morgan et al. (2015)
2 hours	Nonaka et al. (2010); Rein et al. (2014)
4 hours	Stout and Chaminade (2007)
5 hours	Putt et al. (2014)
6 hours	Ohnuma et al. (1997)
16 hours	Stout et al. (2011), 2014
84–175 hours	Hecht et al., 2014; Stout et al. (2015)

conditions given practice times much shorter than what might be expected ethnographically or archaeologically. We thus seek to complement this existing body of work with a study that increases external validity at the expense of experimental tractability. In particular, we prioritize more extended training time, naturalistic face-to-face instruction, objective outcome metrics, and experimental learning objectives based on an explicit archaeological model (Stout and Khreisheh, 2015).

All of these studies, including ours, face the additional challenge of using data from modern human—and other primate (Toth and Schick, 2009)—participants to make inferences about learning in extinct hominin species. In general, the argument has been made that experimental effects observed in modern humans can inform evaluations of the plausibility of different evolutionary scenarios, such as the likelihood that stone toolmaking did (Morgan et al., 2015; Lombao et al., 2017) or did not (Putt et al., 2014; Cataldo et al., 2018) provide selective pressures favoring the evolution of teaching and language. We seek to extend this approach toward the development of a broader inferential framework by characterizing the processes, costs, demands, and material correlates of modern human stone knapping skill acquisition across extended training times. This will serve three main purposes. First, it will provide a reference point for comparative investigations of human cultural and cognitive evolution (Whiten, 2015; Stout and Hecht, 2017). For example, learning processes and demands observed in human handaxe-making can be compared with skill learning in other primate species (e.g., Whiten, 2015; Schuppli et al., 2016; Fragaszy et al., 2017) to identify any derived human features specifically relevant to Paleolithic technology. This includes identification of particular aspects of individual behavioral and cognitive variation that impact handaxe-making skill acquisition across participants and thus constitute especially likely targets for selection acting on technological capacity (cf. Thornton and Lukas, 2012). Second, it will relate observed learning processes and outcomes to variation in objective artifact features to generate expectations for the archaeological record. This includes estimates of the time, effort, and material required to achieve archaeologically-observed competencies (Stout et al., 2014; Lycett et al., 2016; Garcia-Medrano et al., 2018) as well as characteristic features of unskilled performance (Shelley, 1990). Future work can then test specific hypotheses regarding developmental (Högberg, 2018), cultural (Henrich et al., 2010), and other contextual influences (e.g., Eren et al., 2014; Morgan et al., 2015) that might modulate these expectations. Third, the study will make methodological contributions to the objective quantification of knapping skill by benchmarking the training time required to experimentally capture particular aspects of tool-making skill acquisition. In our study, more realistic training times in a substantial participant sample are achieved by limiting the experiment to a single, non-iterated, learning condition. The specific condition we selected is unrestricted teaching by an experienced knapping instructor (the second author). This included a broad range of modern pedagogical techniques, ranging from explicit multimodal (verbal, gestural) instruction, to demonstration, assistance, interactive feedback, and opportunity scaffolding (Kline and Boyd, 2010; Stout and Hecht, 2017). We do not know to what extent this approximates any particular learning contexts out of the range that actually occurred during the Paleolithic, but it does provide a baseline 'unrestricted teaching' condition as a point of comparison. The theoretical (Gärdenfors et al., 2017) and empirical (Morgan et al., 2015) expectation is that more restricted instruction conditions would either impair or fail to affect learning, but not improve it. This condition also has the advantage of not requiring artificial manipulations, such as an injunction to 'teach without talking,' that arguably might produce unnatural behaviors unlike both the present and the past. For our Western, Educated, Industrialized, Rich and Democratic (WEIRD; Henrich et al., 2010) instructor and participants, 'unrestricted teaching' is the naturalistic form of instruction.

Table 2
Details of the skill score rubric and break down and each score component's technological/cognitive domain.

Scoring criteria	Subcluster	Details (scored from 1 to 5 in 0.5 increments)
Stacks	Outcomes	Step fractures associated with the flaking platform.
Thinning		Handaxe cross-sectional thinning
Shaping		Handaxe shaping
Platform preparation	Perceptual motor execution	Application of platform preparation
Striking angle/force		Use of appropriate for and flaking angles
Platform angle		Recognition of suitable flaking platforms
Hammerstone choice		Hammerstone choice relative to task
Strategy	Strategic understanding	Staging of sequential operations in the flaking sequence
Bifacial plane		Creation and management of successful bifacial plane
Abandonment		Suitably timed abandonment of the flaking process

Another key issue we seek to address in this study is the need to develop objective and generalizable methods for quantifying knapping skill variation across individuals and training. Skill itself is a complex concept that had been defined in multiple ways invoking various combinations of experience, acquired ability, natural aptitude, discursive knowledge, and practical execution (Bamforth and Finlay, 2008). Experimental studies have operationalized skill in correspondingly diverse ways ranging from conventional lithic metrics tracking technological abilities, such as biface refinement and symmetry, flake size and shape, or core reduction intensity (e.g., Putt et al., 2014; Morgan et al., 2015; Stout et al., 2015; Lombao et al., 2017), to subjective ratings by experts (Putt et al., 2014; Morgan et al., 2015), coding of video-recorded behavior sequences (Geribás et al., 2010; Lombao et al., 2017), and performance on artificial proxy tasks such as flake prediction (Nonaka et al., 2010; Stout et al., 2015) or motor accuracy (Hecht et al., 2014). These different approaches focus on different aspects of skill and have different strengths, weaknesses, and trade-offs with respect to objectivity, completeness, and potential applicability to archaeological materials. Our aim here is to develop a single, broadly applicable method that combines the strengths of these different approaches to produce an objective, artifact-based, quantification of skill that can equally be used as a global attribute or dissected into a set of interacting subcomponents that might develop at different stages and rates. Critically, the metric we develop is grounded in the instructor's own evaluations of the degree to which participants mastered specific teaching objectives in the training she provided (Table 2). This 'hermeneutic' approach to skill measurement establishes high internal validity. External validity in turn depends upon the degree to which the instructor's teaching objectives match the objectives of actual Paleolithic tool-makers, a point to which we now turn.

1.2. Archaeological framework for the experiment

Our experimental teaching objectives (Table 2) are informed by broad archaeological (e.g., Beyene et al., 2013; Moncel and Ashton, 2018) and experimental (e.g., Schick and Toth, 1993; Edwards, 2001; Winton, 2005; Shipton and Clarkson, 2015) consensus

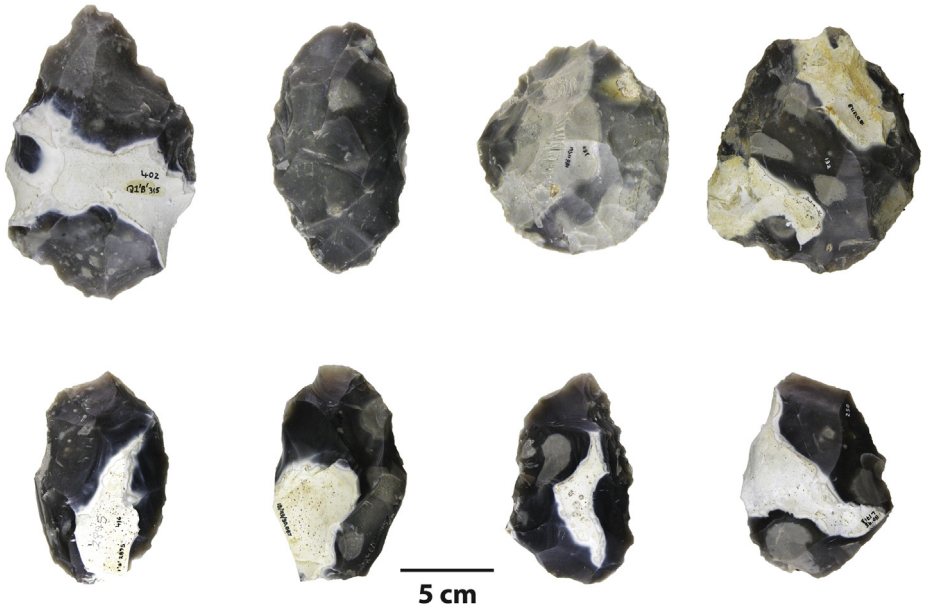

Figure 1. Selection of handaxes from the Boxgrove collection illustrating the assemblage's degree of morphological variability.

regarding the goals and challenges of Late Acheulean handaxe production in general, as well as detailed studies of knapping behaviors at the Middle Pleistocene site of Boxgrove in particular (e.g., Stout et al., 2014; Garcia-Medrano et al., 2018). Whereas there is increasing recognition of the actual technological diversity subsumed by broad typological classifications like 'Late Acheulean' (Lycett and Gowlett, 2008; Iovita and McPherron, 2011), Boxgrove provides a specific, well-studied point of comparison. Boxgrove is widely recognized as representing an extreme example of Acheulean skill expression and is not only one of the oldest handaxe sites in Europe (dated ca. 524–478 ka), but also one of the continent's richest in situ handaxe assemblages (Pitts and Roberts, 1998; Fig. 1). The exceptional richness, preservation, and expression of high-level knapping skills at Boxgrove makes this site an ideal focus for researchers interested in documenting evolving hominin technical capacities. Whereas it is difficult to know if the absence of evidence for a behavior is due to a lack of motivation or opportunity rather than capacity, Boxgove provides a positive demonstration of what Middle Pleistocene knappers were capable of under the right conditions. Several other examples of such high-level handaxe-making skills have been reported from Late Acheulean sites in Africa and continental Europe (Roche, 2005; Iovita et al., 2017; Shipton, 2018). While the technological particulars of the Boxgrove case study will obviously not generalize across the wide diversity of Late Acheulean archaeological occurrences, the evidence of capacity it provides is the most relevant datum for attempts to reconstruct broad patterns of hominin cognitive and cultural evolution (Stout et al., 2011).

Experimentally informed technological studies of cores and flakes from Boxgove document careful shaping and intensive thinning (e.g., Shipton and Clarkson, 2015; Garcia-Medrano et al., 2018) using knapping techniques such as soft-hammer, marginal, percussion and platform preparation that are directly comparable to those employed by modern experimental knappers (Stout et al., 2014). Whereas there is debate over the status of some Acheulean handaxes as intentional products (e.g., Moore and Perston, 2016), the pursuit of particular morphological goals using context-appropriate strategies and techniques is thus well documented at Boxgrove. As in previous work (e.g., Hecht et al., 2014; Stout et al., 2015) we use this technological understanding of Boxgrove as the explicit source of our experimental teaching objectives. To further investigate the external validity of this approach, we report direct comparisons between the handaxes from Boxgrove Quarry 1 Area B, Project D (Q1B/D) and our experimental handaxes using measurements collected by Jan Apel for a previous project (see Stout et al., 2014 for further details).

Our continued focus on Late Acheulean handaxe production is driven by both theoretical and practical considerations. Although 'crude' bifaces persist throughout the record, it is widely agreed that smaller, thinner, more regular and symmetrical forms appear in the later part of the Acheulean (e.g., Isaac, 1989). Archaeologists have long argued (Wynn, 1989; Schick and Toth, 1993; Stout, 2011) that this Early to Late Acheulean transition marks an important increase in hominid cognitive and technological complexity. This is consistent with our previous neuroimaging (Stout et al., 2008; 2011, 2015; Hecht et al., 2014), behavioral (Faisal et al., 2010; Stout et al., 2018), and lithic (Stout et al., 2014) studies of Late Acheulean technology. The Late Acheulean period (~780–400 ka) also coincides with rapid brain size increases in the genus *Homo* (Ruff et al., 1997) and a substantial range expansion and niche diversification (Dennell et al., 2011) including hominid persistence in Africa and Eurasia across major climatic cycles (Stewart and Stringer, 2012). The Early to Middle (<1–0.3 Ma) Pleistocene established the current high amplitude rhythm of extended glacial cycles lasting 100 kyr (Clark et al., 2006), with likely impacts on the distribution and predictability of the resources on which hominins depended (Potts, 1998). Technologically, Late Acheulean bifacial thinning and shaping techniques, including platform preparation and the use of soft hammers spread alongside a host of other technological innovations including the advent of prepared core technology (Tryon et al., 2005), blade production (Johnson and McBrearty, 2010), spear hunting (Thieme, 1997), and hafting (Wilkins et al., 2012). Moreover, biomechanical experiments show that platform preparation as applied during Late Acheulean style handaxe production requires forceful precision-manipulation not necessary when making earlier-Acheulean style handaxes (Key and Dunmore, 2018). These observations motivate our theoretical interest and choice to focus on Late

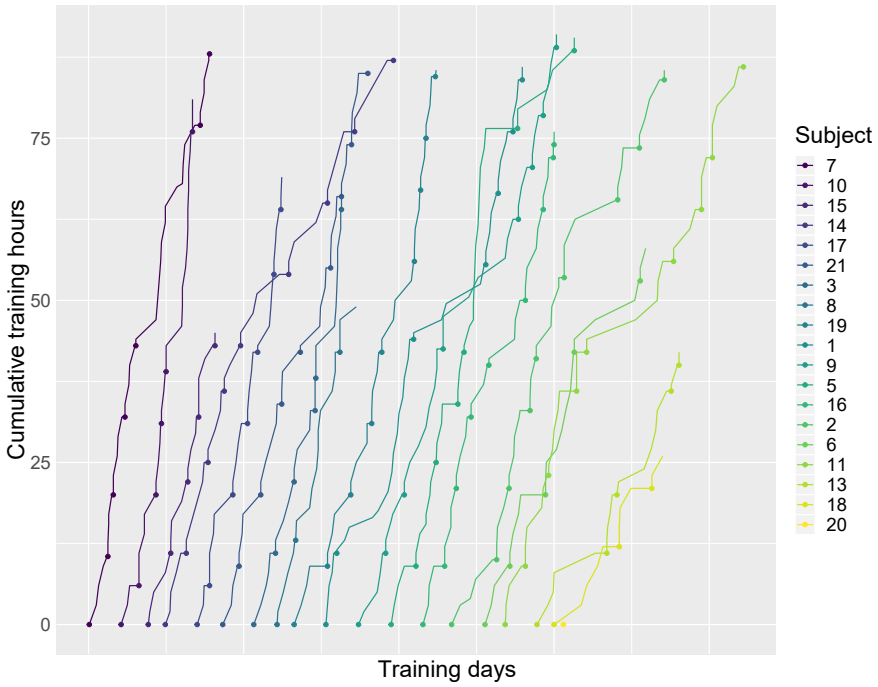

Figure 2. Participants cumulative training hours over the study period. Dots show the location of handaxe assessments across the training period. Participants arranged according to cumulative training rates (decreasing training right from left to right). X-axis signifies training days, but participants are set apart by standard offsets so as to clarify their differences.

Acheulean technological behavior in the present study, although other Paleolithic technologies obviously warrant investigation. Pragmatically, our previous work (Hecht et al., 2014; Stout et al., 2015) established a reasonable expectation that it would be possible to achieve Late Acheulean-relevant levels of trainee skill with resources available to our study.

2. Materials and methods

Seventeen experimental participants (10 female; 48–21 years of age, median = 27) were recruited from Emory University (students and staff) and the surrounding community. Full participation in the study amounted to ~90 hours of which ~80 hours involved training in handaxe production. An additional control group comprising nine participants (6 female, 46–21 years of age, median = 25) participated in the study without receiving training. All participants were right-handed, had no prior knapping experience, and provided written informed consent. The study was approved by Emory University's Internal Review Board (IRB study no: 00067237).

