Archaeology and the Origins of Human Cumulative Culture: A Case Study from the Earliest Oldowan at Gona, Ethiopia — Source link ⧉

Dietrich Stout, Michael J. Rogers, Adrian V. Jaeggi, Sileshi Semaw

Institutions: Emory University, Southern Connecticut State University, University of Zurich

Related papers:

- Early Stone Tools and Cultural Transmission: Resetting the Null Hypothesis

- Ratcheting up the ratchet: on the evolution of cumulative culture

- Experimental evidence for the co-evolution of hominin tool-making teaching and language

- The Archaeology of Teaching and the Evolution of Homo docens

- 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya

Year: 2019

Archaeology and the Origins of Human Cumulative Culture: A Case Study from the Earliest Oldowan at Gona, Ethiopia

Stout, Dietrich ; Rogers, Michael J ; Jaeggi, Adrian V ; Semaw, Sileshi

Archaeology and the Origins of Human Cumulative Culture

A Case Study from the Earliest Oldowan at Gona, Ethiopia

by Dietrich Stout, Michael J. Rogers, Adrian V. Jaeggi, and Sileshi Semaw

Online enhancements: appendix

The capacity of *Homo sapiens* for the intergenerational accumulation of complex technologies, practices, and beliefs is central to contemporary accounts of human distinctiveness. However, the actual antiquity and evolutionary origins of cumulative culture are not known. Here we propose and exemplify a research program for studying the origins of cumulative culture using archaeological evidence. Our stepwise approach disentangles assessment of the observed fidelity of behavior reproduction from inferences regarding required learning mechanisms (e.g., teaching, imitation) and the explanation of larger-scale patterns of change. It is empirically grounded in technological analysis of artifact assemblages using well-validated experimental models. We demonstrate with a case study using a toolmaking replication experiment to assess evidence of behavior copying across three 2.6 Ma Oldowan sites from Gona, Ethiopia. Results fail to reveal any effects of raw material size, shape, quality, or reduction intensity that could explain the observed details of intersite technological variation in terms of individual learning across different local conditions. This supports the view that relatively detailed copying of toolmaking methods was already a feature of Oldowan technological reproduction at ca. 2.6 Ma. We conclude with a discussion of prospects and implications for further research on the evolution of human cumulative culture.

Modern humans live in a culturally constructed niche including complex adaptive technologies, practices, and beliefs accumulated, modified, and improved over generations. The advent of such cumulative culture is thought to have underwritten the remarkable demographic success of modern humans, to have been a key factor in hominin brain expansion, and to have produced many of the cognitive and behavioral characteristics that distinguish our species (Boyd, Richerson, and Henrich 2011; Tomasello 1999). There is thus intense interest in better understanding the evolutionary origins of human cumulative culture.

This has generally been pursued through the comparative study of extant species (e.g., Dean et al. 2014). This method is essential for identifying recurring evolutionary relationships but cannot by itself reveal the contingencies of timing and context that defined the particular path of human evolution (Stout 2018). Reconstructing the origins of human cumulative culture will require archaeological and paleontological evidence of what actually occurred in human evolution, combined with the ethnographic, ethological, and experimental analogies needed to interpret this evidence (Henrich 2015; Stout and Hecht 2017). Here we seek to advance this project through a case study using a toolmaking replication experiment to assess evidence of technological reproduction in the earliest Oldowan from Gona, Afar, Ethiopia.

As with human uniqueness generally, there is spirited disagreement over the degree of continuity between human (cumulative) culture and the behavioral traditions of other animals. One influential view sees human culture as qualitatively distinct in its dependence on evolved capacities for high-fidelity social transmission that is absent in other species (Tennie, Call, and Tomasello 2009; Tomasello 1999). The emergence of distinctly human cumulative culture is thus thought to have been a relatively discrete event requiring "one and only one biological adaptation" (Tomasello 1999:7). A competing perspective

Dietrich Stout is Associate Professor in the Department of Anthropology at Emory University (1557 Dickey Drive, Atlanta, Georgia 30322, USA [dwstout@emory.edu]). **Michael J. Rogers** is Professor in the Department of Anthropology of Southern Connecticut State University (Engleman Hall C027, 501 Crescent Street, New Haven, Connecticut 06515, USA). **Adrian V. Jaeggi** is Head of the Human Ecology Group, Institute of Evolutionary Medicine, University of Zurich (Winterthurerstrasse 190, 8057 Zürich, Switzerland). **Sileshi Semaw** is Director of the Gona Research Project at the Centro Nacional de Investigación sobre la Evolución Humana, Paseo Sierra de Atapuerca, 3 09002 Burgos, España). This paper was submitted 10 VIII 17, accepted 7 IV 18, and electronically published 29 IV 19.

views the animal-human disjunction as one of degree rather than kind, with some rudimentary cultural accumulation evident in other species (Dean et al. 2014; Vale et al. 2017). In this account, humans are exceptional in the degree to which they have enhanced a diverse portfolio of learning strategies shared with other apes (Whiten 2011). This implies a more gradual evolution of cumulative culture, likely involving multiple traits, processes, and events (e.g., Pradhan, Tennie, and van Schaik 2012). Paleolithic archaeology is uniquely positioned to test these opposing views over evolutionary time and across a range of extinct hominins that may have represented conditions absent from the extant comparative sample.

Cumulative culture can be more specifically defined as "the modification, over multiple transmission episodes, of cultural traits (behavioral patterns transmitted through social learning) resulting in an increase in the complexity or efficiency of those traits" (Dean et al. 2014:288), a process also known as the "ratchet effect" (Tomasello 1999). Macroscale increases in maximum observed technological complexity during the Paleolithic might seem to provide prima facie evidence of cumulative culture (Stout 2011), but this has been disputed. One objection is that the pace of change appears too slow to reflect cultural evolution (Richerson and Boyd 2005). However, it is far from clear what pace should be considered indicative of cumulative culture. Modeling indicates that early stages of cumulative culture change can be quite slow (Enquist et al. 2008) and that rates are sensitive to contextual variables such as population size and structure (Powell, Shennan, and Thomas 2009; Pradhan, Tennie, and van Schaik 2012), which may themselves be reciprocally impacted by cultural accumulation (Kolodny, Creanza, and Feldman 2016). These complex dynamics can generate periods of prolonged stasis and punctuated change (Kolodny, Creanza, and Feldman 2016; Morgan 2016), even given cumulative culture capacity.

A second objection is that the definition of cumulative culture presented above is too inclusive. Some researchers prefer to maintain a qualitative animal-human boundary by restricting the term "cumulative" to culture-dependent traits that have already been ratcheted beyond the possibility of individual reinvention (e.g., Reindl et al. 2017). Impossibility is obviously a more challenging standard of evidence and has been difficult to demonstrate for any animal behavior (Vale et al. 2017). Tennie et al. (2016, 2017) apply this standard to Paleolithic archaeology as a thought experiment they term the "island test": only if it seems impossible for an individual isolated from birth (e.g., on a remote island) to reinvent a behavior should it be accepted as evidence of cumulative culture. They conclude that Early Stone Age (ESA) technologies fail this test and that observed patterns of technological change likely resulted from individual learning influenced by "changes in the environment, low-level social learning mechanisms, and the psychology and physiology of the species of interest" (Tennie et al. 2016:127) rather than cultural ratcheting.

Importantly, this stringent approach would reject actual cases of ratcheting in which complexity had not (yet) accumulated past the possibility of individual recapitulation. Such cases are exactly what a gradualist perspective would predict for the initial stages of cumulative culture evolution (e.g., Henrich 2015; Pradhan, Tennie, and van Schaik 2012). By excluding them, the island test favors discontinuity and enforces the theoretical stance that true cumulative culture requires unique human capacities for high-fidelity social transmission via teaching and imitation. The imperative is to avoid false positives in which cultural status is assigned to behaviors reproducible through low-fidelity mechanisms other than teaching or imitation.

In our opinion, the inclusion of particular learning processes as part of the definition of cumulative culture is problematic insofar as it introduces strong theoretical assumptions into the interpretation of archaeological evidence. In fact, relatively little is known about the fidelity and cumulative potential of different social learning processes in the real world. This should be a topic of investigation rather than a source of axiomatic definitions. Even in the lab, there is evidence that supposedly low-fidelity "end-state emulation" (i.e., reverse engineering) is sufficient for cumulative improvement of simple artifacts such as spaghetti towers (Reindl et al. 2017).

More broadly, a taxonomic approach to learning types may not always map easily onto complex, real-world behaviors. For example, so-called object-movement reenactment is considered a "fine-grained form of emulation" (Reindl et al. 2017) rather than imitation because it involves copying the motion of objects rather than bodies. However, it is not clear that copying of particular body, as opposed to tool, movements is critical to high-fidelity technological reproduction (e.g., Rein, Nonaka, and Bril 2014). Indeed, a strict organism-environment distinction is difficult to maintain at a conceptual or neural level (Maravita and Iriki 2004) during the active manipulation of hand-held tools, as occurs, for example, in stone knapping.

Similarly, the attempt to dichotomize process copying from product copying (Tennie, Call, and Tomasello 2009) runs into difficulty with technological behaviors that involve complex sequences of nested subgoals and intermediate target forms and are themselves embedded in a larger goal-oriented behavioral context (Stout 2013). Which of the multiple levels of organization ranging from muscle contractions and joint synergies to net caloric yields and manufactured social identities should mark the boundary between process and product? Exact copying of bodily actions is obviously impossible across individuals with different bodies (e.g., children and adults), and some degree of goal abstraction will always be required (de Vignemont and Haggard 2008). It is possible to design laboratory manipulations that dichotomize bodies from tools and means from ends, but in the real world these exist along a continuum with poorly defined end points.

For these reasons, we have suggested that copying fidelity should be assessed in terms of the degree of structural complexity (Stout 2013) accurately reproduced (Stout 2011) rather than the presence or absence of copying at a particular level of organization or reliance on a particular type of learning. This establishes a clear distinction between fidelity as an empirical phenomenon and the possible mechanisms (social, genetic, en-

vironmental, or otherwise) that may have produced it, allowing for a more incremental and less theory-laden approach to the evolution of cumulative culture.

A Stepwise Approach to Cumulative Culture Origins

Although behavior copying is commonly called social "transmission," this is misleading if it is taken to imply passive reception of rules or recipes for action. In modern human knapping and similar crafts, what is learned is a flexible skill rather than an invariant formula. Such learning requires an extended interaction between social inputs and motivated individual practice (Stout 2013; Stout and Hecht 2017) better described as behavioral "reproduction" (cf. "guided rediscovery"; Ingold 1998) rather than transmission (Nonaka, Bril, and Rein 2010). This dialectic process, exemplified by coaching or apprenticeship, has recently been described by Whiten (2015) as a "helical curriculum." In such a curriculum, reproduction of observed actions during practice is just one potential means to the end of discovering subtle task affordances not directly available to the naïve observer. Other facilitatory influences can include the physical and social context (the learning niche) created by ongoing technological activity as well as affective feedback, attention direction, practice opportunity scaffolding, and even intentional demonstration and instruction. These complexities are increasingly recognized in more diverse and inclusive conceptions of teaching and its evolutionary origins (e.g., Gärdenfors and Högberg 2017).

The complexity of modern human skill learning suggests to us that cumulative culture is unlikely to be a unitary capacity arising from the punctuated appearance of one or two key psychological innovations, and more likely is a complex trait with a correspondingly complex evolutionary history of gradual and/or piecemeal emergence (Henrich 2015; Stout and Hecht 2017). But we might be wrong. The best approach to the archaeological record will thus be one that can recognize continuity or discontinuity from the evidence rather than presupposing one or the other by definition. To this end, we propose a stepwise approach distinguishing assessment of copying fidelity from subsequent inferences regarding social learning processes and attempts to explain diachronic patterns of change or stability (fig. 1). Whereas fidelity can be assessed relatively directly from archaeological reconstructions of technological behavior, identifying past learning processes will require an understanding of the technological implications of different learning conditions that is only now beginning to be developed (reviewed by Stout and Khreisheh 2015). Explaining patterns of diachronic change is an even more distal goal, as it potentially depends on these inferred learning processes as well as other cognitive, demographic, and environmental factors (Tennie et al. 2016).

The first step is to evaluate evidence of behavior reproduction. Although archaeological approaches to social transmission commonly focus on artifact morphology, this is a potentially incomplete and/or misleading proxy for behavior (Lycett and von Cramon-Taubadel 2015) given that similar forms can be produced in diverse ways (Sharon 2009) and that variable forms can result from shared methods (Toth 1985). Ideally, identification of Paleolithic behavior reproduction should rely on the recurrence of specific knapping methods rather than the

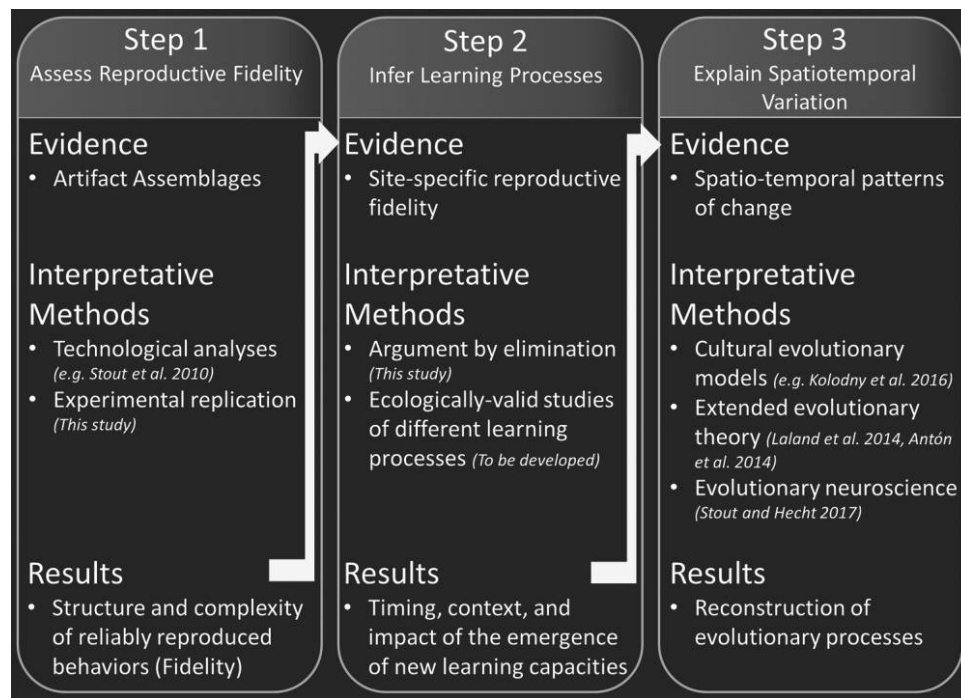| Step 1 Assess Reproductive Fidelity	Step 2 Infer Learning Processes	Step 3 Explain Spatiotemporal Variation
Evidence • Artifact Assemblages	**Evidence** • Site-specific reproductive fidelity	**Evidence** • Spatio-temporal patterns of change
Interpretative Methods • Technological analyses *(e.g. Stout et al. 2010)* • Experimental replication *(This study)*	**Interpretative Methods** • Argument by elimination *(This study)* • Ecologically-valid studies of different learning processes *(To be developed)*	**Interpretative Methods** • Cultural evolutionary models *(e.g. Kolodny et al. 2016)* • Extended evolutionary theory *(Laland et al. 2014, Antón et al. 2014)* • Evolutionary neuroscience *(Stout and Hecht 2017)*
Results • Structure and complexity of reliably reproduced behaviors (Fidelity)	**Results** • Timing, context, and impact of the emergence of new learning capacities	**Results** • Reconstruction of evolutionary processes

Figure 1. A stepwise research program for investigating the evolutionary origins of human cumulative culture. A color version of this figure is available online.

presence, absence, or morphology of tool types. Identification of these methods should in turn be based on well-validated experimental models (Eren et al. 2016). With these in hand, behavior reproduction would be indicated by particular methods' prevalence, and fidelity would be assessed with respect to their faithfully reproduced complexity (Stout 2011).

Actually investigating the processes responsible for behavior reproduction is the second step in the research program and requires identifying testable predictions of alternative explanations for the specific patterns documented in step 1. For example, behavior might be highly constrained by the limited possibilities for action (affordances) provided by objects and environments so that individuals with similar capacities and goals are quite likely to independently rediscover the same solutions. Tennie, Call, and Tomasello (2009) refer to this as a "zone of latent solutions" (ZLS) within which individual learning, niche construction, and low-fidelity copying of goals or outcomes but not detailed means should be sufficient for behavior reproduction. A testable implication is that the prevalence of individually rediscovered behavior details should be determined by the degree of constraint on the solution space. Thus, rediscovery and adoption of a particular detail should be ubiquitous if local conditions (including conspecific behavior; Fragaszy et al. 2013) bias individual learning toward that specific solution. Tennie, Call, and Tomasello (2009) describe this as a behavioral "founder effect." Conversely, if a diversity of more-or-less comparable alternative solutions present themselves, then we should expect commensurate diversity in the particular methods discovered and adopted by individuals. Where this prediction is violated, it is necessary to posit copying at the level of the particular behavior details in question. Importantly, such evidence does not indicate that copied details are necessarily culture dependent (Reindl et al. 2017) in the strong sense of being impossible for individuals to reinvent and thus does not address the question of whether copying was required for a particular technology (Tennie et al. 2017). As outlined above, we think this is the wrong question to ask. Instead our focus is on identifying evidence that detailed copying of particular means did in fact occur, whether or not it was essential.

Here we test the proposal that ESA technological reproduction did not involve social reproduction of detailed means (Tennie et al. 2016, 2017) by assessing the degree to which functional constraints could explain the biased representation of particular knapping methods at Gona. Attempting to explain rates and patterns of ESA technological change (i.e., the third step in the proposed research program) is largely premature at this point, but we do present some speculative scenario building in service of hypothesis generation in the "Discussion" section.

Case Study: Technological Reproduction at Gona

Stout et al. (2010) reported evidence of variation in knapping methods across three initial Oldowan (2.6–2.5 Ma) lithic assemblages from the sites of East Gona (EG) 10 and 12 and

Ounda Gona South (OGS) 7 in the Gona Research Project study area of Ethiopia. The East Gona sites, located in close spatial proximity (within approximately 300 m) and similar depositional contexts (proximal floodplain of the ancestral Awash River), share a preponderance of unifacial reduction. This method (fig. 2*a*, 2*b*), involves removal of flakes from a single core surface or face and is indicated archaeologically by a high percentage of cores bearing a unifacial scar pattern (EG-10 and EG-12: 69% and 78%) and of flakes displaying the original cobble exterior surface (the cortex) on their striking platforms (79% and 81%). Site EG-10 consists of two distinct artifact layers separated by 40 cm of sediment (Semaw 2000), suggesting repeated occupation and some temporal persistence of the unifacial technological preference.

Site OGS-7, located approximately 3 km south-southwest of the EG sites and deposited in a channel bank or margin context, shows a starkly different pattern of predominantly bi-/multifacial reduction. This method (fig. 2*c*, 2*d*) involves core rotation between flake removals and is reflected by a predominance of cores showing bi-/multifacial scar patterns (82%) and flakes with no cortex on their striking platforms (66%). Site OGS-7 is remarkably preserved. Artifacts were recovered from
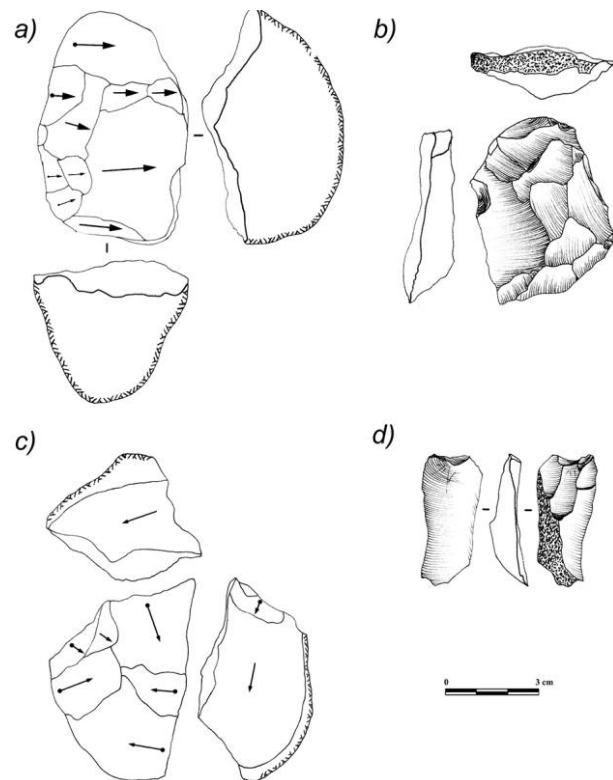

Figure 2. Unifacial core from EG-12 (*a*); cortical platform (type III) flake from EG-10 (*b*); multifacial core from OGS-7 (*c*); noncortical platform (type V) flake from OGS-7 (*d*). Cortex indicated by crosshatching (*a*, *c*) or stippling (*b*, *d*). Illustrations by Dominique Cauche.

a tightly restricted (<10 cm thick) layer located at a local contact between coarse sand and bedded floodplain silts (Semaw et al. 2003); more recent excavations (Rogers et al. 2013) suggest the presence of a second artifact layer separated by 15–20 cm of sediment that displays similar technological characteristics, again implying temporal persistence. As described by Stout et al. (2010), the nearby site of OGS-6 (Semaw et al. 2003) has also yielded a small excavated assemblage consistent with bi-/multifacial knapping (12 whole flakes, none with cortical platforms) from a different depositional setting (floodplain). Thus, there is a spatially patterned and temporally persistent pattern of technological difference at Gona that needs to be explained.

