

1 Testing the motor and cognitive foundations of Paleolithic
2 social transmission

3 Justin Pargeter* Megan Beney Kilgore† Cheng Liu‡ Dietrich Stout§

4 **Abstract**

5 Stone tools provide key evidence of human cognitive evolution but remain difficult to
6 interpret. Toolmaking skill-learning in particular has been understudied even though: 1) the
7 most salient cognitive demands of toolmaking should occur during learning, and 2) variation
8 in learning aptitude would have provided the raw material for any past selection acting on
9 tool making ability. However, we actually know very little about the cognitive prerequisites
10 of learning under different information transmission conditions that may have prevailed
11 during the Paleolithic. This paper presents results from a pilot experimental study to trial new
12 experimental methods for studying the effect of learning conditions and individual differences
13 on Oldowan flake-tool making skill acquisition. We trained 23 participants for 2 hours to make
14 stone flakes under two different instructional conditions (observation only vs. direct active
15 teaching) employing appropriate raw materials, practice time, and real human interaction.
16 Participant performance was evaluated through analysis of the stone artifacts produced.
17 Performance was compared both across experimental groups and with respect to individual
18 participant differences in grip strength, motor accuracy, and cognitive function measured
19 for the study. Our results show aptitude to be associated with fluid intelligence in a verbally
20 instructed group and with a tendency to use social information in an observation-only group.
21 These results have implications for debates surrounding the cumulative nature of human
22 culture, the relative contributions of knowledge and know-how for stone tool making, and
23 the role of evolved psychological mechanisms in “high fidelity” transmission of information,
24 particularly through imitation and teaching.

25 **Keywords:** Oldowan; Stone toolmaking; Social learning; Individual variation; Cognitive
26 aptitudes; Motor skills

27 **Contents**

28 1 Introduction	2
29 1.1 Individual Differences	4
30 1.2 Teaching, Language, and Tool Making	6
31 1.3 Raw materials and knapping skill	9
32 2 Materials and Methods	10
33 2.1 Participants	10
34 2.2 Study Visit	11

*Department of Anthropology, New York University, New York, NY, USA; Palaeo-Research Institute, University of Johannesburg, Auckland Park, South Africa; justin.pargeter@nyu.edu

†Department of Anthropology, Emory University, Atlanta, GA, USA; megan.elizabeth.beney@emory.edu

‡Department of Anthropology, Emory University, Atlanta, GA, USA; raylc1996@outlook.com

§Department of Anthropology, Emory University, Atlanta, GA, USA; dwstout@emory.edu

35	2.3 Individual Difference Measures	11
36	2.4 Stone Tool Making	13
37	2.5 Lithic Analysis	19
38	2.6 Statistical Analyses	21
39	3 Results	21
40	3.1 Principal Component analyses	22
41	3.2 Relationships between Performance Measures	24
42	3.3 Do trained, untrained, and expert knappers perform differently?	25
43	3.4 Does performance change over time?	29
44	3.5 Do individual differences in motor skill and psychometric measures predict flaking performance?	32
45	3.6 Behavioral observations	40
46		
47	4 Discussion	41
48	4.1 Variance Reduction	42
49	4.2 Knapping Behaviors	42
50	4.3 Learning Strategies	44
51	4.4 Limitations and Prospects	46
52	5 Conclusions	47
53	6 Acknowledgments	47
54	References	47

55 **1 Introduction**

56 Stone tools have long been seen as a key source of evidence for understanding human behavioral
 57 and cognitive evolution (Darwin, 1871; Oakley, 1949; Washburn, 1960). Pathbreaking attempts to
 58 infer specific cognitive capacities from this evidence largely focused on the basic requirements of
 59 tool production (Gowlett, 1984; Isaac, 1976; Wynn, 1979; Wynn & Coolidge, 2004). More recently,
 60 increasing attention has been directed to the processes and demands of stone tool making skill
 61 acquisition (Cataldo et al., 2018; Duke & Pargeter, 2015; Geribàs et al., 2010; Hecht, Gutman,
 62 Khreisheh, et al., 2015; Lombao et al., 2017; Morgan et al., 2015; Nonaka et al., 2010; Pargeter et
 63 al., 2020; Pargeter et al., 2019; Putt et al., 2017, 2019; Putt et al., 2014; Roux et al., 1995; Stout et al.,
 64 2005; Stout et al., 2011; Stout, 2002; Stout & Khreisheh, 2015). This is motivated by the expectation
 65 that the most salient cognitive demands of tool making should occur during learning rather than
 66 routine expert performance (Stout & Khreisheh, 2015) and by interest in the relevance of different
 67 social learning mechanisms such as imitation (Rein et al., 2014; Stout et al., 2019), emulation
 68 (Tehrani & Riede, 2008; Wilkins, 2018), and language (Cataldo et al., 2018; Lombao et al., 2017;

69 Morgan et al., 2015; Ohnuma et al., 1997; Putt et al., 2017; Putt et al., 2014) to the reproduction of
70 Paleolithic technologies.

71 Studies investigating these questions have used a range of different experimental designs (e.g.,
72 varying technological goals/instructions, training times, raw materials, live vs. recorded instruc-
73 tion, lithic/skill assessment metrics, pseudo-knapping tasks etc.) and reached disparate con-
74 clusions regarding the neurocognitive and social foundations of skill acquisition. It is plausible
75 that these discordant results reflect actual diversity in how humans acquire and master stone
76 tool making skills. However, this failure of results to generalize across artificial experimental
77 manipulations (cf. Yarkoni, 2020) also raises doubts regarding the external validity (Eren et al.,
78 2016) of conclusions with respect to real-world Paleolithic learning contexts. To address this,
79 we conducted an exploratory study that draws on lessons from previous research in an attempt
80 to balance the pragmatic and theoretical tradeoffs inherent in experimental studies of stone
81 knapping skill acquisition (Pargeter et al., 2019; Stout & Khreisheh, 2015).

82 Learning real-world skills like stone knapping is highly demanding of time and materials and
83 difficult to control experimentally without sacrificing generalizability to real world conditions.
84 Prior efforts have attempted to navigate these challenges by using various combinations of 1)
85 inauthentic raw materials that are less expensive, easier to standardize, and/or easier to knap,
86 2) video-recorded instruction that is uniform across participants and less demanding of experi-
87 menter time, 3) short learning periods, 4) small sample sizes, and 5) single learning conditions.
88 The difficulty of interpreting results from this growing literature led Stout and Khreisheh (2015:
89 870, emphasis original) to call for “studies with sufficient sample sizes to manipulate learning
90 conditions (e.g. instruction, motivation) and assess individual variation (e.g. performance, psy-
91 chometrics, neuroanatomy) that *also* have realistic learning periods.” The current study attempts
92 to strike a viable balance between these demands by investigating early-stage learning of a rela-
93 tively simple technology (least effort, “Oldowan,” flake production (Reti, 2016; Shea, 2016) under
94 two instructional conditions while collecting data on individual differences in strength, coordina-
95 tion, cognition, social learning, self-control, and task engagement. Unlike any previous study, this
96 allows us to address the likelihood that group effects of training conditions might be impacted by
97 interactions with individual participant differences in aptitude, motivation, or learning style.

98 We focus on early stage learning because it has been found to be relatively rapid, variable across
99 individuals, and predictive of later outcomes (Pargeter et al., 2019; Putt et al., 2019; Stout &

¹⁰⁰ Khreisheh, 2015), and thus provides a reasonable expectation of generating meaningful data
¹⁰¹ on skill and learning variation while minimizing training costs. Moreover, understanding the
¹⁰² minimum training times necessary to detect changes in tool making skill will help archaeologists
¹⁰³ design more realistic and cost-effective experiments. To further manage costs, we limited our
¹⁰⁴ study to only two learning conditions (observation only vs. active teaching). This targets a key
¹⁰⁵ controversy in human evolution, namely the origins of teaching and language (Gärdenfors &
¹⁰⁶ Högberg, 2017; Morgan et al., 2015), while avoiding highly artificial manipulations of dubious
¹⁰⁷ relevance to real-world Paleolithic learning. These choices allowed us to invest more in other
¹⁰⁸ aspects of research design that we identified as theoretically important, including measurement
¹⁰⁹ of individual differences in cognition and behavior, inclusion of an in-person, fully interactive
¹¹⁰ teaching condition, and use of naturalistic raw materials. Sample size remained small in this
¹¹¹ internally funded exploratory study but could easily be scaled up at funding levels typical of pre-
¹¹² and post-doctoral research grants in archaeology.

¹¹³ 1.1 Individual Differences

¹¹⁴ “*The many slight differences... being observed in the individuals of the same species inhabiting*
¹¹⁵ *the same confined locality, may be called individual differences... These individual differences are*
¹¹⁶ *of the highest importance to us, for they are often inherited... and they thus afford materials for*
¹¹⁷ *natural selection to act on and accumulate...*” (Darwin, 1859, Chapter 2)

¹¹⁸ Individuals vary in aptitude and learning style for particular skills (Jonassen & Grabowski, 1993)
¹¹⁹ but this has largely been ignored in studies of knapping skill acquisition, which have instead
¹²⁰ focused on group effects of different experimental conditions. There are good pragmatic reasons
¹²¹ for this, as individual difference studies typically require larger sample sizes and additional data
¹²² collection. However, overlooking these distinctions is not ideal since individual differences can
¹²³ provide valuable insight into the mechanisms, development, and evolution of cognition and
¹²⁴ behavior (Boogert et al., 2018). In particular, patterns of association between cognitive traits and
¹²⁵ behavioral performance can be used to test hypotheses about the cognitive demands of learning
¹²⁶ particular skills and the likely targets of natural selection acting on aptitude. More prosaically,
¹²⁷ individual differences can introduce an unexamined and uncontrolled source of variation in
¹²⁸ group level results. This is especially true in the relatively small “samples of convenience” typical
¹²⁹ of experimental archaeology.

130 While testing hypotheses in evolutionary cognitive archaeology remains a considerable challenge
131 (Wynn, 2017), investigation of individual variation in modern research participants represents
132 one promising direction. For any particular behavior of archaeological interest, it is expected that
133 standing variation in modern populations should remain relevant to normal variation in learning
134 aptitude. The presence of trait variation without impact on learning aptitude would provide
135 strong evidence against the plausibility of the proposed evolutionary relationship. An absence
136 of variation (i.e., past fixation and rigorous developmental canalization) is not expected given
137 the known variability of human brains and cognition (Barrett, 2020; Sherwood & Gómez-Robles,
138 2017). Any confirmatory findings of trait-aptitude correspondence would then have the testable
139 implication that humans should be evolutionarily derived along the same dimension (e.g. Hecht,
140 Gutman, Bradley, et al., 2015).

141 To date, a small number of “neuroarchaeological” studies have reported associations between
142 individual knapping performance and brain structure or physiological responses. Hecht et al.
143 (2015) reported training-related changes in white matter integrity (fractional anisotropy [FA])
144 that correlated with individual differences in practice time and striking accuracy change. The
145 regional patterning of FA changes also varied across individuals, with only those individuals
146 who displayed early increases in FA under the right ventral precentral gyrus (premotor cortex
147 involved in movement planning and guidance) showing striking accuracy improvement over the
148 training period. Putt et al. (2019) similarly found that the proportion of flakes to shatter produced
149 by individuals during handaxe making correlated with dorsal precentral gyrus (motor cortex)
150 activation. Pargeter et al. (2020) used a flake prediction paradigm (modeled after Nonaka et
151 al., 2010) to confirm that striking force and accuracy are important determinants of handaxe-
152 making success. These findings all point to the central role of perceptual-motor systems (Stout &
153 Chaminade, 2007) and coordination (Roux et al., 1995) in knapping skill acquisition. In addition,
154 Putt et al. (2019) also found successful flake production to be associated with prefrontal (working
155 memory/cognitive control) activation and Stout et al. (2015) found that prefrontal activation
156 correlated with success at a strategic judgement (platform selection) task which in turn was
157 predictive of success at out-of-scanner handaxe production. Such investigations are thus starting
158 to chart out the more specific contributions of different neural systems to particular aspects of
159 knapping skill acquisition. To date, however, the cognitive/functional interpretation of systems
160 identified in this manner has largely relied on informal reverse inference (reasoning backward
161 from observed activations to inferred mental processes) from published studies of other tasks

¹⁶² that activated the same regions, an approach which is widely regarded as problematic (Poldrack,
¹⁶³ 2011).

¹⁶⁴ Here we take a more direct, psychometric approach to measuring individual differences in
¹⁶⁵ perceptual-motor coordination and cognition. Psychometric instruments (e.g., tasks, question-
¹⁶⁶ naires) are designed to assess variation in cognitive traits and states, such as fluid intelligence,
¹⁶⁷ working memory, attention, motivation, and personality, that have been of theoretical interest to
¹⁶⁸ cognitive archaeologists (e.g., Wynn & Coolidge, 2016). It is thus surprising that they have been
¹⁶⁹ almost entirely neglected in experimental studies of knapping skill. In the only published example
¹⁷⁰ we are aware of, Pargeter et al. (2019) reported significant effects of variation in planning and
¹⁷¹ problem solving (Tower of London test (Shallice et al., 1982)) and cognitive set shifting (Wisconsin
¹⁷² Card Sort test (Grant & Berg, 1948)) on early stage handaxe learning. Of course, cognition is not
¹⁷³ the only thing that can affect knapping performance. Flake prediction experiments highlight the
¹⁷⁴ importance of regulating movement speed/accuracy trade-offs (Nonaka et al., 2010; Pargeter et
¹⁷⁵ al., 2020) and studies of muscle recruitment (Marzke et al., 1998) and manual pressure (Key & Dun-
¹⁷⁶ more, 2018; Williams-Hatala et al., 2018) during knapping highlight basic strength requirements.
¹⁷⁷ Along these lines, Key and Lycett (2019) found that individual differences in hand size, shape, and
¹⁷⁸ especially grip strength were better predictors of force loading during stone tool use than were
¹⁷⁹ attributes of the tools themselves. However, we are unaware of any such studies of biometric
¹⁸⁰ influences on variation in knapping success. Finally, the time and effort demands of knapping
¹⁸¹ skill acquisition suggest that differences in personality (e.g., self-control and “grit” (Pargeter et
¹⁸² al., 2019), motivation (Stout, 2002), and social vs. individual learning strategies (Miu et al., 2020)
¹⁸³ might also affect learning outcomes. We are again unaware of any previous studies that have
¹⁸⁴ assessed such effects. In this study, we assessed all participants with a battery of tests including
¹⁸⁵ grip strength, movement speed/accuracy, spatial working memory, fluid intelligence, self-control,
¹⁸⁶ tendency to use social information, and motivation/engagement with the tool making task. We
¹⁸⁷ were particularly interested in the possibility that these variables might not only impact learning
¹⁸⁸ generally, but might also have different effects under different learning conditions.

¹⁸⁹ 1.2 Teaching, Language, and Tool Making

¹⁹⁰ “A creature that learns to make tools to a complex pre-existing pattern... must have the kind of
¹⁹¹ abstracting mind that would be of high selective value in facilitating the development of the ability

192 *to communicate such skills by the necessary verbal acts.”* (Montagu, 1976: 267)

193 Possible links between tool making and language have been a subject of speculation for nearly
194 150 years (Engles, 2003, p. [1873]), if not longer (Hewes, 1993), although compelling empirical
195 tests have remained elusive. Over 25 years ago, Toth and Schick (1993) suggested that experiments
196 teaching modern participants to make stone tools in verbal and non-verbal conditions could
197 test the importance of language in the social reproduction of Paleolithic technologies. Ohnuma
198 et al. (1997) were the first to implement this suggestion in a study of Levallois flake production,
199 followed by more recent studies of handaxe making (Putt et al., 2017; Putt et al., 2014) and simple
200 flake production (Cataldo et al., 2018; Lombao et al., 2017; Morgan et al., 2015). This reflects
201 recent interest in the hypothesis that language might be an adaptation for teaching (e.g., Laland,
202 2017; Stout & Chaminade, 2012). Teaching and learning demands of Paleolithic tool making
203 would thus provide evidence of selective contexts favoring language evolution (Montagu, 1976;
204 Morgan et al., 2015; Stout, 2010).

205 Toth and Schick (1993) were, however, careful to point out that extinct hominid learning strategies
206 and capacities might differ from modern experimental participants. Even leaving aside potential
207 species differences in social learning (cf. Morgan et al., 2015; Stout et al., 2019), reliance on
208 explicit verbal instruction varies widely across modern human societies (e.g., Boyette & Hewlett,
209 2017). The WEIRD (Western, educated, industrialized, rich, democratic (Henrich et al., 2010))
210 teachers and learners typical of knapping experiments arguably represent an extreme bias toward
211 such instruction. Simply instructing such participants not to speak during an experiment (or to
212 demonstrate but not gesture, etc. (Morgan et al., 2015)) is likely to underestimate the efficacy of
213 non-verbal teaching and learning in cultural contexts where it is more common, let alone in a
214 hypothetical pre-linguistic hominid species.

215 Such concerns are exacerbated in experiments using pre-recorded instructional videos or ex-
216 tremely short training periods. Video does not allow the interactive teaching that is favored even
217 in formal academic knapping classes (e.g., Shea, 2015) and is almost certainly typical of traditional
218 learning contexts (e.g., Stout, 2002). It is not known how video presentation affects the efficacy of
219 teaching generally, or the relative effectiveness of different forms of instruction. Going further,
220 some experiments have manipulated the presence/absence of verbal instruction by presenting
221 the same video with and without sound (Putt et al., 2017) or the sound track without the video
222 (Cataldo et al., 2018). While this provides experimental control, it does not allow the instructor

223 to adjust their multi-modal (Levinson & Holler, 2014) communication strategies as they would
224 naturally do, for example through pointing and pantomime. To simply remove a communication
225 channel without allowing any such adaptation is highly artificial and risks generating results that
226 cannot be generalized beyond the specific context of the experiment (Yarkoni, 2020). Similarly,
227 unnaturally short training periods (e.g., 5-15 minutes (Lombao et al., 2017; Morgan et al., 2015))
228 might misrepresent the relative efficacy of different teaching strategies under more realistic con-
229 ditions (Stout & Khreisheh, 2015; Whiten, 2015). Even the longest training times to date (Pargeter
230 et al., 2019; Stout & Khreisheh, 2015) have not produced knapping skills comparable to relevant
231 archaeological examples, and were achieved by limiting sample size and using only one teaching
232 condition.