Participants retention and motivation were important considerations for this demanding longitudinal study. A large pool of candidates was generated through an intensive recruitment campaign. Candidates were required to submit 500-word statements describing their reasons for wanting to participate in the study. These were used for an initial screening followed by face-to-face interviews. The result was a pool of 17 experimental participants from diverse educational and professional backgrounds. Our participant pool showed a range of training times that is reflected in the statistical confidence/precision we have for later parts of the learning curve. Six participants left the study before the final assessment with the remaining 11 achieving between 74 and 89 hours training (median = 84 hours) (Fig. 2).

2.1. Participants training and assessment

The experiment aimed to test participant's ability to learn the process of Late Acheulean style handaxe production as it was understood by the instructor. Training was provided by verbal instruction and support from the second author, an experienced knapping instructor (e.g., Khreisheh, 2013) with 10 years knapping practice and specific knowledge of Late Acheulean technology including the Boxgrove handaxe assemblage. She was present at all training sessions to provide help and instruction to participants. All training occurred under controlled conditions at the outdoor knapping area of Emory's Paleolithic Technology Lab, with knapping tools and raw materials provided. Sessions lasted from 0.5 to 3.0 hours (mean = 2.7, median = 3.0) and involved from 1 to 5 participants (mean = 1.7, median = 1, mode = 1) depending on practicalities of participants scheduling. We sought to standardize the pace and duration of training to the greatest extent possible, but participants commitments and life events outside the study produced some unavoidable variation (Fig. 2). Participants were instructed to limit their practice to these practice sessions. All tools and materials were kept at the Emory Paleolithic Technology Lab and knappers were not able to take these away with them to practice.

All sessions were video-recorded, and both the instructor and the participants filled out session record/self-assessment forms (types of instruction given/received, evaluation of success, comments) after each session. Future work will use these detailed records to examine individual differences in learning experiences. All participants were instructed in basic knapping techniques including how to select appropriate percussors, initiate flaking on a nodule, maintain the correct flaking gestures and angles, prepare flake platforms, visualize outcomes, deal with raw material imperfections, and correct mistakes. Handaxe-specific instruction included establishment and maintenance of a bifacial plane, cross-sectional thinning, and overall shaping. The importance of

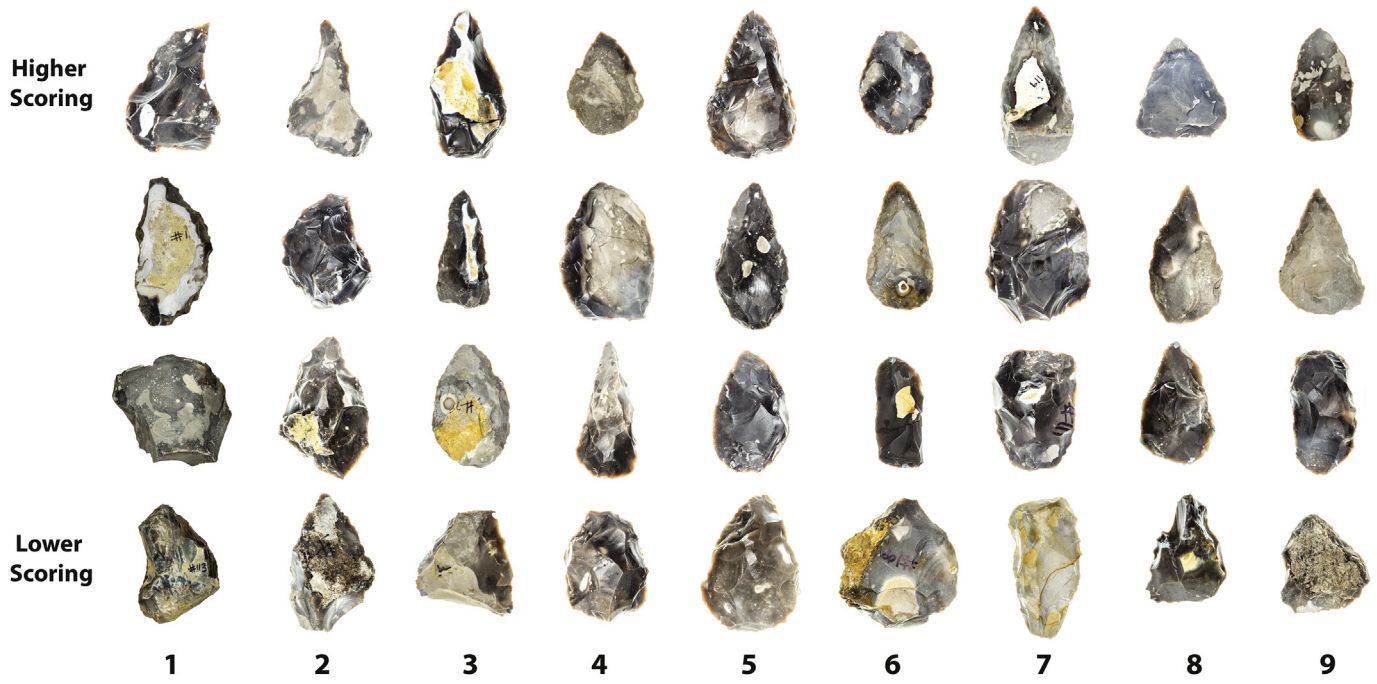

Figure 3. Random selection of novice handaxes from each of the nine assessment periods arranged by descending scores.

producing thin, symmetrical pieces with centered edges was emphasized throughout the training.

Participants were given formal learning assessments at 10 hour increments over the training program, starting prior to any instruction and concluding after the last completed training session. The initial assessment occurred prior to training, but after participants had viewed three 15-minute videos of expert demonstrators producing Late Acheulean style handaxes for other elements of the larger study. Participants were also given four experimental Late Acheulean style handaxes to examine. Participants were never asked to replicate any one specific handaxe. Subsequent assessments occurred after training and reflect participants' attempts to achieve the specific learning objectives imparted by the instructor as outlined above.

Assessments were done individually, and each participant's performance was scored on a 10-point scale (0–5 in 0.5 increments) following a standardized rubric designed to grade technical criteria such as striking angle and force, striking platform preparation, thinning, and the establishment of a bifacial plane (Khreisheh, 2013; Table 2). The skill scores thus assigned were based on actual observed knapping behaviors rather than on physical features of the finished product or its resemblance to any particular 'target' handaxe. At each assessment, participants were provided with a range of spalled flint blanks (purchased from Neolithics.com and sourced from Norfolk, UK) of a similar size and shape from which to select. They were likewise free to choose from a range of hammerstones and antler billets.

After every assessment, the handaxe and all its associated lithic debris were bagged and labeled. Future studies will examine the relationship between these debris and patterns detected in the handaxes. This generated a total sample of 128 handaxes and associated debris. The resulting handaxes were highly variable reflecting the interaction of each participant's individual abilities and learning with inevitable raw material variation and stochastic knapping processes (Fig. 3). Participant scores on behavioral assessments were used as a basis to develop an objective skill metric that can be derived directly from artifact attributes of the kind observable in actual archaeological materials (see Subsection 2.3).

2.2. Handaxe measurements

We recorded nine measurement variables on each of the assessment handaxes ($n = 128$). Table 3 provides details on the measurement variables and the methods used to record them. These variables were chosen to capture elements of handaxe morphological and technological variability that have been proposed to reflect skill (e.g., Schick, 1994) including success in flake production, reduction intensity, imposition of three-dimensional shape and symmetry, and success in maintaining a bifacial edge. Archaeologists use several of these variables (i.e., symmetry indices, unflaked area, shape variables, and flake scar densities) when measuring and comparing prehistoric biface variability (Roe, 1994; Stout, 2002; Lycett, 2008; Machin, 2009; Darmark, 2010; Iovita and McPherron, 2011; Eren et al., 2014; Shipton, 2018; White and Foulds, 2018).

The handaxe measurement data were recorded on photographs taken using a standardized orientation protocol with the handaxe tip placed upwards (Fig. 5). Handaxes were photographed in plan and profile view with a Canon Rebel T3i fitted with a 60 mm macro lens using a photographic stand and adjustable upper and lower light fittings. The camera was positioned directly above the handaxe and kept at a constant height. Photographs were post-processed using Equalight software to adjust for lens and lighting falloff that result from bending light through a lens and its aperture which can affect measurements taken from photographs. Each image was shot with a scale that was then used to rectify the photograph's pixel scale to a real-world measurement scale in Adobe Photoshop. We compared the measurements taken on a sample of the handaxe images versus those derived from the handaxes themselves with a digital caliper and found a ca. 4% measurement error. This error rate compares well with values (ca. 3%) from other studies of inter-observer lithic measurement error

Table 3
Overview of the nine measurement variables recorded on the experimental handaxes and their measurement details.

Variable	Description	Recording software-source
Flake scar density	Flake scars >15 mm in maximum length. Flake scar counts divided by tool mass	Photoshop-rectified 2D photographs
Percent bifacially flaked	Percentage of tool perimeter with alternating bifacial flake scars >15 mm in maximum length	Illustrator-rectified 2D photographs
Unflaked area	Unflaked tool surface area divided by total surface area	Illustrator-rectified 2D photographs
Profile asymmetry index	Degree of tool profile view asymmetry	Flip test on rectified 2D photograph silhouettes
Plan asymmetry index	Degree of tool plan view asymmetry	Flip test on rectified 2D photograph silhouettes
Delta weight	Final tool mass as a percentage of original nodule mass	Measurements taken with a scale
Delta profile thickness CV	Change in tool profile thickness coefficient of variation (CV) relative to starting nodule profile thickness.	ImageJ-rectified 2D photograph silhouettes
PC shape component 1	Shape component 1 extracted from PCA analysis of tool width, thickness, and maximum length measurements taken at 10% increments across the handaxe. PC1 describes the relationship between handaxe length and tip shape.	ImageJ-rectified 2D photograph silhouettes
PC shape component 2	Shape component 2 extracted from PCA analysis of tool width, thickness, and maximum length taken at 10% increments across the handaxe. PC2 describes the relationship between handaxe length and midsection thinning.	ImageJ-rectified 2D photograph silhouettes

with calipers (Fish, 1978; Lyman and VanPool, 2009). The rectified photographs were used along with the actual handaxes to measure the extent of bifacial flaking and to calculate each tool's unflaked area and flake scar density. We relativized these measures by dividing them by the total handaxe perimeter and the tool's total 2D surface area.

We then extracted a series of linear measurements from the photographs for the handaxe shape analysis. The handaxe images were converted to binary black and white format and silhouettes of the tools were extracted in Adobe Photoshop. We wrote a custom ImageJ (Reuden et al., 2017) script to measure width and thickness at 10% increments along the plan and profile handaxe silhouettes starting at the base of each handaxe as well as maximum length along the long axis as defined by the orientation protocol described above. Thereafter, we transformed the linear measurements into shape variables via the geometric mean method (Jungers et al., 1995; Lycett et al., 2006; Lycett, 2009; Eren et al., 2014). This method creates dimensionless, scale-free, variables while preserving shape variation between individual handaxes. These scaled variables were then entered into a principal component analysis (PCA) from which two shape coordinates were extracted (these first two principal component coordinates describe 61% of the total shape variance; Table 4). We used the two PCA coordinates to approximate handaxe shape in the project's multivariate modeling component (see Subsection 2.3). The delta profile thickness coefficient of variation (CV) describes the change in profile thickness variance between starting nodules and finished handaxes. The measurement is derived by calculating the CV of the starting nodule and handaxe thickness and then subtracting the former from the latter.

We measured handaxe profile and plan view symmetry using the freely available flip test software (Hardaker and Dunn, 2005; http://www.fliptest.co.uk). The flip test provides a numerical measure of stone tool symmetry. The test is performed by flipping a photograph of a tool about its vertical axis and measuring the difference between the two superimposed outlines. The 'index of asymmetry' is then calculated as the number of pixels that differ between the original and flipped outlines divided by the tool's maximum length + maximum width squared. Flip test values typically lie between 1.5 and 6.0 with lower values indicating more symmetrical tools. Images used to calculate the index of asymmetry must be scaled to the same height/width ratio and they should ideally have the same pixel resolution ratios for accurate and repeatable results.

2.3. Multivariate modeling: prediction skill scores from handaxe metrics

Our study's main objective was to derive a robust quantitative skill metric for the experimental handaxes. Morgan et al. (2015) approached a similar problem in their experimental study of basic flaking skill by developing a multivariate function approximating the subjective flake quality ratings assigned by three expert coders. Following a similar logic, we used a machine learning approach known as random forest regression (Breiman, 2001) to predict the subjective knapping score assigned to each participant at each assessment using the nine handaxe predictor variables described above. Random forests are a type of additive model that makes predictions by combining decisions from a sequence of base models (trees; Fig. 4). The technique comprises a set of supervised learning algorithms in which collections of decision trees are built from the underlying data. All the base models are constructed independently using a different and random subsample of the larger dataset (Fig. 4).

At the start of each decision tree's construction, each dataset is randomly split into training (2/3) and test (1/3) components through a process of bootstrapping with replacement (cf. Mooney et al., 1993). These two datasets are referred to as in-bag and out-of-bag data respectively. At each fork in the tree, a random subset of predictor variables is selected and used to generate a modeled outcome (skill score) on the in-bag (training) data (Fig. 4). This result is then compared against the same score generated using the out-of-bag (test) data. The average difference between these two score measurements is used to determine the model's overall error rates and the performance of individual predictors. The random forest prediction rate is the unweighted prediction average over a forest of regression trees.

Several factors made random forests the appropriate statistical procedure for this study. Random forest's combined use of bootstrap samples and random predictor draws helps to reduce overall variance inflation in the model building process. The use of random predictor draws at each node in the decision tree also helps to reduce bias driven by any one particular predictor. Random forests are well suited to dealing with multivariate datasets with relatively low sample sizes, they provide intuitive measures to assess variable importance, and they are good at describing complex non-linear relations between predictor and outcome variables.