Step 1: Assessing Fidelity

The first step is to evaluate evidence of behavior reproduction while remaining agnostic about learning processes, social or otherwise. Using the criteria outlined above, the high prevalence of particular reduction strategies at different Gona sites should indicate reproduction of these behavioral details across individuals. We previously (Stout et al. 2010) identified these different reduction strategies at Gona using published technological typologies (i.e., scar patterns) and generalized experimental analogies (Toth 1987). Here we test our technological interpretations in a context-specific knapping experiment designed to see if the inferred methods actually do replicate key features of the Gona assemblages such as core scar patterns and reduction intensity as well as the size, shape, and frequencies of debitage types.

By adopting this experimental approach, we are employing an analogical argument that must be justified (Wylie 1985). As detailed below, we attempted to strengthen this analogy by matching raw materials and knapping methods as closely as possible to those indicated by the archaeology. These efforts serve to enhance what has been termed the source-side credibility (Wylie 1985) or external validity (Eren et al. 2016) of the experimental model. Conversely, we sought to increase the internal validity (Eren et al. 2016) of our experiments by controlling knapper skill, style, and technological goal and by recording initial cobble attributes as well as the number and relative location of all detachments that produced our experimental samples (cf. Reti 2016). As with all early Paleolithic experimental archaeology, however, we face the problem of using modern human performance to model the behavior of extinct hominin species. In principle, it can never be ruled out that subtle differences in technique, skill, unconscious bias, or some other unanticipated factor on the part of either the particular knapper or knappers or modern humans in general could compromise the analogy. It is thus critical to test the actual performance of the experimental model through direct comparison with the archaeology. In other words, the validity of experimental models must be assessed empirically in terms of their success in accurately predicting additional subject-side (Wylie 1985) features of archaeological assemblages. Such recursive

testing is the gold standard for strengthening analogical arguments. We thus present results of this comparison below.

Step 2: Inferring Learning Processes

Rather than behavior copying, different reduction strategies at Gona might reflect locally optimum technological solutions with a high likelihood of independent rediscovery. It would then be necessary to attribute intersite variation in reduction strategy to variation in raw material qualities and/or availability, differing technological goals, or differing hominin anatomical and/or psychological capabilities. Insofar as there is no known direct archaeological indicator of behavior copying versus rediscovery, these possibilities must be addressed through an admittedly flawed (Byrne 2007) process of elimination. Archaeologists will be familiar with this logic from Glynn Isaac's (1984) "method of residuals," which identifies culture as that which is left over when all other possible explanations of variation are exhausted.

This approach is problematic because the actual elimination of all possible alternatives is a logical, not to mention pragmatic, impossibility (see "Discussion"). A more particular problem in the case of technological reproduction is that a high potential for independent rediscovery does not exclude the possibility that a particular behavior was, in practice, reproduced through action copying (Byrne 2007). Thus, an argument, by elimination, has weaknesses going in both directions (confirmation and falsification) but is our only option until experimental research identifies more direct indicators of prehistoric learning processes (e.g., Schillinger, Mesoudi, and Lycett 2015). With these caveats in mind, we can evaluate the relative likelihood of different explanations for intersite variation at Gona.

Culture or biology? First, there is the possibility that biological differences between toolmakers at the various sites in, for example, hand morphology or perceptuomotor capacity, led to different task affordances and thus to different obvious solutions on which individual learners converged. This possibility cannot be dismissed, but it does require the onerous assumption that two biologically distinct hominin populations with differing innate capacities began making the earliest known Oldowan stone tools at roughly the same time and within about 3 km of one another.

The East Gona sites are dated to between 2.6–2.5 Ma (Semaw 1997), while OGS-6 and OGS-7 are even more tightly constrained, to <2.58 Ma (Semaw et al. 2003). Together, they represent the earliest known occurrences of characteristically Oldowan technology. There are no hominin fossils from Gona at ca. 2.5 Ma to support the presence or absence of multiple species; however, early *Homo* has been reported at 2.8 Ma from the nearby Ledi-Geraru research area (Villmoare et al. 2015) and at 2.3 Ma from Hadar (Kimbel et al. 1996), whereas *Australopithecus garhi* is present at 2.5 Ma in the Middle Awash (Asfaw et al. 1999). Unfortunately, none of these finds includes

hand fossils, and only *A. garhi* provides evidence of brain size. The fossil evidence of early hominin hand morphology in general is quite limited and has yet to resolve the antiquity and variability across species of human-like manipulative abilities, although there are indications of pre-*Homo* origins (Kivell 2015).

Turning to the artifacts themselves, we have previously noted that similarities in core reduction intensity and flake morphology across the Gona sites are inconsistent with marked differences in basic competence (Stout et al. 2010), as might perhaps be expected from different taxa such as small-brained *A. garhi* and early *Homo*. None of the sites exhibits the incomplete reduction, heavy core battering, and thin, side-struck flakes characteristic of bonobo knapping (Toth, Schick, and Semaw 2006), whereas all sites demonstrate abilities to identify, exploit, and maintain (sensu Moore 2011; Stout and Chaminade 2007) appropriate core angles for extensive reduction. While it remains possible that some subtle differences in innate aptitude of different hominin groups could have biased individual learning to produce the observed variation in knapping methods, we do not currently consider this to be a parsimonious explanation due to the multiple unsupported assumptions that are required.

Individual reinvention? Another possibility is that intersite differences at Gona are simply the summation of repeated individual adoption of similar responses to different local conditions, such as the nature and local abundance of raw materials or the intended function of tools. This explanation predicts that the different knapping methods should produce different outcomes under these conditions, thus providing some basis for individual toolmakers to reliably favor one over another. In other words, the alternatives should not be functionally neutral. To test this implication, we conducted a knapping experiment using cobbles collected from the local channel gravels (now preserved as cobble conglomerates) that served as the raw material source for Oldowan knappers at Gona (Stout et al. 2005). By systematically varying knapping method (unifacial, bi-/multifacial) across a range of cobble shapes while holding the knapper constant, we tested for technological effects of the different methods on variables including the frequency, type, and morphology of detached pieces per cobble. A similar approach was recently employed by Reti (2016) to study the effects of different Oldowan knapping methods at Olduvai Gorge, Tanzania.

Unfortunately for experimental replicators, the actual technological objectives of Oldowan knappers are not known. The prevailing view is that reduction focused on the production of detached pieces (Toth 1985) and specifically aimed to maximize the production of sharp edges useful for the cutting and/or scraping of animal and plant tissues (Lemorini et al. 2014). It is likely that Oldowan cores were also used as tools (Toth 1985), but there is no evidence that they were intentionally shaped for particular functions. In the Gona assemblages specifically, it would be hard to reconcile the apparent intensity

of reduction (Toth, Schick, and Semaw [2006] estimated about 63% original weight removed at East Gona) with anything other than a primary emphasis on detached piece production. We thus set production of cutting edges on detached pieces as the uniform goal throughout our experiment, while varying knapping methods. Of detached pieces, whole flakes are generally considered to have the highest utility (Toth 1985); however, use-wear evidence indicates fragments were also used (Lemorini et al. 2014), and we made no a priori assumptions regarding the desirability of fragments versus whole flakes. Instead, we evaluated all pieces using published methods for calculating "utility" (i.e., cutting edge length relative to piece size). Although they are theoretically motivated (Braun and Harris 2003) and reflective of archaeological consensus (Morgan et al. 2015), it is of course possible that these measures do not fully capture the goals and preferences of Oldowan knappers. This is a major issue for Early Stone Age archaeology and lies beyond what can reasonably be resolved here. As an initial step in this direction, however, we complemented our utility-based analysis with a data-driven approach using our experimental results to identify archaeologically underrepresented artifact types and then asking if they had distinctive features that might have led to their preferential removal for use elsewhere on the landscape (Schick 1987; Toth, Schick, and Semaw 2006). These desirable features provided an alternative metric of the functional relevance of different knapping methods.

Experimental Materials and Methods

Raw Materials

Cobbles approximating the size and fine-grained composition exploited by the Gona toolmakers (Stout et al. 2005) were selectively collected from local conglomerates. After initial collection, a large sample of cobbles (>150) were visually divided into four shape categories based on relative elongation and thickness: (1) sphere, (2) disk (flattened sphere), (3) rod (elongated sphere), or (4) tablet (elongated disk). During the experiment, the knapper (Dietrich Stout) selected cobbles ($n = 40$) from these categories in an attempt to balance the distribution of size, shapes, and raw material quality across bifacial and unifacial conditions. Knapping method was alternated over consecutive cobble reductions and was not contingent on selected cobble properties. Note that shape categories were used as a pragmatic aid during the experiment, but statistical analyses were conducted on the continuous shape metrics summarized below and detailed in the appendix (available online). As reported in the supplementary methods in the appendix, we found no significant differences in cobble size, shape, or texture between the two experimental conditions; cobble shape distributions are depicted in figure S2 in the appendix (figs. S1–S4 are available online).

Raw material characteristics were recorded and analyzed to ensure there were no systematic differences between the two experimental conditions and that both provided a good ap-

proximation of the archaeological materials. Experimental cobble size and shape were recorded prior to reduction. Size measures were length, breadth, thickness, and weight. The three linear measures were transformed into two dimensionless shape variables by first dividing them by the geometric mean and then performing a principal components analysis. Cobble angularity was estimated from digital photographs, and the most acute available angle (hereafter, Start Angle) was measured using a goniometer. Raw material quality was ranked after each reduction using the same three-point (glassy, smooth, rough) system previously (Stout et al. 2005) applied to the archaeological assemblages.

Comparison between the experimental and archaeological samples was somewhat more difficult. Initial cobble shape data are not available for the archaeology, and starting weight had to be estimated using the scar density index (SDI = scar count/ surface area) of Clarkson (2013). Nevertheless, we found that our experimental sample, while likely not an exact match, provides a useful approximation of the original cobble size distribution of the archaeological sample. Representative cores from the experimental and archaeological samples are shown in figures 3 and 4. Hammerstone selection is discussed in the appendix.
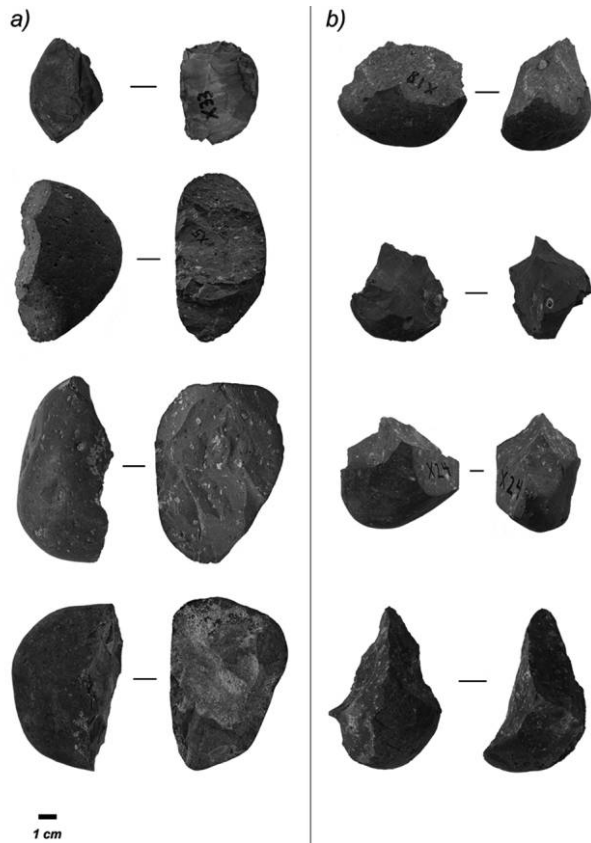

Figure 3. Experimental cores from the unifacial (*a*) and bi-/multifacial (*b*) experimental conditions. A color version of this figure is available online.
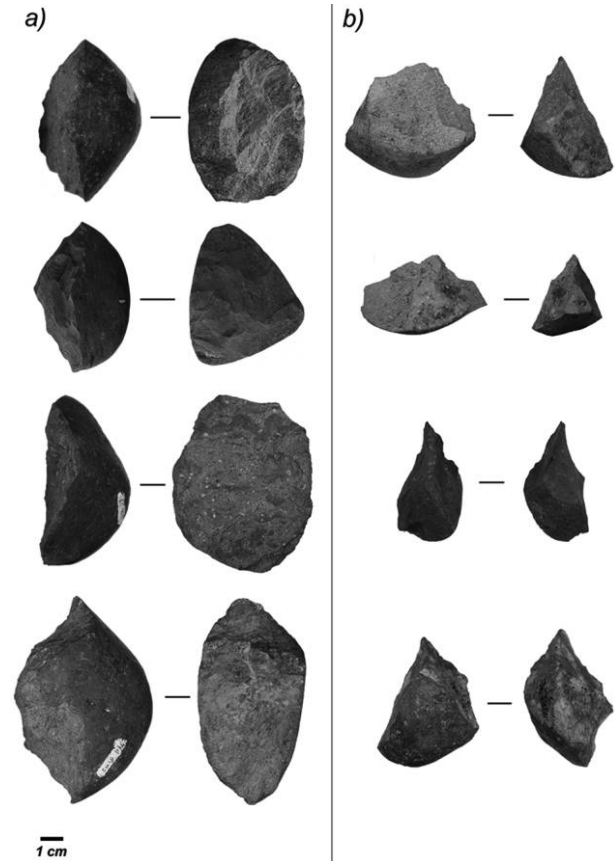

Figure 4. Archaeological cores from East Gona (*a*) and Ounda Gona South (*b*). A color version of this figure is available online.

Experimental Knapping

In order to control for individual variation in knapping skill and style, all reductions were performed by the first author (Dietrich Stout) over the course of two consecutive days. Cobbles were photographed, weighed, and measured before and after reduction. All detached pieces >20 mm were collected and numbered as they were produced. Each blow that produced one or more such pieces was counted (Detachment Number) and associated with the IDs of those pieces. Every such blow was also classified in relation to the previous one, as laterally adjacent (unifacial), alternating (bifacial), noncontiguous (migrating), or initial (first detachment). Cores were reduced to exhaustion using either an exclusively unifacial method (unifacial condition) or one in which bifacial alternation and multifacial migration were allowed (bi-/multifacial condition). The unifacial experiment included 425 productive blows: 93% unifacial and 7% initial. The bi-/multifacial experiment included 403 productive blows: 42% unifacial, 45% bifacial, 8% migrating, and 5% initial. Exhaustion was defined as the point where persistent effort (≥5 blows) failed to generate further removals >20 mm. Knapping method was alternated over consecutive cobble reductions. As reported above, this protocol generated

experimental samples that did not differ in initial cobble size, shape, or texture.

Our two experimental knapping methods represent hypotheses about the methods actually used at Gona. These hypotheses derive from qualitative inspection and evaluation of scar patterns on the Gona cores as well as application of a formal scar pattern typology (Stout et al. 2010). As reported below, we tested these technological hypotheses by directly comparing experimentally generated cores and debitage to the archaeological assemblages.

Lithic Analysis

To permit comparison across conditions and with the archaeological assemblages, we assessed conventional discrete and metric attributes for all detached pieces (classified as Whole Flakes [WF], Split Flakes [SF], Proximal Sections [PS], or Fragments [F]) and for cores before and after reduction. Whole flakes were assigned to technological flake categories (Toth 1987) on the basis of presence of cortex on the platform and dorsal surfaces (types I–VI: +/+, +/partial, +/−, −/+, −/partial, −/−). Core reduction intensity was estimated using the scar density index (SDI) of Clarkson (2013), calculated as (scar count/surface area [mm²]) × 10,000. In the absence of 3-D scans, we calculated surface area as a rectangular prism (2ab + 2ac + 2bc), a formula Clarkson showed to be highly predictive ($r^2 = 0.944$) of 3-D surface area. As our hypothesis concerned the production of sharp edges, we estimated the length of useful edge (<50° [Gurtov and Eren 2014]) on all detached pieces using a goniometer and calipers. Following Morgan et al. (2015) and Putt (2015), we classified any piece with useful edge >0 mm as "Viable." We also calculated two indices of utility for each piece: "Edge-to-Mass" (flake cutting edge/mass$^{1/3}$) (Braun and Harris 2003), and "Utility" (flake cutting edge/flake mass$^{1/3}$) × (1 − exp[−0.31 × (flake maximum dimension − 1.81)]) (Morgan et al. 2015). The first formula uses an exponent to account for the nonlinear relationship between edge and weight. The second formula additionally accounts for the influence of size on utility, using an estimate of this relationship derived from the subjective ratings of three expert coders. In essence, it rewards pieces for high cutting edge and penalizes them for small size (Morgan et al.

2015). In our samples, these two measures were highly correlated (Pearson's $r = 0.819$, $P < .001$) and produced qualitatively similar results (table 2).

Statistical Analyses

To evaluate the association between various possible predictors (knapping method, raw material size, shape and quality, and reduction stage) and technological outcome, we adopted an information-theoretic approach (Burnham and Anderson 2002), which is further described in the supplementary methods. We used the corrected Akaike information criterion (AICc) to rate each possible combination of predictors on the balance between goodness of fit (likelihood of the data given the model) and parsimony (number of parameters). Parameter estimates from alternative models were weighted and averaged to produce the most generalizable inference.

Experimental Results

Validating the Models

It is crucial to our analogy that the experimentally produced assemblages accurately recreate the archaeological assemblages. Here we show that this is indeed the case. Forty cobbles were reduced, producing 40 cores (18 bi-/multifacial), 3 core fragments, and 829 detached pieces >20 mm in maximum dimension. We compared this sample to excavated collections comprising a total of 32 cores and 849 detached pieces >20 mm from East Gona (EG-10 and EG-12) and Ounda Gona South (OGS-7). Our experimental conditions were designed to replicate specific reduction strategies rather than the actual archaeological samples, which reflect mixtures of cores reduced by these methods. To generate archaeological predictions, we modeled artifact type frequencies and mean metric attributes as weighted averages of the values produced by our two experimental conditions (uni- vs. bi-/multifacial: EG model = 0.72 vs. 0.28, OGS model = 0.14 vs. 0.86). Notional artifact type counts were generated from these proportions by setting the total number of artifacts equal to the archaeological samples.

Artifact type frequencies (table 1) predicted by the models closely approximate the actual archaeological samples, par-

Table 1. Artifact type frequencies

	Artifacts					
	Whole flakes (%)	Split flakes (%)	Proximal sections (%)	Fragments (%)	Cores (%)	Total (%)
Experiment:						
Unifacial	124 (27.7)	140 (31.3)	39 (8.7)	122 (27.3)	22 (4.9)	447 (100)
Bi-/multifacial	189 (44.8)	75 (17.8)	32 (7.6)	108 (25.6)	18 (4.3)	422 (100)
Experimental model:						
East Gona	199 (32.5)	168 (27.5)	51 (8.4)	164 (26.8)	29 (4.7)	611 (100)
Ounda Gona South	115 (42.4)	53 (19.7)	21 (7.8)	70 (25.8)	12 (4.4)	271 (100)
Excavated sample:						
East Gona	153 (25.0)	102 (16.7)	4 (.7)	327 (53.5)	25 (4.1)	611 (100)
Ounda Gona South	97 (35.9)	31 (11.5)	2 (.7)	133 (49.3)	7 (2.6)	270 (100)

ticularly with respect to the direction and magnitude of differences between East Gona and Ounda Gona South (fig. 5*a*). Notably, the experimental models consistently predict higher frequencies of whole flakes and identifiable fragments (SF, PS), whereas the archaeological samples contain correspondingly higher frequencies of unidentifiable fragments. The largest deviation is the overestimation (by about 11%) of split flakes from East Gona. Toth, Schick, and Semaw (2006) observed a similar pattern in their replication experiment and suggested that it likely reflects the impact of postdepositional fragmentation and/or the greater difficulty of identifying diagnostic morphology on weathered archaeological specimens. Measures of debitage size and shape support the former interpretation. We found that East Gona split flakes are absolutely larger (fig. 5*b*) and relatively thicker (fig. 5*c*) than predicted, consistent with the selective deletion of small, thin pieces by postdepositional

fragmentation. Furthermore, unidentifiable fragments have the lowest relative thickness among experimental detached pieces (i.e., reflecting the fact that relatively thin flakes are more likely to break during detachment) but the greatest relative thickness among the archaeological samples. This is expected if archaeological fragment counts were inflated substantially by postdepositional fragmentation, which typically involves snaps that reduce length and/or breadth while preserving thickness and thus systematically increase relative thickness. In contrast, such fragmentation is expected to have a random rather than systematic effect on elongation (Length/Breadth), leading to accurate prediction of the mean by the experimental models (fig. 5*d*).

