

The impact of hydraulic processes in Olduvai Beds I and II, Tanzania, through a particle dimension analysis of stone tool assemblages

Ignacio de la Torre^{1*}, Alfonso Benito² & Tomos Proffitt³

¹ Institute of Archaeology, University College London, 31-34, Gordon Square, WC1H 0PY London, United Kingdom

² CENIEH, Paseo Sierra de Atapuerca s/n, 09002 Burgos, Spain

³ RLAHA, School of Archaeology, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, United Kingdom

* Corresponding author: i.torre@ucl.ac.uk

ABSTRACT

The effect of post-depositional processes on the formation of Plio-Pleistocene sites at Olduvai Gorge is the subject of considerable debate, due mainly to its implications for the behavioral interpretation of the Beds I and II assemblages. In light of this debate, here we contribute to the discussion that investigates the role of water flow in site formation at Olduvai. This is achieved by assessing the artifact size and shape ranges of lithic assemblages excavated by Mary Leakey (1971), from both Oldowan (FLK North North Levels 3 and 1, FLK Zinj, FLK North Levels 6-1, Sandy Conglomerate and *Deinotherium*, HWKE Level 1) and early Acheulean (TK Lower and Upper Floor) sites. We apply grain size and shape statistical techniques to these stone tool assemblages in order to classify sites according to patterns in artifact dimensions. These patterns are then compared to those produced during experimental flaking, thus providing a referential framework against which the archaeological assemblages can be interpreted. Artifact size distribution results show pronounced differences between the archaeological and experimental assemblages. Most of the archaeological assemblages are characterized by a bimodal size trend that is opposed to the dominantly unimodal distribution seen in the flaking experiments. The few archaeological assemblages where the distribution is predominantly unimodal (TF Lower Floor, TK Upper Floor and FLK Zinj) also show a significant underrepresentation of smaller artifacts, when compared to the experimental distributions. Overall, the comparison of archaeological materials with experimental results enables a more accurate assessment of the impact of natural processes over the Bed I and II assemblages, and further, it helps to refine our understanding of taphonomic and behavioral contexts for the Oldowan and early Acheulean sites at Olduvai Gorge.

KEYWORDS: Olduvai Gorge; site formation processes; stone tool assemblages; water disturbance; artifact size and shape analysis

INTRODUCTION

Debitage size patterns have been widely used to address fluvial action over East African open-air sites (e.g., Harris, 1978; Potts, 1982; Schick, 1984, 1987, 1991, 1997; Petraglia and Potts, 1994). The ratio of microdebitage (i.e., lithics with length < 20mm) is considered to be especially sensitive to post-depositional disturbance (e.g., Fladmark, 1982), with smaller pieces readily winnowed by water action. Water is probably the most common abiotic agent of post-depositional disturbance in Paleolithic open-air sites (Isaac, 1967; Schick, 1984; Petraglia and Nash, 1987), and its role over the Olduvai assemblages has long been discussed (e.g., Potts, 1982; Pante and Blumenschine, 2010; Benito and de la Torre, 2011). Potts (1988) proposed that hydraulic disturbance could be assessed by considering paleogeography, bone water transport groups, sedimentology, artifact size distributions, edge rounding, and item orientation. Petraglia and Potts (1994) approached the impact of fluvial disturbance at Olduvai by studying edge rounding and size and weight classes of artifacts from a selection of Beds I and II sites. More recently, this same topic has been addressed through the analysis of artifact orientation patterns in the Beds I and II site maps (Benito-Calvo and de la Torre, 2011; de la Torre and Benito-Calvo, 2013).

The present work aims to reassess the effects of water disturbance at the Olduvai sites, by using three key proxies (i.e., size, shape, and weight classes) in archaeological assemblages excavated by Leakey, which are then compared to experimental results. This perspective has been applied successfully to the study of other Early Stone Age (e.g., Harris, 1978; Isaac and Isaac, 1997) and experimental (Schick, 1984) East African assemblages, but has never been applied at Olduvai Gorge. Size classes are considered to be an important factor for the evaluation of Olduvai site integrity (Leakey, 1971; Isaac and Crader, 1981; Potts, 1982), and thus dimensions of bones (Bunn, 1982; Dominguez-Rodrigo et al, 2007; Pante, 2010) and lithics (Petraglia and Potts, 1994; de la Torre and Mora, 2005) have been used to discuss site formation processes. However, apart from Petraglia and Potts (1994), size curves of Olduvai stone tools have not been produced, and not even the most detailed publication on Olduvai Gorge (Leakey, 1971) accounted for the size ranges of smaller objects. To date, a substantial part of Leakey's collections has remained unmeasured, due to the preferential attention given by researchers to the larger sized fraction of the artifact assemblages.