Random forest algorithms have several parameters that can be tuned for improved model performance (reduced out-of-bag error

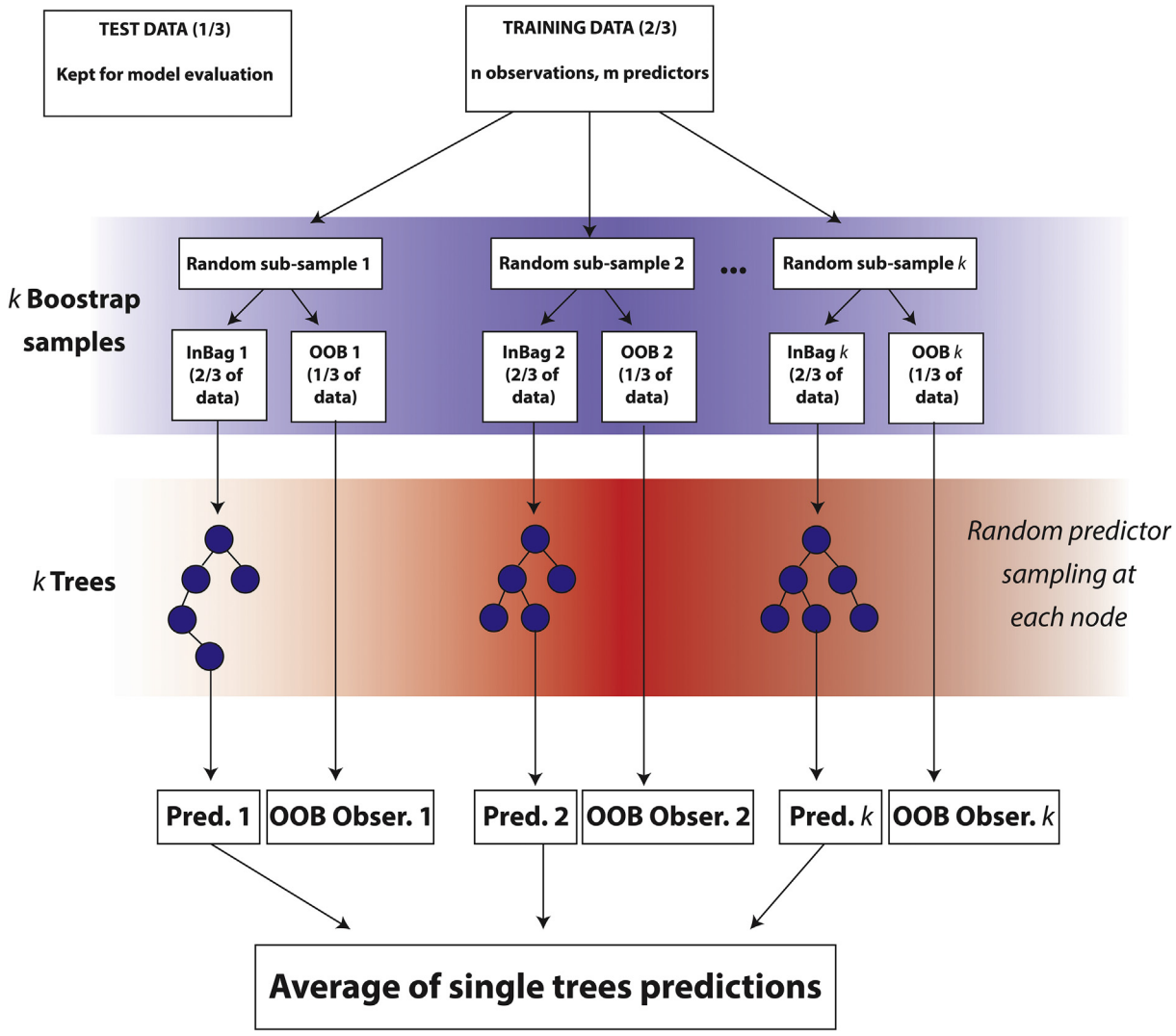

Figure 4. Simplified graphical overview of the random forest modeling process and data operations.

rates). These include the number of trees in the forest (ntrees), the number of random predictor variables chosen at each split in the tree (mtry), and the depth of the trees themselves. We tuned the algorithm's ntrees and mtry components to minimize the model's out-of-bag prediction error rates while leaving the depth of trees at the standard setting (nodes > 1) as standard random forest algorithms are expected to grow full decision trees without pruning.

We followed several steps when building the random forest regression models. First, we assessed the data for overly influential individuals and then divided the handaxe data into test (70%, $n = 83$) and training (30%, $n = 37$) datasets. In each instance, we used Wilcox-rank-sum tests to compare the skill score distributions (our outcome variable) on these two datasets to ensure maximum comparability between training and test data. Second, we ran several iterations of the model with all predictor variables so as to tune the ntrees and mtry model parameters and to minimize out-of-bag prediction error rates. Third, we built a complete model with all nine handaxe predictor variables after which we selected a subset of predictors based on each variable's contribution to overall reduction in model prediction error rates. Fourth, we built a second predictive model using the reduced set of predictor variables. Fifth, we derived a series of model comparison measures (i.e., mean prediction error rates, R^2 and prediction R^2) to compare the models in terms of their fit to the data and their overall skill score prediction accuracies. We also examined each model's predicted values and their respective confidence intervals to assess deviance between instructor ratings and modeled skill scores.

To compare our model's skill score prediction accuracy and relevance to handaxe making more generally, we tested the resulting model on a set of 10 experimental handaxes made by three expert knappers (Fig. 6). The handaxes were made for previous research projects, which similarly aimed to approximate 'Late Acheulean' handaxes explicitly comparable to the Boxgrove assemblage (Faisal et al., 2010; Stout et al., 2011, 2014). We recorded the same nine measurements on the expert handaxes and assigned each handaxe a skill score of five (i.e., maximum) before adding them to the modeled dataset.

2.4. Psychometric tests

To assess potential cognitive correlates of individual variability in handaxe production we compared the modeled handaxe skill scores to a pair of psychometric test scores. Prior to entering the study, participants were given two widely used tests of executive

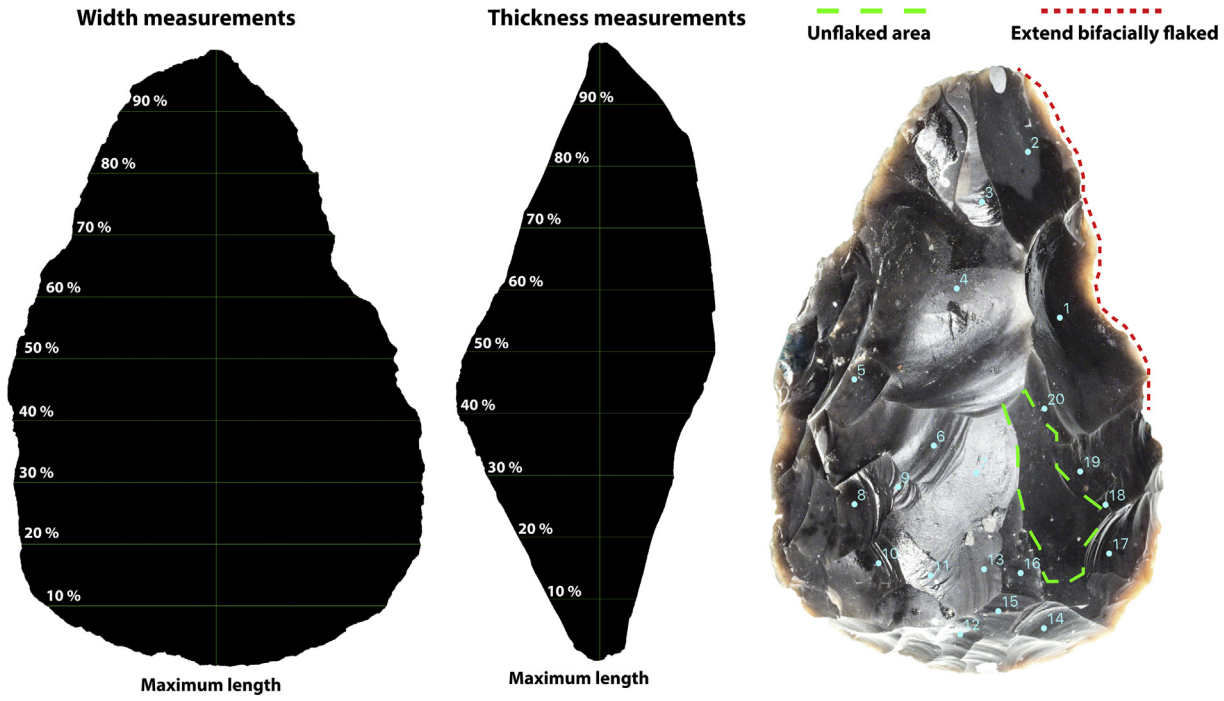

Figure 5. Overview of handaxe measurement protocols following the photogrammetry approach outlined in the text. Blue dots and numbers indicate flake scar counts. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

function: the Tower of London (measuring planning and problem solving; Shallice, 1982) and the Wisconsin Card Sort (measuring 'set shifting' or the ability to display flexibility in the face of changing rules; Grant and Berg, 1948). In the Tower of London (ToL) test, participants reposition 'beads' on 'pegs' to achieve a target configuration in the minimum number of moves. Performance is scored as the number of excess moves made.

For the Wisconsin Card Sort Test (WCST), participants match a series of stimulus cards following various matching rules (match by color, shape, and word). The matching rules shift unannounced throughout the test and participants must notice this change and adjust their behavior accordingly ('task set shifting') rather than persevering with an outdated rule. Participants are scored on the number of such perseverative errors. Both tests were presented under consistent test conditions on a desktop computer in the Paleolithic Technology Lab using the Sanzen Neuropsychological Assessment Tests package (http://neuropsychological-assessment-tests.com).

Neurologically, both of the executive function tests we employed are known to be associated with activity across a distributed 'frontoparietal control network' (Power et al., 2011) that is also consistently activated in studies of stone tool-making (e.g., Stout et al., 2011, 2015). A parametric functional MRI (fMRI) analysis of the ToL task (Wagner et al., 2006) specifically picked out dorsolateral prefrontal cortex (Brodmann Area [BA] 9/46) as being sensitive to task complexity whereas frontopolar cortex (BA 10) showed a more specific response to prospective planning demands. In close agreement with this, an MRI study of a flint knapping judgement task (Stout et al., 2015) found increased functional connectivity in the same two regions for strategic as compared to perceptual-motor judgements. Importantly, performance on these strategic judgments in the scanner correlated with actual handaxe-making success (measured by the 'refinement' index or W/T ratio) outside the scanner. For the WCST, a meta-analysis of neuroimaging studies (Buchsbaum et al., 2005) identified specific contributions of bilateral ventrolateral prefrontal cortex (BA 44/9) for task set switching and right middle (BA 9) and inferior (BA 47) frontal gyrus for response inhibition. Such ventrolateral frontal activity, and in particular the recruitment of right inferior frontal gyrus, is a consistent result of imaging studies of stone tool-making (Stout et al., 2008, 2011, 2018; Putt et al., 2017). It has been suggested (e.g., Stout et al., 2008) that these activations reflect demands on inhibitory and set-shifting processes known to be important for the execution of complex, multi-component behaviors (Dippel and Beste, 2015). The two tests we have selected thus reflect hypotheses regarding the cognitive foundations of handaxe-making skill derived from previous neuroimaging experiments.

Table 4
Overview of two principal component loadings describing 61% of the variability in handaxe width and thickness measurements. Measurements taken with 10% indicating the base and 90% indicating the tip.

Measurement point	PC1	PC2
% variance explained	40	21
width at 10% length	−0.33	0.63
width at 20% length	−0.55	0.63
width at 30% length	−0.76	0.51
width at 40% length	−0.84	0.27
width at 50% length	−0.85	0.02
width at 60% length	−0.81	−0.28
width at 70% length	−0.72	−0.51
width at 80% length	−0.58	−0.67
width at 90% length	−0.45	−0.65
thickness at 10% length	0.41	0.53
thickness at 20% length	0.51	0.55
thickness at 30% length	0.65	0.40
thickness at 40% length	0.70	0.15
thickness at 50% length	0.73	−0.21
thickness at 60% length	0.73	−0.44
thickness at 70% length	0.68	−0.50
thickness at 80% length	0.68	−0.35
thickness at 90% length	0.50	−0.13
maximum length	−0.27	−0.60

Figure 6. Examples of handaxes made by each of the three expert participants.

2.5. Statistical reporting and software

Where possible, we performed all analyses using open source or freely available software. All statistical analyses were performed using the R statistical package (R Core Team, 2013) while all shape measurement data were compiled using open-access scripts in ImageJ (Schindelin et al., 2012). Following recent calls for greater transparency and reproducibility in research (e.g., Marwick, 2017), we include our R code, ImageJ code, and raw data in an open-access repository hosted by the Open Science Framework (Pargeter et al., 2019).

3. Results

Detailed results of all analyses and assessments of the data structure are available through our Open Science Framework data repository listed above (Pargeter et al., 2019). Here we limit discussion to the major findings regarding handaxe measurements and skill acquisition.

3.1. Predicting knapper skill using handaxe measurement variables

This section presents the results of the random forest model building and selection. We began by generating a model using all nine handaxe measurement variables with the trainer's subjective skill rating (on a scale of 1–5 in 0.5 increments) set as the outcome variable. The model was built from 10,000 regression trees with two out of nine measurement variables randomly selected to predict the outcome score at each node in each tree (Table 5).

The complete model has a training data R^2 value of 0.58 and a test data R^2 value of 0.52. The relatively small difference between these two values indicates that the model was not overfit to its training data. The model's mean absolute error (the difference between instructor ratings and modeled scores) is 0.49 measured on the same 0–5 increment scale as the skill score outcome variable. In other words, the model's error is less than the resolution of the original rating scale itself.

Figure 7 depicts the relationship between each of the handaxe measurements and the model's decreased mean squared error rate. Lower values on the x-axis indicate reduced overall contribution to the model's predictive performance. We used a conservation cut-off of 10% to determine variable performance and to eliminate underperforming predictors (Fig. 7). The highest performing metrics were those associated with reduction intensity (flake scar density, delta weight, and percentage unflaked area), shape (PC2 [summarizing tip shape/basal thinning] and plan asymmetry), and the extent of bifacial flaking. Lower performing measurements included shape PC1 (summarizing elongation),

Table 5
Results from the first set of random forest models predicting the subjective skill score from the nine handaxe measurement variables.

Model 1 (all nine predictors)	
r^2 model training set	0.59
r^2 model test set	0.56
n random predictors selected at each node	3
n trees grown	10000
Mean absolute error (scale 0–5)	0.49
Model 2 (six predictors)	
Variables	Percentage bifacially flaked, plan asymmetry, percentage unflaked area, flake scar density, delta weight, shape PC2
r^2 model training set	0.6
r^2 model test set	0.57
n random predictors selected at each node	3
n trees grown	10000
Mean absolute error (scale 0–5)	0.48

Delta profile thickness CV, and profile asymmetry, all contributing less than 10% to reducing the model's overall prediction errors. We removed these variables and built a second model with the six remaining measurements (Table 5). The second reduced and more parsimonious model showed a slightly improved training R^2 value of 0.59 and the same test R^2 value of 0.52. The model's mean absolute error rate dropped marginally to 0.48 with each modeled score remaining within 10% of the actual observed skill ratings. Figure 8 shows the fit between modeled skill scores and the instructor ratings as well as two diagnostic plots for the model's performance. The data show a significant correlation between predicted and instructor skill ratings, normally distributed model residuals and a parabolic prediction confidence interval to deviance value pattern suggesting good model fit and prediction accuracy.

As our objective was not simply to approximate subjective skill scores, but to improve on them using objective and quantifiable artifact data, we tested the behavioral validity of the two scoring methods by comparing scores for each participant's first and last assessments. This comparison enabled us to determine which of the skill scores would better predict a participant's performance across the study. The results in Figure 9 show approximately five times stronger correlation for the model's first and last assessment scores with a significantly different slope ($F\ [2,19] = 3.6$, $p = 0.04$) for the modeled scores compared with the observed skill ratings. This result confirms that the modeled skill scores are better predictors of subsequent participant performance than are subjective skill ratings.

While our random forest models were built on data generated from observations of naïve knappers, it is important to understand how the model would perform on other handaxes made by expert knappers. To test this proposition, we ran ten expert handaxes through the model using the same six predictor variables as we did with the novice knappers (Table 6; Fig. 6). The data show the model assigned all expert handaxes a score above four. Considering that the highest observed score was 4.5 and no naïve knapper scored a five, these values reflect near perfect model scores. When considering the upper 95% confidence interval for these predictions one sees that the majority of handaxes scored above 4.8. The model scored one expert handaxe below 4.5 most likely because their combination of shape idiosyncrasies and some unflaked area were negatively appraised.