233 For these reasons, we sought to explore a middle path between experimental expedience and real-
234 ism by limiting our experiment to two relatively naturalistic learning conditions and a moderate
235 learning period of two hours. As in previous experiments (Hecht, Gutman, Khreisheh, et al., 2015;
236 Pargeter et al., 2019; Stout et al., 2011) the first condition was unrestricted, interactive instruction
237 in small groups, essentially reproducing the “natural” teaching/learning context familiar (cf. Shea,
238 2015) to our WEIRD instructor and student participants. The second condition allowed observa-
239 tion only, with the experimenter visible making flakes but not interacting in any way with learners.
240 This absence of teaching is again a familiar social context for our participants and did not require
241 any novel behaviors from the instructor. It matches the “imitation/emulation” condition of
242 Morgan et al. (2015) although we make no assumptions regarding learning mechanisms. We did
243 not include a “reverse engineering” or “end-state emulation” condition in which only finished
244 products were visible. This has been advocated as an important baseline or control condition
245 (Whiten, 2015) to distinguish observational from individual learning, but is not likely to model any
246 typical Paleolithic learning context nor to stand as an adequate proxy for the cognition of hominid
247 species with different social learning capacities. There is no reason to assume neurocognitive
248 and behavioral processes of reverse-engineering problem solving in modern humans (e.g., Allen
249 et al., 2020) approximate the social learning processes of hominids with more ape-like action
250 observation/imitation capacities (Hecht, Gutman, et al., 2013; Hecht, Murphy, et al., 2013; Stout
251 et al., 2019).

252 We selected a two-hour learning period for both pragmatic and theoretical reasons. Pargeter et al.
253 (Pargeter et al., 2019) found that even ~90 hours of fully interactive instruction and practice was

insufficient to achieve handaxe-making skills comparable to the later Acheulean site of Boxgrove (García-Medrano et al., 2019; Stout et al., 2014), and estimated actual time to mastery as ranging from 121 to 441 hours for different participants. However, they observed the greatest, fastest, and most individually variable skill increases during the first 20 hours of practice. In addition, initial performance was moderately correlated with later achievement. This suggests that studying early-stage learning may be a pragmatic alternative, especially for research investigating individual differences in aptitude. Studies of simple flake production similarly document large initial variation (Stout & Khriesheh, 2015) and rapid early progress (Putt et al., 2019; Stout & Khriesheh, 2015; Stout & Semaw, 2006). We designed the current study to test the utility of studying learning and variation during the first two hours of simple flaking instruction/practice, in hopes of finding a viable compromise between experimental realism and cost.

1.3 Raw materials and knapping skill

Such undertakings – based on raw material which is never standard, and with gestures of percussion that are never perfectly delivered – cannot be reduced to an elementary repetition of gestures... the realization of elaborate knapping activities necessitates a critical monitoring of the situation and of the decisions adopted all through the process. (Pelegrin, 1990: 117)

Lithic raw materials vary in size, shape, and fracture mechanical properties that affect the difficulty of achieving different knapping goals (Eren et al., 2014). Unfortunately, it can be difficult and/or expensive to procure authentic raw materials. Experimental studies of knapping skill have often used proxy materials such as flint (Cataldo et al., 2018; Morgan et al., 2015; Nonaka et al., 2010), limestone (Stout & Semaw, 2006), porcelain (Khriesheh et al., 2013), or heat-treated chert (Putt et al., 2017, 2019; Putt et al., 2014) to model Oldowan and early Acheulean technologies executed in other materials. As well as being more readily available, these proxies are generally easier to knap. This has the benefit of reducing required practice time, but it is unclear how it might affect learning demands more generally or the efficacy of different learning conditions/strategies specifically.

To address this, some studies have attempted to more closely match experimental and archaeological raw material types (Duke & Pargeter, 2015; Pargeter et al., 2019; Stout et al., 2011). However, raw materials vary across individual clasts within as well as between types. This has led to interest in standardizing experimental core morphology (Nonaka et al., 2010) and composition, even if

284 this means using artificial materials such as porcelain (Khreichéh et al., 2013), brick (Geribàs
285 et al., 2010; Lombao et al., 2017), or foam blocks (Schillinger et al., 2014). Such manipulations
286 enhance experimental control and internal validity (Eren et al., 2016) at the expense of external
287 generalizability to actual archaeological conditions. Specifically, they allow more robust results
288 from smaller samples but eliminate a core element of real-world knapping skill: the ability to
289 produce consistent results from variable materials (Pelegrin, 1990; Stout, 2013). For example,
290 Pargeter et al. (2020) found that predicting specific flaking outcomes on actual handaxe preforms
291 was both more difficult and less technologically important than expected from previous work
292 with standardized, frustum-shaped cores (Nonaka et al., 2010). The alternative to control is to
293 incorporate raw material size, shape, and composition as experimental variables (e.g., Stout
294 et al., 2019). This allows consideration of raw material selection and response to variation as
295 aspects of skill but correspondingly increases the sample sizes required to identify patterning.
296 In considering these issues, we again chose to explore a middle path between pragmatism and
297 realism by employing commercially purchased basalt similar to that known from East African
298 Oldowan sites, allowing clast size and shape to vary within set limits, and selecting the particular
299 clasts provided to each participant to approximate the same distribution.

300 **2 Materials and Methods**

301 This research was approved by the Emory Institutional Review Board (IRB00113024). All partici-
302 pants provided written informed consent and completed a video release form (<https://databrary.org/support/irb/release-template.html>).

304 **2.1 Participants**

305 Twenty-four adult participants with no prior stone knapping experience were recruited from
306 the Emory community using paper fliers and e-mail listserv advertisements. We were unable to
307 replace one participant who failed to attend their scheduled session, resulting in a total sample of
308 23. Eleven participants (6 female, 5 male) completed the Untaught condition and 12 (8 female, 4
309 male) completed the Taught condition.

³¹⁰ **2.2 Study Visit**

³¹¹ Participants were asked to visit the Paleolithic Technology Lab at Emory University to complete
³¹² one three-hour session. Participants were scheduled to attend in six groups of four, however one
³¹³ of these groups had only three participants due to a no-show on the day of the experiment. Each
³¹⁴ visit began with the collection of individual differences measures, which took approximately one
³¹⁵ hour. After that, participants undertook 105 minutes (two hours minus a 15-minute break after 1
³¹⁶ hour) of stone tool making practice. This session was video-recorded, and all lithic products were
³¹⁷ collected. After the tool making task, participants completed an “exit questionnaire” comprising
³¹⁸ the Intrinsic Motivation Inventory (see below).

³¹⁹ Participants were compensated for their time with a \$30 gift card. They also had the opportunity
³²⁰ to earn a performance bonus of \$5, \$10, \$15 or \$20 on the gift card. They were told that this
³²¹ bonus would depend on “how well they did” on the last core of their practice session. The actual
³²² performance measure was not specified, but in order to allow on the spot payment a simple
³²³ measure of the percentage of starting weight removed from the final core was used such that: >
³²⁴ 30% earned \$5, > 40% earned \$10, > 50% earned \$15, > 75% earned \$20.

³²⁵ **2.3 Individual Difference Measures**

³²⁶ We used five individual difference measures for this study:

³²⁷ 1) Grip strength was measured in kilograms using an electronic hand dynamometer (Camry
³²⁸ EH101). Strength was measured twice and the higher value recorded. Grip strength is a
³²⁹ simple measure that is well correlated with overall muscular strength (Wind et al., 2010) and
³³⁰ a range of other health and fitness measures (Sasaki et al., 2007). It is hypothesized to be
³³¹ relevant to generating kinetic energy for fracture initiation (Nonaka et al., 2010) as well as
³³² control and support of the hammerstone (Williams-Hatala et al., 2018) and core (Faisal et
³³³ al., 2010; Key & Dunmore, 2015).

³³⁴ 2) Motor accuracy was assessed using a “Fitts Law” reciprocal tapping task. Fitts Law describes
³³⁵ the trade-off between speed and accuracy in human movement, classically measured
³³⁶ by tapping back and forth between two targets of varying size and spacing (Fitts, 1954).
³³⁷ Archaeologists have proposed (Pargeter et al., 2020; Stout, 2002) that management of this
³³⁸ trade-off is critical to the accurate application of appropriate force seen in skilled knapping

339 (Nonaka et al., 2010; Roux et al., 1995). We implemented this test on a Surface Pro tablet
340 running free software (FittsStudy Version 4.2.8, default settings) developed by the Accessible
341 Computing Experiences lab (Jacob O. Wobbrock, director) at the University of Washington
342 (depts.washington.edu/acelab/proj/fittsstudy/index.html). Participants use a touchscreen
343 pen to tap between ribbons on the screen, with average movement time as the performance
344 metric.

- 345 3) Visuospatial working memory is the capacity to “hold in mind,” which researchers have
346 hypothesized to be important in stone toolmaking performance (Coolidge & Wynn, 2005). It
347 also might support a learning process known as ‘chunking,’ in which multiple items or operations
348 are combined into summary chunks stored in long term memory, that is thought to be
349 important in the acquisition of knapping and other skills (Pargeter et al., 2019). We measured
350 visuospatial working memory using a free n-back task (wmp.education.uci.edu/software/)
351 developed by the Working Memory and Plasticity Laboratory at the University of California,
352 Irvine (Susanne Jaeggi, PI) and implemented in E-Prime software on a desktop computer.
353 In this task, participants are asked to remember the position of blue squares presented
354 sequentially on the screen and touch a key when the current position matches that 1, 2,
355 3...n iterations back. Progression to blocks with increasing values of n is contingent on
356 exceeding a threshold success rate. Performance was measured as the highest n achieved.
- 357 4) Fluid intelligence (Cattell, 1963) refers to the capacity to engage in abstract reasoning and
358 problem solving in a way that is minimally dependent on prior experience. It complements
359 “crystallized intelligence” (the ability to apply learned procedures and knowledge) as one of
360 the two factors (g_f, g_c) comprising so-called “general intelligence” (g). Fluid intelligence is
361 closely related to the executive control of attention and manipulation of information held
362 in working memory (Engle, 2018). It is hypothesized to support technological innovation
363 (Coolidge & Wynn, 2005) and/or the intentional learning of new skills (Stout & Khriesheh,
364 2015; Unsworth & Engle, 2005). We measured fluid intelligence using the short version
365 (Bilker et al., 2012) of the classic Raven Progressive Matrices task, which requires participants
366 to complete increasingly difficult pattern matching questions.
- 367 5) The use of social information for learning and decision making varies across individuals
368 and societies (Molleman et al., 2019). Such variation is a key topic for understanding social
369 learning and cultural evolutionary processes (Heyes, 2018; Kendal et al., 2018; Miu et al.,

370 2020) and represents a potential confound for assessing experimental effects of different
371 social learning conditions. We measured participants' tendency to rely on social information
372 vs. their own insights using the Berlin Estimate AdjuStment Task (BEAST) developed by
373 Molleman et al. (2019). In this task, participants are present with large arrays of items on
374 a screen and asked to estimate the number present. They are then provided with another
375 person's estimate and allowed to provide a second estimate. The participants' average
376 adjustment between first and second estimates provides a measure of their propensity to
377 rely on social information.

378 **2.4 Stone Tool Making**

379 After individual difference testing, participants engaged in a 2-hour stone tool making session,
380 with a 15-minute break after 1 hour. Participants were instructed not to seek out additional
381 training or information on stone tool making (i.e., via the internet) during these breaks. Each
382 group of participants was randomly assigned to one of two experimental conditions: no teaching
383 or teaching. In both conditions, participants were first given an opportunity to inspect and
384 handle examples (**Figure 1**) of the kind of stone tools (flakes) they are being asked to produce.
385 They were told that their objective was to produce as many flakes as possible from the materials
386 provided. This meant that even the untaught condition included some minimal instruction
387 (being told the objective), however this was considered to be unavoidable without creating a
388 much more elaborate and naturalistic context in which participants would develop their own
389 technological goals. Such a design would also be expected to increase behavioral variability,
390 demanding correspondingly larger samples of participants to identify patterns and making direct
391 comparisons with the taught condition.



Figure 1: Subjects examining demonstration flakes prior to the experiment. The demonstration flakes were made from the same basalt as used in the experiment with the same knapping technique.

392 **2.4.1 Raw Materials**

393 Each participant was provided with 9 cores for use over the 2-hour experiment. These cores were
394 produced from larger chunks of a fine-grained basalt purchased from neolithics.com by fracturing
395 them with a sledgehammer. This basalt has not been mechanically compared (e.g., rebound
396 hardness, [Braun et al., 2009](#)) to East African basalts, but appears qualitatively similar to finer
397 grained examples from sites around Lake Turkana (D. Braun, pers. comm.) and at Gona. Spalling
398 produced irregular, angular chunks (**Figure 2**) for use in the experiment, weighing between 459g -
399 1876g (mean = 975g). All cores were weighed, measured (Length, Width, Thickness), and painted
400 white so that new fracture surfaces could be discriminated from those created during production.
401 Cores were sorted by shape and weight and then distributed evenly to each participant. As a

402 result, there were no significant difference across participants in the mean weight (ANOVA, df
403 = 22, F=0.3, p = 0.9; Levene test of homogeneity of variance = 1.04, df1=22, df2 = 184, p = 0.4) or
404 shape (Length × Width/Thickness: ANOVA, df = 22, F=0.4, p = 0.9; Levene statistic = .6, df1=22,
405 df2 = 184, p = 0.9) of cores provided. This was also true for cores provided to participants across
406 the two experimental conditions (Taught vs. Untaught mean weight = 1001g vs. 956g, t = 1.24, df =
407 205, p = 0.2, Levene's Test F = 0.6, p = 0.4; mean shape = 221.43 vs. 221.45, t = -0.003, df = 205, p =
408 0.9, Levene's Test F = 3.8, p= 0.05). Participants were, however, allowed to choose which cores to
409 work on so that differences in the weight and shape of cores actually used across participants and
410 conditions could still emerge as a result of selection bias.

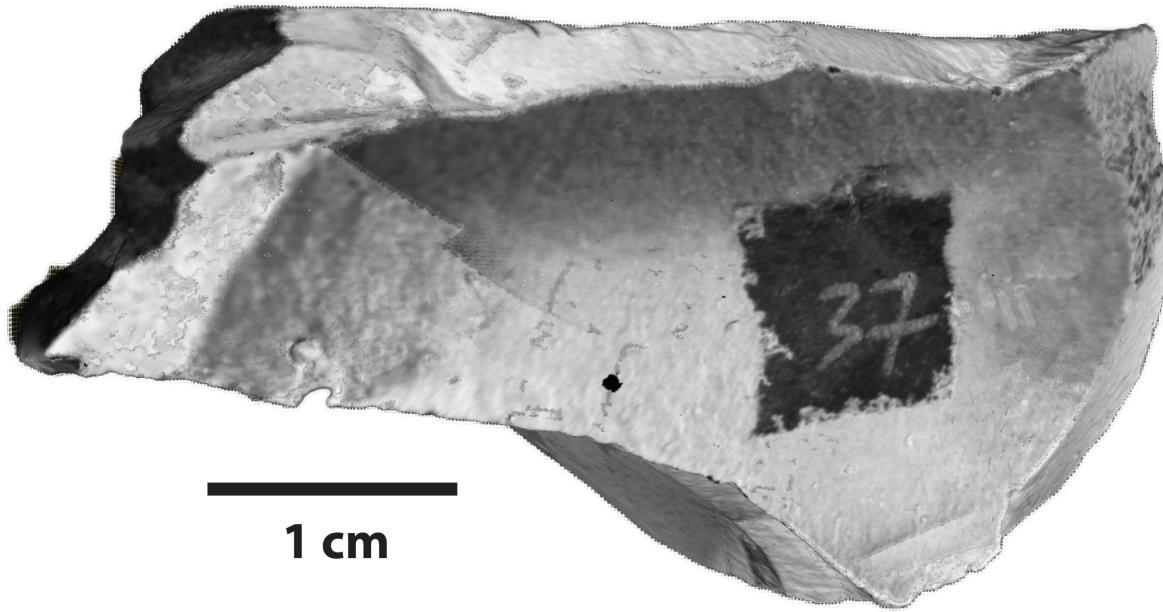


Figure 2: 3D scan of a basalt core prior to knapping. Cores were spray painted white to facilitate subsequent technological analyses of the flake scars and flaking intensity. Scale in the background is in 10mm blocks.

411 Sixty pounds of 3-to-5 inch basalt “Mexican Beach Pebbles” were purchased from a landscaping
412 supply company for use as hammerstones in the experiment. Of these, 90 were selected as
413 suitable for use. These weighed between 213g-1360g (mean = 425) and varied in elongation (L/W
414 = 1.01 to 2.65) and relative thickness ($L \times W / T$ = 90.48 to 283.67). Forty-five stones were placed
415 in the middle of the knapping area ([Figure 3](#)) for participants to freely choose from during the
416 experiment. Broken hammerstones were replaced from the reserve to maintain a consistent
417 number and range of choices. Each hammerstone was numbered and participants’ choices were

⁴¹⁸ recorded along with the number of the core(s) being worked on with a particular hammerstone.

⁴¹⁹ **2.4.2 Experimental Conditions**

⁴²⁰ In both conditions, three researchers were present to record activities and collect materials.
⁴²¹ Participants were seated in a circle ([Shea, 2015](#)) and experiments were video recorded using two
⁴²² cameras. Participants were free to select hammerstones from the common pile and to work on
⁴²³ any or all of their nine assigned cores in any order they preferred ([Figure 3](#)). However, each core
⁴²⁴ and all associateddebitage were collected before participants were allowed to start working on a
⁴²⁵ new core, so it was not possible to partially work and then return to a particular core later. The
⁴²⁶ order of cores used and associated hammerstones were recorded for each participant during the
⁴²⁷ experiment.



Figure 3: Subjects selecting hammerstones.