In this paper, we present new data from the direct analysis of 14 Bed I and II archaeological levels excavated by Leakey (1971) at Olduvai Gorge. Our size and shape class analysis is specifically designed to evaluate sorting caused by water disturbance, and follows our previous efforts on this topic by complementing orientation pattern results (Benito-Calvo and de la Torre, 2011; de la Torre and Benito-Calvo, 2013). More generally, by integrating first-hand analysis of a substantial part of the Leakey's collections with experimental results, this paper aims: 1- to create a referential framework to contextualize the impact of abiotic agents on archaeological assemblages; 2- to contribute to the debate on the formation processes of the Olduvai assemblages and the behavioral meaning of early archaeological sites.

MATERIALS AND METHODS

Despite successive reassessments of the Olduvai collections since the original study by Leakey (1971), only Petraglia and Potts (1994) presented a systematic analysis of artifact dimensions. Petraglia and Potts (1994) sampled the weight and length of stone tools and unmodified rocks

from FLK Zinj, HWKE Level 4 (hereafter HWKL4), MNK Main Site (all levels), FC West Occupation Floor, and TK Upper Floor (TKUF).

For this paper, we reanalyzed six of the sites originally classified by Leakey (1971: 258) as living floors, namely FLKNN Levels 1 and 3 (hereafter FLKNN1 and FLKNN3), FLK Zinj, HWKEL1, TKLF and TKUF (Figure 1 and Table 1), all of them reported to be preserved in very fine grained deposits (clay, tuff or silty/ sandy clays). Leakey (1971) considered living floors to be vertically-constrained assemblages on paleosol/ land surfaces, and assumed that remains were in primary position. Seven assemblages from FLK North were also analyzed, in order to provide a comparison to the allegedly undisturbed living floors: FLK North Levels 6-1 and *Deinotherium* Level are in clays or silty clays (Leakey, 1971: 62 and 85), whilst the FLK North Sandy Conglomerate assemblage was excavated by Leakey within a sandy and gravely deposit. Most of the analyzed assemblages are stratigraphically positioned in Bed I (FLKNN1 and FLKNN3, FLK Zinj, FLK North Levels 6-1), two are in Lower Bed II (FLK North *Deinotherium* Level and HWKEL1), one in Middle Bed II (FLK North Sandy Conglomerate), and two in Upper Bed II (TKLF and TKUF).

Our analysis was conducted in 2002 and 2010, when the Leakey collections were still housed at the National Museum of Kenya in Nairobi (i.e., before they were moved to their present location in the National Museum of Tanzania in Dar es Salaam). Rather than sampling part of the collections or recording some variables only (Potts and Petraglia, 1994; de la Torre and Mora, 2005), we conducted a complete dimensional analysis of all artifacts, which was only constrained by the physical availability of collections at the time of study at the museum; Table 1 shows that, in several instances, virtually all of the artifacts from Leakey's original assemblages were analyzed, whereas for other sites our figures are actually higher than those published by Leakey (1971).

The archaeological material (Figure 2A) was classified as modified or unmodified according to criteria laid out by de la Torre and Mora (2005). Complete dimensions, irrespective of technological axis (maximum length, width, thickness), and weight were recorded for all archaeological lithics with millimeter calipers and a digital balance (0.1 g precision). In total, 13,041 complete and broken artifacts and unmodified rocks were measured. However, given doubts over the contextual relationships between modified and unmodified rocks at Olduvai (de la Torre and Mora, 2005b), our analysis will only consider clearly anthropogenic artifacts (n=12,083) (Table 1).

The ratio between large and small artifacts has been considered a useful proxy when investigating the impact of water flow at Olduvai (Petraglia and Potts, 1994), by considering that the frequency of smaller pieces will decrease as hydraulic disturbance increases. However, there is no consensus on the size threshold to consider a piece as *large* or *small*, and uncertainties are accentuated in the case of technologically-variable assemblages (e.g., Oldowan vs. Acheulean), such as those from Beds I and II. Thus, rather than calculating a large-to-small artifact ratio (Petraglia and Potts, 1994), here we followed Isaac's (1986) distinction between flaked (i.e., cores and retouched tools) and detached pieces (i.e., debitage), which sheds light on the degree of artifact fragmentation, and has proven useful in assessing post-depositional disturbance at Olduvai (de la Torre and Mora, 2005).

Artifact weight is considered a key variable to assess raw material transport strategies at Olduvai (Potts, 1988; de la Torre and Mora, 2005; Blumenschine et al, 2008), and has also been used to address post-depositional disturbance (Potts, 1982; Petraglia and Potts, 1994). In their study of weight distributions, Petraglia and Potts (1994) calculated mean artifact weight.

This is not an entirely satisfactory index when comparing Oldowan and Acheulean assemblages, given that, by definition, the latter will contain larger and heavier artifacts (de la Torre and Mora, 2005), irrespective of the site formation processes involved. Keeping this issue in mind and for the sake of comparison with Petraglia and Potts' (1994) results, we also calculate the mean weight index in our datasets.

With regards to the experimental dataset, nine assemblages were knapped by two individuals through freehand hard hammer percussion, using a quartzite hammerstone. In total, four quartzite blocks, three lava cobbles and two chert nodules (all sourced locally at Olduvai Gorge) were knapped. This selection of raw materials aimed to represent the most common rock types used in Olduvai Beds I and II (Leakey, 1971). Experimental flaking was carried out following protocols set out by Proffitt and de la Torre (2014), where each core was reduced in order to maximize flake production, rather than to produce specific core forms.