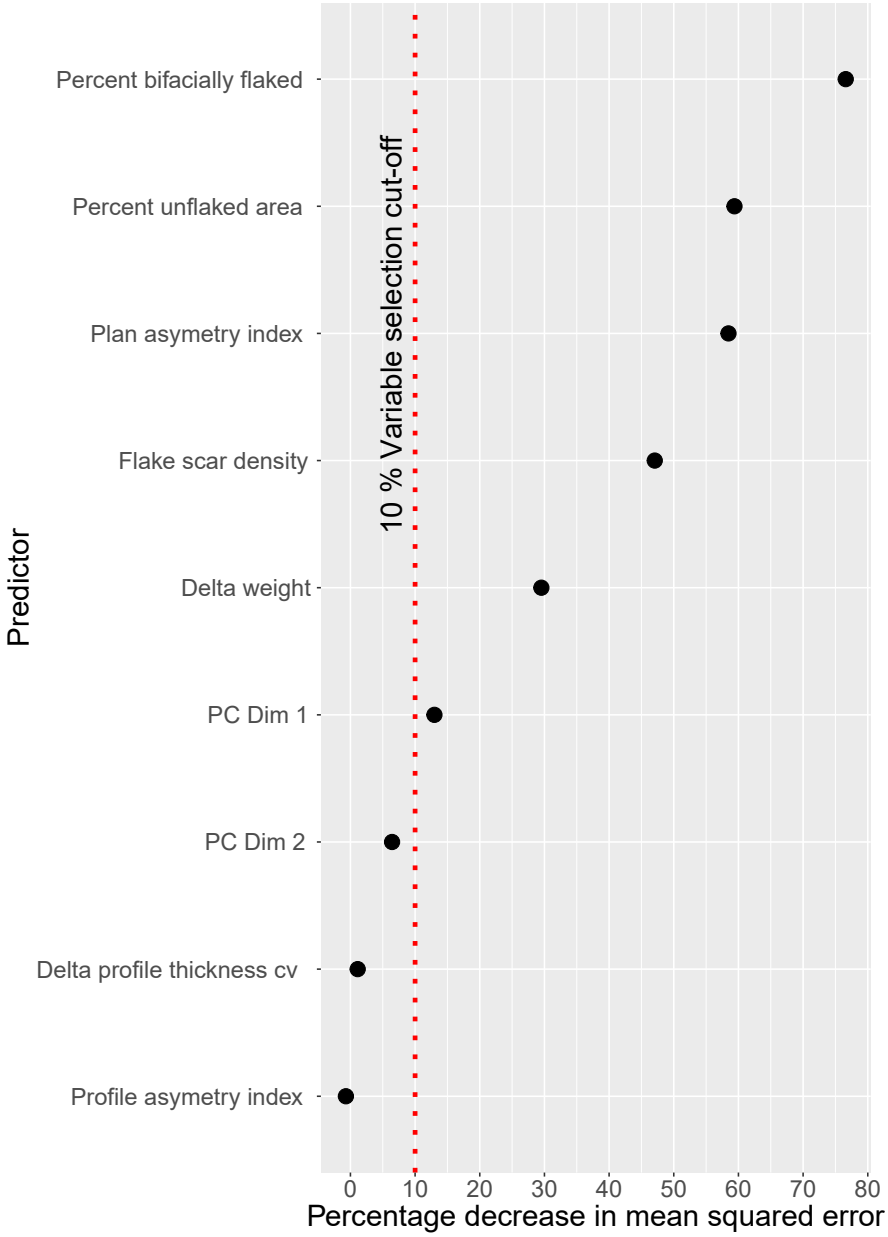

Figure 7. Model predictor comparisons based on their contribution to the percentage decrease in prediction mean squared errors.

3.2. Learning curves

Plotted against the nine assessment intervals, our modeled skill scores can be used to generate a performance curve for the overall participant population (Fig. 11). The shape of this curve follows a characteristic power law known from the learning and psychology literature for over 100 years (Bryan and Harter, 1899), showing that with practice performance generally improves. Power curves show rapid initial improvements followed by asymptotic leveling off as learners reach local performance optima. In our study, initial learning improved rapidly over the first three assessments (~30 hours training), after which we see a decrease in the rate of improvement to the final assessment nine (~90 hours training). Throughout this apparent learning plateau (Gray and Lindstedt, 2017) there is a persistence of occasional outliers with very low model scores. This produces highly skewed distributions with large ranges, similar to the pattern previously observed in an individual learner of 'intermediate' skill (Eren et al., 2011).

Outlier handaxes (those graded lower than the lowest 25% of handaxes) were generally scored lower by our model, which was based solely on finished artifact metrics, than by the instructor, who could base her evaluations on direct observation of participants' behavior during each assessment while considering her impression of their overall progress in the study (Fig. 12). These represent alternative, and potentially complementary, conceptualizations of what it means to evaluate an individual's knapping 'skill-level' on any given day. As we have shown above, our objective, artifact-based method performs relatively well in predicting future performance and rating out-of-sample expert handaxes.

Plotting skill scores across the assessments and examining the handaxes associated with each skill score illustrates how the model generated its scores and penalized different handaxes. For example, Figure 13 shows participants 2 and 19's learning curve and their

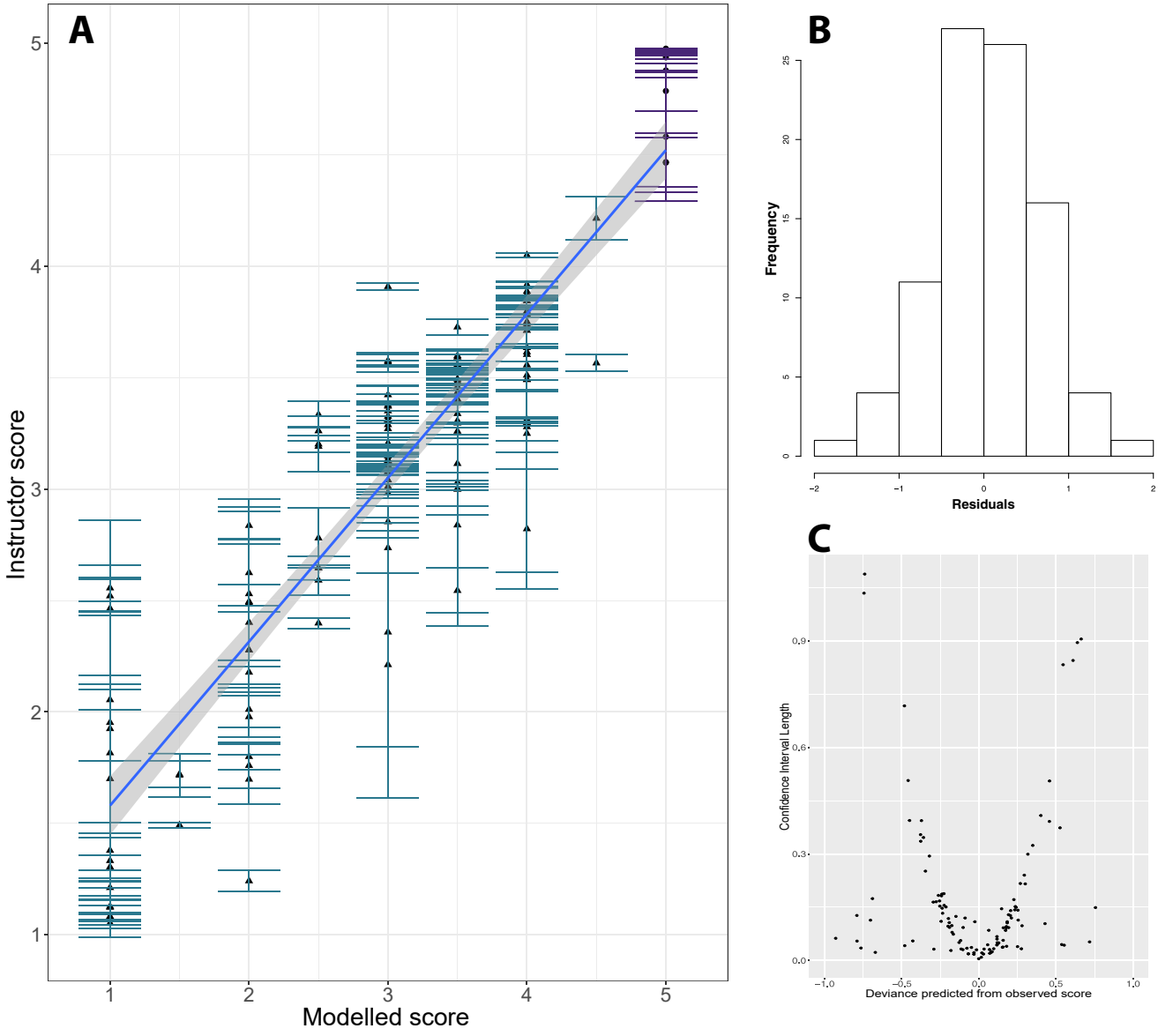

Figure 8. Comparison between modeled skill scores and instructor ratings. Error bars represent 95% confidence intervals; triangles and green error bars represent novices, round symbols and purple error bars represent experts. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

associated handaxes. These handaxes show higher degrees of cortex and lower flake scar counts (i.e., assessment 1) or low rates of bifacial flaking and pieces that were not handaxe-shaped (i.e., assessments 4, 5, and 9) attaining lower scores.

The truncation of our training period at ~90 hours precludes us from saying more about each individual's skill score increases beyond the study interval. Overall, it appears participants were experiencing another period of skill improvement as the study ended. To investigate this trend in more detail, we linearized each individual's learning curve by regressing the square root of the training hours against our modeled skill scores. We used the slope from each linearized learning curve to predict the number of hours each participant would need to reach a perfect score of five. Some participants were close to achieving a score of five when the study ended (i.e., participants 7 and 9), while others would have required upwards of 400 hours training to achieve a perfect score (i.e., participants 5 and 17). The median estimated number of hours to reach a score of five was 225 (range = 121–441).

We have previously shown that handaxes, production debris, and inferred knapping techniques from the same sample of experts knappers considered here are closely comparable to the archaeological assemblage from Boxgrove (Stout et al., 2014). Here we consider handaxe refinement (cross-section width/thickness ratio; Callahan, 1987) as an index of success at bifacial thinning, which was one of our key training goals. This comparison confirms the similarity of expert with archaeological handaxes and the much lower mean refinement achieved by trainees. The results reinforce the point that more extended practice would be required to achieve actual levels of performance documented at Boxgrove, even given maximal social support in terms of raw material and equipment and instruction (Fig. 10).

3.3. Sources of variation in handaxe making and learning outcomes

A next step is to understand possible sources of variation in learning outcomes across participants in our study. One might ask

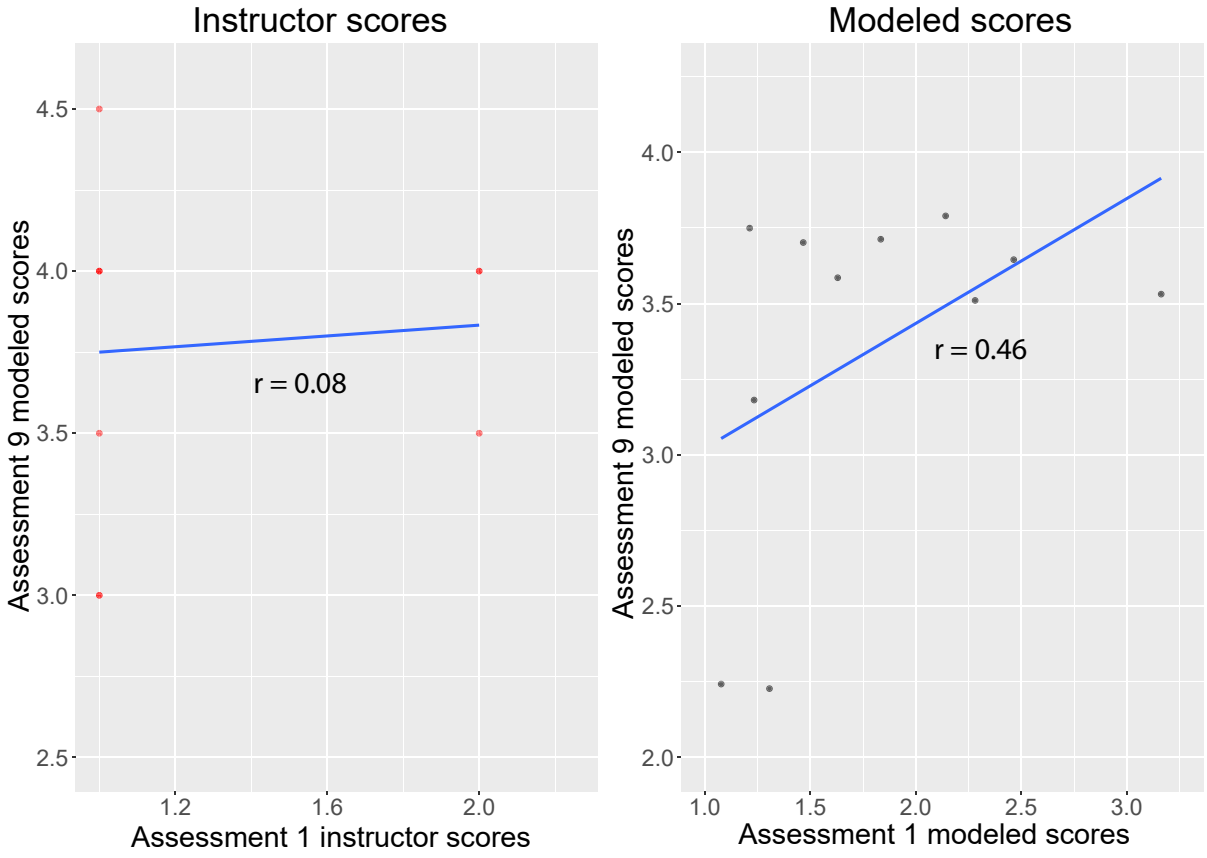

Figure 9. Comparison of modeled and instructor skill scores from the first and last handaxe assessments showing greater predictive power of modeled scores.

what, if any, effect the amount and patterning of practice would have on a participants' handaxe making skills and on subsequent improvements in their skill scores. To examine this question, we plotted the modeled skill scores against the ratio of training hours/day for the 0–30 hours and 40–90 hours training periods. Figure 13 shows the data broken down into earlier stage assessments (1–3) and later stage assessments (4–9) representing two different regions of the study's learning curve (Fig. 11). The results show a statistically significant and positive relationship between score improvements and practice density in the study's later assessments (Fig. 14). This relationship is not present for the earlier assessments.

To further investigate the relationship between handaxe production skill and other aspects of cognitive functioning including planning, problem solving, and the ability to shift between sets of tasks, we compared our skill metrics with the results of the ToL and WCST psychometrics tests. Table 7 summarizes the results of several Bayesian correlation tests comparing the two psychometric evaluations to the modeled skill scores at four intervals across the study. These assessment intervals capture the starting point (assessment 1), the initial period of rapid skill acquisition (assessments 2 and 3), and the first dip in performance (assessment 4). We expect the strongest effect of cognitive functioning to show during the initial, more intensive, periods of learning. To compare each assessment, we examine the central tendency (median) for the posterior distribution of the correlation coefficient (comparable to the r in frequentist approaches), the 90% credible interval, as well as the maximum probability of effect (MPE; the probability that an effect is negative or positive and different from 0). Credible intervals that do not encompass zero are considered strongest.