428 In the untaught condition, a researcher (DS) sat with the participants and made stone tools
429 but remained silent and made no effort to facilitate learning (e.g., through gesture, modified
430 performance, facial expression, attention direction, or verbal instruction). Over the 2-hour
431 period, the researcher completely reduced four cores (one every ~30 minutes). Participants

432 were not restricted from talking to each other, as this would create an unnatural and potentially
433 stressful social context that might affect learning. Participants were asked to avoid any form of
434 communication about the tool making task specifically, and they complied with this request.
435 Participants in this condition thus had the opportunity to observe tool making ([Figure 4](#)) by
436 an expert and/or by other learners, should they choose to do so, but received no intentional
437 instruction.



Figure 4: Subjects observing expert knapper in the untaught condition.

438 In the Taught condition, there were no restrictions on participant interaction and the researcher
439 engaged in direct active teaching ([Kline, 2015](#)) of tool-making techniques through verbal instruc-
440 tion, demonstration, gesture, and shaping of behavior. The instructor has a moderate level of
441 experience teaching basic knapping skills to students in undergraduate archaeology classes and
442 to participants in previous knapping research (e.g., [Stout et al., 2011](#)). The pedagogical strategy
443 employed was based on the instructor's own learning experiences and theoretical interpretations
444 (e.g., [Pargeter et al., 2020](#)), and focused on coaching participants in effective body postures,
445 movement patterns, and grips as well as the assessment of viable core morphology.

446 2.5 Lithic Analysis

447 **Table 1** lists details for the study's various lithic attributes. All finished cores were weighed and
448 measured (L, W, T). Delta weight was calculated as (Start weight-End weight)/Start weight. All
449 detached pieces (DPs) were collected and weighed. We did not sort DPs into types (e.g., whole
450 flakes, fragments) as this would have greatly increased processing time and it is not clear that
451 such distinctions add relevant information regarding utility/desirability beyond that supplied by
452 metrics ([Stout et al., 2019](#)). All DPs larger than 40mm in maximum dimension were photographed
453 and measured. It is conventional in Early Stone Age lithic analysis to employ a 20 mm cut-off.
454 We selected a higher threshold for both pragmatic (analysis time) and theoretical reasons. Flake
455 use experiments have shown that flakes weighing less than 5–10 g or with a surface area below
456 7–10 cm² ([Prascunas, 2007](#)) or with a maximum dimension <50-60 mm ([Key & Lycett, 2014](#))
457 become markedly inefficient for basic cutting tasks. Similarly, data from Oldowan replication
458 experiments ([Stout et al., 2019](#)) show that the utility index (flake cutting edge/flake mass^{1/3}) * (1 -
459 exp[-0.31 * (flake maximum dimension – 1.81)]) developed by Morgan et al. (2015) falls off rapidly
460 below 40mm maximum dimension (Mean Utility < 40mm = 0.508; >=40mm = 0.946; t= 11.99, df =
461 707, p < 0.000). By including weight in our cut-off criteria we also avoid skewing the flake shape
462 distribution by selectively retaining long, thin pieces (i.e., MD > 40, weight < 5g) while discarding
463 rounder pieces of similar (or greater) weight and area.

Table 1: Overview of lithic variables and recording methods used in the study.

Variable	Variable.class	Definition	Recording.method
Delta mass	Ratio	Final core mass as a percentage of original nodule mass. Higher values show more completely flaked nodules.	Digital scale
Mass of large detached pieces / total debitage mass	Ratio	Combined mass of all detached pieces >40mm in maximum dimension and 5g in mass divided by the mass of all material detached from a specific core. Higher values show cores with more detached pieces per knapped material.	Calipers, counts, digital scale
Length	Linear measurement	Measurement along each core and flake's longest axis in mm.	2D photogrammetry
Width	Linear measurement	Measurement taken at 10% increments along a detached piece's longest axis starting at the detached piece's tip (in mm). Same measurement taken orthogonal to a core's length.	2D photogrammetry
Thickness	Linear measurement	Measurement taken at the maximum point in a detached piece's profile and orthogonal to width on the cores.	Calipers

- 464 For measurement, DP length was defined as the longest axis and width as the maximum di-
 465 mension orthogonal to length. Thickness was defined as the maximum dimension orthogonal
 466 to the plane formed by L and W and was measured using calipers. L, W, and plan-view area
 467 measurements were taken from photographs captured using a Canon Rebel T3i fitted with a 60
 468 mm macro lens and attached to a photographic stand with adjustable upper and lower light
 469 fittings. The camera was positioned directly above the flakes and kept at a constant height. DPs
 470 were positioned irrespective of any technological features so that the longest axis was vertical,
 471 and the wider end was placed toward the bottom of the photograph.
- 472 Photographs were post-processed using Equalight software to adjust for lens and lighting falloff
 473 that result from bending light through a lens and its aperture which can affect measurements
 474 taken from photographs. Each image was shot with a scale that was then used to rectify the
 475 photograph's pixel scale to a real-world measurement scale in Adobe Photoshop. Images were
 476 converted to binary black and white format and silhouettes of the tools were extracted in Adobe
 477 Photoshop. We then used a custom ImageJ ([Rueden et al., 2017](#)) script ([Pargeter et al., 2019](#)) to
 478 measure DP length and take nine width measurements at 10% increments of length starting at

479 the base of each DP. We used the built-in ImageJ tool to measure DP area. A “Proportion Larger
480 DPs” was calculated per core as the combined weight of all DPs >40mm in maximum dimension
481 and 5g in weight divided by the weight of all DPs. Higher values show cores with proportionally
482 more large DPs.

483 **2.6 Statistical Analyses**

484 To evaluate the association between psychometric, motor-skill, and training measures and tech-
485 nological outcomes, we adopted an information-theoretic approach ([Burnham & Anderson,](#)
486 [2002](#)). Information-theoretic approaches provide methods for model selection using all possible
487 combinations of variables while avoiding problems associated with significance-threshold step-
488 wise selection. We used the corrected Akaike information criterion (AICc) to rate each possible
489 combination of predictors on the balance between goodness of fit (likelihood of the data given
490 the model) and parsimony (number of parameters). The AICc consists of the log likelihood (i.e.,
491 how well does the model fit the data?) and a penalty term for the number of parameters that
492 must be estimated in the model (i.e., how parsimonious is the model?), with a correction for small
493 sample sizes (AICc converges to the standard AIC at large samples). A lower AICc indicates a
494 more generalizable model and we used it to compare and rank various possible models. Each
495 analysis begins with a full model that includes all predictors of interest. All possible combinations
496 of predictors are then fit, and the resulting models are ranked and weighted based on their AICc.
497 The “best” model is chosen because it has the lowest AICc score.
498 Continuous predictors were centered such that zero represents the sample average, and units are
499 standard deviations. The full model was fitted with the lm function in R 3.2.3, and the glmulti
500 package was used for multi-modal selection and model comparison.

501 **3 Results**

502 Following a recent protocol to enhance the reproducibility and data transparency of archaeo-
503 logical research ([Marwick, 2017](#)), detailed results of all analyses and assessments of the data
504 structure are available in our paper’s supplementary materials and through Github (<https://github.com/Raylc/PaST-pilot>). Here we limit discussion to the major findings regarding
505 flaking performance and individual differences. We were particularly interested in: 1) group

507 level effects of experimental condition, 2) individual differences in aptitude and learning, and 3)
508 potential interactions between learning conditions and individual differences. To address these
509 questions, we employed data reduction (Principal Component Analysis) to derive two summary
510 metrics of flaking performance, compared these factors across the two experimental condi-
511 tions, and built multivariate models examining the relations between our various psychometric
512 measures, subject's motor skill scores, and our two lithic performance factors.

513 **3.1 Principal Component analyses**

514 The following two sections outline factor analyses designed to summarize our main study metrics
515 tracking individual variation in DP sizes and shapes and lithic performance measures.

516 **3.1.1 Detached Piece size and shape**

517 To better understand the relationship between DP shape and training/individual variation, we
518 entered our nine flake linear plan measurements along with maximum flake length and thickness
519 into a principal component analysis (PCA) from which summary coordinates were extracted.
520 Bartlett's Test of Sphericity was significant ($\chi^2 (10) = 4480$, $p < .01$) indicating that the set of
521 variables are adequately related for factor analysis.

522 The analysis yielded three factors explaining a total of 90% of the variance for the entire 11
523 measurement variable set (Table 2). Factor 1 tracks flake size with higher scores indicating larger
524 flakes since all 11 measures load positively on this factor. Factor 2's loadings track the increasing
525 relationship between thickness, length, and flake width. As factor 2 scores increase, flakes
526 get thicker, longer, and narrower, resembling irregular splinters. Factor 3 tracks the relationship
527 between flake proximal and distal width relative to thickness. As factor 3 scores go up, flakes get
528 thinner and narrower at the distal ends and wider at the base. Factor 3 therefore tracks flakes with
529 a typical shape having a thin cross-section, wider base, and narrower tip. We used these three
530 flake shape coordinates to approximate DP size and shape in the project's flake performance
531 factor analysis.

Table 2: Lithic size/shape PCA factor loadings.

Variable	Factor.1	Factor.2	Factor.3
Variance %	78	6.2	5.4
Interpretation	Size	Relative thickness and elongation	Relative thinness and proximal breadth
Flake thickness	0.68	0.58	-0.35
Flake length	0.81	0.31	-0.11
Width-10% (tip)	0.82	-0.31	-0.16
Width-20%	0.91	-0.27	-0.19
Width-30%	0.94	-0.2	-0.16
Width-40%	0.95	-0.07	-0.05
Width-50%	0.96	-0.06	-0.03
Width-60%	0.94	-0.06	0.01
Width-70%	0.95	0.05	0.18
Width-80%	0.92	0.11	0.34
Width-90% (base)	0.8	0.12	0.47

532 3.1.2 Lithic flaking performance measures

533 To better understand the relationship between our various lithic performance measurements and
 534 to reduce data dimensionality, we conducted a second principal component analysis examining
 535 the study's six lithic performance measures (count of large pieces [$>40\text{mm}$ and 5g], mass of large
 536 pieces relative to total detached mass, core delta mass, and the three flake shape factors). All of
 537 these measures were summarized for each core and unique factor scores were calculated from
 538 these core-specific measures. Bartlett's Test of Sphericity was significant ($\chi^2 (6) = 3185$, $p < .01$)
 539 indicating that the set of variables are at least adequately related for factor analysis.

540 The analysis yielded two factors explaining a total of 56% of the variance for the entire set of
 541 variables (Table 3). Factor 1 (hereafter "Quantity") explains 28.7% of the variance and tracks
 542 flaking quantity due to high positive loadings on large DP count and mass ratio and on core
 543 delta mass. Performance factor 2 (hereafter "Quality") covers 27% of the sample variance and
 544 measures flaking quality as reflected in high positive loadings on Shape Factors 1 (size) and 3
 545 (thin, "flake-like" shape) and a negative loading on Shape Factor 2 ("splinter-like" thickness and
 546 elongation). High scores on Quality thus reflect production of larger, relatively thinner, and more
 547 typically flake-shaped vs. splinter-shaped DPs.

Table 3: Overview of principal components results on the six flake performance measures. DP = detached piece.

Variable	Factor.1	Factor.2
Variance %	28.7	27
Interpretation	Flaking quantity (higher values = greater quantity detached)	Flaking quality (higher values = larger, more "flake-shaped" pieces)
Mass of DPs > 40mm and 5g : total mass of all DPs	0.61	-0.48
Number of DPs > 40mm and 5g	0.84	0.4
Core delta mass	0.78	-0.23
Shape PC factor 1 (size)	0.11	0.63
Shape PC factor 2 (thickness and elongation)	0.01	-0.71
Shape PC factor 3 (thinness and base breadth)	0.13	0.55

548 These two factors address flaking performance at the level of individual cores, however we were
 549 also interested in the overall productivity/rate of work of each participant over the entire two
 550 hours. For example, looking at a knappers average Quality and Quantity factor scores would not
 551 differentiate between a participant who spent the entire time exhaustively reducing one core
 552 vs. another participant who did the same to all nine of their allotted cores in the same time. To
 553 capture this aspect of variation we calculated a simple Total Productivity metric as the sum of all
 554 mass a participant removed from cores during the experiment.

555 3.2 Relationships between Performance Measures

556 This approach also allowed us to compare the relationship between Total Productivity, Quantity,
 557 and Quality across our two experimental groups (**Figure 5** and **Figure 6**). As might be expected,
 558 we found that per-core Quantity and Total Productivity are positively correlated in both groups
 559 (**Figure 5a**), although this relationship is twice as strong in the trained ($F[1, 9] = 33$, $p < 0.01$,
 560 Adj. $R^2 = 0.8$) compared to untrained ($F[1, 8] = 8$, $p = 0.02$, Adj. $R^2 = 0.4$) group. Interestingly,
 561 we also found evidence of a negative correlation between Total Productivity and Quality in the
 562 untrained group ($F[1, 8] = 28$, $p = <0.01$, Adj. $R^2 = 0.7$), but no relation in the Trained group (**Figure**
 563 **5b**). A qualitatively similar trend with respect to Quality vs. Quantity (**Figure 5c**) did not achieve
 564 significance ($F[1, 17] = 0.6$, $p = 0.2$).

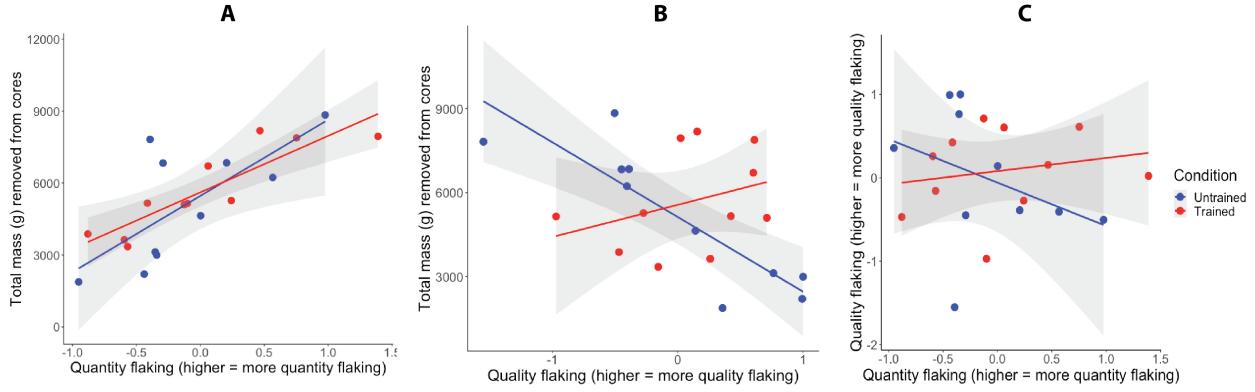


Figure 5: Relationships between flaking quantity, quality, productivity, and the two training conditions. Each dot represents a participant, colors represent training conditions.

565 Thus, it appears that Trained participants achieved higher Total Productivity by increasing average
 566 flaking Quantity across cores and without sacrificing Quality, whereas Untrained participants
 567 found other ways to vary Total Productivity (e.g., number of cores knapped rather than Quantity
 568 per core, see variance Table and Figure) and generally increased productivity at the expense of
 569 Quality. Experimental artifacts illustrating these trade-offs are presented in Figure???.

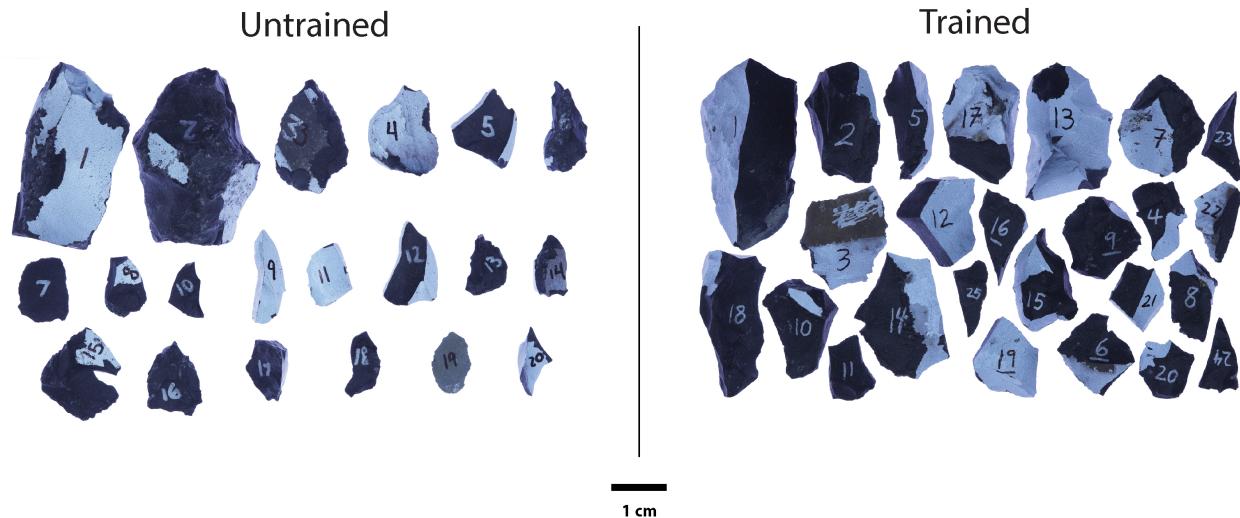


Figure 6: Comparison of untrained and trained detached pieces illustrating the ability of Trained participants to maximize Quality and Productivity at the same time.

570 3.3 Do trained, untrained, and expert knappers perform differently?

571 Here we compare our flaking outcomes (DP size/shape and flaking performance factors) be-
 572 tween the trained and untrained groups. Our expert demonstrator/instructor is included as a
 573 performance benchmark.

574 **Table 4** summarizes the results of ANOVA tests group level difference in central tendency on
575 various performance measures. We found no significant differences between the trained and
576 untrained groups on our flaking Quantity and Quality factors. In contrast, three-way flake size
577 and shape comparisons between our expert knapper and the two novice groups showed that the
578 expert knapper made significantly more large flakes (effect size = 0.14), had a significantly higher
579 core delta mass signal than either of the novice groups (effect size = 0.26), and left significantly
580 smaller finished cores (effect size = 0.27) (**Figure 7**). All three of these results show either medium
581 or large effect sizes. In all three comparisons, the trained group's data distributions tended
582 towards the expert sample although they were not significantly different from the untrained
583 group (**Figure 8**). We also observed a significant difference in shape factor 2 (splinters) driven by
584 the expert's lower values, but with a very low (<0.01) effect size. These results show that mean core
585 reduction intensity and large flake production rates distinguish expert and novice performance
586 whereas novices in experimental groups produced pieces of similar mean size and shape as those
587 of the expert trainer.