With respect to core attributes, we considered measures of size (weight), shape (Length/Breadth; Breadth/Thickness), and reduction intensity (SDI). A multiple ANOVA of these four variables across our experimental and archaeological samples
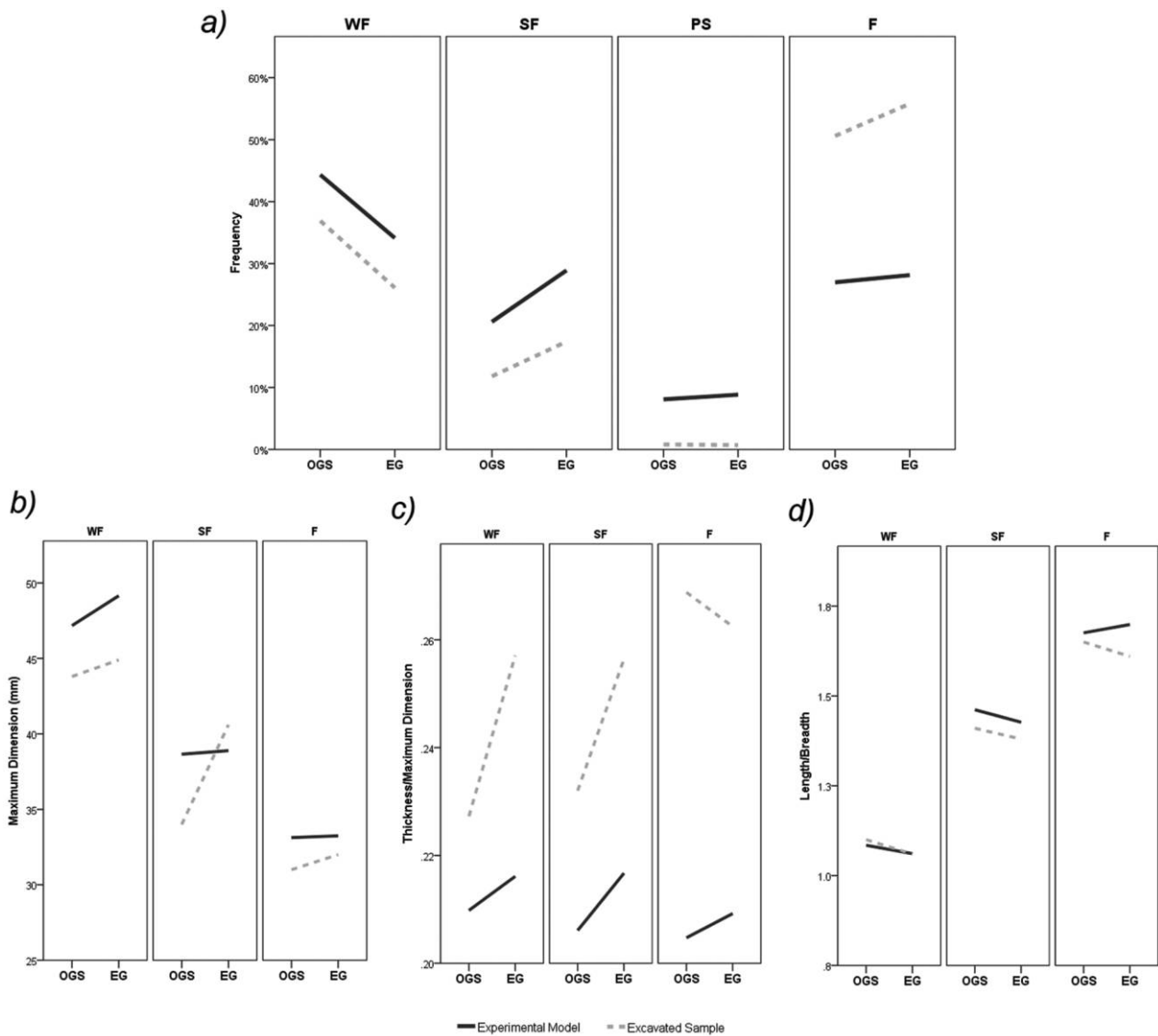

Figure 5. Experimental prediction of artifact type frequencies (*a*), size (*b*), relative thickness (*c*), and shape (*d*). A color version of this figure is available online.

found a significant effect (Wilks's λ = 0.607, $F(12, 166.974)$ = 2.892, P = .001), driven by between-samples effects on SDI (P = .001) and weight (P = .001) but not shape (Length/Breadth P = .001; Breadth/Thickness P = .292). Considering the former (fig. 6), we find that the experimental models accurately predict SDI as well as the relative core size difference between East Gona and Ounda Gona South but overestimate core size generally. This parallels general model overestimation of the size of debitage (fig. 5*b*) and may reflect somewhat larger initial cobble size in the experimental samples.

Importantly, the success of our models in predicting reduction intensity (SDI) for East Gona and Ounda Gona South argues against the possibility that differing knapping methods across these sites simply reflect different points along a continuum of reduction (cf. Toth 1985). As discussed above, both East Gona and Ounda Gona South cores appear extensively reduced (Stout et al. 2010; Toth, Schick, and Semaw 2006). Our experimental results now show that such differences as do exist between the sites match those generated by variation in knapping method across cores reduced to a uniform criterion of exhaustion. This is consistent with evidence of technological flake category (Stout et al. 2010) and debitage type frequencies.

Results thus show that our experimental models are successful in predicting features of the archaeological samples ranging from artifact type frequencies to core reduction intensity. The models are especially successful in predicting the direction and magnitude of differences between East Gona and Ounda Gona South, which are the focus of our research questions. We conclude that an analogy between the technological effects of knapping method in our experiments and in the archaeological sample is warranted.

Effects of Knapping Method

As expected, unifacial flaking produced a higher proportion of cortical platform flakes (especially type II) compared to the noncortical platform flakes (especially type V) produced by bifacial flaking (fig. 7). The difference between flake type distributions was highly significant ($χ^2$ = 220.9, df = 5, n = 307, P < .001). Unifacial flaking also produced a higher proportion of split flakes, and correspondingly fewer whole flakes, compared to bifacial flaking (table 1; $χ^2$ = 34.2, df = 3, n = 829, P < .001). As reported above, this result is echoed by the higher proportion of split flakes in the East Gona versus Ounda Gona South archaeological samples (fig. 5*a*).

Unifacial cores were less completely reduced as indicated by both the scar density index (fig. 6*a*) and the actual proportion of weight removed (table 2). As seen in table 2, apparent trends for unifacial flaking to produce fewer total pieces and fewer Viable pieces per cobble did not achieve significance. However, unifacial flaking did produce a significantly lower proportion of Viable pieces, lower total Utility (measured by summed Utility and Edge-to-Mass values for all detached pieces), and less utility per gram of original cobble (as measured by summed Utility and Edge-to-Mass divided by start weight). This suggests that unifacial flaking is generally less efficient in producing useful pieces and sharp edges from a given cobble. Results presented in tables 1 and 2 further suggest this is due to a tendency for unifacial flaking to produce a higher rate of fragmentation. Importantly, this difference in fragmentation rate is also seen across the archaeological samples. However, these results do not yet take into account possible effects of raw material variation that might result in advantages for unifacial knapping under certain conditions.

Core analysis. To address possible effects of raw material variation, we first considered productivity per core measured as Utility generated per unit of cobble starting weight. As shown in table 2, various alternative measures behave similarly. We selected this particular measure because (1) we find it to be the most theoretically justified and (2) it produced the largest effect on the sample means in table 2, and we wished to maximize the chance of discovering differences between the two knapping methods.

To analyze which factors best predict Utility/Weight (normalized through log transformation) from a given core, we considered a suite of core traits and their interactions with knapping method, resulting in the following full model, wherein each core provides one data point:
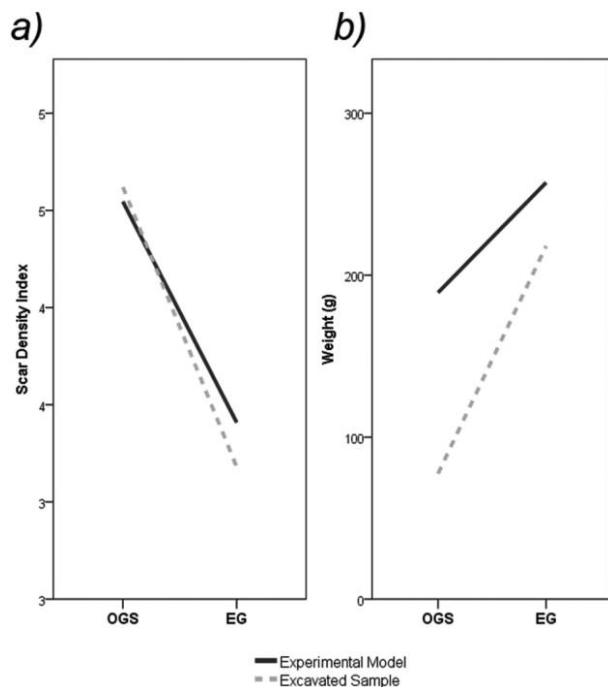

Figure 6. Experimental prediction of core reduction intensity (SDI; *a*) and size (weight; *b*). A color version of this figure is available online.

ln(Utility by Weight) ∼ Method × Raw Material Quality
+ Method × Shape PC1 + Method × Shape PC2
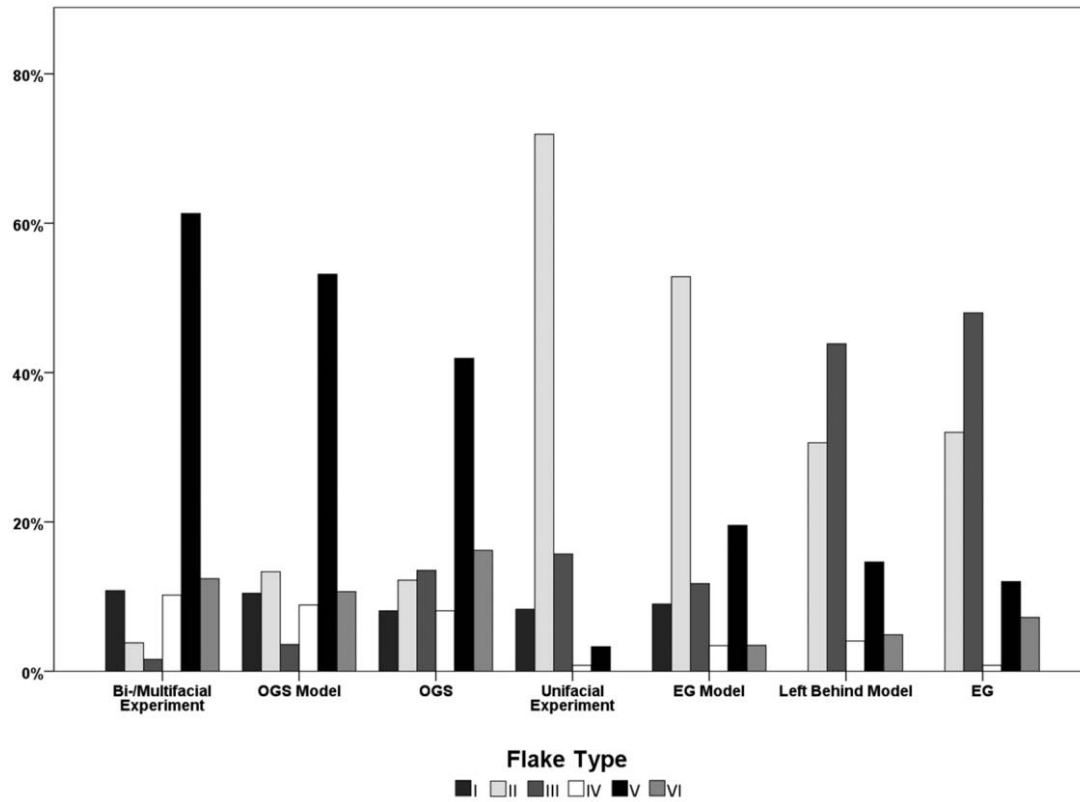+ Method × Solidity + Method × Start Angle.

Figure 7. Technological flake type (defined by presence of cortex on the platform/dorsal surfaces: type I: +/+; type II: +/partial; type III: +/−; type IV: −/+; type V: −/partial; type VI: −/−) frequencies for experimental and archaeological samples, including modeled modifications of the experimental samples.

Continuous predictors (Shape PCs [PC = principal component analysis], Solidity, and Start Angle) were centered such that zero represents the sample average, and units are standard deviations. The full model was fit with the lm function in *R* 3.2.3, and the MuMIn package (Bartoń 2015) was used for automated model comparison. The average model is presented in table 3, wherein baseline refers to Method = unifacial, Raw Material Quality = glassy, with all continuous predictors at the sample average. The other parameters represent deviations from this baseline; for example, the predicted Utility for Method = bi-/multifacial is −3.72 + 0.42. The parameter estimates for the continuous predictors reflect the expected change in utility for 1 standard deviation change in the predictor variable. Significant decreases in utility were found for cores of lower quality and higher start angle but not for the two knapping methods, regardless of core quality, start angle, shape, or solidity. Keeping all continuous predictors at the sample mean, figure 8A shows the predicted effects (±95% CI; CI = confidence interval) of flaking method and raw material quality on Utility, wherein point size and line width are proportional to the number of cores in each category. There are no significant differences in Utility produced by unifacial and bi-/multifacial methods at any level of quality. Knapping method similarly had no effect on the relationship between Start Angle and Utility

Table 2. Reduction statistics by core

	Proportion weight removed	Total pieces produced	Viable pieces produced	Proportion viable	Sum of utility values	Utility/start weight	Sum of edge/mass values	(Edge/mass)/start weight
Unifacial (*n* = 22):								
Mean	.51	19.4	16.0	.80	10.13	.018	222.15	.42
Median	.56	19.0	15.5	.84	8.64	.018	194.09	.34
Bi-/multifacial (*n* = 18):								
Mean	.66	22.4	19.83	.88	15.67	.032	341.58	.69
Median	.67	20.0	18.0	.89	13.67	.026	287.96	.53
t-test sig. (2-tailed)	.013	.327	.192	.012	.028	.012	.049	.042
Mann-Whitney *U*	.039	.338	.180	.010	.022	.011	.032	.014

Table 3. Average model predicting core utility/weight

	Estimate	Adjusted SE	Z	P	Lower 95% CI	Upper 95% CI
Baseline (unifacial; glassy)	−2.91	.64	4.54	<.001	−4.16	−1.65
Deviation from baseline:						
Method = bi-/multifacial	−.04	.71	.06	.95	−1.44	1.36
Quality = smooth	−1.26	.65	1.93	.053	−2.55	.02
Quality = coarse	−1.58	.73	2.17	.030	−3.01	−.16
Start angle (+1 SD)	−.32	.1	2.82	.005	−.55	−.10
Shape PC1 (+1 SD)	.03	.08	.34	.73	−.13	.18
Shape PC2 (+1 SD)	−.01	.05	.12	.91	−.10	.09
Solidity (+1 SD)	.02	.06	.35	.73	−.10	.15
Method = bi-/multifacial; quality = smooth	.47	.74	.64	.52	−.98	1.92
Method = bi-/multifacial; quality = coarse	.57	.87	.65	.52	−1.15	2.28
Method = bi-/multifacial; start angle (+1 SD)	−.02	.10	.24	.81	−.21	.17
Method = bi-/multifacial; shape PC1 (+1 SD)	−.01	.06	.13	.90	−.12	.11
Method = bi-/multifacial; shape PC2 (+1 SD)	−.01	.06	.15	.88	−.12	.10
Method = bi-/multifacial; solidity (+1 SD)	.00	.04	.03	.97	−.08	.07

Note. Cores of lower quality and with greater start angles produce lower utility, but utility does not vary according to knapping method.

(fig. 8*B*) and did not interact with cobble shape PCs or solidity. Indeed, model estimates for all these possible shape effects display narrow confidence intervals centered near 0, with all $P >$ .80. These experimental results thus fail to provide any suggestion that local variation in raw material availability/selection could have favored one knapping method over another.

Debitage viability. We next considered the production of minimally Viable flakes. To analyze what factors—knapping method, material traits, and their interactions—best explained whether a given flake was Viable or not, we began with the following full model:

$$\text{Viable(yes/no)} \sim \text{Method} \times \text{Raw Material Quality}$$
$$+ \text{Method} \times \text{Debitage Type} + \text{Raw Material Quality}$$
$$\times \text{Debitage Type} + \text{Cobble Weight} + \text{Detachment Number}$$
$$+ \text{Method} \times \text{Shape PC1} + \text{Method} \times \text{Shape PC2}$$
$$+ \text{Method} \times \text{Solidity} + \text{Method}$$
$$\times \text{Start angle} + \text{random effect for Cobble.}$$

Again, all continuous predictors were centered. The full model was fit using the MCMCglmm package (Hadfield 2010) in *R* 3.2.3 with a logit link function, slice sampling, a weak

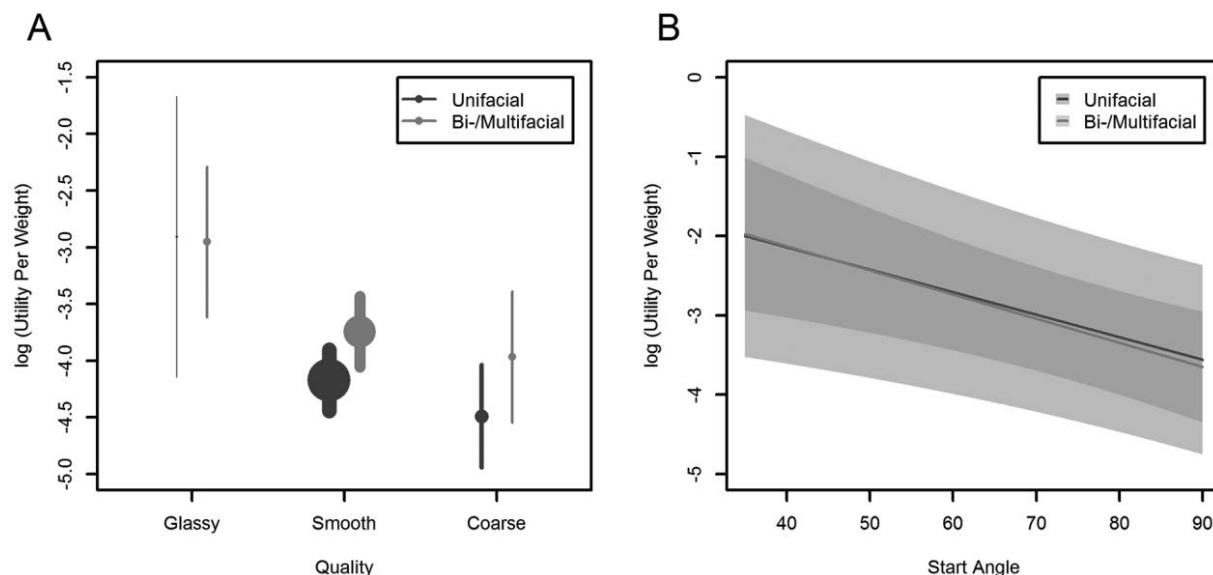
Figure 8. Significant effects of raw material quality (*A*) and start angle (*B*) on utility generated per unit cobble weight. Knapping methods had no significant interaction with either effect. A color version of this figure is available online.

Cauchy prior (variance = 1,000) for the random effect, and residual variance fixed to 10 (Hadfield 2014). Chains were run for 50,000 iterations and sampled at every tenth iteration after a burn-in of 10,000; convergence was satisfactory as assessed visually.

The cobble-level random effect contributed relatively little to explaining viability; compared to the same model without the cobble-level random effect, the full model had essentially the same AICc (371.5 vs. 371.2; comparison of Deviance Information Criteria yielded the same inference); hence we proceeded without the random effect and refit the model using the glm function. The absence of a substantial cobble-level random effect provides reassurance that our selection of cobble form and quality variables has not neglected major sources of variation relevant to Utility.

In the averaged model (table 4), baseline refers to Method = unifacial, Debitage Type = whole flakes, Quality = glassy, with continuous predictors at the sample average. All other rows represent deviations from this baseline, and all parameters are presented as the probability of producing a Viable flake. Debitage Types other than whole flakes significantly lower the probability of a piece being Viable. There was a trend toward higher Viability with increasing shape PC2 (i.e., less elongated cobbles, $P = .092$) and some suggestion of a small general advantage for bi-/multifacial flaking ($P = .244$). It is plausible to speculate that these are real effects that might achieve significance with a sufficiently large sample. Critically, however, there was again no suggestion of an interaction between knapping method and any raw material shape or quality variable (P range = .65–.83) regardless of debitage type ($P = .82$–1.00).

Figure 9*A* presents a visualization of the probability of producing a Viable piece (± 95% CI) as a function of debitage type and knapping method. The probability declines with each change in debitage type, and at each type bi-/multifacial knapping has a slightly higher probability of Viability. Our Viability results indicate that the archaeologically observed variation in knapping method is either functionally neutral or provides a small overall advantage for bi-/multifacial knapping regardless of raw material and debitage type variation. Thus, there is no suggestion of any local conditions that could have favored individual adoption of unifacial flaking at East Gona.