Table 6
Modeled skill scores for expert handaxes.

Knapper	Modeled score	Lower 95% CI	Upper 95% CI
Expert 1.1	4.98	4.98	4.98
Expert 1.2	4.98	4.98	4.98
Expert 1.3	4.97	4.97	4.97
Expert 1.4	4.97	4.97	4.98
Expert 2.1	4.93	4.92	4.94
Expert 2.2	4.65	4.47	4.82
Expert 2.3	4.76	4.64	4.88
Expert 3.1	4.84	4.79	4.88
Expert 3.2	4.95	4.94	4.95
Expert 3.3	4.23	4.21	4.25

The data show a negative relationship (median r = −0.33, 90% CI = −0.49, 0.09, MPE = 94) between the skill metric and ToL excess moves at assessment 2 with the relationship weakening dramatically by assessment 4 (median r = −0.04, 90% CI = −0.34, 0.44, MPE = 57; Table 7). We found no strong relationships after assessment 4. This suggests that as ToL task error rates increase, handaxe production skill scores decrease but only during phases of rapid learning. The WCST score shows the strongest negative correlation with skill scores at assessment 2 (median r = −0.36, 90% CI = −0.69, −0.03, MPE = 94) with weaker relationships before and after that. The data show that as WCST error rates increase handaxe skill scores decrease during the initial training phases. Overall, the data support our initial prediction that starting performance is poorly predicted by the psychometric tests while the subsequent, rapid learning stage shows stronger relationships. As learning evens out (~40 hours), these relationships diminish to be largely replaced by an effect of practice density.

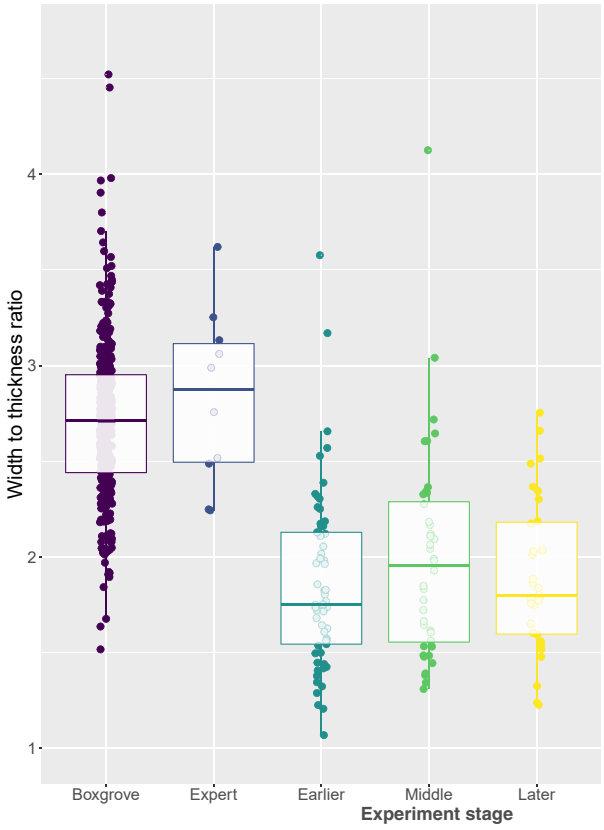

Figure 10. Box-and-whisker plot comparing the width to thickness (handaxe refinement) ratio between this study's experimental handaxes and the Late Acheulean handaxes from Boxgrove. In the plot, the ends of the box are the upper and lower quartiles, so the box spans the interquartile range. The median is marked by a vertical line inside the box. The whiskers are the two lines outside the box that extend to the highest and lowest observations.

4. Discussion

Our skill evaluation methods and training results represent an initial contribution to the broader comparative study of skill acquisition across Paleolithic technologies needed to investigate the complex interactions between tool-making, social organization, cognition, and behavior in human biocultural evolution. Here we have focused on the costs (time and effort) and demands (cognitive and affective) of learning to make Late Acheulean style handaxes because these factors are critical to hypotheses about the patterning of Paleolithic technological change and variation.

Even when a particular technology is 'present' in a population, higher learning costs and/or challenges would be expected to increase the proportion of individuals that, either by choice or accident, fail to acquire the requisite skills. Modeling (e.g., Henrich, 2004; Powell et al., 2009; but see Vaesen et al., 2016) and some archaeological evidence (Roux, 2010) suggest that technologies mastered by smaller (skilled) subsets of the overall population would be increasingly fragile and vulnerable to loss (Shennan, 2013). Indeed, it has been suggested that vulnerability due to small effective population sizes may help to explain the uneven geographic distribution (Schick, 1994; Lycett and Norton, 2010) and technological variation (Nowell and White, 2010) of handaxe production across Middle Pleistocene Eurasia. Recently, it has been shown that the uneven distribution of technical expertise across subpopulations will tend to decrease a population's total equilibrium technological repertoire size (Creanza et al., 2017) and that high learning costs favor patterns of technological 'stasis' reminiscent of the Paleolithic archaeological record (Morgan, 2016). Conversely, such costs will also impose selective pressure for the evolution of enhanced learning capacity following successful innovations (Morgan, 2016). Precisely what needs to be enhanced in order to increase 'learning capacity' is outside the bounds of such models, but it can be addressed through experiments like the current one.

Reasoning along these lines led Stout et al. (2014) to propose that the invention of skill-intensive biface-thinning techniques, including platform preparation, was related to a broader pattern of

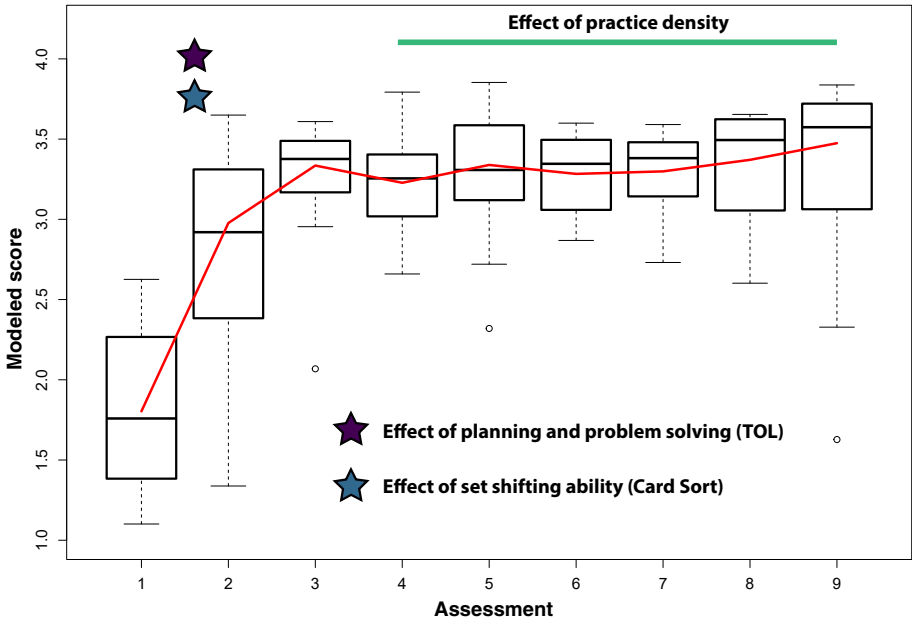

Figure 11. Modeled skill scores compared across the nine handaxe assessments overall learning curve indicated by the smoothed curve. Purple star indicates effect of Wisconsin Card Sort (Card Sort) psychometric measure. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

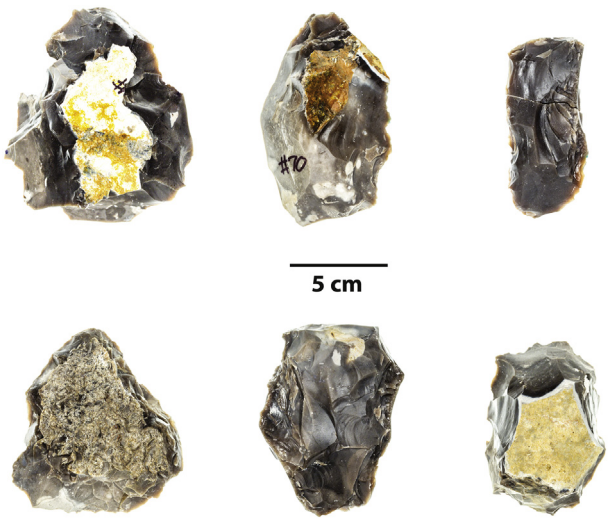

Figure 12. Example of low scoring novice handaxes from assessments 7–9.

accelerated technological and brain size change in the early Middle Pleistocene (~780–400 ka). Key and Dunmore's (2018) 's experimental kinematics data confirm this hypothesis and demonstrate that platform preparation requires forceful precision-manipulative capabilities not needed for earlier Acheulean handaxe production. Similarly, Stout et al. (2019) suggested that a reduction in learning costs associated with the appearance of larger-brained and bodied *Homo erectus* sensu lato at ~1.9 Ma (Antón et al., 2014) might help explain the increased frequency, density, temporal persistence, and geographical and ecological range of tool-making sites after 2.0 Ma (Plummer and Bishop, 2016). This is in contrast to the preceding 1.3 million years (Harmand et al., 2015) during which tool-making appears to have been a rare and discontinuous behavior of marginal net value to hominins (Shea, 2017). Consonant with this, the earliest knapping is heavily reliant on bipolar and passive hammer techniques that are less demanding of manual dexterity (Lewis and Harmand, 2016) and appear to be efficient at relatively low levels of investment in skill learning (Putt, 2015, but see Duke and Pargeter, 2015).

4.1. Skill, learning, and cognition

The proposal that brain size, cognitive capacity, and technological change over human evolution might be linked is not new. However, there is relatively little empirical evidence actually linking specific cognitive capacities and neural substrates to particular stone tool-making techniques and abilities (Shea, 2011). The cognitive underpinnings of knapping skill acquisition in particular are understudied, despite the fact that it is precisely during learning that we expect cognitive demands to be most pronounced (Stout et al., 2015).

Results of the current study document a knapping learning curve that follows a well-known 'power-law of practice' recognized across a wide range of both informal (sewing and cooking) and formal (biology and chess) learning domains (Newell and Rosenbloom, 1981) across humans, monkeys (Brooks et al., 1978), and mice (Shiotsuki et al., 2010). As in these other domains, we find that rapid initial increases in knapping skill are followed by diminishing returns as performance asymptotically approaches a local optimum. Such learning curves are thought to reflect a cognitive process of 'chunking' in which multiple items or operations are combined into summary chunks stored in long term memory. For example, chess experts might encode the position of 15 pieces as a single chunk (e.g., the 'King's Indian Defense'; Gobet and Simon, 1996). Such classic examples typically highlight semantic knowledge, but the same process can be applied to explain motor and perceptual learning across species ranging from pigeons to monkeys (Terrace, 1993). Chunking allows experts to perform complex operations without exceeding the limited attentional resources of working memory, and is expected to produce a power curve if the amount of structure remaining to be summarized decreases as chunk size increases (Newell and Rosenbloom, 1981). Species differences in learning rate (e.g. Brooks et al., 1978) may thus be related to biologically and culturally evolved differences in the memory systems and information compression strategies that support chunk formation (Carruthers, 2013), indicating likely targets for selection acting on skill-learning capacity in human evolution.

Our results support Wynn and Coolidge's (2004) characterization of stone knapping as expert performance in the sense developed by the Long-Term Working Memory Theory (LT-WMT) of Ericsson and Kintsch (1995). LT-WMT and related ideas such as Template Theory (TT; Gobet and Simon, 1996) extend chunking theory to account for the rapidity with which information stored in long term memory can be accessed and manipulated by experts using learned 'retrieval structures' (LT-WMT) or 'templates' (TT). Wynn and Coolidge (2004) focused on the implications for expert knapping, which, being guided by established structures in long term memory, is expected to be relatively undemanding of

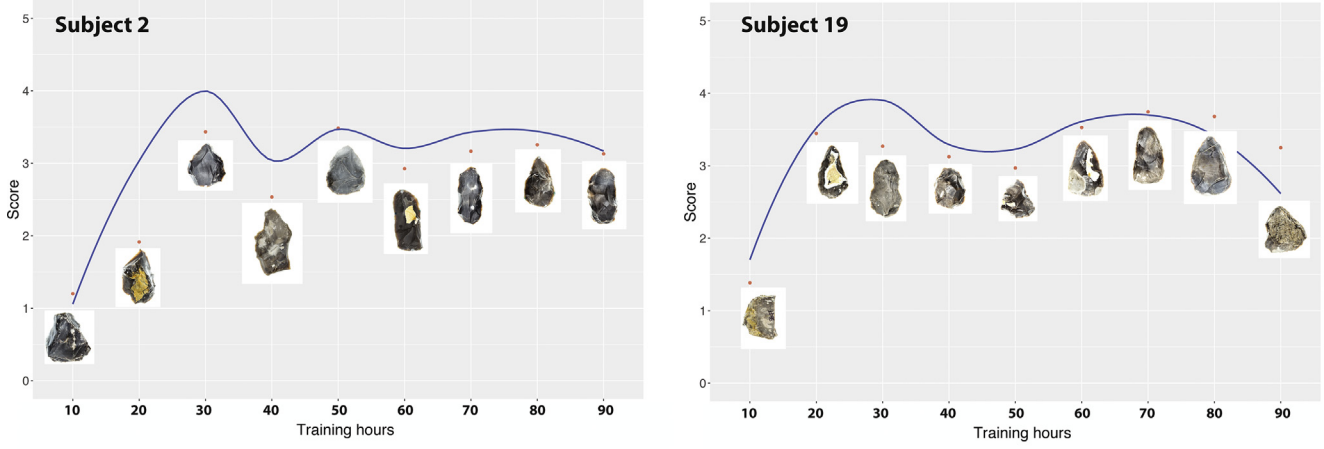

Figure 13. Individual learning curves for participants 2 (left) and 19 (right) showing handaxe variation across their learning trajectories and the corresponding modeled skill scores.