Table 4: Summary group-wise comparison statistics contrasting flake performance metrics by experiment condition and with the expert instructor. All comparisons are three-way except for the two skill PC factors and the total mass removed, which are two-way comparisons between the subject population as these factors were not applied to the expert data sample. Eta2 = ANOVA effect size (>0.1 = medium effect size, >0.2 = large effect size). * = non-parametric permuted p-values to account for unequal sample variances. SS = sum of squared differences, df = degrees of freedom, MS = mean squared difference.

Variable	Parameter	SS	df	MS	F	p	Eta2
Skill PC factor 1 (quantity flaking)*	Group	0.3	1	na	0.8	0.3	na
	Residuals	9.4	20	na			
Skill PC factor 2 (quality flaking)*	Group	<0.1	1	na	<0.1	0.9	na
	Residuals	9	20	na			
Flake factor 1 (size)	Group	22.3	2	11.1	1.3	0.2	na
	Residuals	36741.35	4328	8.5			
Flake factor 2 (relative thickness)	Group	11	2	5.46	8.3	<0.01	<0.01
	Residuals	2842.2	4328	0.66			
Flake factor 3 (basal relative to tip width)*	Group	0.04	2	na	0.9	0.9	na
	Residuals	2404	4328	na			
Flakes > 40mm and 5g*	Group	4967	2	na	23.5	<0.01	0.14
	Residuals	19496	185	na			
Mass of flakes : flaked mass	Group	0.06	2	0.03	2.3	0.1	na
	Residuals	2.4	183	0.01			
Core delta mass*	Group	0.6	2	na	13.7	<0.01	0.26
	Residuals	4.4	184	na			
Total cores used	Group	8.5	1	8.5	2.2	0.1	na
	Residuals	79.4	21	3.7			
Total flaked mass	Group	22623898.1	1	2262389	0.5	0.4	na
	Residuals	39348298797	21	4451571			

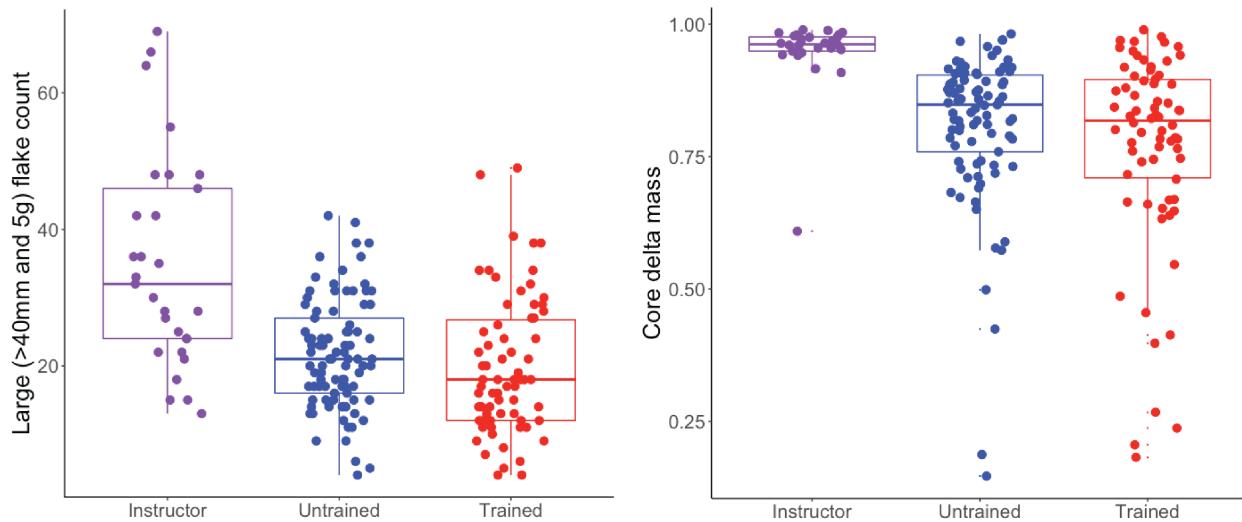


Figure 7: Results showing significant differences between the instructor (expert) and novice flaking performance.

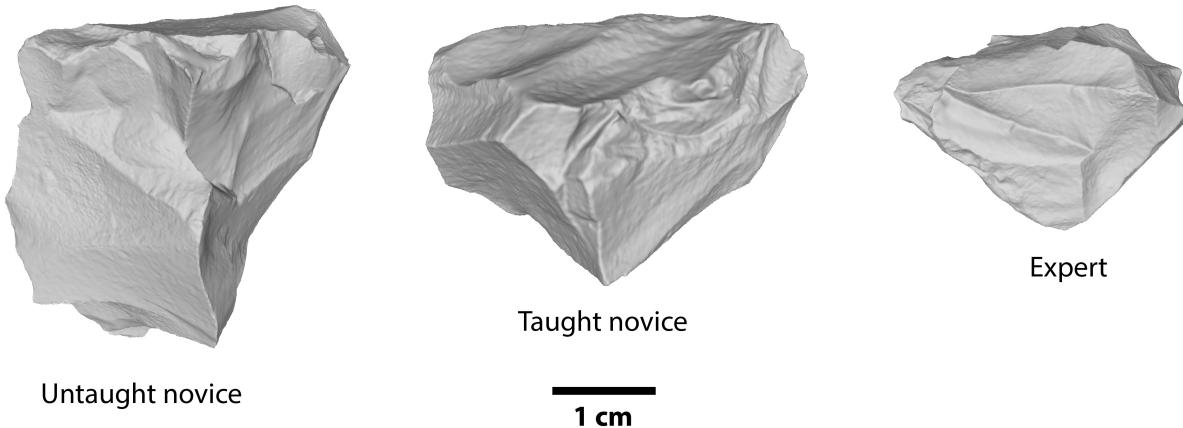


Figure 8: Comparison of untaught, taught, and expert cores.

588 While we did not find significant differences in central tendency between our two experimental
 589 groups, results (**Figure 7**) did indicate lower variance in the trained group. To test whether training
 590 reduced variability in performance outcomes between subjects, we compared variance metrics
 591 between the trained and untrained individuals using the F-test on either core-averaged or flake
 592 specific variances. **Table 5** and **Figure 9** present the results from these comparisons showing
 593 significant variance differences predominantly in flaking Quality, number of large DPs, core delta
 594 mass, and total amount of flaked mass). In most instances, variance in the untrained group
 595 exceeds that of trained individuals by 1.5 to 4.7 times. The most salient effect of instruction was
 596 thus not to shift mean performance but to reduce variability by eliminating the skew (generally
 597 toward poorer outcomes) seen in the untrained group (**Figure 9**), rather than to shift the mean.

Table 5: F-test results comparing variances between trained and untrained subjects across our various flaking performance metrics.

Variable	df	F	95% CI	p	Var.ratio
Skill PC factor 1 (quantity flaking)	10	1.1	0.3 - 4.2	0.8	1.1
Skill PC factor 2 (quality flaking)	10	2.1	1.3 - 3.3	<0.01	2.1
Flake factor 1 (size)	1379	1.1	0.9 - 1.2	0.05	1.1
Flake factor 2 (relative thickness)	1379	0.9	0.8 - 1.0	0.6	0.9
Flake factor 3 (basal relative to tip width)	1379	1.0	0.9 - 1.2	0.06	1.0
Detached pieces > 40mm and 5g	69	1.5	0.9 - 2.4	0.05	1.5
Mass of detached pieces : flaked mass	69	1.3	0.8 - 2.1	0.1	1.3
Core delta mass	69	1.7	1.1 - 2.7	0.01	1.7

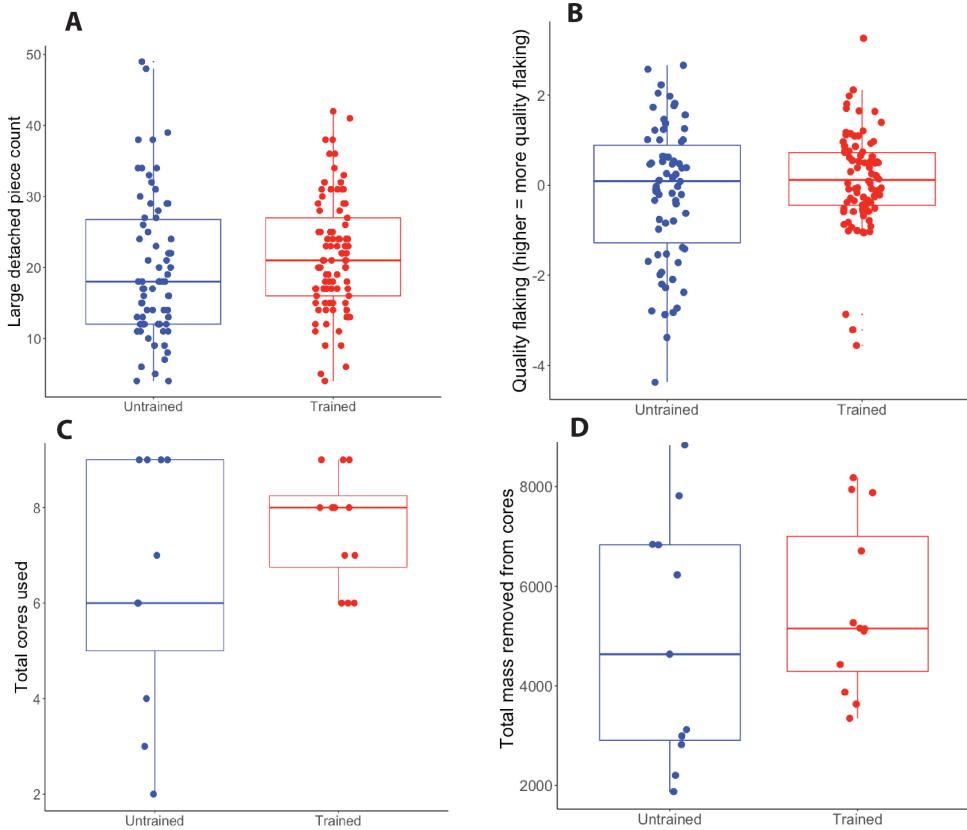


Figure 9: Variance comparisons between trained and untrained individuals across various flaking performance metrics. Note the lower variance of the trained group in all the plots.

598 3.4 Does performance change over time?

599 In addition to comparing overall performance during the two hour experiment, we also wanted to
 600 determine if groups or individuals differed in learning (i.e., performance change) over the period.
 601 For these analyses, we calculated the learning stage as the ordinal number of each core out of the
 602 total number knapped by each subject (i.e., core 2 of 4 or 4 of 8 both equal 50% complete). These
 603 relative core use-order percentages were then binned into 20 percent brackets for core-order
 604 and group-level comparisons. Flaking outcomes were tracked using the two performance factors
 605 (Quality and Quantity). We added the nodule starting mass to track whether training/practice
 606 times impacted raw material selection.

607 **Table 6** shows no significant training effects across the two performance measures either as
 608 grouped data or between individuals (**Figure 10** and **Figure 11**). This result demonstrated that
 609 flaking outcomes did not change dramatically across the study interval. This lack of significant

learning effects is confirmed by an inspection of individual learning curves (Figure). The one significant main training effect related to core starting mass (with a strong effect size = 0.25). On average, core starting masses start low and increase, showing that participants selected smaller nodules first. As the smaller nodules in their allotment were depleted, participants were left to knap larger, less preferred nodules. This preference for smaller cores is somewhat less pronounced in the untrained group, as indicated by a small main effect of learning condition and generally higher starting nodule masses for the untrained group (Figure??).

Table 6: Summary group-wise comparison statistics contrasting flake performance metrics by experiment condition and relative core order. All comparisons are three-way. Eta2 = ANOVA effect size (>0.1 = medium effect size, >0.2 = large effect size). SS = sum of squared differences, df = degrees of freedom, MS = mean squared difference.

Variable	Parameter	SS	df	MS	F	p	Eta2
Skill factor 1 (quantity flaking)	Condition	1.80	1	1.80	1.00	0.3	na
	Relative core order	4.90	4	1.20	0.70	0.6	na
	Condition*Relative core order	1.00	4	0.20	0.10	1	na
	Residuals	263.00	148	1.80	NA		
Skill factor 2 (quality flaking)	Condition	3.40	1	3.40	2.20	0.1	na
	Relative core order	9.00	4	2.20	1.40	0.2	na
	Condition*Relative core order	14.80	4	3.70	2.40	0.1	na
	Residuals	231.70	148	1.60	NA		
Core starting mass	Condition	0.30	1	0.30	4.72	0.03	0.03
	Relative core order	3.27	4	0.82	12.70	<0.01	0.25
	Condition*Relative core order	0.30	4	0.07	1.17	0.32	0.03
	Residuals	9.58	149	0.06	NA		

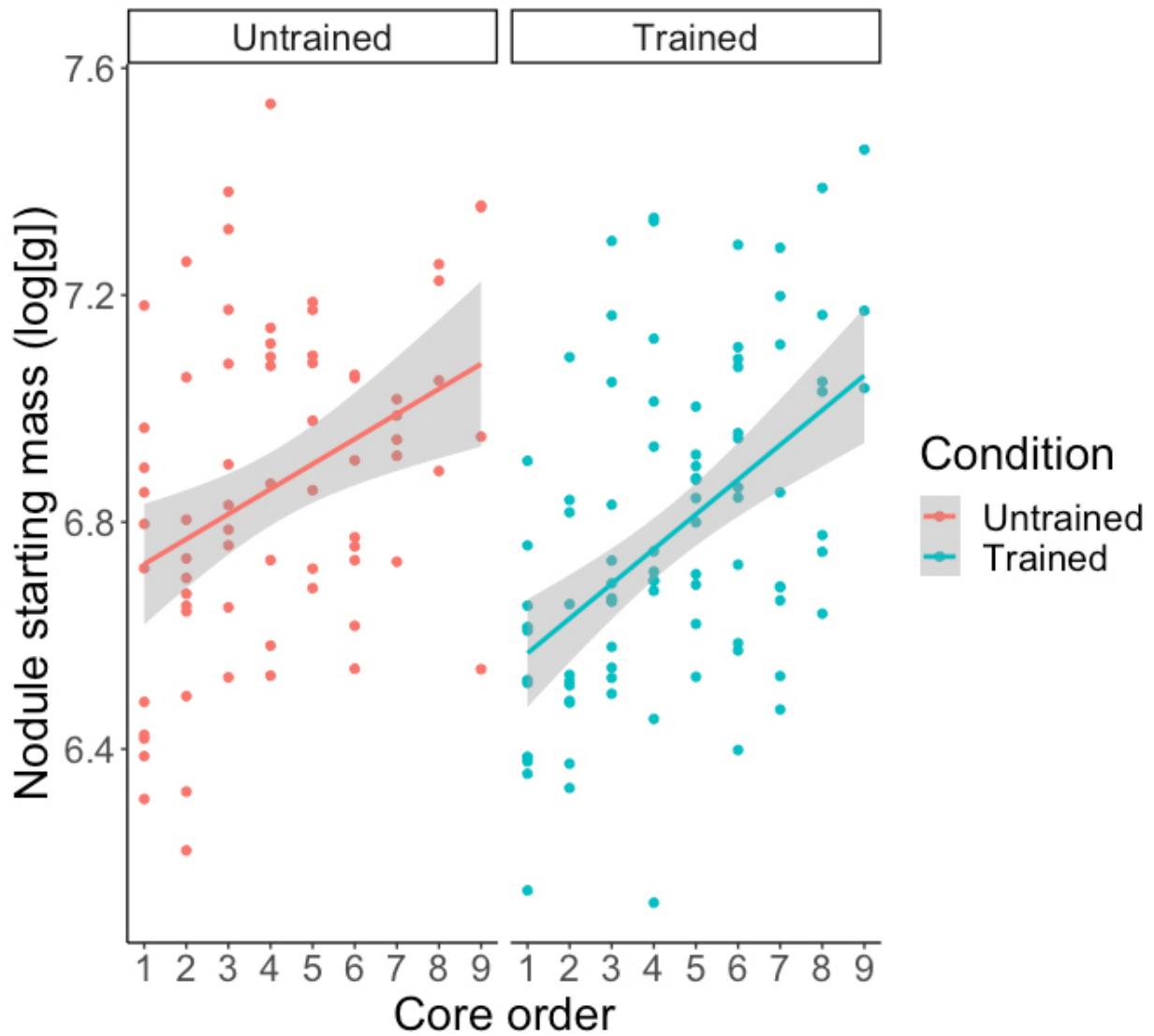


Figure 10: Comparisons of nodule starting mass across the study period by training condition. Results show a significant relationship between nodule starting mass and length of time in the study.

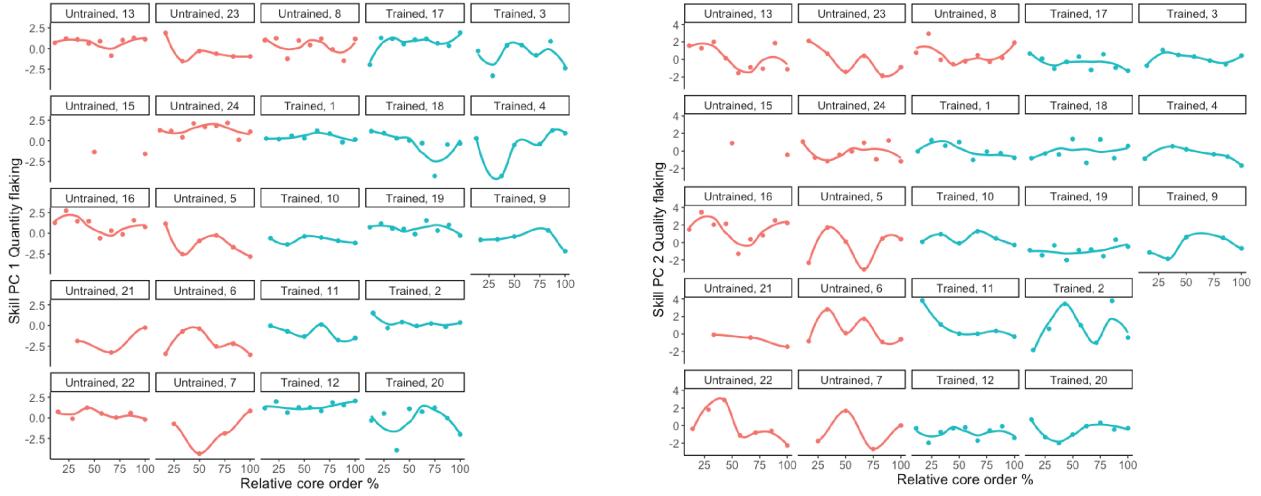


Figure 11: Individual learning curves derived from each subject's relative core use order and the two flake performance factors.