Utility of viable debitage. Finally, we considered variation in the Utility of Viable pieces. The Utility distribution is highly

Table 4. Average model predicting the probability that detached pieces are viable

	Estimate	Adjusted SE	Z	P	Lower 95%	Upper 95%
Baseline (unifacial; glassy; whole flakes)	2.64 (.93)	.41	6.43	<.001	1.83	3.44
Deviation from baseline:						
Method = bi-/multifacial	.53 (.96)	.46	1.17	.244	−.36 (.91)	1.43 (.98)
Debitage = split flakes	−.86 (.86)	.43	2.00	.046	−1.71 (.75)	−.02 (.92)
Debitage = proximal sections	−1.24 (.80)	.54	2.28	.022	−2.31 (.59)	−.18 (.92)
Debitage = fragments	−1.88 (.68)	.40	4.66	<.001	−2.66 (.52)	−1.09 (.81)
Shape PC1 (+1 SD)	−.06 (.93)	.13	.48	.63	−.32 (.85)	.19 (.97)
Shape PC2 (+1 SD)	.23 (.95)	.14	1.69	.092	−.04 (.89)	.50 (.98)
Solidity (+1 SD)	−.16 (.92)	.15	1.04	.30	−.46 (.84)	.14 (.97)
Core weight (+1 SD)	−.07 (.93)	.11	.61	.54	−.28 (.85)	.15 (.97)
Detachment (+1 SD)	.05 (.94)	.10	.54	.59	−.14 (.87)	.25 (.97)
Start angle (+1 SD)	.04 (.94)	.11	.39	.70	−.17 (.87)	.26 (.97)
Quality = smooth	−.01 (.93)	.37	.03	.97	−.74 (.88)	.71 (.96)
Quality = Coarse	−.03 (.93)	.47	.06	.95	−.95 (.84)	.90 (.97)
Method = bi-/multifacial; solidity (+1 SD)	−.09 (.95)	.19	.45	.65	−.47 (.88)	.29 (.98)
Method = bi-/multifacial; shape PC2 (+1 SD)	−.05 (.97)	.15	.35	.73	−.34 (.92)	.24 (.99)
Method = bi-/multifacial; start angle (+1 SD)	−.03 (.96)	.12	.26	.80	−.26 (.91)	.20 (.98)
Method = bi-/multifacial; shape PC1 (+1 SD)	.02 (.96)	.11	.22	.83	−.20 (.90)	.25 (.98)
Method = bi-/multifacial; quality = smooth (+1 SD)	−.11 (.95)	.42	.26	.79	−.92 (.92)	.71 (.97)
Method = bi-/multifacial; quality = rough (+1 SD)	−.10 (.95)	.40	.24	.81	−.88 (.89)	.69 (.98)
Quality = smooth; debitage = split flakes	.07 (.86)	.36	.19	.85	−.64 (.80)	.77 (.91)
Quality = coarse; debitage = split flakes	−.04 (.85)	.40	.09	.93	−.82 (.75)	.75 (.91)
Quality = smooth; debitage = proximal sections	.01 (.80)	.36	.04	.97	−.70 (.66)	.73 (.89)
Quality = coarse; debitage = proximal sections	−.22 (.76)	.95	.23	.82	−2.07 (.46)	1.63 (.92)
Quality = smooth; debitage = fragments	.07 (.69)	.34	.19	.85	−.61 (.60)	.74 (.77)
Quality = coarse; debitage = fragments	−.03 (.67)	.38	.09	.93	−.78 (.54)	.71 (.78)
Method = bi-/multifacial; debitage = split flakes	.00 (.91)	.14	.00	1.00	−.27 (.81)	.27 (.96)
Method = bi-/multifacial; debitage = proximal sections	.00 (.87)	.18	.01	.99	−.35 (.68)	.35 (.96)
Method = bi-/multifacial; debitage = fragments	.01 (.79)	.12	.04	.97	−.24 (.64)	.25 (.88)

Note. Fragments have a significantly lower probability of producing viable pieces, independent of knapping method. Note that parameter estimates are in logit space, and transformed values in natural probability units are given in parentheses. For instance, the probability of a glassy whole flake producing a viable flake with bi-/multifacial knapping is the logistic of (2.64 + .53) = .96.
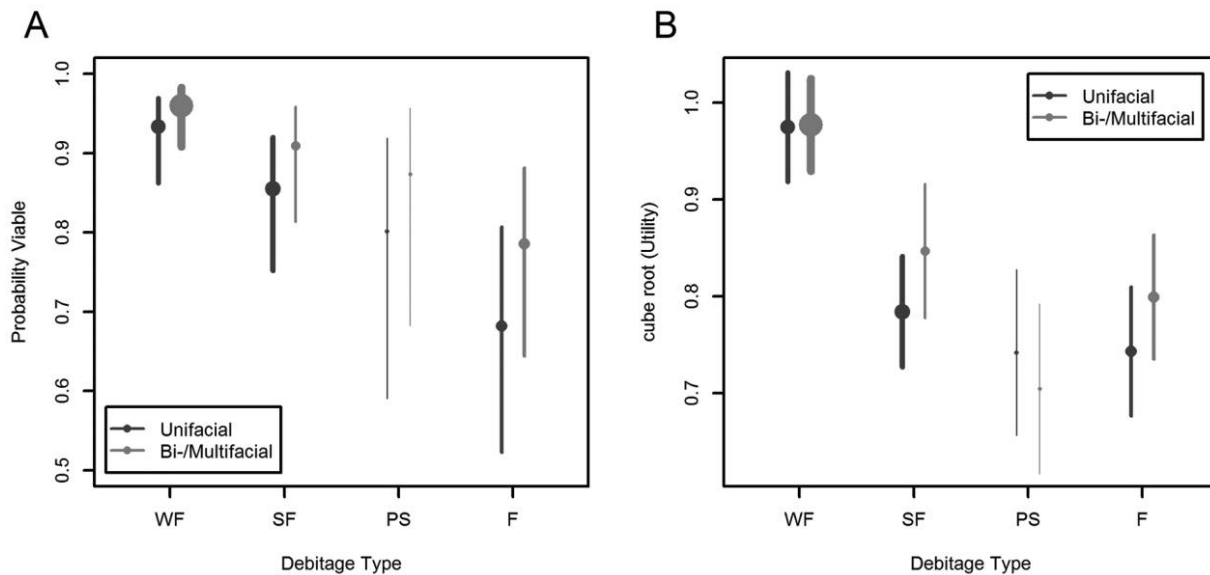
Figure 9. Significant effects of debitage type on viability (*A*) and utility (*B*). There were no significant effects of or interactions with knapping method. A color version of this figure is available online.

skewed; exploratory analyses showed it to be most effectively normalized through a cube root transformation (Shapiro-Wilk = 0.997, df = 709, P = .177). We began with the same full model. The cobble-level random effect again contributed very little to explaining variation in Utility; the same model fit without the random effect had a substantially lower AICc (−145 vs. −309), and we focus on this simpler model from now on.

In the averaged model (table 5), debitage types other than whole flakes produced lower Utility. There were trends for heavier (P = .081) and higher quality (P = .085) cores to produce pieces with higher Utility and for higher start angles (P = .053) to result in lower Utility (fig. S3). There was, however, no indication of an effect of knapping method (P = .94) nor any suggestion of an interaction with raw material variables (P = .49–.99). These results once more indicate that the technological variation at Gona was functionally neutral and independent of any effects of hypothetical variation in local raw materials.

"Missing" flakes and hominin preferences. As shown in figure 7, our experimental model yielded a very close approximation of the actual Ounda Gona South flake category distribution. This supports the validity of our experimental model as well as our previous (Stout et al. 2010) conclusion that there is an equal representation of all reduction stages at OGS-7.

In contrast, the experimental model underestimates the frequency of type III flakes in the East Gona sample. A previous replication experiment by Toth, Schick, and Semaw (2006) produced the same discrepancy, which the authors successfully

modeled as a product of predominantly late-stage reduction combined with the selective removal of the most useful (based on expert appraisal) flakes. Our experiment is limited by a relatively small sample size for late-stage unifacial whole flakes but similarly suggests that the relative frequency of type III flakes increases in later reduction (fig. S4). In contrast to Toth, Schick, and Semaw (2006), however, we found that accounting for this effect failed to recreate the greater frequency of type III versus type II flakes and dramatically overestimated the representation of type V flakes (26% vs. 12%). We thus considered selective removal as an alternative explanation.

To generate an objective criterion for preferential removal of type II versus type III flakes, we ran a stepwise discriminant function analysis (DFA) on the unifacial experimental sample with flake type (II vs. III) as the grouping variable and a range of size and shape measures as the dependent variables (Weight, Maximum Dimension, Length, Breadth, Thickness, Edge Length, Platform Thickness, Platform Breadth, and Utility, as well as Length, Breadth, and Thickness divided by their geometric mean). The resulting function (0.042 × Length + 0.186 × Thickness − 0.087 × Platform Breadth − 1.749), hereafter referred to as "Favor," characterizes type II flakes as absolutely longer and thicker, with narrow platforms. It successfully classifies 77% of cases. Applying the function to all experimental flakes and using −1.0 (visually selected from distributions) as the cut point above which flakes are removed results in the deletion of 79% of whole flakes and produces a "left behind" model quite similar to the distribution observed at East Gona (fig. 7). Importantly, this also produces more accurate estimates of types I and V flake representations, despite the fact that our function was built to

Table 5. Average model predicting the utility of viable detached pieces

	Estimate	Adjusted SE	Z	P	Lower 95%	Upper 95%
Baseline (unifacial; glassy; whole flakes)	.97	.03	33.70	<.001	.92	1.03
Deviations from baseline:						
Method = bi-/multifacial	.00	.03	.07	.94	−.05	.06
Quality = smooth	−.02	.03	.69	.49	−.07	.03
Quality = coarse	−.07	.04	1.72	.085	−.15	.01
Core weight (+1 SD)	.02	.01	1.75	.081	.00	.04
Debitage = split flakes	−.19	.03	6.33	<.001	−.25	−.13
Debitage = proximal sections	−.23	.04	5.85	<.001	−.31	−.16
Debitage = fragments	−.23	.03	6.92	<.001	−.30	−.17
Detachment (+1 SD)	−.01	.01	1.21	.23	−.03	.01
Start angle (+1 SD)	−.02	.01	1.93	.053	−.05	.00
Shape PC2 (+1 SD)	.00	.01	.56	.58	−.02	.01
Solidity (+1 SD)	.00	.01	.32	.75	−.02	.01
Shape PC1 (+1 SD)	.00	.01	.01	.99	−.01	.01
Method = bi-/multifacial; debitage = split flakes	.06	.05	1.29	.20	−.03	.15
Method = bi-/multifacial; debitage = proximal sections	−.04	.05	.73	.47	−.15	.07
Method = bi-/multifacial; debitage = fragments	.05	.04	1.22	.22	−.03	.14
Method = bi-/multifacial; start angle (+1 SD)	−.01	.02	.69	.49	−.04	.02
Method = bi-/multifacial; shape PC2 (+1 SD)	.00	.01	.25	.81	−.02	.01
Method = bi-/multifacial; solidity (+1 SD)	.00	.01	.24	.81	−.02	.01
Method = bi-/multifacial; quality = smooth	.00	.02	.07	.94	−.03	.03
Method = bi-/multifacial; quality = coarse	.00	.02	.20	.84	−.04	.05
Method = bi-/multifacial; shape PC1 (+1 SD)	.00	.01	.01	.99	−.01	.01
Quality = smooth; debitage = split flakes	.00	.01	.14	.89	−.03	.03
Quality = coarse; debitage = split flakes	.00	.01	.00	1.00	−.02	.02
Quality = smooth; debitage = proximal sections	.00	.02	.12	.90	−.03	.03
Quality = coarse; debitage = proximal sections	.00	.03	.09	.93	−.05	.05
Quality = smooth; debitage = fragments	.00	.01	.13	.90	−.03	.03
Quality = coarse; debitage = fragments	.00	.02	.14	.89	−.04	.04

Note. Debitage types other than whole flakes, cores of coarse quality, heavier weight, and greater start angle have lower utility.

distinguish types II versus III only. This suggests that Favor really is capturing hominin-valued flake qualities independent of technological flake categories.

To see if Favor could produce an otherwise unrecognized advantage for one or another knapping method, we began with the same full model employed in the analysis of debitage Utility. The cobble-level random effect again contributed relatively little to explaining variation in Favor; the same model fit without the random effect had a substantially lower AICc (891 vs. 865), and we focus on this simpler model from now on.

Neither knapping methods nor any material traits or their interactions significantly influenced a flake's Favor, although later-stage detachments exhibit a trend toward lower Favor (*P* = .79; table S1, available online). The Favor function includes Platform Breadth, which is not relevant to debitage types other than Whole Flakes. However, Split Flakes, Proximal Sections, and Fragments all tend to have lesser lengths and thicknesses than Whole Flakes (by post hoc Tukey tests, all *P* = .001) and presumably would have been less preferred on that basis. As we have seen, they also have lower Utility and Viability rates (fig. 9). Thus, there is little to suggest that types other than whole flakes would have been preferred products, and again no indication of functional difference between knapping methods.

Discussion

Results of our experiment do not support the hypothesis that the reproduction of particular behavioral forms (knapping methods) at Gona was accomplished solely through individual learning. Explaining the observed pattern of within-site homogeneity and between-site difference in terms of repeated individual rediscovery of particular technical solutions appropriate to different contexts would require that the different knapping methods actually yield differential benefits across these contexts. To test this expectation, we conducted a wide range of comparisons between knapping methods, taking into account variation in raw material size, shape, and quality. An absence of substantial cobble-level random effects indicates that our raw material variables successfully capture the variation relevant to our outcome metrics, which include theory-driven measures of reduction intensity and detached piece utility as well as a data-driven measure of Favor derived from the apparent preference of Gona toolmakers for long, thick flakes.

Contradicting the expectations of the individual-learning hypothesis, we found no significant differences between the two knapping methods for any outcome, regardless of any measurable features of the material. While this inference is of course

contingent on our sample and statistical models, it is important to reiterate that information-theoretic model comparison and averaging was designed to maximize out-of-sample predictive accuracy, and thus generalizability, of statistical models (Burnham and Anderson 2002). This procedure makes the best use of the available data, and our inferences of no difference between the two methods are not based on a single, arbitrarily selected model that might fail to control for crucial predictors. If anything, our averaged models generally predicted small advantages for bi-/multifacial knapping irrespective of raw material variation, but with 95% confidence intervals that did not exclude values of no difference between the two methods (see figs. 8, 9). While it is possible that a future study with a larger sample would identify a significant advantage for bi-/multifacial knapping in general, our results contain no suggestion of any interaction between knapping methods and raw material variation of the kind that would be needed to explain intersite variation at Gona.

We did not collect data on energy expenditure or productivity per unit time; however, we have assessed efficiency in terms of utility per gram of raw material (relevant to transport and handling costs including time and effort) and the proportion of Viable pieces generated (i.e., rate of payoff from effort invested in percussion), both of which fail to provide any evidence of unifacial advantage. Furthermore, at the level of individual knapping actions, unifacial flaking produced an average of only 0.83 viable flakes and 0.524 total utility per blow, whereas bi-/multifacial flaking produced 0.92 and 0.738, respectively. Any putative unifacial time and effort savings from minimizing the (relatively rapid) process of core rotation thus seem likely to be offset by greater costs in terms of raw material economy and a lower rate of useful flake generation per blow.

Our results thus offer some mixed support for the possibility of a small across-the-board advantage for bi-/multifacial flaking, but they fail to provide any basis for explaining the predominance of unifacial flaking at two out of three Gona sites, in terms of reliable individual rediscovery of a locally advantageous technical solution. While we would certainly not argue that Oldowan technology was beyond the ability of an individual to reinvent in a lifetime, we nevertheless contend that the observed recurrence of functionally underdetermined knapping variants indicates social reproduction of these particular action details.

There are three caveats to this conclusion. First, we do not actually know how many individuals contributed to each assemblage (see Claudio Tennie's reply in Tennie et al. 2017). However, even if we posit a minimal group size of two and consider only two alternative flaking methods as possible (i.e., 50-50 chance, equivalently conceptualized as 100% individual commitment to one or another method or an unbiased distribution of individual mixtures of methods), we find that the observed frequencies are unlikely to occur by chance (OGS-7: $z = 0.901$, $P = .18$; EG-10: $z = 0.536$, $P = .30$; EG-12: $z =$

0.79, $P = .21$). Furthermore, the likelihood of observing all three sites together is only 1% ($0.18 \times 0.30 \times 0.21 = 0.011$). Increasing the number of individuals or recognizing that alternative reduction strategies reported from other Oldowan sites were possible solutions despite being poorly represented at Gona (Stout et al. 2010) only serves to make the observed pattern even more unlikely (e.g., for $n = 4$ individuals: $0.10 \times 0.22 \times 0.13 = 0.003$). We thus conclude that individual toolmakers at Gona typically reproduced not only the basic means (hard-hammer percussion) and overall outcome (efficient flake production; see discussion below) of knapping but also particular methods for achieving this outcome.

Second, our conclusion depends on an analogy between our modern experiments and ancient knapping at Gona. Such analogical inference, though fallible, plays an essential role in archaeology and can be rendered more reliable through appropriate design and validation (Eren et al. 2016; Wylie 1985). We validated our experimental models against the archaeological samples, finding that they successfully predicted assemblage characteristics ranging from flake fragmentation rates to core reduction intensity. Both positive and negative (dis-)analogies were observed and in each case were explicable in terms of known casual relations (e.g., excessive force and fragmentation, differential postdepositional breakage). Furthermore, our analyses of the experimental samples indicated that the variables most strongly influencing Utility across methods (core reduction intensity, flake fragmentation rate) are precisely those that our models most successfully predicted in the archaeological samples. We conclude that the experimental analogy is warranted.

Finally, we have employed an argument by elimination. As a form of inductive inference, the progressive elimination of alternatives can only ever lead to increasing confidence, not certainty. Indeed, it is logically impossible to test or even to enumerate every possible alternative condition or objective that might provide an advantage for unifacial flaking. For example, it is possible that East Gona toolmakers were motivated to maximize the production of split flakes for an unknown reason and that this resulted in preference for the fracture-prone unifacial method. However, our experiment did exhaustively test implications of the current consensus that Oldowan technology sought to maximize the production of cutting edge. Pending a major revision of this consensus, we assign a high level of confidence to our technological conclusions. Similarly, it remains possible that intersite differences are explained by biological differences in the anatomy or psychology of the toolmakers. As discussed above, however, there is no independent evidence to suggest this possibility and it requires a wide array of untested auxiliary assumptions. Last, there is the possibility that various different knapping methods already existed as evolved tendencies in the motor repertoires of Gona hominins and were simply released through social response facilitation (e.g., like yawning) rather than acquired through copying. Although something similar has recently been proposed for

Acheulean technology (Corbey et al. 2016), it is difficult to see how multiple, highly specific yet functionally neutral, alternative behavioral programs could have been constructed by natural selection in the earliest known Oldowan knappers. Perhaps the most plausible mechanism would be some form of Baldwinian genetic assimilation (Stout and Hecht 2017; Weber and Depew 2003), but this would itself presuppose an earlier stage in which behaviors were learned rather than innate. In sum, we thus consider social learning to be the most convincing and parsimonious explanation for the evidence currently available. Researchers will, however, reasonably differ on the degree of confidence they require to accept the presence of detailed behavior copying in the Oldowan, in part due to broader theoretical commitments reviewed in the introduction. To promote debate on this important topic, we will (briefly) develop our interpretation of the evidence and speculate about its evolutionary implications.

Interpretation and Implications

Even "simple" flake production is a challenging perceptual-motor skill that must be acquired through practice (Nonaka, Bril, and Rein 2010). For many modern humans, basic competence can be achieved within a few hours (Putt 2015; Stout and Khreisheh 2015; Stout and Semaw 2006), but greater expertise is required to reproduce the more patterned and exhaustive flaking seen at many Oldowan sites (Stout et al. 2008). It has been argued that core maintenance techniques required for such flaking are difficult and subtle enough that their reliable reproduction required intentional demonstration and imitation (Gärdenfors and Högberg 2017), and there is experimental evidence that teaching facilitates rapid learning of basic knapping (Morgan et al. 2015). However, an alternative view is that core maintenance is enabled by the same basic skills that allow relatively large individual flake removals (Moore 2011; Stout and Chaminade 2007) and thus is readily discoverable over realistic practice times. The latter is consistent with current experimental results, which successfully modeled the OGS-7 assemblage using a least-effort flaking strategy (see below), but it remains to be experimentally demonstrated how reliably motivated modern humans would rediscover effective core maintenance given longer practice times.

In contrast, evidence available from apes and monkeys indicates that they can produce examples of Oldowan-like artifacts (including by accident; Proffitt et al. 2016), but have yet to generate Oldowan-like assemblages indicative of skilled flaking and core maintenance even given years of practice, demonstration, motivation, assistance, and feedback (Toth, Schick, and Semaw 2006). Humans, meanwhile, have relatively little difficulty mastering chimpanzee-like nut cracking (Bril, Parry, and Dietrich 2015), and 2- to 3.5-year-old children take less than 3 minutes to spontaneously solve tasks approximating wild chimpanzee and orangutan tool-use behaviors (Reindl

et al. 2016). This is particularly striking given that children up to 8 years of age are very poor at tool innovation relative to adults (Cutting et al. 2014). We might reasonably expect Plio-Pleistocene hominins to fall somewhere between these extremes in aptitude, which raises the intriguing possibility that knapping skills easily reinvented by modern human adults might not have been so obvious for early Oldowan toolmakers at Gona (Tennie et al. 2016).

Tool making at OGS-7, as successfully modeled by our experiments, corresponds to what Early Stone Age archaeologists call least-effort flaking: the efficient production of flakes through hard-hammer percussion without stylistic constraint or strategic elaboration. Such ad hoc knapping is a simple process of finding a viable platform, removing a flake, and then repeating the procedure. It is generally regarded as a null model (Reti 2016) for the most obvious and natural form of flaking. In fact, when a modern knapper identifies all viable platforms on a core, even complete randomization of actual platform selection will typically result in a bifacial flaking pattern (Moore and Perston 2016). All of this, however, assumes human-like capacities for the perception and exploitation of core affordances. Least-effort flaking is demonstrably not an obvious solution for modern bonobos (Toth, Schick, and Semaw 2006) and may not have been for the Gona toolmakers either. Indeed, least-effort knapping by modern humans is known to be especially demanding of visual-motor and attentional systems (Stout and Chaminade 2007) that have evolved substantially since the human-chimpanzee split (Stout and Hecht 2017). Inverting the widespread assumption that imitation is more psychologically and phylogenetically advanced than individual learning, early Oldowan hominins may actually have been better at copying such behaviors than rediscovering them independently.

This would be consistent with evidence that chimpanzees will copy large-scale body movements and object manipulations with sufficient fidelity to maintain stable behavioral traditions (Whiten and van de Waal 2016, but see Tennie, Call, and Tomasello [2012] for an alternative view) but appear to have a limited capacity and/or willingness to search for nonobvious functional properties of objects (Povinelli and Frey 2016). In a helical curriculum, copying actions (e.g., core rotation and inspection) in this way can provide valuable aid to discovery of object affordances, casual relations, and task structure. Importantly with respect to the critique of Tennie, Call, and Tomasello (2012), we suggest that substantively new actions can thus be learned by iterative practice and refinement of a loose initial approximation constructed entirely from familiar action elements applied to a new context (Byrne 1999; Stout 2013). In fact, it seems likely that all nontrivial skills rely to some degree on this kind of iterative learning and refinement.

This suggests an evolutionary scenario similar to those presented by Pradhan, Tennie, and van Schaik (2012) and Henrich (2015), in which early Oldowan technology was located at or near the limits of contemporary hominin innovative capacities

and thus was heavily reliant on socially mediated learning, including (as we have argued here) reproduction of detailed behavioral means. This implies high learning costs (Morgan 2016) in terms of time, effort, and possible failure of skill acquisition. These costs would impose a ceiling on the ratcheting of technological complexity (Pradhan, Tennie, and van Schaik 2012) and reduce chances of reinvention if the skill were stochastically lost in small and poorly connected hominin communities (Powell, Shennan, and Thomas 2009). However, if social learning nevertheless did maintain toolmaking in a population (and toolmaking carried fitness benefits), high learning costs would then generate selective pressure favoring cognitive or perceptual-motor adaptations to make social learning easier or more reliable (Morgan 2016). To the extent that such adaptations are also expected to enhance individual learning (Heyes 2012, and review in Stout and Hecht [2017]), this would decrease obligate reliance on social inputs and make individual rediscovery, modification, and innovation of technological behaviors more likely.