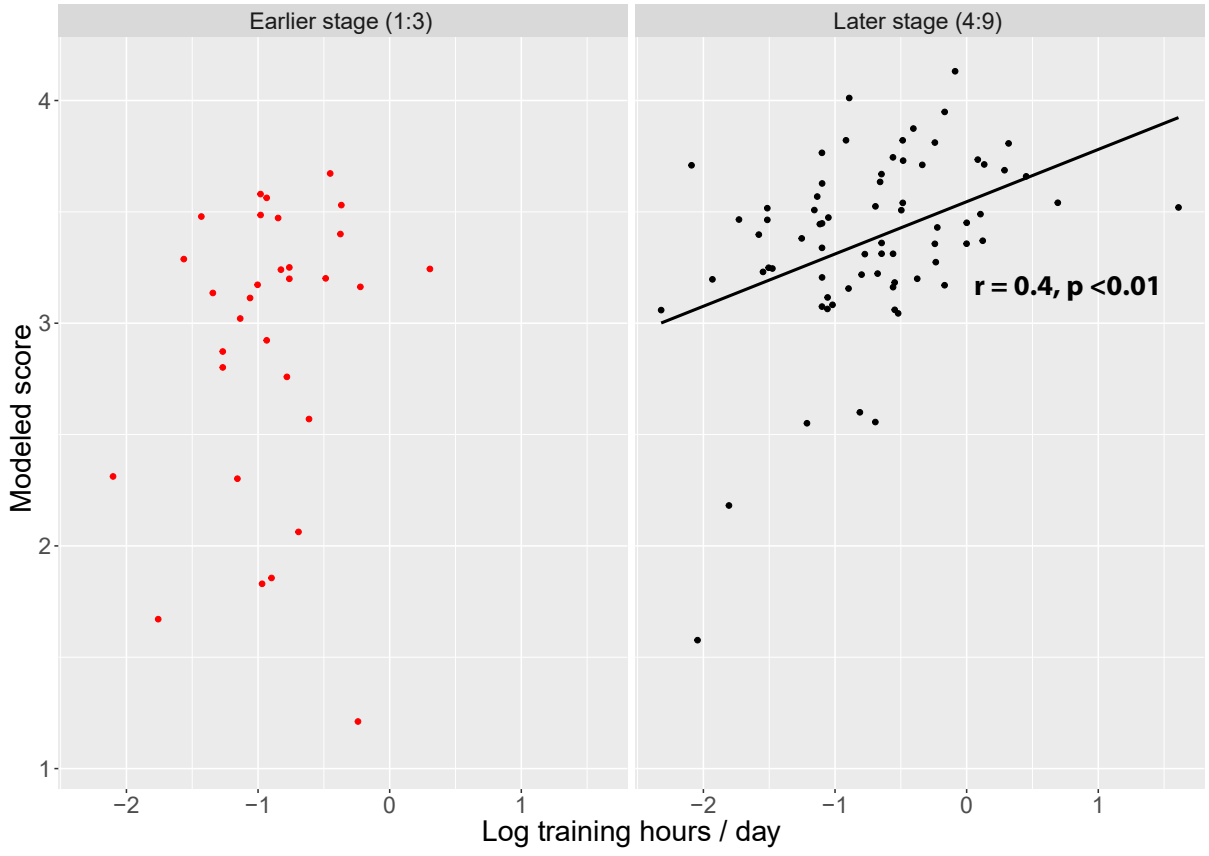

Figure 14. Comparison of modeled skill scores and training density (training hours/day).

the executive functions of working memory. As mentioned above, however, the complementary implication is that these functions are expected to be taxed more heavily during the acquisition of expertise. For example, Guida et al. (2012) provided a two-stage account of skill acquisition (from chunk creation to chunk retrieval, and from chunk retrieval to knowledge structure retrieval), which accords well with neurophysiological evidence of early working memory demands transitioning to functional reorganization in long term memory areas at higher levels of expertise.

Plateaus, dips, and leaps Reliance upon knowledge structures also implies the potential for more complex learning trajectories including plateaus, dips, and leaps in performance. As explained by Gray and Lindstedt (2017), performance plateaus occur when incremental chunk-based learning asymptotes on a suboptimal task structure. For example, steady practice can allow visually-guided typists to plateau at rates of 30–40 words per minute (wpm), whereas touch-typists will reach 60–70 wpm. A visual typist transitioning to touch-typing is expected to experience a temporary dip in performance with the unfamiliar method that is subsequently erased by a power-law leap to the higher performance asymptote associated with the new task structure.

Our results show evidence of just such a suboptimal plateau in handaxe-making across assessments 4–7, with median scores stabilizing just below 3.5. These aggregate results likely subsume substantial individual variation in plateau onset and duration beyond the resolution of our 9 assessment-per-participants data set but do clearly indicate that some such plateau is characteristic of learners in our study. An apparent up-tick in median scores over assessments 8 and 9 suggests that a majority of our participants have initiated a 'leap' to more effective knapping methods by this point. With respect to cognition, leaps (both initial and subsequent) are expected to correspond to cognitively demanding periods of chunk formation, whereas plateaus would reflect periods of long-term memory structure consolidation through dedicated practice (Guida et al., 2012).

Across participants, we observe substantial variation in the precise shape of the learning curve. It is an open question to what extent this shape variation reflects real differences in the process of skill development (Gray and Lindstedt, 2017) vs. the inevitable 'noise' produced via random factors such as raw material variability or knapper performance on a given day. Indeed, inconsistent performance is an expected expression of poorly consolidated skills in novice knappers (Eren et al., 2011).

Table 7
Bayesian correlation results for comparisons of the two psychometric tests and modeled handaxe skill scores.

Test	Assessment	Median r	CI low	CI high	MPE
Tower of London	1	−0.20	−0.49	0.09	79
	2	−0.33	−0.67	0.00	94
	3	−0.31	−0.67	0.05	92
	4	−0.04	−0.34	0.44	57
Wisconsin Card Sort	1	−0.15	−0.14	0.45	86
	2	−0.36	−0.69	−0.03	94
	3	−0.28	−0.63	0.09	88
	4	−0.21	−0.18	0.61	80

Abbreviations: CI = 90% credible interval; MPE = maximum probability of effect.

Executive function As expected from a chunk-based learning account, we found that individual differences in performance during the initial leap from assessment 1 to 3 were correlated with differences in executive function (as measured by our psychometric tests), whereas performance variation during the subsequent learning plateau from assessment 4 through 9 was associated with practice density (hours/day). Although it is likely that some participants experienced another leap or leaps during this period, which might also have been influenced by executive function, the fact that leaps did not occur in synchrony across participants means that such effects would be very difficult to detect from our group data.

The relationship between ToL performance and early-stage knapping success in the current study adds to a growing body of evidence that flexible planning abilities play a role supporting the acquisition of knapping skills (Stout et al., 2015). Similarly, the observed WCST correlation with early-stage knapping success supports the hypothesis that ventrolateral frontal cortex responses documented in neuroimaging studies of handaxe production (Stout et al., 2008, 2011, 2018; Hecht et al., 2014; Putt et al., 2017) reflect demands on inhibitory and set-shifting processes known to be important for the execution of complex, multi-component behaviors. Together with the ToL results and comparative evidence of human enhancements to executive function (Carruthers, 2013; MacLean et al., 2014), this highlights particular cognitive processes and neural systems that would have been likely targets for any selective pressures acting on Paleolithic tool-making aptitude.

Self-control and grit The time and effort needed to acquire knapping skills also have potentially important cognitive and affective implications. Learning to knap not only takes significant amounts of deliberate practice, but would appear to involve a mixture of (presumably rewarding) rapid progress with more extended asymptotic plateaus of diminishing returns and even the occasional need to accept temporary dips in performance while transitioning to more advanced methods.

Our study was designed to extrinsically impose a standardized pace and duration of training and did not included direct measures of participant self-control and grit (Duckworth and Gross, 2014). Nevertheless, unavoidable variation due to participant commitments and life events outside the study revealed a significant effect of practice density on handaxe performance. In a more naturalistic learning context, intrinsic motivation would be critical to maintain practice density even during extended periods of little performance gain, and when immediately useful tools are seldom produced. This accords well with ethnographic observations (Stout, 2002) and emphasizes the importance of motivation, persistence, and self-control in knapping skill acquisition. Future experimental studies might more directly operationalize and investigate these self-regulatory factors.

Deliberate practice, in contrast to incidental learning through everyday activity, is explicitly directed toward improving performance (Ericsson et al., 1993). Such practice requires effort and sustained attention and is neither inherently enjoyable nor immediately rewarding in the short term. Prolonged motivation can be particularly challenging for novices engaged in tasks, such as stone knapping, where the relationship between actions and outcomes can be difficult to interpret or visualize (Stout, 2013; Lycett and Eren, 2019). Sustained deliberate practice thus requires both the immediate discipline to focus on the task at hand and the perseverance to stick with practice over the long haul, despite setbacks and frustrations. Duckworth and Gross (2014) referred to these two determinants of success as 'self-control' and 'grit.' The short-term self-control required to resist impulses may be an effective predictor of some positive life outcomes in contemporary society (Hampton et al., 2018), and is thought to rely on the same kinds of prefrontal executive control discussed above with respect to the WCST. Across species, such self-control is strongly associated with brain size and diet breadth (MacLean et al., 2014).

Grit has been less studied, but is also associated with positive outcomes. In one study, the effect of grit on National Spelling Bee rankings was fully mediated by the greater number of hours of deliberate practice accumulated by grittier competitors. Ericsson et al. (1993) suggested that such commitment to practice is motivated by an understanding of long-term consequences, and a recent study found that individual differences in grit were related to resting state activity in dorsomedial prefrontal cortex (Wang et al., 2017). This region is a key node in the so-called 'default mode' network, which sits at the apex of large-scale cortical connectivity gradients (Margulies et al., 2016) and is involved in the introspection, prospection, and planning that Suddendorf and Corballis (2007) referred to as 'mental time travel.' While neither this network nor its functions are entirely unique to humans, both do appear to be substantially elaborated. Among other things, the resulting capacity to vividly relive past failures and visualize future rewards is a critical component of cognitive strategies for self-regulation in the pursuit of long-term goals (Wang et al., 2017). Modern human learners also benefit from teachers, cultural norms, and other social support structures that provide and reinforce such individual strategies (e.g., Stout, 2002). By better understanding the learning demands of particular Paleolithic technologies, we can hope to gain insight into the evolutionary emergence of these critical cognitive and social pillars of the human technological niche. In particular, it would be useful for future studies to evaluate participant self-control and grit using published scales (e.g., de Ridder et al., 2012; Duckworth and Gross, 2014).

4.2. Approximating learning time costs

The current study provides an approximate estimate for time-to-acquisition of later Acheulean handaxe-making expertise for modern humans under an 'unrestricted teaching' condition (121–441 hours of deliberate practice). Results corroborate previous experimental evidence that Paleolithic stone tool-making is a demanding technical skill that can require years to master, even given substantial social support and explicit instruction (Eren et al., 2011; Stout et al., 2015). Our quantitative estimates for time-to-mastery are specific to the case of refined handaxes made on spalled flint, but they do provide at least an approximate reference point for more general application. Other particular cases might be loosely estimated to be more or less demanding than the one examined here, and some less refined flint handaxe assemblages might even be compared directly to earlier stages in our experimental learning curve. However, more precise estimates for different technological variants and raw materials will clearly require additional experiments.

Further experiments will also be needed to assess learning under less supportive training conditions, but available evidence (Morgan et al., 2015) and theory (Vygotsky, 1978; Ericsson et al., 1993) provide no reason to expect reduced support to produce enhanced learning. Lycett and Eren (2019) argued for the merits of deliberate and dedicated pedagogical strategies to overcome the potential for misdirection during the observational learning of knapping (with specific reference to handaxe production). The costs of such extended learning and teaching are potentially important variables influencing human decision-making and social relations. For example, various authors have considered the interlocking relationships between learning demands, technological innovation, adoption, and reproduction, craft specialization, social inequality, and the formation of institutions like guilds and apprenticeship (e.g., Roux, 1990, 2010; Ziman, 2003).

Paleolithic foragers in particular would have had to balance the costs and benefits of making and maintaining technology against investments in reproductive effort, finding food, and avoiding predators (Hames, 1992). Previous considerations of different technologies' costs and benefits have focused largely on functional constraints, production time, and questions related to foraging efficiency (e.g., Ugan et al., 2003; Mackay and Marwick, 2011; Stevens and McElreath, 2015). Rarely have these models considered the costs of skill acquisition. Although our study's observation of ~200 hours of deliberate practice for refined Late Acheulean style handaxe production might not have been particularly onerous if spread over months or years (but note that density effects indicate diminishing returns for low-frequency practice), this number includes only active knapping practice with provided materials. Actual Paleolithic learning would require additional investments in the procurement/preparation of raw materials (e.g., spalling) and knapping tools (e.g., billet production) either by or for learners (e.g., Stout, 2002). Such costs would increase further in contexts where raw materials are unpredictable, and/or of poor quality, insufficient size, or inappropriate shape (Hayden, 1989). Our own study consumed more than 1000 kg of flint and we might expect this to be roughly doubled if we had carried on until all participants achieved expertise. As with more widely recognized production and procurement costs (i.e., Mackay and Marwick, 2011) learning costs would be expected to inform hominin choices and influence spatiotemporal distribution of particular lithic technologies.

4.3. Future research directions

It is an inconvenient fact that the behavioral complexity and extensive learning requirements that make stone knapping skill acquisition an interesting object of study in the first place are the very features that make it so challenging to study. One useful solution that continues to warrant further pursuit is to identify research questions that can be effectively addressed using more tractable 'model' tasks (e.g., Lycett et al., 2016) or a combination of such tasks and the more naturalistic methods explored in this paper. However, many other interesting questions, especially with respect to learning costs and cognitive demands, may be specific to particular technologies and lithic media. One outstanding issue worth pursuing would be the comparison of learning curves for different kinds of lithic technologies (e.g., Oldowan flaking versus different variations of Acheulean handaxe making). Such investigations will require more naturalistic methods like those developed here. In particular such research must combine (1) the lengthy learning periods required to actually replicate Paleolithic performance levels with (2) adequate numbers of participants to produce robust results across multiple experimental conditions and (3) robust methods for measuring learning behaviors and outcomes (Miton and Charbonneau, 2018). Points (1) and (2) are largely pragmatic issues relating to research effort and support, whereas (3) is a substantial methodological challenge that we have sought to address in the current study. The handaxes produced in our study show a wide range of morphological variability and, despite the best intentions of the knappers, many of the tool forms produced would probably not be classified as handaxes in archaeological assemblages. In fact, some of the pieces produced in our experiment show similar morphology to bifaces made in knapping experiments with randomized flake removal (e.g., Moore and Perston, 2016), although it is possible they would differentiate from Moore and Perston's (2016) on other dimensions (e.g., size and number of flake scars, evidence of battering and stepping, invasiveness, etc.) If confirmed in future comparative studies, such metrics might provide criteria for distinguishing unskilled failures to achieve a handaxe from random outcomes of skilled flake production without the intention to shape a biface. In any case, these observations strongly suggest that archaeological studies of skill variation, material culture variability, cultural transmission, and the intentions behind these patterns should be careful to consider complete assemblages rather than only recognized artifact types.

5. Conclusions

The teaching and learning of complex skills is a central pillar of the human technological niche (Stout and Hecht, 2017) that may depend on uniquely human cognitive capacities (Tomasello, 1999; Gärdenfors et al., 2017). While the archaeological record can potentially inform us about the evolutionary history and foundations of these exceptional technological learning abilities, this potential has been limited by the absence of realistic middle range evidence to ground interpretations. To address this, we adopted an experimental approach, training modern participants over realistic periods to make stone tools and collecting behavioral and lithic data. Our results show that it is possible to accurately quantify handaxe making skill using a multivariate dataset comprising conventional measurements recoverable on archaeological handaxes. This skill metric can then be used to trace group-level learning patterns indicative of underlying cognitive processes as well individual differences in aptitude of the kind potentially visible to natural selection. Finally, these individual differences in aptitude can be related to differences in cognitive control capacity as measured by conventional psychometric tests. Our results also provide a practical benchmark for experimental design by documenting the training time required to capture particular aspects of tool-making skill acquisition. For example, we have shown that even ~90 hours of practice with extensive pedagogical support was insufficient for our modern human participants to achieve expertise comparable to that attested in some Middle Pleistocene archaeological collections. This highlights the need for future studies to consider the social and individual mechanisms supporting the motivation, persistence, and self-control needed for such protracted and effortful investments in skill acquisition when assessing the cognitive and evolutionary implications of Paleolithic stone tools.