617 3.5 Do individual differences in motor skill and psychometric measures predict 618 flaking performance?

619 One of the experiment's primary goals was to test if measures of individual perceptual-motor
620 and cognitive variation predict success in stone flaking across different training conditions. To
621 address this goal, we built three multivariate models examining the relations between training
622 conditions, individual difference measures, and our three lithic performance measures (overall
623 productivity and average per-core Quantity and Quality). These models enabled us to determine
624 which of the psychometric and motor skill factors are better predictors of a participant's flaking
625 performance in the study.

626 We considered all possible interactions between five individual difference measures, core size,
627 training condition, and the three performance measures, with each subject providing one data
628 point. Each model's continuous predictors (highest n-back level, Raven's Progressive Matrix
629 score, BEAST score, starting nodule mass, Fitts score, and grip strength) were centered such that
630 zero represents the sample average, and units are standard deviations. Our two motor skill and
631 strength measures (grip strength and Fitt's performance scores) are also strongly correlated (F
632 $[1,19] = 15, p < 0.01, R^2 = 0.41$). However, these two measures track complementary components
633 of athleticism (strength vs. speed/accuracy tradeoffs) and so we decided to include both in the
634 model selection process.

635 The full models were fitted with the lm function in R 3.2.3, and we used the Glmulti package's

636 automated model selection algorithm to select the best performing model (lowest AICc score)
637 (see methods for further details on the multimodal selection process). All three models follow the
638 same complete model statement as follows:

639 *Flaking performance variable* ~ Training condition + Highest n-back level + Raven's Progressive
640 Matrix score + BEAST score + Fitt's score + Grip strength

641 For our two per-core performance factors (Quantity and Quality) it is also relevant to consider how
642 individual core features may have affected performance. We found no evidence of individual or
643 group level practice effects over the two hours, so we did not include core order in the models. We
644 did, however, find that subjects selected progressively larger nodules throughout the experiment.
645 It is thus important to understand whether nodule variability had any impact on our flaking
646 results. Because starting nodule size (mass) and shape were strongly correlated ($F [1,157] = 186$, p
647 < 0.01 , $R^2 = 0.54$) we included nodule mass as a covariate to control for any variance in flaking
648 performance that may be driven by nodule differences.

649 **3.5.1 Model 1: Individual differences and overall productivity**

650 Our first model examined variance in overall flaking productivity measured by each subject's
651 combined flaked mass (nodule starting mass - core final mass). This provides a basic measure of
652 variation in individuals' success detaching pieces and reducing cores from a standardized (see
653 Methods) raw material supply. From the same candidate pool size of 55893 possible multivariate
654 models, the best performing model returned an AICc value of -18 (Average AIC = -13). This model
655 comprised the following statement with two main and three interaction effects:

656 *Total flaked mass* ~ Training condition + Grip strength + RPM × Highest n-back level + Fitts
657 score × BEAST score + Grip strength × RPM

658 This model explains a statistically significant and substantial proportion of variance in flaking
659 productivity ($R^2 = 0.84$, $F (6, 14) = 12.7$, $p < 0.01$, adj. $R^2 = 0.77$). A model residuals normality
660 test shows no significant differences with the normal distribution ($p = 0.72$) indicating that this
661 relationship is linear. A Breusch-Pagan test showed no evidence for heteroskedasticity (BP = 2, df
662 = 6, $p = 0.8$).

Table 7: Summary and estimates for the overall flaking productivity model.

Covariate	Estimate	95% CI	t(12)	p
(Intercept)	-0.7	[-1.06, -0.33]	-4.11	0
Condition[Trained]*BEAST	-1.3	[-1.86, -0.73]	-4.91	<0.01
Condition[Trained]*Highest nBack	-1.3	[-1.82, -0.75]	-5.15	<0.01
Grip strength	0.9	[0.59, 1.21]	6.24	<0.01
Training condition	0.7	[0.21, 1.22]	3.02	0.01
Visuospatial nBack	0.7	[0.29, 1.17]	3.57	<0.01
BEAST	0.6	[0.17, 0.96]	3.08	0.01

663 **Table 7** presents this model's coefficients and summary outputs, wherein baseline refers to the
 664 untrained condition with all continuous predictors at the sample average. The parameter esti-
 665 mates for the continuous predictors reflect the expected change in utility for 1 standard deviation
 666 change in the predictor variable. We found significant ($p < 0.05$) and substantial (Standardized
 667 Estimate ≥ 0.50 , i.e. a 50% change in variable) main effects of Grip Strength, Visuospatial nBack,
 668 and BEAST. The main effect of Grip Strength (**Figure 12**), irrespective of learning condition, in-
 669 dicates the basic importance of strength in generating higher production rates among naive
 670 knappers at least when efficiency and quality are not considered.

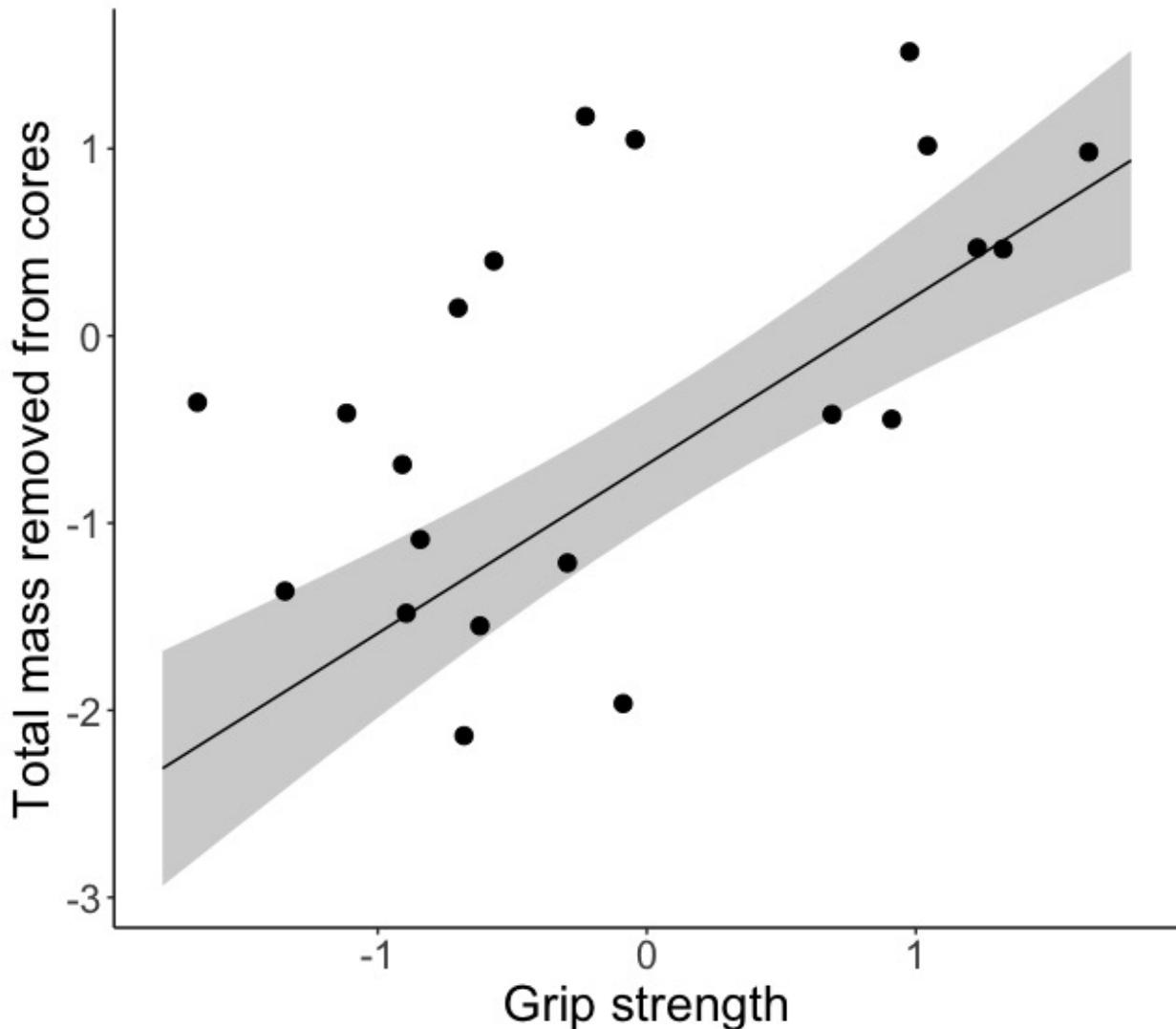


Figure 12: Significant main effect of grip strength on overall flaking productivity.

671 Effects of visuospatial working memory capacity and social information use are more complicated,
 672 as indicated by strong interactions with learning condition ([Figure 13](#)). In each case, higher scores
 673 were associated with better performance in the uninstructed group but worse performance in
 674 the instructed group. Positive effects in the uninstructed group were as expected, given the
 675 hypothesized importance of spatial cognition (Coolidge and Wynn 2005) and social learning
 676 (Morgan et al. 2015) in the acquisition of knapping skills. Negative effects in the trained group are
 677 unexpected but presumably reflect differences in learning strategies adopted under instruction.

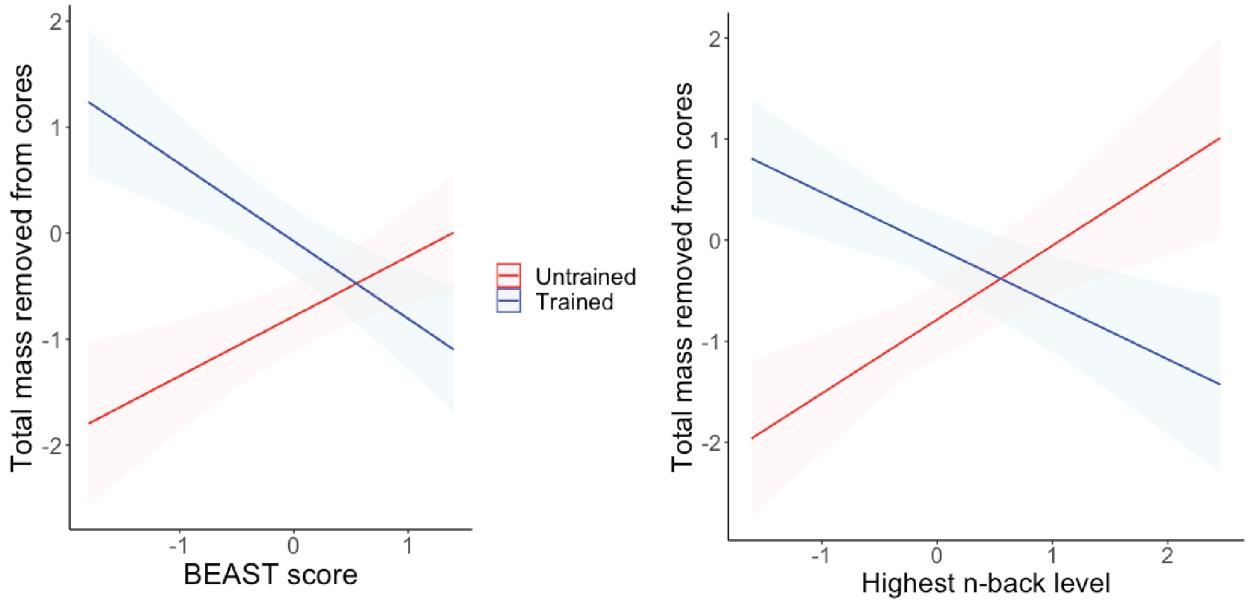


Figure 13: Two significant interaction effects of training condition, social information, and visuo-spatial working memory on overall flaking productivity.

678 3.5.2 Model 2: Individual differences and quality flaking

679 The second full model examined the variance in average flaking Quantity per core. It thus
 680 complements our first model assessing overall productivity by testing for differences in reduction
 681 intensity at the level of individual cores. From a candidate pool of 55893 possible multivariate
 682 models, the best performing model returned an AICc value of 32 (Average AIC = 44). This model
 683 comprised the following statement with three main and four interaction effects:

684 *Quantity ~ Highest n-back level + BEAST score + Fitt's score + Grip strength + Training*
 685 *condition × Highest n-back level + Training condition × BEAST score + Training condition × Grip*
 686 *strength + Nodule mass (as control)*

687 This model explains a statistically significant and substantial proportion of variance in quantity
 688 flaking ($R^2 = 0.7$, $F(8, 12) = 3.6$, $p = 0.02$, adj. $R^2 = 0.5$). A model residuals normality test shows no
 689 significant differences with the normal distribution ($p = 0.38$) indicating that this relationship
 690 (as required) is linear. A Breusch-Pagan test showed no evidence for heteroskedasticity (whether
 691 variance for all observations in our data set are the same) ($BP = 4.4$, $df = 8$, $p = 0.8$).

692 **Table 8** presents this model's coefficients and summary outputs, following the same format as
 693 Table.

Table 8: Summary and estimates for Quantity flaking model.

Covariate	Estimate	95% CI	t(12)	p
(Intercept)	0.0	[-0.20, 0.24]	0.2	0.85
Training Condition[Trained]*BEAST score	-0.9	[-1.37, -0.42]	-4.1	<0.01
Training Condition[Trained]*Highest n-back level	-0.7	[-1.23, -0.26]	-3.3	<0.01
Training Condition[Trained]*Grip strength	0.7	[0.08, 1.27]	2.5	0.03
BEAST score	0.3	[-0.02, 0.66]	2.1	0.06
Fitts score	-0.3	[-0.59, 0.02]	-2.0	0.06
Highest n-back level	0.2	[-0.19, 0.58]	1.1	0.29
Grip Strength	-0.1	[-0.48, 0.29]	-0.6	0.6
Nodule mass	-0.1	[-0.35, 0.08]	-1.3	0.21

694 The Quantity model roughly paralleled results for Total Production, yielding substantial and
 695 significant interactions between training condition, n-back level, BEAST scores, and grip strength.
 696 As with Total Production, higher visuospatial n-back levels and BEAST scores were associated with
 697 lower Quantity scores in the trained group but higher or unchanged Quantity in the untrained
 698 group (**Figure 14**).

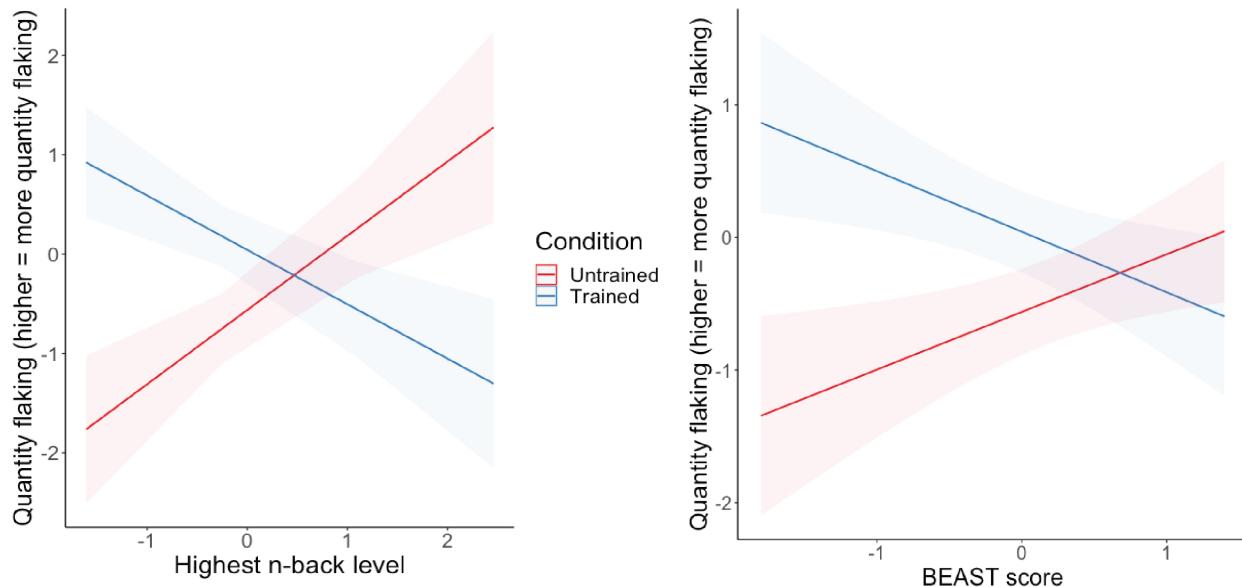


Figure 14: Two significant interaction effects of training condition, visuo-spatial working memory, and social information, on overall flaking quality.

699 Unlike Total Productivity, the effect of Grip Strength on per-core Quantity was mediated by an
 700 interaction with learning condition (**Figure 15**). Thus, high Grip Strength enabled individuals
 701 in both groups to produce more total debitage, but only Instructed individuals translated Grip

⁷⁰² Strength into more intense reduction of individual cores, including not only delta mass, but also
⁷⁰³ number and proportion of larger pieces.