Such a biocultural feedback process would make sense of the fact that Oldowan tools predate evidence of significant brain expansion by hundreds of thousands of years, during which their occurrence was extremely patchy, discontinuous, and lacking in evidence of progressive change (Antón, Potts, and Aiello 2014; Plummer and Bishop 2016). It is only after approximately 2.0 Ma that Oldowan toolmaking becomes geographically widespread, with sites that occur in a wider variety of habitats, are more temporally persistent, and display both longer raw material transport distances and a more regular appearance of least-effort flaking methods (Plummer and Bishop 2016; Stout et al. 2010). This is closely coincident with the first appearance of larger-brained and larger-bodied *Homo erectus* by about 1.9 Ma and is rapidly followed by the invention of early Acheulean technology by 1.76 Ma (Antón, Potts, and Aiello 2014). It is not known what occasioned this shift in tempo at ca. 2.0 Ma. An enhancement to social learning is one obvious possibility, as are externally driven habitat shifts (Antón, Potts, and Aiello 2014; Henrich 2015). As an alternative or addition to these hypotheses, we suggest that already existing capacities for high-fidelity technological reproduction at 2.6 Ma supported cultural niche construction (Laland et al. 2015) that eventually stimulated Baldwinian behavior-led (e.g., Hecht et al. 2014) biological evolution of enhanced cognitive or perceptual-motor capacities for skill acquisition (Henrich 2015). By decreasing the cost of maintaining Oldowan technology, this would have expanded the range of ecological contexts in which it provided a net energetic benefit. By enhancing individual learning and innovative potential, it would also have enhanced technological persistence, flexibility, and innovation leading to further change and adaptation. One prediction of this hypothesis is that Oldowan technology should display a temporal trend toward increasing within-assemblage technological diversity and flexible adaptation of knapping methods across a greater range of raw materials. This prediction has some initial support (e.g., Braun et al. 2009) but requires further investigation.

Conclusion

In our view, the origins of human cumulative culture are best approached from an agnostic stance that does not presuppose continuity or discontinuity nor assume the necessity of particular learning types. To this end, we advocate a stepwise research program that disentangles archaeological evidence of reproductive fidelity from the identification of learning processes and the explanation of larger-scale patterns of change. This program emphasizes the real-world complexity of skill learning and the fact that rates and patterns of cultural evolution are historically contingent products of dynamic interactions between diverse influences. It thus calls for a bottom-up approach grounded in detailed empirical study of particular archaeological cases rather than top-down interpretation of broad syntheses. This further implies that experimental studies of technological reproduction (Morgan et al. 2015; Putt, Woods, and Franciscus 2014; Schillinger, Mesoudi, and Lycett 2015) will be most profitable if they employ behavioral models that have been validated against actual archaeological occurrences.

To demonstrate this approach, we presented a case study of technological variation at Gona employing an archaeologically validated experimental model to test predictions about social learning processes. This provided support for the view that copying of detailed knapping methods was already a feature of Oldowan technological reproduction at ca. 2.6 Ma. Such copying was not, however, associated with evidence of increased rates of technological change or ratcheting, suggesting that explanations for the emergence of human cumulative culture should not focus narrowly on social learning mechanisms but should instead address a range of factors also including individual cognitive and perceptual-motor capacities, levels of sociability, the internal dynamics of cultural evolution, and biocultural coevolutionary processes. In particular, we suggested a scenario in which cultural niche construction enabled by socially supported skill acquisition led to a new selective context favoring cognitive and perceptual-motor adaptations for enhanced learning that then fostered further technological flexibility and innovation.

Acknowledgments

Comments

Gili Greenbaum, Noa Truskanov, Nicole Creanza, Shimon Edelman, Marcus W. Feldman, and Oren Kolodny
Department of Biology, Stanford University, Stanford, California 94305, USA (gili.greenbaum@gmail.com, orenkolodny@gmail.com; Greenbaum, Feldman, Kolodny)/Institute of Biology, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland (Truskanov)/Department of Biological Sciences, Vanderbilt University, 465 21st Avenue South, Room 4280A, Nashville, Tennessee 37240, USA (Creanza)/Department of Psychology, Cornell University, 232 Uris Hall, Ithaca, New York 14853-7601, USA (Edelman)/Department of Ecology, Evolution and Behavior, Hebrew University of Jerusalem, Jerusalem, Israel (Kolodny). 3 X 18

Separating Social Learning from Technical Skill Provides a New Perspective on the Record of Hominin Tool Use

Stout, Rogers, Jaeggi, and Semaw conduct a thorough, rigorous, and convincing experimental exploration of two major methods of Oldowan tool production. They conclude that at three 2.6-million-year-old sites, toolmaking was not invented repeatedly and independently by individuals but, rather, was most likely acquired through learning from conspecifics. In our view, the primary importance of the study is in its tour de force demonstration of the applicability of the experimental paradigm to the study of prehistoric hominin use of stone tools.

One way to place this work in a broader context would be to appraise its goal from the perspective of evolutionary biology. The authors examine variation in toolmaking products at several sites and ask whether tools within each site are similar due to homology in their method of production—the methods used by different knappers share a common origin—or analogy—the method was reinvented independently by each knapper in response to common selective pressures. Their study compares two knapping methods in detail and finds that neither seems better suited to local conditions, suggesting homology as a parsimonious explanation. In evolutionary terms, Stout et al. argue that the observed variation in knapping methods is selectively neutral, by showing that unifacial and bifacial knapping led to similarly effective products at a similar duration of time and expenditure of effort when using raw material from any of the sites. They then evaluate the relative likelihood of phylogenies of toolmaking methods that represent convergent evolution and those that represent common ancestry. In showing that the stone tool variation is selectively neutral, Stout et al. also demonstrate by extension that this variation could be informative about the phylogenetic relationships among stone tools. Neutral variants can be particularly reliable for phylogenetic inference, because they are likely to be uncorrelated with one another, and because modeling the stochasticity of the evolutionary process is more tractable without the additional effects of selection.

The importance of neutral variation for inference about genetic phylogenies (Cavalli-Svorza and Edwards 1967; Felsen-

stein 2004; Steel 2016) has long been appreciated. More recently, phylogenetic inference methods have also been applied to cultural traits under the influence of both cultural and biological selection (Gray, Greenhill, and Ross 2007; Mace and Holden 2005; Mace, Holden, and Shennan 2005; Mendoza Straffon 2016). For language phylogenies in particular, researchers have constructed well-resolved cultural phylogenies using multiple traits, and Bayesian analysis of these phylogenies has enabled specific hypotheses about the age and original location of particular linguistic clades to be tested (Bouckaert et al. 2012; Gray and Atkinson 2003; Gray et al. 2009; Lee and Hasegawa 2011). The approach proposed by Stout et al. might be further strengthened by adopting the well-developed methods of phylogenetic inference from evolutionary biology.

Cultural transmission of behavioral sequences is not unique to humans. Behavioral ecologists have recently designed tasks that target neutral variation of behavioral sequences in order to disentangle social learning from independent arrival at solutions. Thus, for example, sparrow fledglings have been shown to learn from their mothers not only where to find food, which would be expected to have a selective advantage, but also the specific method of reaching it: they can choose to peck through an artificial leaf or shift it aside, alternatives that are similar in their cost and effectiveness (Truskanov and Lotem 2017). Learning of behavioral sequences has been observed in many other animals, such as black rats (Aisner and Terkel 1992), great tits (Aplin et al. 2013, 2015), and nonhuman primates (Voelkl and Huber 2000; Whiten 1998; Whiten, Horner, and de Waal 2005; Whiten et al. 1999). That many organisms acquire behavioral sequences through social learning makes the notion that early hominins routinely learned behavioral sequences from one another very plausible. In fact, from this perspective, it would have been surprising if the authors' exploration had concluded otherwise.

Studies of behavioral ecology also suggest that social learning is common across the animal kingdom but that elaborate tool manipulation is much less so (Bentley-Condit and Smith 2010; Galef and Giraldeau 2001; Hoppitt and Laland 2008; Shettleworth 2010). Therefore, one might argue that, from an evolutionary perspective, social learning of behavioral sequences likely preceded tool manipulation, and the former may rely on somewhat less specific cognitive abilities. This supports an interesting and important suggestion made by Stout et al., that social learning and tool manipulation, fundamental aspects of hominin cultural evolution, be viewed as distinct, and perhaps initially uncorrelated, traits. They speculate that to the early hominins at Gona, copying unifacial or bifacial knapping verbatim was easier than learning the general idea and then individually reaching an effective method via trial and error. This is a nontrivial point, particularly because, as the authors point out, for modern human adults the opposite is likely to be true. This difference may skew our intuition about the likelihood of analogy, as opposed to homology, for production methods at different sites and by different individuals. The authors' view of social learning as a process in-

volving both social and individual learning, which may draw upon different skill sets, resembles similar approaches that were recently suggested in the behavioral ecology literature (Galef 2013, 2015; Truskanov and Lotem 2017; Truskanov and Prat 2018).

Stout et al. suggest that preexisting learning capacities allowed the initial behavioral interaction of hominins with tool manipulation, leading to a Baldwin effect situation in which social learning made it possible for natural selection to act directly on genetic or cultural variation underlying toolmaking skills (see also van Schaik and Burkart 2011). We find this a reasonable hypothesis that suggests a mechanism for one of the key evolutionary developments in hominins. It also provides a nonintuitive prediction, namely, that we should expect a "decrease in the variation" in quality or precision of tool products over time, as selection acts to improve the performance of toolmaking, while at the same time we should expect an "increase in the subtle variation" in methods used to manufacture these tools, as individual trial and error becomes more prominent in the learning process relative to exclusively verbatim copying. This prediction can be evaluated against the archaeological record, and analysis of variation in tool production methods and tool quality from this perspective may provide insights into human evolution.

Although the Baldwin effect is plausible in these circumstances, we suggest that the behavioral plasticity afforded by social learning may also have acted in the opposite direction, creating an effect similar to the "reverse Baldwin effect" (Deacon 2003), in which efficient social learning masks independent innovation from being culturally or genetically selected, slowing down the evolutionary process. If indeed the ability of early hominins to learn behavioral sequences was greater than their ability to innovate in tool manipulation, the efficacy of precise tool production through copying would have reduced the likelihood of innovation of other tools or other methods through trial and error once a behavioral goal, such as the manufacturing of a particular tool, had been achieved. This reduction may explain what seems to be a puzzling discrepancy between the complexity of certain material cultures, such as the Acheulean (Ayebare et al. 2011), and the long period in which no major technological innovation took place, innovation that we otherwise might expect from hominins capable of executing the complex behavioral sequences needed to produce Acheulean tools. The early arrival at a fairly complex sequence of behavior that achieves a well-defined goal (e.g., producing a tool) may cause entrenchment of that sequence in the population, even if it is suboptimal, and hinder the emergence of alternative, perhaps more efficient, solutions. Some studies in behavioral ecology have attempted to address such questions (e.g., Hrubesch, Preuschoft, and van Schaik 2009; Marshall-Pescini and Whiten 2008); further experiments with apes or humans at different ages may provide insight into how such processes might have occurred in early human evolution.

The two opposing hypotheses—that preexisting social learning abilities could have accelerated or impeded technological cultural evolution—may provide a productive avenue for investigating critical steps in hominin evolution. These processes may have left their mark on the archaeological record. A Baldwin effect speedup or a reverse-Baldwin slowdown could have acted both on the rate of adaptive evolution in the genetics that underlie motor and cognitive aspects of toolmaking and on the rate of cultural evolution independent of its underlying genetics. As suggested above, further study of these effects may help explain long periods of surprisingly little cultural change in stone tool technology as well as periods of sudden rapid change. Importantly, the two hypotheses are mutually exclusive only in a certain context; it is quite possible that the proposed Baldwin effect operated in some contexts and its reverse in others, particularly during different cultural and genetic evolutionary phases along the hominin lineage. Because the baseline for these rates (in the absence of any Baldwin-type effect) is unknown, the archaeological record can, most likely, only be interpreted from this perspective by a comparison of relative rates across periods or regions.

Acknowledgments

Anders Högberg, Marlize Lombard, and Peter Gärdenfors
School of Cultural Studies, Archaeology, Linnaeus University, SE-391 82 Kalmar, Sweden (anders.hogberg@lnu.se; Högberg)/ Centre for Anthropological Research, University of Johannesburg, PO Box 524, Auckland Park, ZA-2006, South Africa (Högberg, Lombard)/Cognitive Science, Department of Philosophy, Lund University, SE-22100 Lund, Sweden (Gärdenfors)/Stellenbosch Institute for Advanced Studies, Wallenberg Research Centre, Stellenbosch University, Stellenbosch, South Africa (Högberg, Lombard, Gärdenfors). 7 IX 18

Cumulative Culture: An Integrated Perspective

Notions of a few revolutionary leaps in hominin evolution ultimately resulting in *Homo sapiens* behavior (Coolidge and Wynn 2009; Klein 2000) are increasingly being challenged and replaced by the understanding that cultural evolution is cumulative, developing gradually over extended periods (Garofoli 2016; Haidle et al. 2015). Part of the "leap paradigm" has been that there are no substantial differences between chimpanzee and Oldowan tool behavior (Wynn et al. 2011). It is in this context that we welcome the effort by Stout, Rogers, Jaeggi, and Semaw to provide a detailed empirical approach to sub-

stantiate the gradual-progression position in which they argue that Oldowan technology cannot be "accomplished solely through individual learning."

According to some critics, few experiments can convincingly assess nonextant hominin tool behavior (Coolidge and Wynn 2016). The experiments presented by Stout et al. are, however, a thorough attempt to reconstruct past hominin tool behavior. Their commitment to design experiments that can be reproduced and elaborated on by others is commendable. We further appreciate their efforts to reduce variables in their experimental setup, but we find the separation between functional and cultural explanations restricting in the effort to understand the emergence of new learning capacities within the hominin lineage (fig. 1).

Although Stout et al. acknowledge that the functional criterion of knapping with the aim to obtain as much cutting edge as possible out of a core is insufficient to explore the full repertoire of Oldowan knapping, they continue to base their experiments and arguments on it. Thus, although they argue for a bottom-up approach, they still rely on a top-down "theory of function" (Arthur 2018). Further, their argument by elimination builds on the assumption that "culture is what is left over when all other possible explanations of variation are exhausted" (Arthur 2018:12). Among other things, this line of thinking makes it problematic to claim that chimpanzees have culture.

Their approach works well until the argument by elimination in the second step of figure 1. In our opinion, to proceed with their point regarding "ecologically valid studies of different learning processes" in step 2, other theoretical models on the evolution of learning processes and cumulative culture are necessary already here instead of being left to step 3. Thus, as an alternative to the either-or approach regarding bottom-up/top-down, we suggest a continuous integration of theory and data. Any model of cultural evolution has to explain the cause of changes in cultural capacities such as learning and social transmission of information. We next turn to two such models that might be useful in the context of the arguments presented by Stout et al.

By reconstructing modes of teaching from the archaeological record, Gärdenfors and Högberg (2017) came to the conclusion that the learning functions of copying, evaluative feedback, and, in particular, demonstration, were involved in intergenerational knowledge transmission during the Oldowan at Lokalalei 2C at ca. 2.34 Ma. The latter modes of teaching build on different forms of mind reading that involve much more than copying, because a teacher must understand that the learner does not have the necessary knapping knowledge to produce a tool (Gärdenfors and Högberg 2017). Working with material from the sites of Gona dated to ca. 2.6 Ma, Stout et al. focus on copying. There are, however, marked differences between Gona and Lokalalei 2C in terms of technological complexity. As opposed to Gona, there is evidence for core maintenance at Lokalalei 2C. This indicates variation, and possibly progression, in the processes of social transmission during the Oldowan.

We find it interesting to position the results obtained by Stout et al. between the observations made by Gärdenfors and Högberg regarding the Lokalalei 2C Oldowan and the pre-Oldowan of Lomekwi 3 with an age estimate of 3.3 Ma (Harmand et al. 2015; Lewis and Harmand 2016). For the bipolar technique practiced at Lomekwi 3, Lombard, Högberg, and Haidle (forthcoming) suggest that at least the nonintentional teaching modes of enhancement and intentional evaluative feedback were already in play. This would require that a practiced knapper draws attention to the different objects and aspects that have to be controlled. At a minimum, this necessitates joint attention and possibly also joint intention (Gärdenfors and Högberg 2017). According to their analysis, this type of teaching is absent in chimpanzee nut-cracking culture, which can be accomplished with joint attention only. Lombard, Högberg, and Haidle (forthcoming) therefore find a distinction already between the learning processes in chimpanzee nut cracking and Lomekwi 3 bipolar knapping. Cumulatively, the Stout et al., Gärdenfors and Högberg, and Lombard, Högberg, and Haidle et al. studies bolster the view of a long, varied, and gradual process of cultural accumulation along the hominin line, which links to Shea's (2017) argument for the development of early technologies from occasional to habitual.

The evolution of cumulative culture depends on the transmission of information that is inherent in models of the evolution of cultural capacities. Our second example is the eight-grade model of Haidle et al. (2015), who hypothesize that chimpanzee culture is limited to basic social information structures, whereas the Oldowan technology represents the modular transmission of sets of cultural units. Such modularity is, for example, represented in using tools to make tools. The study by Stout et al. feeds into such an integrated theoretical approach and serves to further distinguish between nonhominin and early hominin technologies by contributing experimental data. We thus agree that "the emergence of human cumulative culture should not focus narrowly on social learning mechanisms but should instead address a range of factors also including individual cognitive and perceptual-motor capacities, levels of sociability, the internal dynamics of cultural evolution, and biocultural coevolutionary processes." Here we draw attention to the model presented by Haidle et al. (2015) that considers all of these factors in an attempt to explore and explain the evolution and expansion of cultural capacities in hominins and other animals.

Finally, Stout et al. entertain the idea that each generation might have reinvented the skill of knapping early tools. However, for an invention to become "culture," as it did during the Oldowan, it requires that the technology is socially accepted and consistently performed throughout groups and generations (Högberg 2009). The skill thus transitions from "I can do it" to "we do it." We therefore argue that key to development in hominin cultural evolution was not individual tool production and use in itself but the traditions that emerged when the technology used to make sharp flakes became intentionally taught between groups and generations.

Erella Hovers

Institute of Archaeology, Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91905, Israel (hovers@mail.huji.ac.il). 17 X 18

Stout, Rogers, Jaeggi, and Semaw make a contribution to the debate about the antiquity of cumulative culture by showing through an experimental research program that particular knapping procedures (the particular actions on matter) in the 2.6–2.5 Ma archaeological assemblages from Gona were functionally neutral. For the archaeologist, a major strength of this contribution is the structuring of a "bottom-up" research approach that builds on the strengths of prehistoric archaeology. The paper focuses on details of the variation of perceptual-motor skills of the ancient stone toolmakers, as reflected by artifacts, rather than on the more nebulous concepts of social learning. Stout et al. then infer that the observed intra- and inter-assemblage variation is consistent with social learning through detailed copying rather than individual reinvention. This interpretation of the Gona assemblages is reasonable. Social learning exists in many nonhuman species (e.g., Galef 2012), and it is not surprising that hominins possessed capacity for it. Indeed, the emergence of cultural traditions (within-group socially learned and persistent behaviors) is a parsimonious explanation for the patterns so neatly laid out by the experimental and archaeological analyses. How this informs us of the antiquity of cumulative culture, considered a hallmark of our species (e.g., Henrich 2015), remains unclear.

While the Oldowan is not the earliest stone tool technology known to us (Harmand et al. 2015; Lewis and Harmand 2016; see Panger et al. 2002), it is revolutionary for archaeologists because it marks the beginning of persistent, archaeologically visible technological behavior, in turn enabling for the first time identifying, observing, and interpreting variation in lithic technological procedures. The successful knapping and use of Oldowan stone tools indeed required understanding of mechanical forces and geometric relationship and their interaction (e.g., Delagnes and Roche 2005; Goldman-Neuman and Hovers 2012; Hovers 2009; Stout 2011, fig. 1; Stout et al. 2010). Still, the "invention" of stone toolmaking by freehand direct percussion enlisted preexisting cognitive as well as anatomical characteristics involved in technological tasks (Read and van der Leeuw 2008) and their reconfiguration as a novel behavior, while other behaviors continued from the last common ancestor.[1] The procedural elements of Oldowan knapping—even when using bi-/mutifacial reduction—involved basic skills (Moore 2011; Moore and Perston 2016; Stout 2011; Stout and Chaminade 2007) and may have been used in unrelated, possibly even "primitive" tasks outside of the domain of lithic material culture. As Stout et al. note, the revolutionary effect that Oldowan technology had on human cultural (and biological) evolution occurred gradually in a Baldwinian mode—much

after the onset of the Oldowan (see, e.g., Morgan et al. 2015; Tocheri et al. 2008; Wynn et al. 2011).

It is also important to consider that flaking stone may not have been the exclusive way of obtaining functional edges or the preferred option under all circumstances, even if social learning offset the costs and risks involved (Morgan 2016). Less costly but still functional cutting edges could be obtained by, for example, using sharp-edged, naturally fragmented stones. This is especially reasonable if toolmakers of some of the Oldowan assemblages differed from *Homo*, given that the hand anatomy of other contemporaneous hominin genera was not suited for freehand knapping (Rolian, Lieberman, and Zermeno 2011). A plausible hypothesis is that detailed copying was not restricted to stone knapping and therefore could be operationalized toward this particular goal when conditions favored this behavior. Even if Oldowan toolmakers were cognitively constrained and at the limit of innovative capacities (e.g., Andersson 2011; Pradhan, Tennie, and van Schaik 2012), some if not all the motor and perception skills could be readily called upon to execute a socially learned procedure or "reinvent, through individual learning by trial and error, the limited repertoire of gestures and spatial relationships involved in Oldowan flaking." Such "mundane creativity" (see Hovers 2012) would fall within the "zone of latent solutions" (Tennie et al. 2016) for the limited range of problems that Oldowan stone tools aimed to solve.