Acknowledgments

This work was supported by funding from the National Science Foundation of the USA (grants SMA-1328567 & DRL-1631563), the John Templeton Foundation (grant 47994), and the Emory University Research Council. We thank Jan Apel for providing metric data on the Boxgrove Q1B/D handaxes. J.P. also wishes to acknowledge his wife's forbearance during the drafting of the manuscript.

Supplementary Online Material

Supplementary online material to this article can be found online at https://doi.org/10.1016/j.jhevol.2019.05.010.

References

Antón, S.C., Potts, R., Aiello, L.C., 2014. Evolution of early *Homo*: An integrated biological perspective. Science 345, 1236828.
Archer, W., Pop, C.M., Rezek, Z., Schlager, S., Lin, S.C., Weiss, M., Dogandžić, T., Desta, D., McPherron, S.P., 2018. A geometric morphometric relationship

predicts stone flake shape and size variability. Archaeological and Anthropological Sciences 10, 1991–2003.
Bamforth, D.B., Finlay, N., 2008. Introduction: Archaeological approaches to lithic production skill and craft learning. Journal of Archaeological Method and Theory 15(1), 1–27.
Beyene, Y., Katoh, S., WoldeGabriel, G., Hart, W.K., Uto, K., Sudo, M., Kondo, M., Hyodo, M., Renne, P.R., Suwa, G., Asfaw, B., 2013. The characteristics and chronology of the earliest Acheulean at Konso, Ethiopia. Proceedings of the National Academy of Sciences 110(5), 1584–1591.
Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.
Bril, B., Rein, R., Nonaka, T., Wenban-Smith, F., Dietrich, G., 2010. The role of expertise in tool use: skill differences in functional action adaptations to task constraints. Journal of Experimental Psychology: Human Perception and Performance 36, 825–839.
Brooks, V.B., Reed, D.J., Eastman, M.J., 1978. Learning of pursuit visuo-motor tracking by monkeys. Physiology & Behavior 21, 887–892.
Bruner, E., Fedato, A., Silva-Gago, M., Alonso-Alcalde, R., Terradillos-Bernal, M., Fernández-Durantes, M.Á., Martín-Guerra, E., 2018. Cognitive archeology, body cognition, and hand-tool interaction. Progress in Brain Research 238, 325–345.
Bryan, L.W., Harter, N., 1899. Studies on the telegraphic language: The acquisition of a hierarchy of habits. Psychological Review 6, 345–375.
Buchsbaum, B.R., Greer, S., Chang, W.L., Berman, K.F., 2005. Meta-analysis of neuroimaging studies of the Wisconsin Card-Sorting task and component processes. Human Brain Mapping 25, 35–45.
Callahan, E., 1987. Primitive Technology: Practical Guidelines for Making Stone Tools, Pottery, Basketry, etc. The Aboriginal Way. Piltdown Productions, Lynchburg.
Carruthers, P., 2013. Evolution of working memory. Proceedings of the National Academy of Sciences 110(Supplement 2), 10371–10378.
Cataldo, D.M., Migliano, A.B., Vinicius, L., 2018. Speech, stone tool-making and the evolution of language. PLoS One 13, e0191071.
Clark, P.U., Archer, D., Pollard, D., Blum, J.D., Rial, J.A., Brovkin, V., Mix, A.C., Pisias, N.G., Roy, M., 2006. The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO_2. Quaternary Science Reviews 25, 3150–3184.
Creanza, N., Kolodny, O., Feldman, M.W., 2017. Cultural evolutionary theory: How culture evolves and why it matters. Proceedings of the National Academy of Sciences USA 114, 7782–7789.
Darmark, K., 2010. Measuring skill in the production of bifacial pressure flaked points: a multivariate approach using the flip-test. Journal of Archaeological Science 37, 2308–2315.
Dennell, R.W., Martinón-Torres, M., Bermúdez de Castro, J.M., 2011. Hominin variability, climatic instability and population demography in Middle Pleistocene Europe. Quaternary Science Reviews 30, 1511–1524.
de Ridder, D.T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F.M., Baumeister, R.F., 2012. Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. Personality and Social Psychology Review 16, 76–99.
Dippel, G., Beste, C., 2015. A causal role of the right inferior frontal cortex in implementing strategies for multi-component behaviour. Nature Communications 6, 6587.
Duckworth, A., Gross, J.J., 2014. Self-control and grit: Related but separable determinants of success. Current Directions in Psychological Science 23, 319–325.
Duke, H., Pargeter, J., 2015. Weaving simple solutions to complex problems: An experimental study of skill in bipolar cobble-splitting. Lithic Technology 4, 349–366.
Edwards, S., 2001. A modern knapper's assessment of the technical skills of the Late Acheulean biface workers at Kalambo Falls. In: Clark, J.D., Cormack, S. Chin, Kleindienst, M.R. (Eds.), Kalambo Falls Prehistoric Site. Cambridge University Press, Cambridge, pp. 605–611.
Eren, M.I., Bradley, B., Sampson, C.G., 2011. Middle Paleolithic skill level and the individual knapper: An experiment. American Antiquity 76, 229–251.
Eren, M.I., Roos, C.I., Story, B.A., von Cramon-Taubadel, N., Lycett, S.J., 2014. The role of raw material differences in stone tool shape variation: an experimental assessment. Journal of Archaeological Science 49, 472–487.
Eren, M.I., Lycett, S.J., Patten, R.J., Buchanan, B., Pargeter, J., O'Brien, M.J., 2016. Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-driven Archaeology. Ethnoarchaeology 8(2), 103–136.
Ericsson, K.A., Krampe, R.T., Tesch-Romer, C., 1993. The role of deliberate practice in the acquisition of expert performance. Psychological Review 100, 363–406.
Ericsson, K.A., Kintsch, W., 1995. Long-term working memory. Psychological Review 102, 211.
Faisal, A., Stout, D., Apel, J., Bradley, B., 2010. The manipulative complexity of Lower Paleolithic stone toolmaking. PLoS One 5, e13718.
Fish, P.R., 1978. Consistency in archaeological measurement and classification: a pilot study. American Antiquity 43, 86–89.
Flenniken, J.J., 1984. The past, present, and future of flintknapping: An anthropological perspective. Annual Review of Anthropology 13, 187–203.
Fragaszy, D.M., Eshchar, Y., Visalberghi, E., Resende, B., Laity, K., Izar, P., 2017. Synchronized practice helps bearded capuchin monkeys learn to extend attention while learning a tradition. Proceedings of the National Academy of Sciences USA 114, 7798–7805.
García-Medrano, P., Ollé, A., Ashton, N., Roberts, M.B., 2018. The mental template in handaxe manufacture: new insights into Acheulean lithic technological behavior at Boxgrove, Sussex, UK. Journal of Archaeological Method and Theory 1–27.
Gärdenfors, P., Högberg, A., 2017. The archaeology of teaching and the evolution of *Homo docens*. Current Anthropology 58(2), 000-000.
Gärdenfors, P., Högberg, A., Donald, M., Haidle, M.N., Gärdenfors, P., Högberg, A., 2017. The archaeology of teaching and the evolution of *Homo docens*. Current Anthropology 58, 188–208.
Geribàs, N., Mosquera, M., Vergès, J.M., 2010. What novice knappers have to learn to become expert stone toolmakers. Journal of Archaeological Science 37, 2857–2870.
Gobet, F., Simon, H.A., 1996. Templates in chess memory: A mechanism for recalling several boards. Cognitive Psychology 31, 1–40.
González-Forero, M., Gardner, A., 2018. Inference of ecological and social drivers of human brain-size evolution. Nature 557, 554–557.
Grant, D.A., Berg, E., 1948. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. Journal of Experimental Psychology 38, 404–411.
Gray, W.D., Lindstedt, J.K., 2017. Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. Cognitive Science 41, 1838–1870.
Guida, A., Gobet, F., Tardieu, H., Nicolas, S., 2012. How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: a two-stage framework. Brain and Cognition 79, 221–244.
Hames, R., 1992. Time allocation. In: Smith, E.A., Winterhalder, B. (Eds.), Evolutionary Ecology and Human Behavior. de Gruyter, New York, pp. 203–235.
Hampton, W.H., Asadi, N., Olson, I.R., 2018. Good things for those who wait: Predictive modeling highlights importance of delay discounting for income attainment. Frontiers in Psychology 9, 1545.
Hardaker, T., Dunn, S., 2005. The Flip Test-a new statistical measure for quantifying symmetry in stone tools. Antiquity 79, 306–307.
Harmand, S., Lewis, J.E., Feibel, C.S., Lepre, C.J., Prat, S., Lenoble, A., Boes, X., Quinn, R.L., Brenet, M., Arroyo, A., Taylor, N., Clement, S., Daver, G., Brugal, J.-P., Leakey, L., Mortlock, R.A., Wright, J.D., Lokorodi, S., Kirwa, C., Kent, D.V., Roche, H., 2015. 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. Nature 521, 310–315.
Hayden, B., 1989. From chopper to celt: The evolution of resharpening techniques. In: Torrence, R. (Ed.), Time, Energy and Stone Tools. Cambridge University Press, Cambridge, pp. 33–43.
Hecht, E.E., Gutman, D.A., Khreisheh, N., Taylor, S.V., Kilner, J., Faisal, A.A., Bradley, B.A., Chaminade, T., Stout, D., 2014. Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. Brain Structure and Function 1–17.
Hecht, E.E., Gutman, D.A., Khreisheh, N., Taylor, S.V., Kilner, J., Faisal, A.A., Bradley, B.A., Chaminade, T., Stout, D., 2015. Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. Brain Structure and Function 220, 2315–2331.
Henrich, J.P., 2004. Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses–the Tasmanian sase. American Antiquity 69, 197–214.
Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? Behavioral and Brain Sciences 33, 61–83.
Henrich, J.P., 2016. The Secret of Our Success : How Culture is Driving Human Evolution, Domesticating our Species, and Making us Smarter. Princeton University Press, Princeton.
Hill, K., Barton, M., Hurtado, A.M., 2009. The emergence of human uniqueness: Characters underlying behavioral modernity. Evolutionary Anthropology 18, 187–200.
Högberg, A., 2018. Approaches to children's knapping in lithic technology studies. Revista de Arqueologia 31, 58–74.
Iovita, R., McPherron, S.P., 2011. The handaxe reloaded: A morphometric reassessment of Acheulian and Middle Paleolithic handaxes. Journal of Human Evolution 61, 61–74.
Iovita, R., Tuvi-Arad, I., Moncel, M.-H., Despriee, J., Voinchet, P., Bahain, J.-J., 2017. High handaxe symmetry at the beginning of the European Acheulian: The data from la Noira (France) in context. PLoS One 12, e0177063.
Isler, K., Van Schaik, C.P., 2014. How humans evolved large brains: Comparative evidence. Evolutionary Anthropology 23, 65–75.
Isaac, G., 1989. Chronology and tempo of cultural change during the Pleistocene (1972). In: Isaac, B. (Ed.), The archaeology of human origins: papers by Glynn Isaac. Cambridge University Press, Cambridge, pp. 37–76.
Johnson, C.R., McBrearty, S., 2010. 500,000 year old blades from the Kapthurin Formation, Kenya. Journal of Human Evolution 58, 193–200.
Jungers, W.L., Falsetti, A.B., Wall, C.E., 1995. Shape, relative size, and size-adjustments in morphometrics. American Journal of Physical Anthropology 38, 137–161.
Kempe, M., Lycett, S., Mesoudi, A., 2012. An experimental test of the accumulated copying error model of cultural mutation for Acheulean handaxe size. PLoS One 7(11), e48333.
Key, A.J., Dunmore, C.J., 2018. Manual restrictions on Palaeolithic technological behaviours. PeerJ 6, e5399.
Kaplan, H., Hill, K., Lancaster, J., Hurtado, A.M., 2000. A theory of human life history evolution: Diet, intelligence, and longevity. Evolutionary Anthropology 9, 156–185.