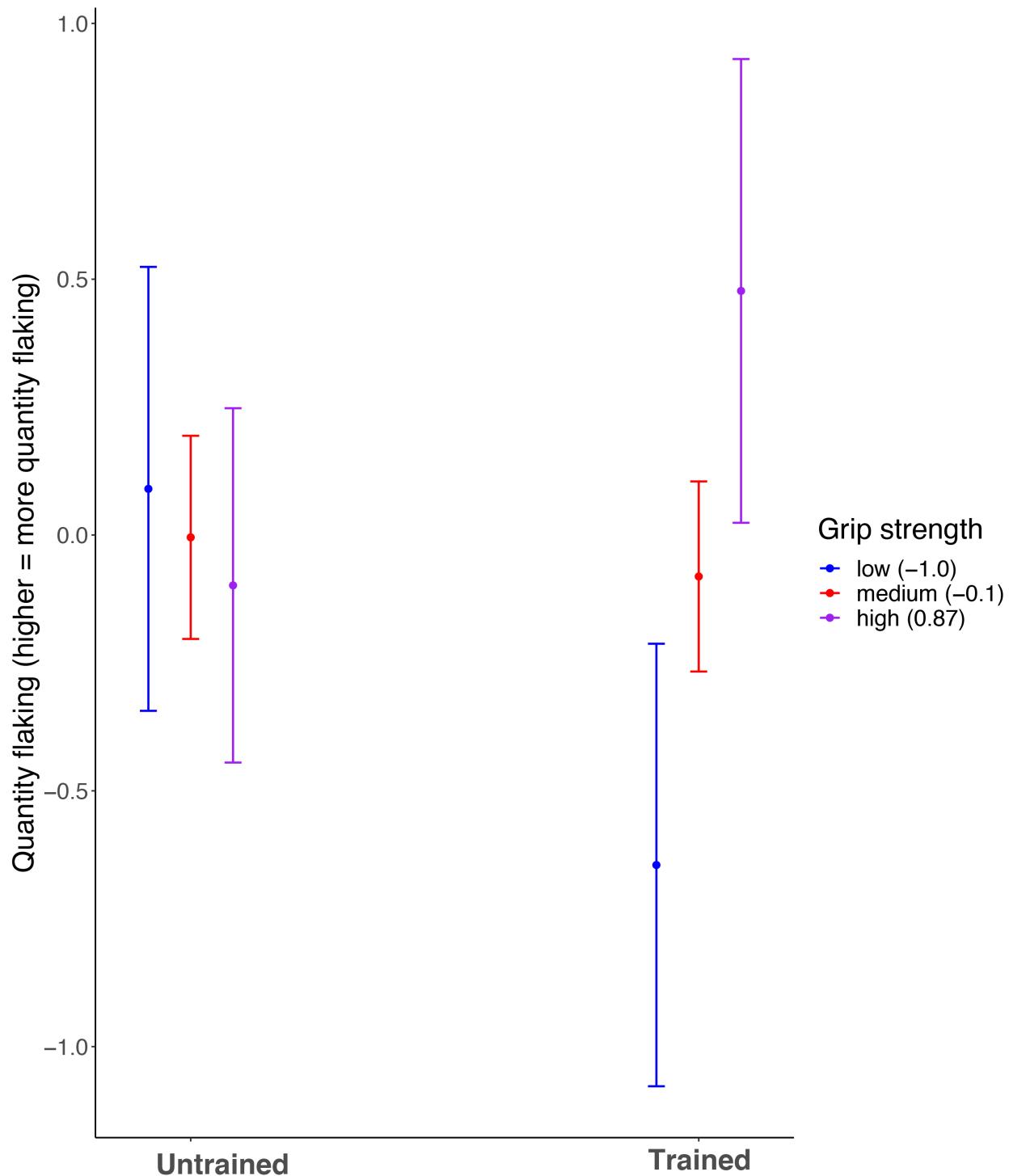


Figure 15: Three significant interaction effects of training condition, visuo-spatial working memory and social information, and training and grip strength use on quantity flaking.

704 **3.5.3 Model 3: Individual differences and quality flaking**

705 Our third model examining variance in Quality follows the same complete model statement we
 706 used for Quantity. From the same candidate pool size of 55893 possible multivariate models, the
 707 best performing model returned an AICc value of 32 (Average AIC = 39). This model comprised
 708 the following statement with three main and four interaction effects:

709 *Quality flaking ~ Highest n-back level + Fitt's score + Grip strength + Fitt's score × BEAST score + Grip*
 710 *strength × BEAST score + Grip strength × Fitt's score + Training condition × Grip strength + Nodule*
 711 *mass (as control)*

712 This model explains a statistically significant and substantial proportion of variance in Quality
 713 ($R^2 = 0.75$, $F(8, 12) = 4.6$, $p < 0.01$, adj. $R^2 = 0.6$) in the absence of any main training effects. A
 714 model residuals normality test shows no significant differences with the normal distribution (p
 715 = 0.41) indicating that this relationship is linear. A Breusch-Pagan test showed no evidence for
 716 heteroskedasticity (BP = 7, df = 8, p = 0.5).

Table 9: Summary and estimates for the Quality flaking model.

Covariate	Estimate	95% CI	t(12)	p
(Intercept)	-0.1	[-0.31, 0.17]	-0.62	0.55
Training condition[Trained]*Fitts score	-0.5	[-1.15, 0.06]	-1.97	0.07
Grip strength	-0.4	[-0.86, 0.07]	-1.84	0.09
BEAST score*RPM	0.3	[0.06, 0.55]	2.70	0.02
Highest n-back level*Fitts score	-0.3	[-0.57, -0.09]	-2.98	0.01
Highest n-back level	-0.3	[-0.54, 0.01]	-2.10	0.06
Fitts score	0.2	[-0.30, 0.75]	0.93	0.37
Nodule starting mass	0.1	[-0.06, 0.33]	1.48	0.17
Grip strength*Training condition[Trained]	0.2	[-0.44, 0.85]	0.68	0.51

717 **Table 9** presents this model's coefficients and summary outputs following the same data format as
 718 **Table 7 & Table 8**. The results show no effects that were both significant and substantial (Estimate
 719 ≥ 0.5).

720 The Quality model did produce two statistically ($p < 0.05$) significant interaction effects (RPM *
 721 BEAST & Fitts * n-back). However, these interactions had relatively small effects on Quality (<0.5)
 722 and we believe that interpreting these results from our small, exploratory study would be inap-
 723 propriate. **Figure 16** shows the uneven distribution of data points for these interactions, which
 724 suggests vulnerability to leveraging effects of a small number of extreme value combinations.

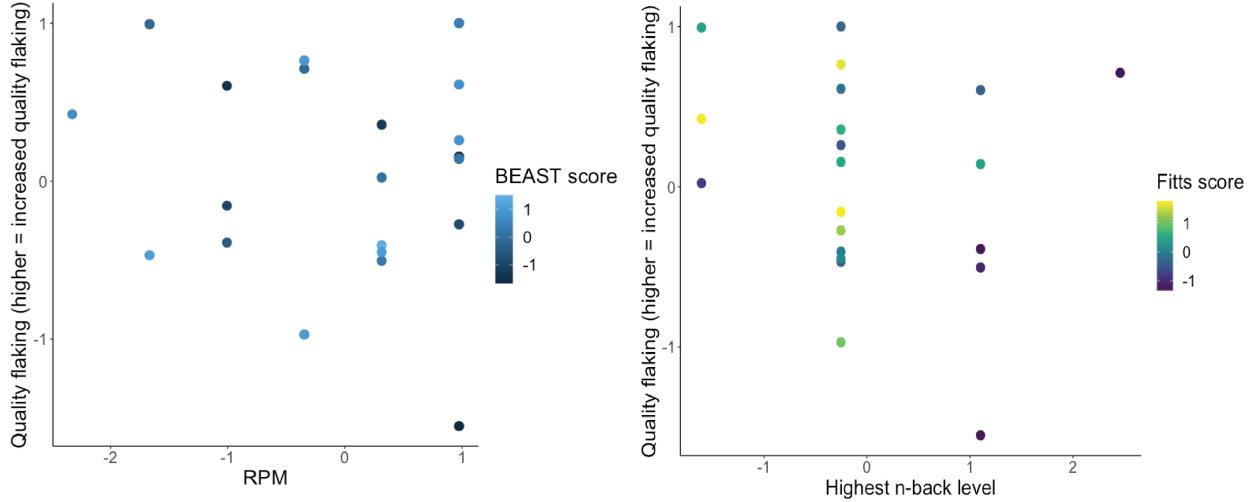


Figure 16: Significant interactions with low estimated effect sizes from the quality flaking model.

725 Without a larger sample, it is not possible to determine if these are anomalous outliers or simply
 726 represent a poorly sampled part of the broader population.

727 3.6 Behavioral observations

728 We designed this exploratory study primarily to trial experimental design elements such as train-
 729 ing time, conditions, and raw materials and to collect preliminary data on the effect of individual
 730 differences and training on knapping outcomes. We thus focused on collecting quantitative
 731 psychometric and lithic data. However, we also considered that quantifying participant knapping
 732 behaviors as well as products could be important for future studies. To support methods develop-
 733 ment in this regard, we made ad hoc notes on observed behaviors during the experiments and
 734 video-recorded all experiments to enable later, more systematic analyses yet to be completed.
 735 However, even casual behavior observation was sufficient to reveal an unexpected effect. Whereas
 736 all trained participants copied the general posture and technique of the expert (free hand knap-
 737 ping seated in a chair) fully half (6) of the uninstructed participants experimented with or even
 738 knapped all of their cores using the floor as a support (**Figure 17**). Three of these participants
 739 were in the same session, which is also the only group composed of just three individuals. In this
 740 group, knapping on the ground appears to have been transmitted from one participant to the
 741 other two based on appearance order and the point of gaze of participants.

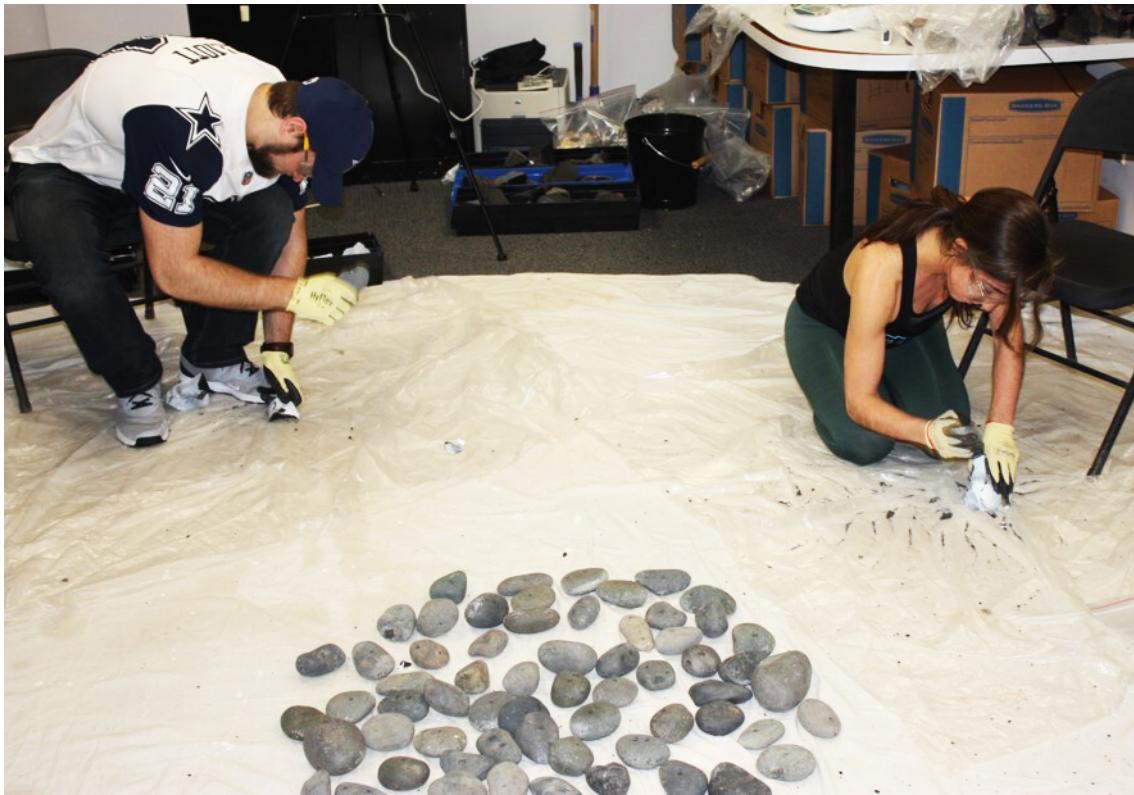


Figure 17: Novices in the untaught condition developing and transmitting a form of bipolar technology involving core reduction on the floor.

742 4 Discussion

743 The most salient finding of our exploratory study is that the presence/absence of teaching clearly
744 impacted knapping performance but did so in nuanced and individually variable ways that have
745 not been explored in previous studies. In fact, we did not observe any significant differences
746 in mean performance between our experimental conditions. Some non-significant tendencies
747 toward enhanced Instructed group performance suggest that a larger participant sample might
748 detect significant effects, but also that the size of any such effects would likely remain small. This
749 could reasonably lead to the conclusion that teaching does not substantially facilitate early stage
750 knapping skill acquisition ([Ohnuma et al., 1997](#); cf. [Putt et al., 2014](#)). Looking closer, however, we
751 found a number of important teaching effects.

752 **4.1 Variance Reduction**

753 In our experiment, the strongest effects of teaching were to reduce variance ([Figure 9](#), [Table 5](#))
754 rather than shift mean values. In particular, teaching acted as a “safety net” that homogenized
755 performance by reducing the frequency of extremely poor outcomes (i.e., learning failures). This
756 finding provides additional support for the hypothesis that teaching would have increased the
757 reliability of Oldowan skill reproduction ([Morgan et al., 2015](#)) while simultaneously corroborating
758 the view that basic flaking competence can be achieved in its absence ([Tennie et al., 2017](#)). Our
759 results thus do not imply that teaching was required or even present during Oldowan times (but
760 see [Gärdenfors & Höglberg \(2017\)](#)), but rather serve to reinforce the plausibility of co-evolutionary
761 scenarios positing the cost/reliability of technological skill acquisition as a selection pressure
762 favoring the evolution of teaching and language ([Morgan et al., 2015](#); [Stout, 2010](#); [Stout & Hecht,
763 2017](#)).

764 **4.2 Knapping Behaviors**

765 The current study complements the transmission chain design of Morgan et al. ([2015](#)) by finding
766 similar effects in more naturalistic learning contexts. Our design further allowed us to examine
767 individual variation to better understand how teaching produces its effects. Whereas transmission
768 chains are optimized to investigate iterative learning effects (but see [Caldwell et al., 2020](#)), they
769 necessarily involve a different instructor/model for each participant. We sacrificed this iterative
770 component in order to consider how the presence/absence of teaching affected the behavior of
771 individuals under otherwise standardized learning conditions.

772 We found that a key impact of teaching was to alter basic flake production strategies, as reflected
773 in the relationship between Total Productivity and detached piece Quality (Fig.Xb). Whereas
774 Untrained participants achieved greater Productivity at the expense of Quality and vice versa,
775 these dimensions were unrelated across Trained participants. Thus, even though Untrained
776 participants achieved the highest values on each metric, only trained individuals managed to
777 maximize both simultaneously. Indeed, a core function of teaching is to reduce the search space
778 that learners must explore and increase the likelihood of discovering globally as opposed to
779 locally optimum solutions (cf. [Hinton & Nowlan, 1996](#); [Stout, 2013](#)). In our study, Untrained
780 individuals explored a greater range of basic behavioral variations not seen in the Trained group,
781 including knapping on the floor, concentrating on working just a few cores (2-4, Figure) over the

782 practice period, and showing less constrained nodule size preferences (Figure). It is notable that
783 this variation occurred even in the presence of an observable expert example, suggesting it may
784 be interesting for future experiments to address the impact of social context, expectations, and
785 relationships on observational learning strategies ([Kendal et al., 2018](#)).

786 We also found a strong positive effect of Grip Strength on Total Productivity independent of
787 learning condition (Figure). While it is tempting to interpret this with respect to the demands for
788 hand strength specifically, it is important to remember that grip strength is strongly correlated
789 with total muscle strength ([Wind et al., 2010](#)) and overall fitness ([Sasaki et al., 2007](#)). Thus, it is
790 best taken to indicate some importance of fitness generally in increasing the rate and intensity
791 of core reduction by naïve knappers, potentially affecting rate of work and the kinetic energy
792 of the swing as well as the handling of core and hammerstone. It thus provides further support
793 for hypotheses positing stone tool making as a selection pressure on the functional anatomy
794 of hand, arm, and shoulder (e.g., [Williams-Hatala et al., 2018](#)), but initially appears orthogonal
795 to variations in learning condition and knapping behaviors in our study. However, we also
796 found that the effect of Grip Strength on per-core knapping Quantity is dependent on teaching
797 (Figure). The absence of this effect in the uninstructed group reflects the weaker association
798 between Total Productivity and per-core Quality across these participants (Fig.Xa) and shows
799 Grip Strength increased uninstructed Productivity specifically by allowing them to knap more
800 cores rather than to reduce individual cores more heavily. In keeping with this, uninstructed
801 Grip Strength is positively correlated with Total Cores knapped ($R^2 = 0.54$, $p = 0.01$). Conversely,
802 strength allowed instructed participants to increase their average Quantity per core without
803 affecting the total number of cores knapped ($R^2 = 0.18$, $p = 0.165$). Thus, strength appears to
804 have achieved its effects on core reduction rate and intensity in different ways, depending on
805 teaching. This difference is likely related to the homogenizing effect of teaching on knapping
806 rate (all instructed participants knapped 6 or more cores) and methods. Subjectively, knapping
807 behaviors of uninstructed participants often appeared more physically demanding (e.g., greater
808 number of non-productive blows, rapid and unregulated battering) which would imply different
809 demands on both strength and aerobic fitness ([Mateos et al., 2019](#); [Williams-Hatala et al., 2021](#)).
810 However, this remains to be systematically investigated.