As archaeologists, we also need to consider context (Hovers 2012; and see the comment by Ignacio de la Torre in Tennie et al. 2017). The archaeological record of deep time is notoriously incomplete and fragmented. With this caveat in mind, one has to wonder whether social learning among small groups that were dispersed over large temporal (slightly less than a million years) and spatial (eastern and southern Africa) dimensions could give rise to cumulative culture. While we are far from understanding the demography of Oldowan toolmakers, it is consensually accepted that they lived in small groups and may not have reached the effective cultural population size (either demographically or due to reduced connectivity) necessary to sustain a cumulative cultural system (Kolodny, Creanza, and Feldman 2015).

Detailed copying as well as reinvention through individual trial and error likely occurred at different temporal, social, and demographic scales throughout the nearly million years of the Oldowan. I agree with Stout et al. that explanations for the emergence of cumulative culture must incorporate a range of factors at the individual level, degree of sociality, and the internal dynamics of cultural evolution (Kolodny, Creanza, and Feldman 2015, 2016). It is for this reason that the present study does not lead to a better understanding of cultural "ratcheting" (Tennie, Call, and Tomasello 2009) or of whether the Oldowan can be considered as its time of emergence except in the very broad sense. The effects of the various modes of information transmission are not easily pried apart archaeologically, and hypotheses such as those structured by Stout et al. may not be amenable to rigorous testing beyond pattern recognition in the archaeological record. Nevertheless, Stout

---

1. Campbell Rolian and Susana Carvalho, "Tool Use and Manufacture in the Last Common Ancestor of *Pan* and *Homo*," unpublished manuscript.

et al. have taken an important step toward bringing closer bottom-up and top-down accounts of the evolution of human culture.

Jay S. Reti
Anthropology Department, University of California, Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA (jreti@ucsc .edu). 1 X 18

Stout, Rogers, Jaeggi, and Semaw draw important and necessary attention to the earliest forms of cultural transmission in the hominin lineage. While Oldowan technology is commonly discussed in terms of raw material economics and transport, functionality, and production, it is only recently that we have begun the discussion of cultural transmission within the Oldowan. Creating experimental models for the Gona material allows for a glimpse of what early cumulative culture, likely in the form of behavioral reproduction, looks like. Stout et al. provide a discussion that outlines the various implications of cultural transmission in the Oldowan, and these discussions offer valuable perspectives on the limitations and motivations for such early cultural transmission.

Focus on flake production in the form of unifacial cores from East Gona versus bifacial cores at Ounda Gona South provides an effective case study for testing variation in behavioral reproduction among Gona hominins. Stout et al. model the effects of raw material economics, flake viability, and core morphology as potential reasons for differences in stone tool production strategies. Their results indicate that production behavior (unifacial vs. bifacial core reduction) does not yield significantly different assemblages in terms of efficiency. It stands to reason, then, that aspects of behavioral reproduction or cultural transmission could be the reason for such variation.

The authors also make a refreshing explanation concerning how early forms of cultural transmission may have operated in the hominin lineage: "copying [detailed knapping methods in the Oldowan] was not, however, associated with evidence of increased rates of technological change or ratcheting, suggesting that explanations for the emergence of human cumulative culture should not focus narrowly on social learning mechanisms but should instead address a range of factors." Thinking about cumulative culture outside of the normal "ratcheting effect" assumptions helps to frame Oldowan production patterns as unique unto themselves, leading to cultural niche construction but not necessarily immediately more complex technologies. Such conclusions help to frame perceived periods of technological stasis as important evolutionary events and justify further research to identify and quantify subtle patterns of technological variability throughout the Oldowan.

A logical outgrowth of these conclusions is to compare production patterns found at Gona with other Oldowan assemblages, both temporally and geographically. Are the cultural patterns that Stout et al. identify unique to Gona, its raw materials, and the environment in which they were created? Or is the pattern recognized at Gona indicative of a broader behavioral trend among Oldowan producing hominins? Might these behaviors be seen even earlier in time, with assemblages such as Lomekwi (an important site and relevant line of evidence that was not discussed in this article [Harmand et al. 2015])?

My own research focuses on regional variation within Oldowan assemblages at Olduvai Gorge (Reti 2016) and Koobi Fora (Reti 2013), and I am encouraged to see Stout et al. framing the Gona material in a cultural light. I immediately think of statistically comparing diverse assemblages (both experimental and archaeological) from Gona, Lokalalei, Koobi Fora, Kanjera, and Olduvai Gorge in order to assess the diversity of cultural transmission, production techniques, and economic strategies. In order to do this, we must work together to standardize experimental practices and data sets and be open to data sharing of these data sets and assemblages. It is only with such data sharing that the broader patterns of early human evolutionary cultural transmission will be identified. I hypothesize that the broader pattern of cultural transmission in the Oldowan is due primarily to economic factors of raw material quality, transport, and related costs. However, there are many possible ways to mitigate these economic factors, and Oldowan producing hominins across East Africa may have addressed these problems using different culturally constructed niche systems. Broader comparisons between Oldowan technologies will allow evolutionary archaeology to quantify these behavioral responses to raw material economics and to compare hominin strategies of stone procurement and production.

By drawing attention to questions of cultural transmission, Stout et al. elaborate the argument made by many others that the Oldowan represents a complex and plastic adaptive skill set. It is my hope that this article begins a dynamic and collaborative discussion among Paleolithic archaeologists in order to address (1) the broader implications of cultural transmission beginning in the Oldowan, (2) differential patterns of cultural transmission in the Oldowan, (3) how cultural niche construction in the Oldowan ultimately leads to technological change over time, (4) primate comparisons of Stout et al.'s methods to address the uniqueness of behavioral reproduction at Gona versus primate models and older archaeological assemblages such as those from Lomekwi, and (5) standardization of experimental models in Paleolithic archaeology for the purpose of producing comparable experimental data sets.

Claudio Tennie
Department for Early Prehistory and Quaternary Ecology, University of Tübingen, 72070 Tübingen, Germany (claudio.tennie@ uni-tuebingen.de). 10 IX 18

The Zone of Latent Solutions Account Remains the Most Parsimonious Explanation for Early Stone Tools

The main new claims deducible from Stout, Rogers, Jaeggi, and Semaw are (*a*) that core rotation versus nonrotation was

influenced by some variant of social learning ("rotation culture"); (*b*) that action copying has to be the specific variant in such cases; and (*c*) that *a* and *b* together disprove Tennie et al.'s (2016, 2017) zone of latent solutions (ZLS) account for early stone tools.

Only claim *a* ("rotation culture") finds initial support. Early Stone Age toolmakers thus join many species in showing variants of social learning. They also join nonhuman great apes (henceforth, apes) in showing population-level differences not readily traceable to genetic and/or environmental differences—that is, culture (Whiten et al. 1999). But this leaves open which variant(s) of social learning, and it also does not exclude latent solutions—not least because the ZLS hypothesis is designed as a possible explanation for exactly such patterns (Tennie, Call, and Tomasello 2009; opposing claims *b* and *c*).

Contra Stout et al., the latent solution founder effect does not consist of independent individual responses to environmental pressures.[2] As explained by Tennie, Call, and Tomasello (2009; also Bandini and Tennie 2017), at the beginning of the latent solution founder effect, behaviors from within the ZLS are shown by one or more individuals in a population that does not yet express any latent solution to the particular problem. Which behavior is chosen can indeed depend on chance; for example, the first individual(s) may stumble upon tool material A rather than B and thus express latent solution A (see the multiple independent similar innovations in Hobaiter et al. [2014]). Next, others are socially biased toward developing a similar latent solution—for example, as an individual consequence of attending more to A. Therefore, perhaps counterintuitively (and against claim *b*), cultural patterns do not require action copying (or action teaching).[3]

Stout et al.'s approach over-infers copying due to a focus on similarity. Consider an uncontroversial case: if you see me yawn, you might also yawn—in a similar form (using similar sounds/actions). Yet, typically, you did not copy my "yawn form" at all. Stout et al.'s approach, however, judges your yawn a "behavior reproduction"—a phrase that connotes copying. The ZLS hypothesis instead clearly distinguishes between the "transmission of a form"[4] and "social influences on the frequency of a form"[5] (Bandini and Tennie 2017). While the former requires copying, the latter may or may not involve any copying but can also produce cultures (Bandini and Tennie 2017 and references therein). Hence, data in support of claim *a* cannot—by itself—support claims *b* or *c*.

We can already infer that action copying is not necessary for rotation culture (contra claim *b*) because Stout et al. recreated the underlying actions without ever having observed the actions of the original makers. Instead, the social learning variant we should infer is "object movement re-enactment" (OMR; Custance, Whiten, and Freden 1999).[6] This inference is most parsimonious also because OMR underlies ape learning in so-called two-target tasks where there is a similar requirement:[7] to recreate one of two object movements (Tennie, Call, and Tomasello 2006; OMR pinpointed in Hopper et al. [2008]). Crucially, all recreations (here core movement vs. no core movement) can be latent solutions.[8] Our account indeed predicts that both variants of rotation culture will also be found in culturally unconnected populations (similar to the ape cultures so far examined; Tennie et al. 2016, 2017).

Because the cultural models they cite report stasis as an outcome of copying, Stout et al. conclude that any variant of copying can fit observed stasis. However, those models excluded the fine-grained copying used within the verbal cultural model of Stout et al. This copying variant has unavoidable copying error (see Eerkens and Lipo 2005). For tasks involving proportional error, even this copying can lead to stasis (Hamilton and Buchanan 2009).[9] But Stout et al.'s specific model (approximately based on action details) lacks proportional error and therefore fails to fit overall stasis (Tennie et al. 2016, 2017).

Stout et al. also claim that emulation alone can lead to cumulative culture, but when we tested this in children, it did not (Reindl and Tennie 2018). They also claim that the distinction between emulation and imitation is meaningless and should be collapsed, but then why do untrained apes emulate pure environmental results (Hopper et al. 2008) but fail to imitate pure actions (Clay and Tennie 2017)? Why do they not solve difficult tasks better after seeing both actions and results underlying the solution than when they merely see the underlying results (Tennie, Call, and Tomasello 2010)? Why is training required for apes to enable action copying, and why does this lead to brain changes that are linked to action copying (Pope et al. 2018)? The best answer is that apes are not good natural imitators (action copiers). Instead, they are emulators. Humans emulate and imitate (Tennie, Call, and Tomasello 2009),[10] and these simultaneous copying skills enable special forms of error correction that can increase copying fidelity beyond the level required to escape the ZLS (Acerbi and Tennie 2016; cf. Lewis and Laland 2012).

I am pleased that archaeologists examine when human-like culture first arose. However, for the reasons above and in Tennie et al. (2016, 2017), I am still of the opinion that the latent solutions account remains the most parsimonious hypothesis for Early Stone Age toolmakers: not least because it does not involve detailed copying—and thus predicts the observed stasis.[11]

---

2. This is notwithstanding the fact that, where environmental and/or genetic biases do exist, they can influence the likely direction of the latent solution founder effect.

3. In the example I used here, "stimulus enhancement" sufficed.

4. Or "fidelity" in yawning contagion = ~0.

5. In yawning contagion = high.

6. The clue is in the name.

7. Often erroneously called "two action" tasks.

8. And so, not every case of copying escapes the ZLS (contra claim C).

9. Because its mean and variance decrease across generations. Many thanks to Luke Premo, who opened my eyes on this.

10. In humans this is perhaps the outcome of another, continuing triple inheritance feedback loop; cf. Heyes (2018).

11. But also because it includes exaptations.

Andrew Whiten
Centre for Social Learning and Cognitive Evolution, School of Psychology and Neuroscience, University of St. Andrews, St. Andrews KY16 9JP, United Kingdom (a.whiten@st-andrews.ac.uk). 27 IX 18

Culture in the Stone Age

Stout, Rogers, Jaeggi, and Semaw present a sophisticated and revealing empirical approach to reconstructing the nature of early hominin lithic cultures. They offer meticulous methods, well-reasoned conclusions and a thoughtful stepwise rationale for their study. In the first two steps of this, the authors offer compelling evidence for significant differences in the artifacts at the EG-10 and EG-12 versus OGS-7 sites, and the experimentally verified, corresponding flaking techniques. As the authors acknowledge, their further conclusions about the likelihood that such differences in technique were socially learned rest on arguments by elimination, with inherent risks of too-hasty falsification or, alternatively, hypothesis confirmation. That said, the working conclusion that the differences in technique were likely culturally based is compelling and offers strong justification for further investment in experiments dissecting the social learning processes at work, building on the few existing experimental explorations by our own group (Morgan et al. 2015) and others such as Putt et al. (2015).

Applying to the Oldowan what we have learned of chimpanzee cultural variation across Africa, Whiten, Schick, and Toth (2009) concluded that early hominin communities sep-

arated by around 700 km could be expected to have about half their technological traditions in common and half not. Against these kinds of distances, it might seem implausible that sites like those at Gona, separated by just 3 km, would generate quite different cultural modes of lithic technique. However, a range of types of evidence shows that the percussive behavior of chimpanzees most similar to knapping—nut-cracking with natural hammers—is culturally transmitted (Whiten 2015), and elements of these traditions may differ over just a few kilometers. Luncz and Boesch (2014) showed that one community of chimpanzees in the Taï Forest prefer to employ stone tools all year, unlike two neighboring groups that exhibit seasonal shifts to wooden versus stone hammers. These authors add evidence that dispersing females conform to what is the locally common approach. We may never know if such effects occurred in Oldowan toolmakers, but these comparative data are compatible with the differences between the Gona communities being culturally based.

Turning to focus on the next step, the identification of social learning processes, the authors cite the model of culturally enabled skill development that I dubbed a "helical curriculum" by analogy with the "spiral curriculum" concept of education theory, where a topic is revisited repeatedly at rising levels of sophistication as the learner accumulates knowledge (Whiten 2015). Primatological observations suggest this process underlies the development of percussive tool use and probably occurs in other culturally transmitted skills that are challenging to master. In a conceptual graphic of the model (fig. 10a), along
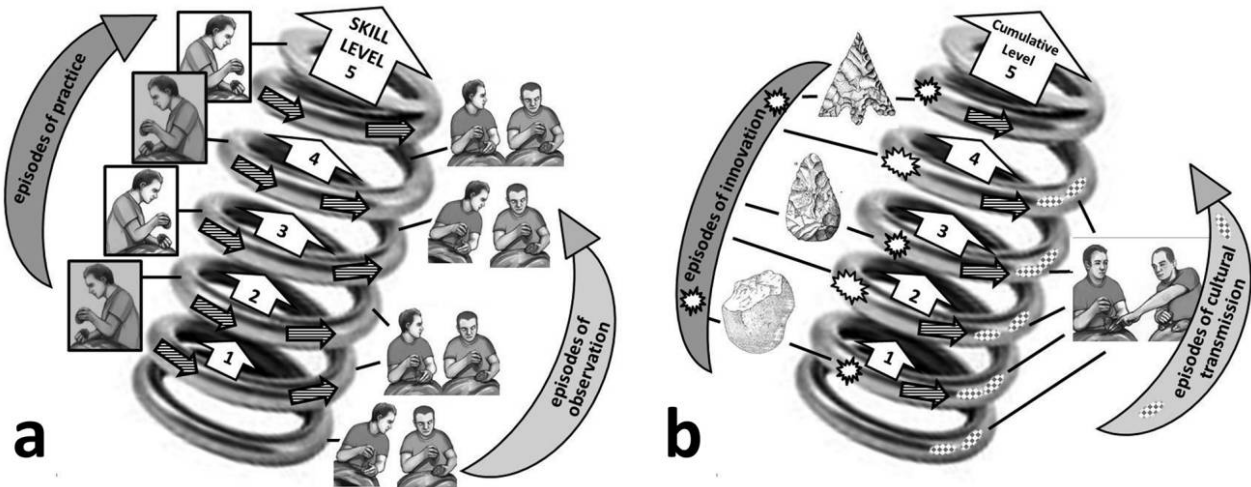

Figure 10. Conceptual graphic illustration of helical curriculum model of skill development and counterparts in cumulative cultural evolution. *a*, "Helical curriculum" model (after Whiten 2015). Each turn in the helix is marked by a succession of episodes of social learning from a model and individual learning, including practice and exploration, each facilitating progress in the next episode of the other, with skill, hence rising through notional levels 1–5. *b*, Comparable processes in cumulative cultural evolution, likewise driven by alternation of episodes of social learning from previous generations and further elaborations to this by rare individuals. Here each turn of the helix represents a generation, with lithic sophistication rising through levels 1–5. The illustration is of gestural teaching, but different forms of social learning could be in play during different historical phases. The processes conceptualized in figure 1*a* would actually run along the thread of the helix in figure 2*b*, as such ontogenetic helixes mark the generations represented in the evolutionary helix. See text for further explanation and discussion of *a* and *b*.

each turn of the developmental helix there are episodes of social learning, followed by individual practice and exploration. Another turn is then marked by further social learning, in which the personal-level learning that has taken place allows the learner to extract more information than was possible earlier. If further research supports the reality of this model, it follows that it will be important to facilitate its operation in the "ecologically valid studies of different learning processes (to be developed)" advocated in Stout et al.'s figure 1, which I agree should play a key role in the broad endeavor covered by the figure.

As Stout et al. note, this iterative model may apply to the development of "all nontrivial skills," particularly where observational learning can provide only outline information on what to do, because some key elements of the skill are simply not visible. In fact, flint knapping is a paradigmatic illustration of this, in which how the raw material will behave (fracture) is not intuitively obvious (arguably counterintuitive), as evident in the inept efforts of novices attempting to knap a hand axe informed only by seeing an already-worked example (Geribas and Verges 2010). The repeated alternation between social and personal learning in the helix ameliorates such problems because as the learner initially relatively blindly tries to copy what the model did, they discover the significance of some of the opaque elements like fracture dynamics.

There may be significant resemblances here with a phenomenon recently much studied in developmental psychology: over-imitation (Whiten et al. 2009). In over-imitation, young children copy much of what they see a trusted model perform, including elements whose causal relevance is opaque or even downright implausible. Lyons, Young, and Keil (2007) christened this over-imitation and suggested it is an automatic social learning process that aids learning the use of the numerous causally opaque artifacts in a child's world, because as the child imitates, she discovers some of the underlying causal relations that were not visible. Over-imitation turns out to be more contextually flexible than the automaticity these authors suggested, but they probably identified an important functional role for this process, that Whiten, Horner, and de Waal (2005) had characterized as a "copy all, refine later" strategy. Over-imitation has been documented in adults as well as in children (Whiten et al. 2016) and in a substantial diversity of cultures (e.g., Nielsen and Tomaselli 2010). The complementary roles of social, followed by individual, learning map closely to those discussed in the paragraph above, under the heading of longer-term ontogenetic skill learning, and suggest some potentially shared fundamental principles of cultural transmission of complex behavior (see Shipton and Nielsen [2015] for fuller discussion of over-imitation in relation to the Oldowan and Acheulian).

The final step in the authors' scheme brings us from this to the macrolevel and cumulative culture. I endorse the authors' view that it is not productive to define cumulative culture in terms of the supposed social learning processes necessary to it. Cumulative culture is best defined simply as culture that cu-

mulatively builds on what came before, and we can then tackle empirical questions about underlying social learning processes. However, I suggest that the helical cycles discussed for the microlevel processes above recur in important guises at the macrolevel of cumulative culture, and accordingly I explore this through the comparable figure 10*b*.

One final question: Why no mention of the prior Lomekwian period?

Reply

We are honored by the quality of commentary on our article, and encouraged by numerous productive suggestions for the road forward. There is room for reasonable disagreement on this topic, and we are gratified that this is captured by the commentaries. Some see our interpretation as clearly wrong (Tennie), others as convincing (Reti) but unsurprising (Greenbaum et al., Hovers, Whiten). Some even feel it does not go far enough (Högberg, Lombard, and Gärdenfors).

A Zone of Latent Solutions (ZLS) Account?

Tennie asserts that the technological pattern observed at Gona is explicable in terms of a "latent solutions founder effect." Tennie, Call, and Tomasello (2009:2407) explain this effect as follows: "if an individual in a given chimpanzee population, by chance, invents a way to crack nuts with a wooden stick, then the others in its group—by virtue of their exposure to sticks and open nuts in close proximity—will be exposed to learning conditions favouring the individual discovery of stick use." The actual details of stick use are thus filled in by individual learning, with any similarities in "behavioral form" being due to biased/constrained learning conditions (e.g., the affordances of sticks). This is precisely the possibility we addressed with our experiment by testing whether bias toward a particular raw material form would be sufficient to elicit the observed flaking patterns. We found no support for this hypothesis. Insofar as Tennie does not dispute these findings, the disagreement seems largely to be about terminology (see my comment, Dietrich Stout in Tennie et al. 2017:661–662).

Tennie agrees that social learning is implicated at Gona but argues that the particular variant is object movement reenactment (OMR). Because OMR is classified as a form of emulation (Custance, Whiten, and Freden 1999; Whiten et al. 2009), Tennie assumes it is a low-fidelity transmission mechanism indicative of ZLS learning. We question the assumption that copying object movements is an inaccurate way to learn how to use objects. Research going back to the 1930s demonstrates the intuitive point that skilled tool use is characterized by accurate reproduction of working point (e.g., hammer head) rather than body part trajectories (Bernstein 1996). Nor

does the assumption follow from the original identification of OMR as a distinct learning process (Custance, Whiten, and Freden 1999) as this distinction focused on cognitive mechanisms rather than reproductive fidelity. The assumption instead appears to reflect confusion with the quite distinct conception of emulation learning developed by Michael Tomasello.