Khreisheh, N., 2013. The acquisition of skill in early flaked stone technologies: An experimental study. Ph.D. Dissertation, Exeter University.
Kline, M.A., Boyd, R., 2010. Population size predicts technological complexity in Oceania. Proceedings of the Royal Society B 277, 2559–2564.
Lewis, J.E., Harmand, S., 2016. An earlier origin for stone tool making: implications for cognitive evolution and the transition to Homo. Philosophical Transactions of the Royal Society B 371, 20150233.
Lin, S.C., Rezek, Z., Dibble, H.L., 2018. Experimental design and experimental inference in stone artifact archaeology. Journal of Archaeological Method and Theory 25(3), 663–688.
Lombao, D., Guardiola, M., Mosquera, M., 2017. Teaching to make stone tools: new experimental evidence supporting a technological hypothesis for the origins of language. Scientific Reports 7, 14394.
Lombard, M., 2015. Hunting and hunting technologies as proxy for teaching and learning during the Stone Age of southern Africa. Cambridge Archaeological Journal 25, 877–887.
Lycett, S.J., 2008. Acheulean variation and selection: does handaxe symmetry fit neutral expectations? Journal of Archaeological Science 35, 2640–2648.
Lycett, S.J., 2009. Quantifying transitions: Morphometric approaches to Palaeolithic variability and technological change. In: Camps, M., Chauhan, P. (Eds.), Sourcebook of Paleolithic Transitions. Springer, New York, pp. 72–92.
Lycett, S.J., Bae, C.J., 2010. The Movius Line controversy: the state of the debate. World Archaeology 42, 521–544.
Lycett, S.J., Eren, M.I., 2013. Levallois economics: An examination of 'waste' production in experimentally produced Levallois reduction sequences. Journal of Archaeological Science 40, 2384–2392.
Lycett, S.J., Gowlett, J.A.J., 2008. On questions surrounding the Acheulean 'tradition'. World Archaeology 40, 295–315.
Lycett, S.J., Norton, C.J., 2010. A demographic model for Palaeolithic technological evolution: the case of East Asia and the Movius Line. Quaternary International 211, 55–65.
Lycett, S.J., von Cramon-Taubadel, N., Foley, R.A., 2006. A crossbeam co-ordinate caliper for the morphometric analysis of lithic nuclei: a description, test and empirical examples of application. Journal of Archaeological Science 33, 847–861.
Lycett, S.J., Schillinger, K., Eren, M.I., von Cramon-Taubadel, N., Mesoudi, A., 2016. Factors affecting Acheulean handaxe variation: Experimental insights, microevolutionary processes, and macroevolutionary outcomes. Quaternary International 411, 386–401.
Lycett, S.J., Eren, M.I., 2019. Built-in misdirection: On the difficulties of learning to knap. Lithic Technology 44, 8–21.
Lyman, R.L., VanPool, T.L., 2009. Metric data in archaeology: a study of intra-analyst and inter-analyst variation. American Antiquity 74, 485–504.
Machin, A., 2009. The role of the individual agent in Acheulean biface variability: a multi-factorial model. Journal of Social Archaeology 9, 35–58.
Mackay, A., Marwick, B., 2011. Costs and benefits in technological decision making under variable conditions: examples from the Late Pleistocene in southern Africa. In: Marwick, B., Mackay, A. (Eds.), Keeping your Edge: Recent Approaches to the Organisation of Stone Artefact Technology. Archeopress, Oxford, pp. 119–134.
MacLean, E.L., Hare, B., Nunn, C.L., Addessi, E., Amici, F., Anderson, R.C., Aureli, F., Baker, J.M., Bania, A.E., Barnard, A.M., 2014. The evolution of self-control. Proceedings of the National Academy of Sciences USA 111, E2140–E2148.
Magnani, M., Rezek, Z., Lin, S.C., Chan, A., Dibble, H.L., 2014. Flake variation in relation to the application of force. Journal of Archaeological Science 46, 37–49.
Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences USA 113, 12574–12579.
Marwick, B., 2017. Computational reproducibility in archaeological research: Basic principles and a case study of their implementation. Journal of Archaeological Method and Theory 24, 424–450.
Miton, H., Charbonneau, M., 2018. Cumulative culture in the laboratory: methodological and theoretical challenges. Proceedings of the Royal Society B 285, 20180677.
Moncel, M.H., Ashton, N., 2018. From 800 to 500 ka in Western Europe. The oldest evidence of Acheuleans in their technological, chronological, and geographical framework. In: Gallotti, R., Mussi, M. (Eds.), The Emergence of the Acheulean in East Africa and Beyond. Springer, New York, pp. 215–235.
Mooney, C.Z., Duval, R.D., Duvall, R., 1993. Bootstrapping: A Nonparametric Approach to Statistical Inference. Sage Publications, London.
Moore, M.W., Perston, Y., 2016. Experimental insights into the cognitive significance of early stone tools. PLoS One 11(7), e0158803.
Morgan, T.J., 2016. Testing the cognitive and cultural niche theories of human evolution. Current Anthropology 57, 370–377.
Morgan, T.J., Uomini, N.T., Rendell, L.E., Chouinard-Thuly, L., Street, S.E., Lewis, H.M., Cross, C.P., Evans, C., Kearney, R., De la Torre, I., 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. Nature Communications 6, 6029.
Newell, A., Rosenbloom, P.S., 1981. Mechanisms of skill acquisition and the law of practice. In: Andersen, J.R. (Ed.), Cognitive Skills and Their Acquisition. Erlbaum, Hillsdale, pp. 1–55.
Nonaka, T., Bril, B., Rein, R., 2010. How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. Journal of Human Evolution 59, 155–167.
Nowell, A., White, M., 2010. Growing up in the Middle Pleistocene: Life history strategies and their relationship to Acheulian industries. In: Nowell, A., Davidson, I. (Eds.), Stone tools and the evolution of human cognition. University Press of Colorado, Boulder, Colorado, pp. 67–82.
Ohnuma, K., Aoki, K., Akazawa, T., 1997. Transmission of tool-making through verbal and non-verbal communication: Preliminary experiments in Levallois flake production. Anthropological Science 105, 159–168.
Pargeter, J., Khreisheh, N., Stout, D., 2019. Data repository for: Understanding stone tool-making skill acquisition: Experimental methods and evolutionary implications. https://doi.org/10.17605/OSF.IO/H5C8T.
Pitts, M.W., Roberts, M., 1998. Fairweather Eden: Life Half a Million Years Ago as Revealed by the Excavations at Boxgrove. Fromm International, New York.
Plummer, T.W., Bishop, L.C., 2016. Oldowan hominin behavior and ecology at Kanjera South, Kenya. Journal of Anthropological Sciences 94, 29–40.
Potts, R., 1998. Variability selection and hominid evolution. Evolutionary Anthropology 7, 81–96.
Powell, A., Shennan, S., Thomas, M.G., 2009. Late Pleistocene Demography and the Appearance of Modern Human Behavior. Science 324(5932), 1298–1301.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., 2011. Functional network organization of the human brain. Neuron 72, 665–678.
Putt, S.S., 2015. The origins of stone tool reduction and the transition to knapping: An experimental approach. Journal of Archaeological Science: Reports 2, 51–60.
Putt, S.S., Woods, A.D., Franciscus, R.G., 2014. The role of verbal interaction during experimental bifacial stone tool manufacture. Lithic Technology 39, 96–112.
Putt, S.S., Wijeakumar, S., Franciscus, R.G., Spencer, J.P., 2017. The functional brain networks that underlie Early Stone Age tool manufacture. Nature Human Behaviour 1, 0102.
R Core Team, 2013. R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna.
Roche, H., 2005. From simple flaking to shaping : Stone knapping evolution among early hominins. In: Brill, B., Roux, V. (Eds.), Stone Knapping: The Necessary Conditions for a Uniquely Hominid Behaviour. Cambridge University, McDonald Institute, pp. 35–52.
Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., Eliceiri, K.W., 2017. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics 18, 529.
Rein, R., Nonaka, T., Bril, B., 2014. Movement pattern variability in stone knapping: implications for the development of percussive traditions. PLoS One 9, e113567.
Režek, Ž., Dibble, H.L., McPherron, S.P., Braun, D.R., Lin, S.C., 2018. Two million years of flaking stone and the evolutionary efficiency of stone tool technology. Nature Ecology & Evolution 2, 628.
Roe, D.A., 1994. A metrical analysis of selected sets of handaxes and cleavers from Olduvai Gorge. In: Leakey, M.D., Roe, D.A. (Eds.), Olduvai Gorge, Volume 5: Excavations in Beds III, IV and the Masek Beds, 1968-1971. Cambridge University Press, Cambridge.
Roux, V., 1990. The psychological analysis of technical activities: a contribution to the study of craft specialization. Archaeological Review from Cambridge 9, 142–153.
Roux, V., 2010. Technological innovations and developmental trajectories: Social factors as evolutionary forces. In: O'Brien, M.J., Shennan, S. (Eds.), Innovation in Cultural Systems. The MIT Press, Cambridge, pp. 217–234.
Roux, V., Bril, B., Dietrich, G., 1995. Skills and learning difficulties involved in stone knapping. World Archaeology 27, 63–87.
Ruff, C.B., Trinkaus, E., Holliday, T.W., 1997. Body mass and encephalization in Pleistocene *Homo*. Nature 387, 173–176.
Schick, K.D., 1994. The Movius line reconsidered: Perspectives on the early Paleolithic of eastern Asia. In: Robert, S.C., Ciochon, R.L. (Eds.), Integrative Paths to the Past: Paleoanthropological Advances in Honor of F. Clark Howell. Prentice-Hall, Englewood Cliffs, pp. 569–595.
Schick, K.D., Toth, N.P., 1993. Making Silent Stones Speak: Human Evolution and the Dawn of Technology. Simon and Schuster, New York.
Schillinger, K., Mesoudi, A., Lycett, S.J., 2014. Considering the role of time budgets on copy-error rates in material culture traditions: An experimental assessment. PloS One 9, e97157.
Schillinger, K., Mesoudi, A., Lycett, S.J., 2015. The impact of imitative versus emulative learning mechanisms on artifactual variation: implications for the evolution of material culture. Evolution and Human Behavior 36, 446–455.
Schillinger, K., Mesoudi, A., Lycett, S.J., 2016. Copying error, evolution, and phylogenetic signal in artifactual traditions: An experimental approach using "model artifacts". Journal of Archaeological Science 70, 23–34.
Schillinger, K., Mesoudi, A., Lycett, S.J., 2017. Differences in manufacturing traditions and assemblage-level patterns: the origins of cultural differences in archaeological data. Journal of archaeological Method and Theory 24, 640–658.
Shallice, T., 1982. Specific impairments of planning. Philosophical Transactions of the Royal Society B 298, 199–209.
Shea, J.J., 2011. *Homo sapiens* is as *Homo sapiens* was: Behavioral variability vs. "behavioral modernity" in Paleolithic archaeology. Current Anthropology 52, 1–35.
Shea, J.J., 2017. Stone Tools in Human Evolution: Behavioral Differences among Technological Primates. Cambridge University Press, Cambridge.
Shelley, P.H., 1990. Variation in Lithic Assemblages: An Experiment. Journal of Field Archaeology 17(2), 187–193.

Shennan, S., 2013. Demographic continuities and discontinuities in Neolithic Europe: evidence, methods and implications. Journal of Archaeological Method and Theory 20, 300–311.
Shiotsuki, H., Yoshimi, K., Shimo, Y., Funayama, M., Takamatsu, Y., Ikeda, K., Takahashi, R., Kitazawa, S., Hattori, N., 2010. A rotarod test for evaluation of motor skill learning. Journal of Neuroscience Methods 189, 180–185.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676.
Shipton, C., 2018. Biface knapping skill in the East African Acheulean: Progressive trends and random walks. African Archaeological Review 35, 107–131.
Shipton, C., Clarkson, C., 2015. Flake scar density and handaxe reduction intensity. Journal of Archaeological Science: Reports 2, 169–175.
Schuppli, C., Meulman, E.J., Forss, S.I., Aprilinayati, F., Van Noordwijk, M.A., Van Schaik, C.P., 2016. Observational social learning and socially induced practice of routine skills in immature wild orang-utans. Animal Behaviour 119, 87–98.
Stevens, N.E., McElreath, R., 2015. When are two tools better than one? Mortars, millingslabs, and the California acorn economy. Journal of Anthropological Archaeology 37, 100–111.
Stewart, J.R., Stringer, C.B., 2012. Human evolution out of Africa: The role of refugia and climate change. Science 335, 1317–1321.
Stout, D., 2002. Skill and cognition in stone tool production: An ethnographic case study from Irian Jaya. Current Anthropology 45, 693–722.
Stout, D., 2013. Neuroscience of technology. In: Richerson, P.J., Christiansen, M. (Eds.), Cultural Evolution: Society, Technology, Language, and Religion. MIT Press, Cambridge, pp. 157–173.
Stout, D., Chaminade, T., 2007. The evolutionary neuroscience of tool making. Neuropsychologia 45, 1091–1100.
Stout, D., Hecht, E.E., 2017. Evolutionary neuroscience of cumulative culture. Proceedings of the National Academy of Sciences USA 114, 7861–7868.
Stout, D., Khreisheh, N., 2015. Skill learning and human brain evolution: an experimental approach. Cambridge Archaeological Journal 25, 867–875.
Stout, D., Toth, N., Schick, K.D., Chaminade, T., 2008. Neural correlates of Early Stone Age tool-making: technology, language and cognition in human evolution. Philosophical Transactions of the Royal Society of London B 363, 1939–1949.
Stout, D., Passingham, R., Frith, C., Apel, J., Chaminade, T., 2011. Technology, expertise and social cognition in human evolution. European Journal of Neuroscience 33, 1328–1338.
Stout, D., Apel, J., Commander, J., Roberts, M., 2014. Late Acheulean technology and cognition at Boxgrove, UK. Journal of Archaeological Science 41, 576–590.
Stout, D., Hecht, E., Khreisheh, N., Bradley, B., Chaminade, T., 2015. Cognitive demands of Lower Paleolithic toolmaking. PLoS One 10, e0121804.
Stout, D., Chaminade, T., Thomik, A., Apel, J., Faisal, A.A., 2018. Grammars of action in human behavior and evolution. bioRxiv. https://doi.org/10.1101/281543.
Stout, D., Rogers, M.J., Jaeggi, A.V., Semaw, S., 2019. Archaeology and the origins of human cumulative culture: A case study from the earliest Oldowan at Gona, Ethiopia. Current Anthropology 60(3). https://doi.org/10.1086/703173.
Suddendorf, T., Corballis, M.C., 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? Behavioral and Brain Sciences 30(03), 299–313.
Tennie, C., Braun, D.R., Premo, L.S., McPherron, S.P., 2016. The Island Test for Cumulative Culture in the Paleolithic. In: Haidle, N.M., Conard, J.N., Bolus, M. (Eds.), The Nature of Culture: Based on an Interdisciplinary Symposium 'The Nature of Culture', Tübingen, Germany. Springer Netherlands, Dordrecht, pp. 121–133.
Tennie, C., Premo, L.S., Braun, D.R., McPherron, S.P., 2017. Early stone tools and cultural transmission: Resetting the null hypothesis. Current Anthropology 58, 652–654.
Terrace, H., 1993. The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. Psychological Science 4(3), 162–169.
Thieme, H., 1997. Lower Palaeolithic hunting spears from Germany. Nature 385, 807–810.
Thomas, D.H., 1986. Points on points: A reply to Flenniken and Raymond. American Antiquity 51, 619–627.
Tomasello, M., 1999. The human adaptation for culture. Annual Review of Anthropology 28, 509–529.
Toth, N., Schick, K., 2009. The Oldowan: the tool making of early hominins and chimpanzees compared. Annual Review of Anthropology 38, 289–305.
Thornton, A., Lukas, D., 2012. Individual variation in cognitive performance: developmental and evolutionary perspectives. Philosophical Transactions of the Royal Society of London B 367, 2773–2783.
Tryon, C.A., McBrearty, S., Texier, P.-J., 2005. Levallois lithic technology from the Kapthurin Formation, Kenya: Acheulian origin and Middle Stone Age diversity. African Archaeological Review 22, 199–229.
Ugan, A., Bright, J., Rogers, A., 2003. When is technology worth the trouble? Journal of Archaeological Science 30, 1315–1329.
Vaesen, K., Collard, M., Cosgrove, R., Roebroeks, W., 2016. Population size does not explain past changes in cultural complexity. Proceedings of the National Academy of Sciences USA 113, E2241–E2247.
Vygotsky, L.S., 1978. Mind in Society: The Development of Higher Psychological Process. Harvard University Press, Cambridge.
Wagner, G., Koch, K., Reichenbach, J.R., Sauer, H., Schlösser, R.G., 2006. The special involvement of the rostrolateral prefrontal cortex in planning abilities: an event-related fMRI study with the Tower of London paradigm. Neuropsychologia 44, 2337–2347.
Wang, S., Zhou, M., Chen, T., Yang, X., Chen, G., Wang, M., Gong, Q., 2017. Grit and the brain: spontaneous activity of the dorsomedial prefrontal cortex mediates the relationship between the trait grit and academic performance. Social Cognitive and Affective Neuroscience 12, 452–460.
White, M., Foulds, F., 2018. Symmetry is its own reward: on the character and significance of Acheulean handaxe symmetry in the Middle Pleistocene. Antiquity 92, 304–319.
Whiten, A., 2015. Experimental studies illuminate the cultural transmission of percussive technologies in *Homo* and *Pan*. Philosophical Transactions of the Royal Society of London B 370, 20140359.
Wilkins, J., Schoville, B.J., Brown, K.S., Chazan, M., 2012. Evidence for early hafted hunting technology. Science 338, 942–946.
Williams-Hatala, E.M., Hatala, K.G., Gordon, M., Key, A., Kasper, M., Kivell, T.L., 2018. The manual pressures of stone tool behaviors and their implications for the evolution of the human hand. Journal of Human Evolution 119, 14–26.
Winton, V., 2005. An investigation of knapping-skill development in the manufacture of Palaeolithic handaxes. In: Roux, V., Bril, B. (Eds.), Stone Knapping: The Necessary Conditions For a Uniquely Hominin Behaviour. Cambridge University Press, Cambridge, pp. 109–116.
Wynn, T., 1989. The Evolution of Spatial Competence. University of Illinois Press, Chicago.
Wynn, T., Coolidge, F.L., 2004. The expert Neandertal mind. Journal of Human Evolution 46, 467–487.
Wynn, T., Coolidge, F.L., 2016. Archeological insights into hominin cognitive evolution. Evolutionary Anthropology 25, 200–213.
Ziman, J., 2003. Technological Innovation as an Evolutionary Process. Cambridge University Press.