811 In this respect, it is also important to note that we do not know how well the knapping objectives
812 and strategies communicated by the expert in our experiment correspond to actual Oldowan

813 goals and behaviors. The instructor has successfully replicated assemblage-level patterning at
814 Gona ([Stout et al., 2019](#)) but Oldowan behavior is variable across space and time (e.g., [Braun et al.,](#)
815 [2019](#)) and alternative knapping methods might maximize different values (productivity, quality,
816 effort), especially in novices ([Putt, 2015](#)) ([Putt 2015](#)). Nevertheless, the effect of instruction to
817 constrain behavioral exploration and homogenize outcomes is clear. We expect that this effect
818 would generalize to the teaching of alternative knapping goals and behaviors, although this
819 remains to be tested.

820 **4.3 Learning Strategies**

821 One major goal of this experiment was to test the viability of a moderate, two-hour, learning period
822 for studies of skill acquisition. Unfortunately, we found that this duration was insufficient to
823 capture learning effects for Oldowan-like flake production. The lack of performance change over
824 the period (Figure) cannot be attributed to a ceiling effect (i.e., rapid task mastery at the outset of
825 the practice period) as participants remained well below expert levels and continued to display
826 the high within-individual variability typical of naïve/novice knapping ([Eren et al., 2011](#); [Pargeter](#)
827 [et al., 2019](#)). This negative result was unexpected but is broadly consistent with evidence that Early
828 Stone Age flaking, while conceptually simple, requires substantial practice for perceptual-motor
829 skill development ([Nonaka et al., 2010](#); [Pargeter et al., 2020](#); [Stout & Khriesheh, 2015](#)). Future
830 investigations of learning variation across individuals and/or experimental conditions may thus
831 need to incorporate longer practice periods to capture skill acquisition processes. In theory,
832 much shorter knapping trials might be used to assess the variation in initial performance across
833 individuals and under different conditions that is captured in our study. However, the presence of
834 substantial core-to-core variation within individuals cautions against overly brief experiments
835 that might not provide a representative sample. Greater durations also allow for the expression of
836 different learning strategies over time, even in the absence of directional performance change.

837 At a basic level, learners of any new task must balance investment in task exploration vs. ex-
838 ploitation of knowledge and skills already in hand ([Sutton & Barto, 2018](#)). Premature exploitation
839 risks settling for a sub-optimal local solution whereas continued exploration sacrifices more
840 immediate payoffs. Managing this trade-off is especially challenging for complex, real-world
841 tasks like stone knapping, and is thought to depend on the interplay of uncertainty and reward
842 expectation ([Wilson et al., 2021](#)). Teaching and social learning generally have the potential to

843 provide low-cost information about task structure and payoffs (Kendal et al., 2018; Rendell et al.,
844 2010), which if adopted, would be expected to affect exploration/exploitation decisions. Such
845 adoption is itself known to be influenced by individual cognitive differences, for example if higher
846 fluid intelligence allows observers to better understand observed tasks (Vostroknutov et al., 2018)
847 or if individuals vary in their tendency to use and value social information (Molleman et al., 2019;
848 Toelch et al., 2014).

849 In our study, we did not observe any effect of fluid intelligence (RPM) on knapping outcomes but
850 did find strong interactions of learning condition with participant visuospatial working memory
851 and social information use tendency (Figure). As expected, uninstructed individuals with higher
852 scores on these dimensions displayed higher Total Productivity and average per-core flaking
853 Quantity (although the effect on n-Back on Quantity did not achieve significance). We attribute
854 these effects to increased ability to hold relevant morphological/spatial information in mind and a
855 tendency to benefit from observing successful strategies of others, including the expert model. In
856 contrast, instructed individuals with higher scores tended to have lower Productivity and Quantity.
857 We interpret this unexpected effect to an increased tendency to privilege exploratory learning
858 behavior over exploitation. In particular, we suggest that trained participants might knap more
859 slowly and less productively if higher working memory capacity inclined them to experiment
860 more with morphological/spatial variables highlighted by the instructor or if a predisposition to
861 use social information use caused them to invest greater time and effort attending to and trying
862 out observed actions and/or instructions. These suggestions remain to be tested by further work.
863 Unfortunately, the training period in our current experiment was insufficient to capture learning
864 effects and so we have no evidence of the effects of these individual differences and putative
865 exploration/exploitation tradeoffs to the ultimate achievement of expertise. A similar negative
866 effect of instruction on knapping outcomes during early stage learning was reported by Putt et
867 al. (2014), and has been interpreted to reflect learners experimenting with advanced techniques
868 before they have the perceptual-motor skill to execute them (Stout & Khriesheh, 2015; Whiten,
869 2015). Such effects might be further explored with more detailed behavioral data, as opposed to
870 purely lithic data, and with longer learning periods.

871 **4.4 Limitations and Prospects**

872 Although our exploratory study produced a number of robust results with respect to the effects of
873 instruction and individual differences on lithic products, it is clearly limited by a small sample
874 size, short training duration, and lack of detailed quantification of observed behaviors. These
875 are limitations that can hopefully be addressed in future studies building on the methods and
876 evidence presented here. For example, it is notable that our study failed to document any reliable
877 effects on knapping Quality. Obviously, this might reflect an actual lack of such effects, but it
878 may also indicate a need for more sensitive measures and/or increased sample size and training
879 duration to identify subtle or delayed effects. One aspect of our attempt to balance pragmatic
880 costs and benefits in our study was to test the efficacy of relatively limited lithic analysis. More
881 detailed ongoing analyses of core morphology and debitage features (e.g., typology, cutting
882 edge length, platform dimensions) may yet reveal a more reliable signal of knapping quality.
883 Results of the Quality model in particular also seem to suffer from the uneven distribution and
884 discrete rather than continuous nature of scores on our RPM and n-Back tests. Concerns about
885 the sampling of variation on these dimensions could be addressed with larger samples or by
886 pre-screening participants to ensure more even representation. Alternative psychometric tests
887 (e.g., full rather than short version of the RPM) might also provide more sensitive and continuous
888 measures.

889 Another major limitation that our study shares with all other published experiments on knapping
890 skill acquisition is that we do not address variation in social and cultural context or in teaching
891 style. Currently, we have little basis other than personal experience/tradition (Callahan, 1979;
892 Shea, 2015; Whittaker, 1994) and theoretical speculation (Stout, 2013; Whiten, 2015) from which to
893 assess which pedagogical techniques are most effective even in WEIRD contexts. No study to date
894 has considered how variation in teacher skill (Shea, 2015) or social relationship to participants
895 might impact learning under different conditions. To properly address these questions would
896 require a major research program, including both cross-cultural comparative studies (Barrett,
897 2020) and more naturalistic study designs. While costly, such research would produce results
898 of broad relevance to anthropologists, biologists, psychologists, and sociologists interested in
899 teaching and learning.

900 **5 Conclusions**

901 **6 Acknowledgments**

902 **References**

- 903 Allen, K. R., Smith, K. A., & Tenenbaum, J. B. (2020). Rapid trial-and-error learning with simulation
904 supports flexible tool use and physical reasoning. *Proceedings of the National Academy of
905 Sciences*, 117(47), 29302–29310. <https://doi.org/10.1073/pnas.1912341117>
- 906 Barrett, H. C. (2020). Towards a Cognitive Science of the Human: Cross-Cultural Approaches and
907 Their Urgency. *Trends in Cognitive Sciences*, 24(8), 620–638. [https://doi.org/10.1016/j.tics.202 0.05.007](https://doi.org/10.1016/j.tics.202
908 0.05.007)
- 909 Bilker, W. B., Hansen, J. A., Brensinger, C. M., Richard, J., Gur, R. E., & Gur, R. C. (2012). Develop-
910 opment of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test.
911 *Assessment*, 19(3), 354–369. <https://doi.org/10.1177/1073191112446655>
- 912 Boogert, N. J., Madden, J. R., Morand-Ferron, J., & Thornton, A. (2018). Measuring and under-
913 standing individual differences in cognition. *Philosophical Transactions of the Royal Society B:
914 Biological Sciences*, 373(1756), 20170280. <https://doi.org/10.1098/rstb.2017.0280>
- 915 Boyette, A. H., & Hewlett, B. S. (2017). Autonomy, Equality, and Teaching among Aka Foragers and
916 Ngandu Farmers of the Congo Basin. *Human Nature*, 28(3), 289–322. [https://doi.org/10.1007/s12110-017-9294-y](https://doi.org/10.1007/
917 s12110-017-9294-y)
- 918 Braun, D. R., Aldeias, V., Archer, W., Arrowsmith, J. R., Baraki, N., Campisano, C. J., Deino, A.
919 L., DiMaggio, E. N., Dupont-Nivet, G., Engda, B., Feary, D. A., Garello, D. I., Kerfelew, Z.,
920 McPherron, S. P., Patterson, D. B., Reeves, J. S., Thompson, J. C., & Reed, K. E. (2019). Earliest
921 known Oldowan artifacts at >2.58 Ma from Ledi-Geraru, Ethiopia, highlight early technological
922 diversity. *Proceedings of the National Academy of Sciences*, 116(24), 11712–11717. <https://doi.org/10.1073/pnas.1820177116>
- 924 Braun, D. R., Plummer, T., Ferraro, J. V., Ditchfield, P., & Bishop, L. C. (2009). Raw material quality
925 and Oldowan hominin toolstone preferences: Evidence from Kanjera South, Kenya. *Journal of
926 Archaeological Science*, 36(7), 1605–1614. <https://doi.org/10.1016/j.jas.2009.03.025>

- 927 Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical
928 Information-Theoretic Approach* (2nd ed.). Springer-Verlag. <https://doi.org/10.1007/b97636>
- 929 Caldwell, C. A., Atkinson, M., Blakey, K. H., Dunstone, J., Kean, D., Mackintosh, G., Renner, E., &
930 Wilks, C. E. H. (2020). Experimental assessment of capacities for cumulative culture: Review
931 and evaluation of methods. *WIREs Cognitive Science*, 11(1), e1516. [https://doi.org/10.1002/
932 wcs.1516](https://doi.org/10.1002/wcs.1516)
- 933 Callahan, E. (1979). The basics of biface knapping in the eastern fluted point tradition: A manual
934 for flintknappers and lithic analysts. *Archaeology of Eastern North America*, 7(1), 1–180.
935 <https://www.jstor.org/stable/40914177>
- 936 Cataldo, D. M., Migliano, A. B., & Vinicius, L. (2018). Speech, stone tool-making and the evolution
937 of language. *PLOS ONE*, 13(1), e0191071. <https://doi.org/10.1371/journal.pone.0191071>
- 938 Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of
939 Educational Psychology*, 54(1), 1–22. <https://doi.org/10.1037/h0046743>
- 940 Coolidge, F. L., & Wynn, T. (2005). Working Memory, its Executive Functions, and the Emergence
941 of Modern Thinking. *Cambridge Archaeological Journal*, 15(1), 5–26. [https://doi.org/10.1017/S0959774305000016](https://doi.org/10.1017/
942 S0959774305000016)
- 943 Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or, The Preservation of
944 Favoured Races in the Struggle for Life* (1st ed.). John Murray.
- 945 Darwin, C. (1871). *The descent of man, and selection in relation to sex* (1st ed.). John Murray.
- 946 Duke, H., & Pargeter, J. (2015). Weaving simple solutions to complex problems: An experimental
947 study of skill in bipolar cobble-splitting. *Lithic Technology*, 40(4), 349–365. [https://doi.org/10.1179/2051618515Y.0000000016](https://doi.org/10.
948 .1179/2051618515Y.0000000016)
- 949 Engle, R. W. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives on Psycho-
950 logical Science*, 13(2), 190–193. <https://doi.org/10.1177/1745691617720478>
- 951 Engles, F. (2003). The part played by labour in the transition from ape to man. In R. C. Scharff &
952 V. Dusek (Eds.), *Philosophy of Technology – The Technological Condition: An Anthology* (pp.
953 71–77). Blackwell.
- 954 Eren, M. I., Bradley, B. A., & Sampson, C. G. (2011). Middle Paleolithic Skill Level and the Individual

- 955 Knapper: An Experiment. *American Antiquity*, 76(2), 229–251. <https://doi.org/10.7183/0002-7316.76.2.229>
- 956
- 957 Eren, M. I., Lycett, S. J., Patten, R. J., Buchanan, B., Pargeter, J., & O'Brien, M. J. (2016). Test, model,
958 and method validation: The role of experimental stone artifact replication in hypothesis-
959 driven archaeology. *Ethnoarchaeology: Journal of Archaeological, Ethnographic and Experi-
960 mental Studies*, 8(2), 103–136. <https://doi.org/10.1080/19442890.2016.1213972>
- 961 Eren, M. I., Roos, C. I., Story, B. A., von Cramon-Taubadel, N., & Lycett, S. J. (2014). The role of raw
962 material differences in stone tool shape variation: an experimental assessment. *Journal of
963 Archaeological Science*, 49, 472–487. <https://doi.org/10.1016/j.jas.2014.05.034>
- 964 Faisal, A., Stout, D., Apel, J., & Bradley, B. (2010). The Manipulative Complexity of Lower Paleolithic
965 Stone Toolmaking. *PLOS ONE*, 5(11), e13718. <https://doi.org/10.1371/journal.pone.0013718>
- 966 Fitts, P. M. (1954). The information capacity of the human motor system in controlling the
967 amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391. <https://doi.org/10.1037/h0055392>
- 968
- 969 García-Medrano, P., Ollé, A., Ashton, N., & Roberts, M. B. (2019). The Mental Template in Handaxe
970 Manufacture: New Insights into Acheulean Lithic Technological Behavior at Boxgrove, Sussex,
971 UK. *Journal of Archaeological Method and Theory*, 26(1), 396–422. <https://doi.org/10.1007/s10816-018-9376-0>
- 972
- 973 Gärdenfors, P., & Högberg, A. (2017). The archaeology of teaching and the evolution of homo
974 docens. *Current Anthropology*, 58(2), 188–208. <https://doi.org/10.1086/691178>
- 975 Geribàs, N., Mosquera, M., & Vergès, J. M. (2010). What novice knappers have to learn to become
976 expert stone toolmakers. *Journal of Archaeological Science*, 37(11), 2857–2870. <https://doi.org/10.1016/j.jas.2010.06.026>
- 977
- 978 Gowlett, J. A. J. (1984). Mental abilities of early man: A look at some hard evidence. *Higher
979 Education Quarterly*, 38(3), 199–220. <https://doi.org/10.1111/j.1468-2273.1984.tb01387.x>
- 980
- 981 Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting
982 to new responses in a weigl-type card-sorting problem. *Journal of Experimental Psychology*,
983 38(4), 404–411. <https://doi.org/10.1037/h0059831>
- 984
- 985 Hecht, E. E., Gutman, D. A., Bradley, B. A., Preuss, T. M., & Stout, D. (2015). Virtual dissection and

- 984 comparative connectivity of the superior longitudinal fasciculus in chimpanzees and humans.
985 *NeuroImage*, 108, 124–137. <https://doi.org/10.1016/j.neuroimage.2014.12.039>
- 986 Hecht, E. E., Gutman, D. A., Khreisheh, N., Taylor, S. V., Kilner, J. M., Faisal, A. A., Bradley, B. A.,
987 Chaminade, T., & Stout, D. (2015). Acquisition of Paleolithic toolmaking abilities involves
988 structural remodeling to inferior frontoparietal regions. *Brain Structure & Function*, 220(4),
989 2315–2331. <https://doi.org/10.1007/s00429-014-0789-6>
- 990 Hecht, E. E., Gutman, D. A., Preuss, T. M., Sanchez, M. M., Parr, L. A., & Rilling, J. K. (2013). Process
991 versus product in social learning: Comparative diffusion tensor imaging of neural systems
992 for action executionobservation matching in macaques, chimpanzees, and humans. *Cerebral*
993 *Cortex*, 23(5), 1014–1024. <https://doi.org/10.1093/cercor/bhs097>
- 994 Hecht, E. E., Murphy, L. E., Gutman, D. A., Votaw, J. R., Schuster, D. M., Preuss, T. M., Orban,
995 G. A., Stout, D., & Parr, L. A. (2013). Differences in neural activation for object-directed
996 grasping in chimpanzees and humans. *The Journal of Neuroscience*, 33(35), 14117–14134.
997 <https://doi.org/10.1523/JNEUROSCI.2172-13.2013>
- 998 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302),
999 29–29. <https://doi.org/10.1038/466029a>
- 1000 Hewes, G. W. (1993). A history of speculation on the relation between tools and language. In K.
1001 R. Gibson & T. Ingold (Eds.), *Tools, language and cognition in human evolution* (pp. 20–31).
1002 Cambridge University Press.
- 1003 Heyes, C. (2018). Enquire within: Cultural evolution and cognitive science. *Philosophical Transac-*
1004 *tions of the Royal Society B: Biological Sciences*, 373(1743), 20170051. <https://doi.org/10.1098/rstb.2017.0051>
- 1006 Hinton, G. E., & Nowlan, S. J. (1996). How learning can guide evolution. In R. K. Belew & M.
1007 Mitchell (Eds.), *Adaptive individuals in evolving populations: Models and algorithms* (pp.
1008 447–454). Addison-Wesley Publishing Company.
- 1009 Isaac, G. L. (1976). Stages of Cultural Elaboration in the Pleistocene: Possible Archaeological
1010 Indicators of the Development of Language Capabilities. *Annals of the New York Academy of*
1011 *Sciences*, 280(1), 275–288. <https://doi.org/10.1111/j.1749-6632.1976.tb25494.x>
- 1012 Jonassen, D. H., & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and*