Tomasello (1996:321) described emulation as an animal "learning some affordance or change of state of the inanimate world as a result of the behavior of another animal, and then using what it has learned to devise its own behavioral strategies." Emulation in this sense fits well with the example of a latent solutions finder effect quoted above and is low fidelity because particular behavioral strategies are reinvented by each individual. In contrast, Tomasello defined "imitation" as learning organized around the reproduction of action intentions. This is thought to enable higher-fidelity copying, including the reproduction of intentional actions that the observer does not understand causally (Tomasello 1998). These implications regarding transmission fidelity do not, however, transfer to the alternative taxonomic system that includes OMR (e.g., Whiten et al. 2004). Here, Tomasello's emulation would be termed "affordance learning," whereas "emulation" includes the copying of action goals or outcomes and thus substantially overlaps with Tomasello's sense of imitation (to which Tomasello [1996, 1998] objected). The definition of "imitation" is correspondingly narrowed from behavior copying generally to reproduction of bodily movements specifically (Tomasello's [1996] "mimicry"), thus creating space for OMR as an intermediate learning process. The restriction to bodily movements is intended to align "imitation" with requirements for visual-motor cross-modal processing that are posited to be more cognitively demanding than the visual-visual matching hypothesized to account for emulation (including OMR; Custance, Whiten, and Freden 1999). This scheme defines imitation and emulation as two poles along a behavioral continuum with no particular implications regarding relative fidelity (Whiten et al. 2009), which is the point we make in our article. We do not advocate one taxonomy over the other but note that blending the two produces confusion.

For example, Tennie, Call, and Tomasello (2009) test the ZLS hypothesis using a task in which participants learn from observation to form a string into a loop in order to retrieve a baited platform. This is pretty clearly imitation in Tomasello's sense, and children's success was taken as evidence of a uniquely human capacity for process copying outside the ZLS. However, it might alternatively be described as OMR or even end-state emulation since it is actually the string configuration, rather than specific finger movements, that must be reproduced. Indeed, this is the stance Tennie takes in attributing core manipulation patterns at Gona to OMR. This is an atypical use of OMR, which normally refers to large-scale object trajectories (e.g., movement of the platform rather than the string in Tennie, Call, and Tomasello [2009]) rather than in-hand manipulation. This usage raises interesting issues regarding the neurocognitive mechanisms of "in-hand" OMR, which may not be simple or unimodal (Arbib et al. 2009; Stout 2013) but has no implications for reproductive fidelity or cumulative potential.

The argument we present does not rely on any assumptions about the social learning capacities of nonhuman apes. Nevertheless, Greenbaum et al., Hovers, and Whiten all suggest that our conclusions are unsurprising given comparative expectations for a chimpanzee-human common ancestor. Interestingly, Tennie accepts that apes can learn to imitate although they find it difficult and are not inclined to do so. This does not seem like a large gulf to be covered by incremental evolutionary change along the hominin line, and it aligns quite well with the extended evolutionary account we elaborate elsewhere (Stout and Hecht 2017). Methodologically, we are concerned that a taxonomic approach to designing social learning experiments can draw attention away from other relevant details of the tasks and contexts being studied and may lie at the root of some of the more persistent controversies in the comparative psychology literature. This is why we call for an emphasis on ecologically valid experiments moving forward, a suggestion strongly supported by Whiten.

Tennie also suggests that archaeologists' success in reverse engineering ancient technologies argues against social transmission of action details in the past. We already presented an extended argument against this criterion of "possibility," but it is interesting to speculate about reverse engineering end-state emulation (Whiten et al. 2004) as a mechanism. As we found no functional differences between alternate tool forms that might bias copying, any such reverse engineering effort can only be seen as an attempt to reproduce the perceived (and causally opaque) morphological goals of the knapper. In other words, this would be what Tomasello terms "imitation" and archaeologists have called "the imposition of arbitrary form" (Holloway 1969) and hailed as a hallmark of human culture (Isaac 1976). Perhaps the theoretical exchange between comparative psychology and archaeology can profitably flow in both directions.

Skill Reproduction

Whiten describes knapping as a "paradigmatic" example of an opaque skill requiring repeated alternation of social (imitation/goal emulation) and individual (trial and error) learning (see also Isaac 1986; Stout 2013). Greenbaum et al. call attention to similar ideas in recent behavioral ecology literature. This includes formal modeling from Truskanov and Prat (2018) showing that, in realistically variable environments, the inclusion of individual trial-and-error learning during cultural transmission increases fidelity, whereas exact copying is fragile. We went so far as to argue that exact action copying is literally impossible (not to mention undesirable) in the real world and that some degree of goal-directed approximation is always necessary (de Vignemont and Haggard 2008). Cer-

tainly this is the case with stone knapping, in which consistent results must be produced from variable materials and skill acquisition requires extended individual practice (Stout 2013). This account of skill reproduction evokes Tomasello's intention-based concept of imitation while emphasizing that intentions exist across multiple levels of abstraction (Byrne and Russon 1998; Stout 2011). Formal (Mesoudi and O'Brien 2008; Truskanov and Prat 2018) and verbal (Stout 2013) models suggest that such multilevel organization is critical to both the learnability and the evolvability of cultural traits.

Due to the abstraction and flexibility of the behavior elements being copied (e.g., unifacial method), evolutionary models cited by Tennie that address the accumulation of small perceptual errors in functionally neutral metric attributes of artifact morphology are not relevant. We instead point to the multifaceted "quantitative genetic" approach of Lycett and von Cramon-Taubadel (2015) that recognizes the complex relations between knapping behaviors and observed artifact form as influenced by production methods, reuse, and discard as well as functional and raw material constraints. Thus, while Greenbaum et al. advocate the application of phylogenetic methods to Oldowan variation (as has been done for later time periods; e.g., O'Brien, Darwent, and Lyman 2001), we are concerned that substantial questions about the proper units of analysis for Early Stone Age technology remain to be addressed. Some combination of the site comparison approach cited by Whiten (Whiten, Schick, and Toth 2009) with the standardized, multisite collaborative research program advocated by Reti may be a promising direction.

Cumulative Culture Origins

As Hovers notes, our article does not demonstrate the presence or absence of cultural ratcheting in the Oldowan. This is beyond current evidence and methods. Instead, we sought to "propose and exemplify a research program" for investigating such questions. Greenbaum et al., Högberg, Lombard, and Gärdenfors, and Reti endorse our gradualistic approach, which sees cumulative culture as a complex phenomenon with multiple contributing elements interacting in complicated ways. Reti highlights economic factors as drivers of technological change, stability, and variation, while Greenbaum et al. discuss the potential for social learning to either accelerate (Baldwin effect) or slow (reverse Baldwin) the biological evolution of toolmaking skills. Both comments align with our scenario's focus on the costs and benefits of investment in skill learning, which are themselves contingent on bioculturally evolving learning aptitudes, adaptive strategies, life-history, and technological systems (Isler and van Schaik 2014; Kolodny, Creanza, and Feldman 2015; Stout 2011; Stout and Hecht 2017). We favor the hypothesis that social and individual learning overlap cognitively and developmentally and so tend to evolve together (Heyes 2012; Stout and Hecht 2017; van Schaik and Burkart 2011). This would argue against a reverse Baldwin effect insofar as innovative potential might

be linked to other capacities that remain visible to selection. More broadly, it suggests a kind of "biocultural ratchet effect" producing unidirectional or equifinal evolutionary responses to diverse and shifting selection pressures. As suggested by Greenbaum et al., the interaction between social and individual learning is a key question for further biological and cultural evolutionary research. A discrete "time of emergence" (Hovers) for cumulative culture may or may not ultimately be identifiable in the record, but Whiten's suggestion that cumulative culture should be defined "simply as culture that cumulatively builds on what came before" at least provides an empirically tractable criterion.

Archaeological Evidence of Teaching and Learning?

In various publications (Stout 2002, 2005; Stout and Hecht 2017; Stout and Khreisheh 2015), we have explored the extent to which the demands of knapping skill acquisition might imply social support ranging from incidental "learning niche" construction to direct, active teaching. We are sympathetic to the suggestion from Högberg, Lombard, and Gärdenfors that evaluative feedback and intentional demonstration were present in the Oldowan but are concerned that this remains difficult to establish with confidence. As we argued in the article, competent flaking plus simple heuristics can support exhaustive and systematic core reduction (Moore 2011; Moore and Perston 2016), as seen at both Gona (Stout et al. 2010) and Lokalalei 2C (Delagnes and Roche 2005). Further experimental work is needed to evaluate the likelihood of developing these heuristics under different learning conditions ("ecologically valid studies" in fig. 1, step 2). As suggested by Reti, this experimental program needs to be integrated with systematic, quantitative comparisons across sites using standardized methodology. Obviously, this is a major undertaking but should be our objective.

For similar reasons, we did not have much to say about Lomekwian toolmaking (thus disappointing Högberg, Lombard, and Gärdenfors, Reti, and Whiten). As Hovers explains, we simply have far more information about Oldowan toolmaking. We might speculate that the bipolar and passive hammer techniques described at Lomekwi 3 (Harmand et al. 2015; Lewis and Harmand 2016) allow expedient flake production (Putt 2015) with lower investments in skill acquisition. This would be consistent with our suggestion that the rarity of stone tool sites prior to 2.0 Ma may be related to high learning costs and low marginal values for lithic technology. However, robust interpretation awaits publication of more extensive technological analyses, experimental replication, and expanded excavations. The latter might also lay to rest any lingering concerns about artifact context (Domínguez-Rodrigo and Alcalá 2016).

We share the concerns of Högberg, Lombard, and Gärdenfors and Whiten regarding the limitations of argument by elimination, which we consider to be the "least worst" solution currently available. However, we are encouraged by broad

support for our bottom-up approach (Högberg, Lombard, and Gärdenfors, Hovers, and Reti). To clarify, we did not state that the goal of efficient edge production was insufficient to explain Oldowan technology (Högberg, Lombard, and Gärdenfors) but simply that this is not known with certainty. There is substantial bottom-up evidence supporting the inference, and our analysis of "Favor" was an attempt to add to this. More generally, we agree with Greenbaum et al., Reti, and Whiten that a broad-based empirical research program is the best way forward.

—Dietrich Stout, Michael J. Rogers, Adrian V. Jaeggi, and Sileshi Semaw

References Cited

Acerbi, A., and Claudio Tennie. 2016. The role of redundant information in cultural transmission and cultural stabilization. *Journal of Comparative Psychology* 130(1):62–70. [CT]

Aisner, R., and J. Terkel. 1992. Ontogeny of pine cone opening behaviour in the black rat, *Rattus rattus. Animal Behaviour* 44:327–336. [GG/NT/NC/SE/MWF/OK]

Andersson, Claes. 2011. Paleolithic punctuations and equilibria: did retention rather than invention limit technological evolution? *PaleoAnthropology* 2011:243–259. [EH]

Antón, Susan C., Richard Potts, and Leslie C. Aiello. 2014. Evolution of early *Homo*: an integrated biological perspective. *Science* 345(6192):1236828.

Aplin, L. M., D. R. Farine, J. Morand-Ferron, A. Cockburn, A. Thornton, and B. C. Sheldon. 2015. Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* 518:538–541. [GG/NT/NC/SE/MWF/OK]

Aplin, L. M., B. C. Sheldon, and J. Morand-Ferron. 2013. Milk bottles revisited: social learning and individual variation in the blue tit, *Cyanistes caeruleus. Animal Behaviour* 85:1225–1232. [GG/NT/NC/SE/MWF/OK]

Arbib, Michael A., James Bonaiuto, Stéphane Jacobs, and Scott Frey. 2009. Tool use and the distalization of the end-effector. *Psychological Research* 73(4):441–462.

Arthur, Kathryn Weedman. 2018. *The lives of stone tools: crafting the status, skill, and identity of flintknappers*. Tucson: University of Arizona Press. [AH/ML/PG]

Asfaw, Berhane, Tim White, Owen Lovejoy, Bruce Latimer, Scott Simpson, and Gen Suwa. 1999. *Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* 284(5414):629–635.

Ayebare, S., D. Moyer, A. J. Plumptre, and Y. Wang. 2011. Chronology of the Acheulean to Middle Stone Age transition in eastern Africa. *Science* 360:95–98. [GG/NT/NC/SE/MWF/OK]

Bandini, E., and Claudio Tennie. 2017. Spontaneous reoccurrence of "scooping," a wild tool-use behaviour, in naïve chimpanzees. *PeerJ* 5(e3814):1–26. [CT]

Bartoń, Kamil. 2015. *MuMIn: multimodel inference*. https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf.

Bentley-Condit, V. K., and E. O. Smith. 2010. Animal tool use: current definitions and an updated comprehensive catalog. *Behaviour* 147:185–221. [GG/NT/NC/SE/MWF/OK]

Bernstein, Nicholai A. 1996. On dexterity and its development. In *Dexterity and its development*. M. L. Latash, ed. Pp. 3–235. Mahwah, NJ: Erlbaum.

Bouckaert, R., P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337:957–960. [GG/NT/NC/SE/MWF/OK]

Boyd, Robert, Peter J. Richerson, and Joseph Henrich. 2011. The cultural niche: why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences of the USA* 108(suppl. 2):10918–10925.

Braun, David R., and J. W. K. Harris. 2003. Technological developments in the Oldowan of Koobi Fora: innovative techniques for artifact analysis. In *Oldowan: rather more than smashing stones*. J. M. Moreno, R. M. Torcal, and I. de la Torre, eds. Bellaterra: Universitat Autonoma de Barcelona.

Braun, David R., Thomas Plummer, Peter Ditchfield, Laura C. Bishop, and Joseph Ferraro. 2009. Oldowan technology and raw material variability at Kanjera South. In *Interdisciplinary approaches to the Oldowan*. Erella Hovers and David R. Braun, eds. Dordrecht: Springer Science+Business Media B.V.

Bril, Blandine, Ross Parry, and Gilles Dietrich. 2015. How similar are nut-cracking and stone-flaking? a functional approach to percussive technology. *Philosophical Transactions of the Royal Society B* 370(1682):20140355.

Burnham, Kenneth P., and David R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.

Byrne, Richard. 1999. Imitation without intentionality: using string parsing to copy the organization of behaviour. *Animal Cognition* 2(2):63–72.

———. 2007. Culture in great apes: using intricate complexity in feeding skills to trace the evolutionary origin of human technical prowess. *Philosophical Transactions of the Royal Society B* 362:577–585.

Byrne, Richard, and Anne E. Russon. 1998. Learning by imitation: a hierarchical approach. *Behavioral and Brain Sciences* 21:667–721.

Cavalli-Svorza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21:550–570. [GG/NT/NC/SE/MWF/OK]

Clarkson, Chris. 2013. Measuring core reduction using 3D flake scar density: a test case of changing core reduction at Klasies River Mouth, South Africa. *Journal of Archaeological Science* 40(12):4348–4357.

Clay, Z., and Claudio Tennie. 2017. Is overimitation a uniquely human phenomenon? insights from human children as compared to bonobos. *Child Development* 10.1111/cdev.12857. [CT]

Coolidge, Frederick L., and Thomas Wynn. 2009. *The rise of Homo sapiens: the evolution of modern thinking*. Oxford: Oxford University Press. [AH/ML/PG]

———. 2016. An introduction to cognitive archaeology. *Current Directions in Psychological Science* 25(6):386–392. [AH/ML/PG]

Corbey, Raymond, Adam Jagich, Krist Vaesen, and Mark Collard. 2016. The Acheulean handaxe: more like a bird's song than a Beatles' tune? *Evolutionary Anthropology: Issues, News, and Reviews* 25(1):6–19.

Custance, Deborah, Andrew Whiten, and Tamar Fredman. 1999. Social learning of an artificial fruit task in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology* 113(1):13–23.

Cutting, Nicola, Ian A. Apperly, Jackie Chappell, and Sarah R. Beck. 2014. The puzzling difficulty of tool innovation: why can't children piece their knowledge together? *Journal of Experimental Child Psychology* 125:110–117.

Deacon, T. W. 2003. Multilevel selection in a complex adaptive system: the problem of language origins. In *Evolution and learning: the Baldwin Effect reconsidered*. W. H. Bruce and D. J. Depew, eds. Pp. 81–106. Cambridge, MA: MIT Press. [GG/NT/NC/SE/MWF/OK]

Dean, Lewis G., Gill L. Vale, Kevin N. Laland, Emma Flynn, and Rachel L. Kendal. 2014. Human cumulative culture: a comparative perspective. *Biological Reviews* 89(2):284–301.

Delagnes, Anne, and Helene Roche. 2005. Late Pliocene hominid knapping skills: the case of Lokalalei 2C, West Turkana, Kenya. *Journal of Human Evolution* 48(5):435–472.

de Vignemont, Frederique, and Patrick Haggard. 2008. Action observation and execution: what is shared? *Social Neuroscience* 3(3–4):421–433.

Domínguez-Rodrigo, Manuel, and Luis Alcalá. 2016. 3.3-million-year-old stone tools and butchery traces? more evidence needed. *PaleoAnthropology* 2016:46–53.

Eerkens, J. W., and C. P. Lipo. 2005. Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology* 24(4):316–334. [CT]

Enquist, M., S. Ghirlanda, A. Jarrick, and C. A. Wachtmeister. 2008. Why does human culture increase exponentially? *Theoretical Population Biology* 74(1):46–55.

Eren, Metin I., Stephen J. Lycett, Robert J. Patten, Briggs Buchanan, Justin Pargeter, and Michael J. O'Brien. 2016. Test, model, and method validation: the role of experimental stone artifact replication in hypothesis-driven archaeology. *Ethnoarchaeology* 8(2):103–136.

Felsenstein, J. 2004. *Inferring phylogenies*. Sunderland, MA: Sinauer. [GG/NT/NC/SE/MWF/OK]

Fragaszy, Dorothy M., D. Biro, Y. Eshchar, T. Humle, P. Izar, B. Resende, and E. Visalberghi. 2013. The fourth dimension of tool use: temporally enduring artefacts aid primates learning to use tools. *Philosophical Transactions of the Royal Society B* 368(1630):20120410.

Galef, Bennett G. 2012. Social learning and traditions in animals: evidence, definitions, and relationship to human culture. *Wiley Interdisciplinary Reviews: Cognitive Science* 3(6):581–592. [EH]

———. 2013. Imitation and local enhancement: detrimental effects of consensus definitions on analyses of social learning in animals. *Behavioural Processes* 100:123–130. [GG/NT/NC/SE/MWF/OK]

———. 2015. Laboratory studies of imitation/field studies of tradition: towards a synthesis in animal social learning. *Behavioural Processes* 112:114–119. [GG/NT/NC/SE/MWF/OK]

Galef, Bennett G., Jr., and L.-A. Giraldeau. 2001. Social influences on foraging in vertebrates: causal mechanisms and adaptive functions. *Animal Behaviour* 61:3–15. [GG/NT/NC/SE/MWF/OK]

Gärdenfors, Peter, and Anders Högberg. 2017. The archaeology of teaching and the evolution of *Homo docens*. *Current Anthropology* 58(2):188–208. [AH/ML/PG]

Garofoli, D. 2016. Metaplasticit-*ies*: material engagement meets mutational enhancement. In *Embodiment in evolution and culture*. G. Etzelmüller and C. Tewes, eds. Pp. 307–336. Tübingen: Mohr Siebeck. [AH/ML/PG]

Geribas, N., and M. M. Verges. 2010. What novice knappers have to learn to become expert stone toolmakers. *Journal of Archaeological Science* 37:2857–2870. [AW]

Goldman-Neuman, Talia, and Erella Hovers. 2012. Raw material selectivity in Late Pliocene Oldowan sites in the Makaamitalu Basin, Hadar, Ethiopia. *Journal of Human Evolution* 62:353–366. [EH]

Gray, R. D., and Q. D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435–439. [GG/NT/NC/SE/MWF/OK]

Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science* 323:479–483. [GG/NT/NC/SE/MWF/OK]

Gray, R. D., S. J. Greenhill, and R. M. Ross. 2007. The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory* 2:360–375. [GG/NT/NC/SE/MWF/OK]

Gurtov, Alia N., and Metin I. Eren. 2014. Lower Paleolithic bipolar reduction and hominin selection of quartz at Olduvai Gorge, Tanzania: what's the connection? *Quaternary International* 322:285–291.

Hadfield, Jarrod D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* 33(2):1–22.

———. 2014. *MCMCglmm course notes* (accessed August 7, 2014). https://cran.r-project.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf.

Haidle, Miriam N., Michael Bolus, Mark Collard, Nicholas J. Conard, Duilio Garofoli, Marlize Lombard, April Nowell, Claudio Tennie, and Andrew Whiten. 2015. The nature of culture: an eight-grade model for the evolution and expansion of cultural capacities in hominins and other animals. *Journal of Anthropological Sciences* 93:43–70. doi:10.4436/jass.93011. [AH/ML/PG]

Hamilton, M. J., and B. Buchanan. 2009. The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: quantifying Clovis evolutionary dynamics. *Journal of Anthropological Archaeology* 28(1):55–69. [CT]

Harmand, Sonia, Jason E. Lewis, Craig S. Feibel, Christopher J. Lepre, Sandrine Prat, Arnaud Lenoble, Xavier Boës, et al. 2015. 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature* 521 (7552):310–315.

Hecht, Erin E., D. A. Gutman, N. Khreisheh, S. V. Taylor, J. Kilner, A. A. Faisal, B. A. Bradley, T. Chaminade, and D. Stout. 2014. Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Structure and Function* 220(4):2315–2331.

Henrich, Joseph. 2015. *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press.

Heyes, Cecilia. 2012. What's social about social learning? *Journal of Comparative Psychology* 126(2):193.