- 1013 *instruction*. Lawrence Erlbaum.
- 1014 Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social
1015 Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, 22(7),
1016 651–665. <https://doi.org/10.1016/j.tics.2018.04.003>
- 1017 Key, A. J. M., & Dunmore, C. J. (2015). The evolution of the hominin thumb and the influence
1018 exerted by the non-dominant hand during stone tool production. *Journal of Human Evolution*,
1019 78, 60–69. <https://doi.org/10.1016/j.jhevol.2014.08.006>
- 1020 Key, A. J. M., & Dunmore, C. J. (2018). Manual restrictions on Palaeolithic technological behaviours.
1021 *PeerJ*, 6, e5399. <https://doi.org/10.7717/peerj.5399>
- 1022 Key, A. J. M., & Lycett, S. J. (2014). Are bigger flakes always better? An experimental assessment of
1023 flake size variation on cutting efficiency and loading. *Journal of Archaeological Science*, 41,
1024 140–146. <https://doi.org/10.1016/j.jas.2013.07.033>
- 1025 Key, A. J. M., & Lycett, S. J. (2019). Biometric variables predict stone tool functional performance
1026 more effectively than tool-form attributes: a case study in handaxe loading capabilities.
1027 *Archaeometry*, 61(3), 539–555. <https://doi.org/10.1111/arcm.12439>
- 1028 Khreisheh, N. N., Davies, D., & Bradley, B. A. (2013). Extending Experimental Control: The Use of
1029 Porcelain in Flaked Stone Experimentation. *Advances in Archaeological Practice*, 1(1), 38–46.
1030 <https://doi.org/10.7183/2326-3768.1.1.37>
- 1031 Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of
1032 teaching behavior in humans and other animals. *The Behavioral and Brain Sciences*, 38, e31.
1033 <https://doi.org/10.1017/S0140525X14000090>
- 1034 Laland, K. N. (2017). The origins of language in teaching. *Psychonomic Bulletin & Review*, 24(1),
1035 225–231. <https://doi.org/10.3758/s13423-016-1077-7>
- 1036 Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philoso-
1037 sophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130302. <https://doi.org/10.1098/rstb.2013.0302>
- 1039 Lombao, D., Guardiola, M., & Mosquera, M. (2017). Teaching to make stone tools: new experi-
1040 mental evidence supporting a technological hypothesis for the origins of language. *Scientific
1041 Reports*, 7(1), 1–14. <https://doi.org/10.1038/s41598-017-14322-y>

- 1042 Marwick, B. (2017). Computational Reproducibility in Archaeological Research: Basic Principles
1043 and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory*,
1044 24(2), 424–450. <https://doi.org/10.1007/s10816-015-9272-9>
- 1045 Marzke, M. W., Toth, N., Schick, K., Reece, S., Steinberg, B., Hunt, K., Linscheid, R. L., & An, K.-N.
1046 (1998). EMG study of hand muscle recruitment during hard hammer percussion manufacture
1047 of Oldowan tools. *American Journal of Physical Anthropology*, 105(3), 315–332. <https://doi.or>
1048 [g/10.1002/\(SICI\)1096-8644\(199803\)105:3%3C315::AID-AJPA3%3E3.0.CO;2-Q](g/10.1002/(SICI)1096-8644(199803)105:3%3C315::AID-AJPA3%3E3.0.CO;2-Q)
- 1049 Mateos, A., Terradillos-Bernal, M., & Rodríguez, J. (2019). Energy Cost of Stone Knapping. *Journal*
1050 *of Archaeological Method and Theory*, 26(2), 561–580. <https://doi.org/10.1007/s10816-018-9382-2>
- 1052 Miu, E., Gulley, N., Laland, K. N., & Rendell, L. (2020). Flexible learning, rather than inveterate
1053 innovation or copying, drives cumulative knowledge gain. *Science Advances*, 6(23), eaaz0286.
1054 <https://doi.org/10.1126/sciadv.aaz0286>
- 1055 Molleman, L., Kurvers, R. H. J. M., & van den Bos, W. (2019). Unleashing the BEAST: a brief
1056 measure of human social information use. *Evolution and Human Behavior*, 40(5), 492–499.
1057 <https://doi.org/10.1016/j.evolhumbehav.2019.06.005>
- 1058 Montagu, A. (1976). Toolmaking, Hunting, and the Origin of Language. *Annals of the New York*
1059 *Academy of Sciences*, 280(1), 266–274. <https://doi.org/10.1111/j.1749-6632.1976.tb25493.x>
- 1060 Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M.,
1061 Cross, C. P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., & Laland, K. N. (2015). Experi-
1062 mental evidence for the co-evolution of hominin tool-making teaching and language. *Nature*
1063 *Communications*, 6(1), 6029. <https://doi.org/10.1038/ncomms7029>
- 1064 Nonaka, T., Bril, B., & Rein, R. (2010). How do stone knappers predict and control the outcome
1065 of flaking? Implications for understanding early stone tool technology. *Journal of Human*
1066 *Evolution*, 59(2), 155–167. <https://doi.org/10.1016/j.jhevol.2010.04.006>
- 1067 Oakley, K. P. (1949). *Man the toolmaker*. Trustees of the British Museum.
- 1068 Ohnuma, K., Aoki, K., & Akazawa, A. T. (1997). Transmission of tool-making through verbal
1069 and non-verbal commu-nication: Preliminary experiments in levallois flake production.
1070 *Anthropological Science*, 105(3), 159–168. <https://doi.org/10.1537/ase.105.159>

- 1071 Pargeter, J., Khreisheh, N., Shea, J. J., & Stout, D. (2020). Knowledge vs. know-how? Dissecting
1072 the foundations of stone knapping skill. *Journal of Human Evolution*, 145, 102807. <https://doi.org/10.1016/j.jhevol.2020.102807>
- 1074 Pargeter, J., Khreisheh, N., & Stout, D. (2019). Understanding stone tool-making skill acquisition:
1075 Experimental methods and evolutionary implications. *Journal of Human Evolution*, 133,
1076 146–166. <https://doi.org/10.1016/j.jhevol.2019.05.010>
- 1077 Pelegrin, J. (1990). Prehistoric Lithic Technology : Some Aspects of Research. *Archaeological
1078 Review from Cambridge*, 9(1), 116–125. [/paper/Prehistoric-Lithic-Technology-%3A-Some-
1079 Aspects-of-Pelegrin/5e02fc2a5280ac128727275ab6b833756e6a6056](#)
- 1080 Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to
1081 large-scale decoding. *Neuron*, 72(5), 692–697. <https://doi.org/10.1016/j.neuron.2011.11.001>
- 1082 Prascunas, M. M. (2007). Bifacial Cores and Flake Production Efficiency: An Experimental Test of
1083 Technological Assumptions. *American Antiquity*, 72(2), 334–348. [https://doi.org/10.2307/40 035817](https://doi.org/10.2307/40
1084 035817)
- 1085 Putt, S. S. (2015). The origins of stone tool reduction and the transition to knapping: An experi-
1086 mental approach. *Journal of Archaeological Science: Reports*, 2, 51–60. [https://doi.org/10.101 6/j.jasrep.2015.01.004](https://doi.org/10.101
1087 6/j.jasrep.2015.01.004)
- 1088 Putt, S. S., Wijekumar, S., Franciscus, R. G., & Spencer, J. P. (2017). The functional brain networks
1089 that underlie Early Stone Age tool manufacture. *Nature Human Behaviour*, 1(6), 1–8. <https://doi.org/10.1038/s41562-017-0102>
- 1091 Putt, S. S., Wijekumar, S., & Spencer, J. P. (2019). Prefrontal cortex activation supports the
1092 emergence of early stone age toolmaking skill. *NeuroImage*, 199, 57–69. [https://doi.org/10.1016/j.neuroimage.2019.05.056](https://doi.org/10.1
1093 016/j.neuroimage.2019.05.056)
- 1094 Putt, S. S., Woods, A. D., & Franciscus, R. G. (2014). The role of verbal interaction during
1095 experimental bifacial stone tool manufacture. *Lithic Technology*, 39(2), 96–112. [https://doi.org/10.1179/0197726114Z.00000000036](https://doi.org/10.1
1096 01179/0197726114Z.00000000036)
- 1097 Rein, R., Nonaka, T., & Bril, B. (2014). Movement Pattern Variability in Stone Knapping: Im-
1098 plications for the Development of Percussive Traditions. *PLOS ONE*, 9(11), e113567. <https://doi.org/10.1371/journal.pone.0113567>

- 1100 Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L.,
1101 Ghirlanda, S., Lillicrap, T., & Laland, K. N. (2010). Why Copy Others? Insights from the
1102 Social Learning Strategies Tournament. *Science*, 328(5975), 208–213. <https://doi.org/10.1126/science.1184719>
- 1104 Reti, J. S. (2016). Quantifying Oldowan Stone Tool Production at Olduvai Gorge, Tanzania. *PLOS
1105 ONE*, 11(1), e0147352. <https://doi.org/10.1371/journal.pone.0147352>
- 1106 Roux, V., Bril, B., & Dietrich, G. (1995). Skills and learning difficulties involved in stone knapping:
1107 The case of stone-bead knapping in khambhat, india. *World Archaeology*, 27(1), 63–87. <https://doi.org/10.1080/00438243.1995.9980293>
- 1109 Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., & Eliceiri, K. W.
1110 (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*,
1111 18(1), 529. <https://doi.org/10.1186/s12859-017-1934-z>
- 1112 Sasaki, H., Kasagi, F., Yamada, M., & Fujita, S. (2007). Grip Strength Predicts Cause-Specific
1113 Mortality in Middle-Aged and Elderly Persons. *The American Journal of Medicine*, 120(4),
1114 337–342. <https://doi.org/10.1016/j.amjmed.2006.04.018>
- 1115 Schillinger, K., Mesoudi, A., & Lycett, S. J. (2014). Copying Error and the Cultural Evolution
1116 of “Additive” vs. “Reductive” Material Traditions: An Experimental Assessment. *American
1117 Antiquity*, 79(1), 128–143. <https://doi.org/10.7183/0002-7316.79.1.128>
- 1118 Shallice, T., Broadbent, D. E., & Weiskrantz, L. (1982). Specific impairments of planning. *Philosophical
1119 Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089), 199–209.
1120 <https://doi.org/10.1098/rstb.1982.0082>
- 1121 Shea, J. J. (2015). Making and using stone tools: Advice for learners and teachers and insights for
1122 archaeologists. *Lithic Technology*, 40(3), 231–248. [https://doi.org/10.1179/2051618515Y.000000011](https://doi.org/10.1179/2051618515Y.0000
1123 000011)
- 1124 Shea, J. J. (2016). *Stone tools in human evolution: Behavioral differences among technological
1125 primates*. Cambridge University Press. <https://doi.org/10.1017/9781316389355>
- 1126 Sherwood, C. C., & Gómez-Robles, A. (2017). Brain plasticity and human evolution. *Annual Review
1127 of Anthropology*, 46(1), 399–419. <https://doi.org/10.1146/annurev-anthro-102215-100009>
- 1128 Stout, D. (2002). Skill and cognition in stone tool production: An ethnographic case study from

- 1129 irian jaya. *Current Anthropology*, 43(5), 693–722. <https://doi.org/10.1086/342638>
- 1130 Stout, D. (2010). Possible relations between language and technology in human evolution. In A.
1131 Nowell & I. Davidson (Eds.), *Stone tools and the evolution of human cognition* (pp. 159–184).
1132 University Press of Colorado.
- 1133 Stout, D. (2013). Neuroscience of technology. In P. J. Richerson & M. H. Christiansen (Eds.),
1134 *Cultural evolution: Society, technology, language, and religion* (pp. 157–173). The MIT Press.
- 1135 Stout, D., Apel, J., Commander, J., & Roberts, M. (2014). Late Acheulean technology and cognition
1136 at Boxgrove, UK. *Journal of Archaeological Science*, 41, 576–590. <https://doi.org/10.1016/j.jas.2013.10.001>
- 1138 Stout, D., & Chaminade, T. (2007). The evolutionary neuroscience of tool making. *Neuropsycholo-*
1139 *gia*, 45(5), 1091–1100. <https://doi.org/10.1016/j.neuropsychologia.2006.09.014>
- 1140 Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution.
1141 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1585), 75–87. <https://doi.org/10.1098/rstb.2011.0099>
- 1143 Stout, D., & Hecht, E. E. (2017). Evolutionary neuroscience of cumulative culture. *Proceedings of*
1144 *the National Academy of Sciences*, 114(30), 7861–7868. <https://doi.org/10.1073/pnas.1620738114>
- 1146 Stout, D., Hecht, E., Khriesheh, N., Bradley, B., & Chaminade, T. (2015). Cognitive Demands of
1147 Lower Paleolithic Toolmaking. *PLOS ONE*, 10(4), e0121804. <https://doi.org/10.1371/journal.pone.0121804>
- 1149 Stout, D., & Khriesheh, N. (2015). Skill Learning and Human Brain Evolution: An Experimental
1150 Approach. *Cambridge Archaeological Journal*, 25(4), 867–875. <https://doi.org/10.1017/S0959774315000359>
- 1152 Stout, D., Passingham, R., Frith, C., Apel, J., & Chaminade, T. (2011). Technology, expertise and
1153 social cognition in human evolution. *The European Journal of Neuroscience*, 33(7), 1328–1338.
1154 <https://doi.org/10.1111/j.1460-9568.2011.07619.x>
- 1155 Stout, D., Quade, J., Semaw, S., Rogers, M. J., & Levin, N. E. (2005). Raw material selectivity of the
1156 earliest stone toolmakers at Gona, Afar, Ethiopia. *Journal of Human Evolution*, 48(4), 365–380.
1157 <https://doi.org/10.1016/j.jhevol.2004.10.006>

- 1158 Stout, D., Rogers, M. J., Jaeggi, A. V., & Semaw, S. (2019). Archaeology and the origins of hu-
1159 man cumulative culture: A case study from the earliest oldowan at gona, ethiopia. *Current*
1160 *Anthropology*, 60(3), 309–340. <https://doi.org/10.1086/703173>
- 1161 Stout, D., & Semaw, S. (2006). Knapping skill of the earliest stone toolmakers: Insights from the
1162 study of modern human novices. In N. Toth & K. Schick (Eds.), *The Oldowan: Case studies into*
1163 *the earliest Stone Age* (pp. 307–320). Stone Age Institute Press.
- 1164 Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). The MIT
1165 Press.
- 1166 Tehrani, J. J., & Riede, F. (2008). Towards an archaeology of pedagogy: Learning, teaching and
1167 the generation of material culture traditions. *World Archaeology*, 40(3), 316–331. <https://doi.org/10.1080/00438240802261267>
- 1169 Tennie, C., Premo, L. S., Braun, D. R., & McPherron, S. P. (2017). Early stone tools and cultural
1170 transmission: Resetting the null hypothesis. *Current Anthropology*, 58(5), 652–672. <https://doi.org/10.1086/693846>
- 1172 Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency
1173 and flexibility in human social information use. *Proceedings of the Royal Society B: Biological*
1174 *Sciences*, 281(1776), 20132864. <https://doi.org/10.1098/rspb.2013.2864>
- 1175 Toth, N., & Schick, K. (1993). Early stone industries and inferences regarding language and
1176 cognition. In K. R. Gibson & T. Ingold (Eds.), *Tools, language and cognition in human evolution*
1177 (pp. 346–362). Cambridge University Press.
- 1178 Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and
1179 learning: Evidence from the serial reaction time task. *Memory & Cognition*, 33(2), 213–220.
1180 <https://doi.org/10.3758/BF03195310>
- 1181 Vostroknutov, A., Polonio, L., & Coricelli, G. (2018). The Role of Intelligence in Social Learning.
1182 *Scientific Reports*, 8(1), 6896. <https://doi.org/10.1038/s41598-018-25289-9>
- 1183 Washburn, S. L. (1960). Tools and human evolution. *Scientific American*, 203(3), 62–75. <https://doi.org/10.1038/scientificamerican0960-62>
- 1185 Whiten, A. (2015). Experimental studies illuminate the cultural transmission of percussive tech-
1186 nologies in homo and pan. *Philosophical Transactions of the Royal Society B: Biological*

- 1187 *Sciences*, 370(1682), 20140359. <https://doi.org/10.1098/rstb.2014.0359>
- 1188 Whittaker, J. C. (1994). *Flintknapping: Making and Understanding Stone Tools*. University of Texas
1189 Press.
- 1190 Wilkins, J. (2018). The Point is the Point: Emulative social learning and weapon manufacture
1191 in the Middle Stone Age of South Africa. In M. J. O'Brien, B. Buchanan, & M. I. Eren (Eds.),
1192 *Convergent Evolution in Stone-Tool Technology* (pp. 153–174). The MIT Press.
- 1193 Williams-Hatala, E. M., Hatala, K. G., Gordon, M., Key, A., Kasper, M., & Kivell, T. L. (2018). The
1194 manual pressures of stone tool behaviors and their implications for the evolution of the human
1195 hand. *Journal of Human Evolution*, 119, 14–26. <https://doi.org/10.1016/j.jhevol.2018.02.008>
- 1196 Williams-Hatala, E. M., Hatala, K. G., Key, A., Dunmore, C. J., Kasper, M., Gordon, M., & Kivell, T.
1197 L. (2021). Kinetics of stone tool production among novice and expert tool makers. *American
1198 Journal of Physical Anthropology*, 174(4), 714–727. <https://doi.org/10.1002/ajpa.24159>
- 1199 Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploita-
1200 tion with information and randomization. *Current Opinion in Behavioral Sciences*, 38, 49–56.
1201 <https://doi.org/10.1016/j.cobeha.2020.10.001>
- 1202 Wind, A. E., Takken, T., Helders, P. J. M., & Engelbert, R. H. H. (2010). Is grip strength a predictor for
1203 total muscle strength in healthy children, adolescents, and young adults? *European Journal of
1204 Pediatrics*, 169(3), 281–287. <https://doi.org/10.1007/s00431-009-1010-4>
- 1205 Wynn, T. (1979). The intelligence of later acheulean hominids. *Man*, 14(3), 371–391. <https://doi.org/10.2307/2801865>
- 1206
- 1207 Wynn, T. (2017). Evolutionary cognitive archaeology. In T. Wynn & F. Coolidge (Eds.), *Cognitive
1208 models in Palaeolithic archaeology* (pp. 1–20). Oxford University Press.
- 1209 Wynn, T., & Coolidge, F. L. (2004). The expert Neandertal mind. *Journal of Human Evolution*,
1210 46(4), 467–487. <https://doi.org/10.1016/j.jhevol.2004.01.005>
- 1211 Wynn, T., & Coolidge, F. L. (2016). Archeological insights into hominin cognitive evolution.
1212 *Evolutionary Anthropology: Issues, News, and Reviews*, 25(4), 200–213. [https://doi.org/10.1002/evan.21496](https://doi.org/10.100
1213 2/evan.21496)
- 1214 Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 1–37. <https://doi.or>

1215

[g/10.1017/S0140525X20001685](https://doi.org/10.1017/S0140525X20001685)