———. 2018. *Cognitive gadgets: the cultural evolution of thinking*. New York: Harvard University Press. [CT]

Hobaiter, C., T. Poisot, K. Zuberbühler, W. Hoppitt, and T. Gruber. 2014. Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biology* 12(9):e1001960. [CT]

Högberg, Anders. 2009. *Lithic in the Scandinavian Late Bronze Age: sociotechnical change and persistence*. BAR International Series 1932. Oxford: Archaeopress. [AH/ML/PG]

Holloway, Ralph. 1969. Culture: a human domain. *Current Anthropology* 10:395–412.

Hopper, L. M., S. P. Lambeth, S. J. Schapiro, and Andrew Whiten. 2008. Observational learning in chimpanzees and children studied through "ghost" conditions. *Proceedings of the Royal Society B* 275(1636):835–840. [CT]

Hoppitt. W., and Kevin N. Laland. 2008. Social processes influencing learning in animals: a review of the evidence. *Advances in the Study of Behavior* 38:105–165. [GG/NT/NC/SE/MWF/OK]

Hovers, Erella. 2009. Learning from mistakes: flaking accidents and knapping skills in the assemblage of A. L. 894 (Hadar, Ethiopia). In *The cutting edge: new approaches to the archaeology of human origins*. K. Schick and N. Toth, eds. Pp. 137–150. Gosport: Stone Age Institute. [EH]

———. 2012. Invention, re-invention, innovations: makings of the Oldowan. In *Origins of human creativity and innovation*. S. Elias, ed. Pp. 51–68. Amsterdam: Elsevier. [EH]

Hrubesch, C., S. Preuschoft, and Carel van Schaik. 2009. Skill mastery inhibits adoption of observed alternative solutions among chimpanzees (*Pan troglodytes*). *Animal Cognition* 12:209. [GG/NT/NC/SE/MWF/OK]

Ingold, Tim. 1998. From complementarity to obviation: on dissolving the boundaries between social and biological anthropology, archaeology and psychology. *Zeitschrift für Ethnologie* 123:21–52.

Isaac, Glynn L. 1976. Stages of cultural elaboration in the Pleistocene: possible archaeological indicators of the development of language capabilities. *Annals of the New York Academy of Sciences* 280(1):275–288.

———. 1984. The archaeology of human origins: studies of the Lower Pleistocene in East Africa, 1971–1981. *Advances in World Archaeology* 3:1–86.

———. 1986. Foundation stones: early artefacts as indicators of activities and abilities. In *Stone Age rehistory: studies in honor of Charles McBurney*. G. N. Bailey and P. Callow, eds. Pp. 221–241. London: Cambridge University Press.

Isler, Karin, and Carel P. van Schaik. 2014. How humans evolved large brains: comparative evidence. *Evolutionary Anthropology: Issues, News, and Reviews* 23(2):65–75.

Kimbel, William H., Robert C. Walter, Donald C. Johanson, Kaye E. Reed, James L. Aronson, Zelalem Assefa, Curtis W. Marean, G. G. Eck, R. Bobe, and Erella Hovers. 1996. Late Pliocene *Homo* and Oldowan tools from the Hadar Formation (Kada Hadar Member), Ethiopia. *Journal of Human Evolution* 31(6):549–561.

Kivell, Tracy L. 2015. Evidence in hand: recent discoveries and the early evolution of human manual manipulation. *Philosophical Transactions of the Royal Society B* 370(1682):20150105.

Klein, Richard G. 2000. Archaeology and the evolution of human behavior. *Evolutionary Anthropology Issues News and Reviews* 9(1):17–36. [AH/ML/PG]

Kolodny, Oren, Nicole Creanza, and Marcus W. Feldman. 2015. Evolution in leaps: the punctuated accumulation and loss of cultural innovations. *Proceedings of the National Academy of Sciences of the USA* 112(49):6762–6769. [EH]

———. 2016. Game-changing innovations: how culture can change the parameters of its own evolution and induce abrupt cultural shifts. *PLoS Computational Biology* 12(12):e1005302.

Laland, Kevin N., Tobias Uller, Marcus W. Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, and John Odling-Smee. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B* 282(1813):20151019.

Lee, S., and T. Hasegawa. 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B* 278:3662–3669. [GG/NT/NC/SE/MWF/OK]

Lemorini, Cristina, Thomas W. Plummer, David R. Braun, Alyssa N. Crittenden, Peter W. Ditchfield, Laura C. Bishop, Fritz Hertel, James S. Oliver, Frank W. Marlowe, and Margaret J. Schoeninger. 2014. Old stones' song: use-wear experiments and analysis of the Oldowan quartz and quartzite assemblage from Kanjera South (Kenya). *Journal of Human Evolution* 72:10–25.

Lewis, H. M., and Kevin N. Laland. 2012. Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B* 367(1599):2171–2180. [CT]

Lewis, Jason E., and Sonia Harmand. 2016. An earlier origin for stone tool making: implications for cognitive evolution and the transition to *Homo*. *Philosophical Transactions of the Royal Society B* 371(1698):20150233.

Lombard, Marlize, Anders Högberg, and Miriam N. Haidle. Forthcoming. Cognition: from capuchin rock pounding to Lomekwian flake production. *Cambridge Archaeological Journal*, doi:10.1017/S0959774318000550. [AH/ML/PG]

Luncz, L. V., and C. Boesch. 2014. Tradition over trend: neighboring chimpanzee communities maintain differences in cultural behaviour despite frequent immigration of adult females. *American Journal of Primatology* 76:649–657. [AW]

Lycett, Stephen J., and Noreen von Cramon-Taubadel. 2015. Toward a "quantitative genetic" approach to lithic variation. *Journal of Archaeological Method and Theory* 22(2):646–675.

Lyons, D. E., A. G. Young, and F. C. Keil. 2007. The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the USA* 104:19751–19756. [AW]

Mace, R., and C. J. Holden. 2005. A phylogenetic approach to cultural evolution. *Trends in Ecology and Evolution* 20(3):116–121. [GG/NT/NC/SE/MWF/OK]

Mace, R., C. J. Holden, and S. Shennan. 2005. *The evolution of cultural diversity: a phylogenetic approach*. Walnut Creek, CA: Left Coast. [GG/NT/NC/SE/MWF/OK]

Maravita, Angelo, and Atsushi Iriki. 2004. Tools for the body (schema). *Trends in Cognitive Sciences* 8(2):79–86.

Marshall-Pescini, S., and Andrew Whiten. 2008. Chimpanzees (*Pan troglodytes*) and the question of cumulative culture: an experimental approach. *Animal Cognition* 11:449–456. [GG/NT/NC/SE/MWF/OK]

Mendoza Straffon, L. 2016. *Cultural phylogenetics: concepts and applications in archaeology*. Dordrecht: Springer International. [GG/NT/NC/SE/MWF/OK]

Mesoudi, Alex, and Michael J. O'Brien. 2008. The learning and transmission of hierarchical cultural recipes. *Biological Theory* 3(1):63–72.

Moore, Mark W. 2011. The design space of stone flaking: implications for cognitive evolution. *World Archaeology* 43(4):702–715.

Moore, Mark W., and Yinika Perston. 2016. Experimental insights into the cognitive significance of early stone tools. *PLoS ONE* 11(7):e0158803.

Morgan, Thomas J. H. 2016. Testing the cognitive and cultural niche theories of human evolution. *Current Anthropology* 57(3):370–377. [EH, CT]

Morgan, Thomas J. H., Natalie T. Uomini, L. E. Rendell, L. Chouinard-Thuly, S. E. Street, H. M. Lewis, C. P. Cross, et al. 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications* 6:6029.

Nielsen M., and K. Tomaselli. 2010. Over-imitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science* 21:729–736. [AW]

Nonaka, Tetsushi, Blandine Bril, and Robert Rein. 2010. How do stone knappers predict and control the outcome of flaking? implications for understanding early stone tool technology. *Journal of Human Evolution* 59(2):155–167.

O'Brien, Michael J., John Darwent, and R. Lee Lyman. 2001. Cladistics is useful for reconstructing archaeological phylogenies: palaeoindian points from the southeastern United States. *Journal of Archaeological Science* 28(10):1115–1136.

Panger, Melissa A., Alison S. Brooks, Brian G. Richmond, and Bernard Wood. 2002. Older than the Oldowan? rethinking the emergence of hominin tool use. *Evolutionary Anthropology: Issues, News, and Reviews* 11(6):235–245. [EH]

Plummer, Thomas W., and Laura C Bishop. 2016. Oldowan hominin behavior and ecology at Kanjera South, Kenya. *Journal of Anthropological Sciences* 94:29–40.

Pope, S. M., J. P. Taglialatela, S. A. Skiba, and W. D. Hopkins. 2018. Changes in frontoparietotemporal connectivity following do-as-I-do imitation training in chimpanzees (*Pan troglodytes*). *Journal of Cognitive Neuroscience* 30(3):421–431. [CT]

Povinelli, Daniel J., and Scott H. Frey. 2016. Constraints on the exploitation of the functional properties of objects in expert tool-using chimpanzees (*Pan troglodytes*). *Cortex* 82:11–23.

Powell, Adam, Stephen Shennan, and Mark G. Thomas. 2009. Late Pleistocene demography and the appearance of modern human behavior. *Science* 324(5932):1298–1301.

Pradhan, Gauri R., Claudio Tennie, and Carel P. van Schaik. 2012. Social organization and the evolution of cumulative technology in apes and hominins. *Journal of Human Evolution* 63(1):180–190.

Proffitt, Tomos, Lydia V. Luncz, Tiago Falótico, Eduardo B. Ottoni, Ignacio de la Torre, and Michael Haslam. 2016. Wild monkeys flake stone tools. *Nature* 539(7627):85–88.

Putt, Shelby S. 2015. The origins of stone tool reduction and the transition to knapping: an experimental approach. *Journal of Archaeological Science: Reports* 2:51–60.

Putt, Shelby S., Alexander D. Woods, and Robert G. Franciscus. 2014. The role of verbal interaction during experimental bifacial stone tool manufacture. *Lithic Technology* 39(2):96–112.

Read, Dwight, and Sander van der Leeuw. 2008. Biology is only part of the story. . . . *Philosophical Transactions of the Royal Society B* 363(1499):1959–1968. [EH]

Rein, Robert, Tetsushi Nonaka, and Blandine Bril. 2014. Movement pattern variability in stone knapping: implications for the development of percussive traditions. *PLoS ONE* 9(11):e113567.

Reindl, E., I. A. Apperly, S. R. Beck, and Claudio Tennie. 2017. Young children copy cumulative technological design in the absence of action information. *Scientific Reports* 7, article 1788.

Reindl, E., S. R. Beck, I. A. Apperly, and Claudio Tennie. 2016. Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B* 283(1825).

Reindl', E., and Claudio Tennie. 2018. Young children fail to generate an additive ratchet effect in an open-ended construction task. *PloS ONE* 13(6):e0197828. [CT]

Reti, Jay S. 2013. Methods for determining differential behaviors in stone tool production and application to the Oldowan of Olduvai Gorge, Tanzania and Koobi Fora, Kenya. PhD dissertation, Rutgers University. [JSR]

———. 2016. Quantifying Oldowan stone tool production at Olduvai Gorge, Tanzania. *PLoS ONE* 11(1):e0147352.

Richerson, Peter J., and Robert Boyd. 2005. *Not by genes alone: how culture transformed human evolution*. Chicago: University of Chicago Press.

Rogers, Michael, Sileshi Semaw, Dietrich Stout, and Naomi E. Levin. 2013. New excavations at the 2.6-million-year-old OGS-7 site, Gona, Ethiopia. *PaleoAnthropology* 2013:A33.

Rolian, Campbell, Daniel E. Lieberman, and John Paul Zermeno. 2011. Hand biomechanics during simulated stone tool use. *Journal of Human Evolution* 61(1):26–41. [EH]

Schick, Kathy D. 1987. Modeling the formation of Early Stone Age artifact concentrations. *Journal of Human Evolution* 16:789–807.

Schillinger, Kerstin, Alex Mesoudi, and Stephen J. Lycett. 2015. The impact of imitative versus emulative learning mechanisms on artifactual variation: implications for the evolution of material culture. *Evolution and Human Behavior* 36(6):446–455.

Semaw, Sileshi. 1997. Late Pliocene archaeology of the Gona River deposits, Afar, Ethiopia. PhD dissertation, Rutgers University.

———. 2000. The world's oldest stone artefacts from Gona, Ethiopia: their implications for understanding stone technology and patterns of human evolution 2.6–1.5 million years ago. *Journal of Archaeological Science* 27:1197–1214.

Semaw, Sileshi, Michael J. Roger, Jay Quade, Paul R. Renne, Robert F. Butler, Manuel Dominguez-Rodrigo, Dietrich Stout, William S. Hart, Travis Pickering, and Scott W. Simpson. 2003. 2.6-million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *Journal of Human Evolution* 45:169–177.

Sharon, Gonen. 2009. Acheulian giant core technology: a worldwide perspective. *Current Anthropology* 50(3):335–367.

Shea, John J. 2017. Occasional, obligatory, and habitual stone tool use in hominin evolution. *Evolutionary Anthropology* 26:200–217. [AH/ML/PG]

Shettleworth, S. J. 2010. *Cognition, evolution, and behavior*. Oxford: Oxford University Press. [GG/NT/NC/SE/MWF/OK]

Shipton, C., and M. Nielsen. 2015. Before cumulative culture. *Human Nature* 26:331–345. [AW]

Steel, M. 2016. *Phylogeny: discrete and random processes in evolution*. Philadelphia: SIAM-Society for Industrial and Applied Mathematics. [GG/NT/NC/SE/MWF/OK]

Stout, Dietrich. 2002. Skill and cognition in stone tool production: an ethnographic case study from Irian Jaya. *Current Anthropology* 45(3):693–722.

———. 2005. The social and cultural context of stone-knapping skill acquisition. In *Stone knapping: the necessary conditions for a uniquely hominin behaviour*. V. Roux and B. Bril, eds. Pp. 331–340. Cambridge: McDonald Institute for Archaeological Research.

———. 2011. Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B* 366(1567):1050–1059.

———. 2013. Neuroscience of technology. In *Cultural evolution: society, technology, language, and religion*. P. J. Richerson and M. Christiansen, eds. Pp. 157–173. Cambridge, MA: MIT Press.

———. 2018. Human evolution: history or science? In *Rethinking human evolution*. J. H. Schwartz, ed. Pp. 297–317. Cambridge, MA: MIT Press.

Stout, Dietrich, and Thierry Chaminade. 2007. The evolutionary neuroscience of tool making. *Neuropsychologia* 45(5):1091–1100.

Stout, Dietrich, and Erin E. Hecht. 2017. Evolutionary neuroscience of cumulative culture. *Proceedings of the National Academy of Sciences of the USA* 114(30):7861–7868.

Stout, Dietrich, and Nada Khreisheh. 2015. Skill learning and human brain evolution: an experimental approach. *Cambridge Archaeological Journal* 25(04):867–875.

Stout, Dietrich, Jay Quade, Sileshi Semaw, Michael Rogers, and Naomi Levin. 2005. Raw material selectivity of the earliest stone toolmakers at Gona, Afar, Ethiopia. *Journal of Human Evolution* 48(4):365–380.

Stout, Dietrich, and Sileshi Semaw. 2006. Knapping skill of the earliest stone toolmakers: insights from the study of modern human novices. In *The Oldowan: case studies into the earliest Stone Age*. N. Toth and K. Schick, eds. Pp. 307–320. Gosport, IN: Stone Age Institute.

Stout, Dietrich, Sileshi Semaw, Michael J. Rogers, and Dominique Cauche. 2010. Technological variation in the earliest Oldowan from Gona, Afar, Ethiopia. *Journal of Human Evolution* 58(6):474–491.

Stout, Dietrich, Nicholas Toth, Kathy D. Schick, and Thierry Chaminade. 2008. Neural correlates of Early Stone Age tool-making: technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society of London B* 363:1939–1949.

Tennie, Claudio, David R. Braun, Luke S. Premo, and Shannon McPherron. 2016. The island test for cumulative culture in the Paleolithic. In *The nature of culture*. M. N. Haidle, N. J. Conard, and M. Bolus, eds. Pp. 121–133. Dordrecht: Springer.

Tennie, Claudio, Josep Call, and Michael Tomasello. 2006. Push or pull: imitation vs. emulation in great apes and human children. *Ethology* 112(12):1159–1169. [CT]

———. 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B* 364(1528):2405–2415.

———. 2010. Evidence for emulation in chimpanzees in social settings using the floating peanut task. *PLoS ONE* 5(5):e10544. [CT]

———. 2012. Untrained chimpanzees (*Pan troglodytes schweinfurthii*) fail to imitate novel actions. *PLoS ONE* 7(8):e41548.

Tennie, Claudio, L. S. Premo, David R. Braun, and Shannon P. McPherron. 2017. Early stone tools and cultural transmission: resetting the null hypothesis. *Current Anthropology* 58(5):652–672.

Tocheri, Matthew W., Caley M. Orr, Marc C. Jacofsky, and Mary W. Marzke. 2008. The evolutionary history of the hominin hand since the last common ancestor of *Pan* and *Homo*. *Journal of Anatomy* 212(4):544–562. [EH]

Tomasello, Michael. 1996. Do apes ape? In *Social learning in animals: the roots of culture*. C. M. Heyes and B. G. Galef, eds. Pp. 319–346. San Diego, CA: Academic Press.

———. 1999. *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

———. 1998. Emulation learning and cultural learning. *Behavioral and Brain Sciences* 21(5):703–704.

Toth, Nicholas. 1985. The Oldowan reassessed: a close look at Early Stone artifacts. *Journal of Archaeological Science* 12(2):101–120.

———. 1987. Behavioral inferences from Early Stone artifact assemblages: an experimental model. *Journal of Human Evolution* 16(7–8):763–787.

Toth, Nicholas, Kathy D. Schick, and Sileshi Semaw. 2006. A comparative study of the stone tool-making skills of *Pan*, *Australopithecus*, and *Homo sapiens*. In *The Oldowan: case studies into the earliest Stone Age*. N. Toth and K. D. Schick, eds. Pp. 155–222. Gosport, IN: Stone Age Institute.

Truskanov, Noa, and A. Lotem. 2017. Trial-and-error copying of demonstrated actions reveals how fledglings learn to "imitate" their mothers. *Proceedings of the Royal Society B* 284:20162744. http://dx.doi.org/10.1098/rspb.2016.2744. [GG/NT/NC/SE/MWF/OK]

Truskanov, Noa, and Yosef Prat. 2018. Cultural transmission in an ever-changing world: trial-and-error copying may be more robust than precise imitation. *Philosophical Transactions of the Royal Society B* 373(1743).

Vale, Gillian L., Sarah J. Davis, Susan P. Lambeth, Steven J. Schapiro, and Andrew Whiten. 2017. Acquisition of a socially learned tool use sequence in chimpanzees: implications for cumulative culture. *Evolution and Human Behavior* 38(5):635–644.

van Schaik, Carel P., and Judith M. Burkart. 2011. Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B* 366(1567):1008–1016.

Villmoare, Brian, William H. Kimbel, Chalachew Seyoum, Christopher J. Campisano, Erin N. DiMaggio, John Rowan, David R. Braun, J. Ramón Arrowsmith, and Kaye E. Reed. 2015. Early *Homo* at 2.8 Ma from Ledi-Geraru, Afar, Ethiopia. *Science* 347(6228):1352–1355.

Voelkl, Bernhard, and Ludwig Huber. 2000. True imitation in marmosets. *Animal Behaviour* 60:195–202. [GG/NT/NC/SE/MWF/OK]

Weber, Bruce H., and David J. Depew, eds. 2003. *Evolution and learning: the Baldwin effect reconsidered*. Cambridge, MA: MIT Press.

Whiten, Andrew. 1998. Imitation of the sequential structure of actions by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 112:270–281. [GG/NT/NC/SE/MWF/OK]

———. 2011. The scope of culture in chimpanzees, humans and ancestral apes. *Philosophical Transactions of the Royal Society B* 366(1567):997–1007.

———. 2015. Experimental studies illuminate the cultural transmission of percussive technologies in *Homo* and *Pan*. *Philosophical Transactions of the Royal Society B* 370(1682):20140359.

Whiten, Andrew, G. Allen, S. Devlin, N. Kseib, N. Raw, and N. McGuigan. 2016. Social learning in the real-world: "over-imitation" occurs in both children and adults unaware of participation in an experiment and independently of social interaction. *PLoS ONE* 11:e0159920. [AW]

Whiten, Andrew, Jane Goodall, W. C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. E. G. Tutin, Richard W. Wrangham, and C. Boesch. 1999. Cultures in chimpanzees. *Nature* 399:682–685. [GG/NT/NC/SE/MWF/OK, CT]

Whiten, Andrew, Victoria Horner, and F. B. M. de Waal. 2005. Conformity to cultural norms of tool use in chimpanzees. *Nature* 437:737–740. [GG/NT/NC/SE/MWF/OK]

Whiten, Andrew, Victoria Horner, Carla A. Litchfield, and Sarah Marshall-Pescini. 2004. How do apes ape? *Animal Learning and Behavior* 32(1):36–52.

Whiten, Andrew, Nicola McGuigan, Sarah Marshall-Pescini, and Lydia M. Hopper. 2009. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B* 364(1528):2417–2428.

Whiten, Andrew, Kathy Schick, and Nicholas Toth. 2009. The evolution and cultural transmission of percussive technology: integrating evidence from palaeoanthropology and primatology. *Journal of Human Evolution* 57(4):420–435.

Whiten, Andrew, and Erica van de Waal. 2016. Social learning, culture and the "socio-cultural brain" of human and non-human primates. *Neuroscience and Biobehavioral Reviews* 82:58–75.

Wylie, Alison. 1985. The reaction against analogy. *Advances in Archaeological Method and Theory* 8:63–111.

Wynn, Thomas, R. Adriana Hernandez-Aguilar, Linda F. Marchant, and William C. McGrew. 2011. "An ape's view of the Oldowan" revisited. *Evolutionary Anthropology* 20(5):181–197. [AH/ML/PG, EH]