All knapping experiments were carried out on tarpaulin, which allowed the total recovery of artifacts (Table 2). All experimental pieces larger than 10mm were measured using calipers. All microdebitage (<10mm) was sorted by size, using a series of nesting sieves, with decreasing mesh apertures (5.6mm, 4mm, 2.8mm, 2mm, 1mm, 0.05mm, 0.025mm). A series of automated measurements, including maximum length, width, perimeter, and area, were taken for 0.5-10mm microdebitage pieces using high quality macro photographs analyzed by the open source seed phenotyping program SmartGrain (Tanabata et al, 2012). Microdebitage was photographed collectively in groups of up to 700 individual pieces. Photographs were then loaded into SmartGrain, with measurements being calibrated from a known distance included in each image (Figure 2B). Each piece was automatically detected, labeled and measured using SmartGrain's inbuilt functions. All data were exported as CSV files and added to the hand measurements for pieces >10mm. This resulted in a total measured dataset of 12,388 pieces.

Artifact size distribution was analyzed using relative frequency and cumulative histograms, as customary in particle size analysis in sedimentology (e.g., Syvitski, 1991; Boggs, 2011). Histograms were calculated for length, width, thickness and weight class frequencies of archaeological assemblages, and for length and width in experimental sets. Particle shape of archaeological assemblages was calculated using standard triangular methods (Sneed and Folk, 1958) and new orthogonal proposals (Tarriño, 2015). Data was normalized using the relative percentage of each assemblage. Calculations were made using Spreadsheet Software and Triplot (Graham and Midgley, 2000). The analysis was carried out for each individual assemblage as a whole, and per raw materials in each assemblage, in order to refine comparisons between the archaeological and experimental datasets.

RESULTS

Archaeological assemblages

Flaked:detached artifact ratios (Figure 3A) show that both levels at TK contain the highest frequency of detached pieces (98.2% in TKLF and 97.1% in TKUF), followed by FLK Zinj (96.2%) (see Table 3). At the opposite end, FLKNN1 only contains 30.7% of detached artifacts, followed by FLKN Deino (50.0%). In the rest of the assemblages,debitage relative frequencies oscillate between 88.5% (FLKN1-2) and 60.0% (FLKN4) (see Figure 3A).

Following Petraglia and Potts (1994), Figure 3B shows the mean weight index of the archaeological assemblages under study. Mismatches (e.g., Petraglia and Potts' mean weights for FLK Zinj and TKUF are 29g and 33g respectively, thus differing from results in Table 3) are probably due to the size of the sample (entire assemblages were analyzed in this paper in

contrast to sampling by Petraglia and Potts), and to the removal of unmodified pieces from our calculations. Still, results are consistent with the flaked:detached ratio described above, with FLKNN1 and FLKN Deino on one end, and FLK Zinj and the two TK levels at the other (Figure 3B). The very low mean weight in TK levels is significant, given the presence of substantially heavier artifacts such as handaxes, large cores and anvils –see technological breakdowns in Leakey (1971), and de la Torre and Mora (2005)–, which nevertheless do not bias the overall low mean weight of these assemblages.

Weight classes are in agreement with mean weight index results: more than 95% of artifacts from FLK Zinj, TKUF and TKLF weigh <100g (table 4), whereas this weight class is below 60% at FLK SC, FLK Deino, FLKN4 and FLKNN1 (see also Figure 3C). Further, classing of lighter artifacts (those weighting <50g) is also consistent with regards to assemblages with the smallest fraction (Table 5); FLK Zinj shows the largest percentage of 1-10g artifacts (94.8%), followed by TKLF (86%) and TK UF (85%). On the opposite end, FLKN4, FLK Deino and FLKN SC contain <60% of artifacts below 10g class (Figure 3D).

Table 6 shows length classes of the entire archaeological sample, and Figure 3E and Figure 3F rank sites according to frequencies of smaller artifacts. Sites with a distribution of artifacts in the 20-40mm length class (Figure 3E) do not conform with the pattern observed in flaked:detached ratios and weight classes (i.e., TKLF, TK UF and FLK Zinj ranking first), whereas the smallest length class (<20mm) distribution (Figure 3F) is almost identical to that of variables explored in Figure 3A-D: FLK Zinj presents the highest ratio of <20mm artifacts (51.9%), closely followed by TKLF (50.8%) and TKUF (46.5%), as opposed to FLKNN1, FLKN4, FLKN SC and FLKN Deino, where percentages (0%, 0%, 1.1% and 4%, respectively) are negligible.

Whereas most earlier studies of artifact size distributions have focused on the variable length only, our recording of all artifact dimensions (length, width and thickness) enables shape analysis and a three-dimensional view of each item. Sneed and Folk's (1958) classing (Table 7) indicates a strong predominance of bladed (23.0%) and platy (20.1%) artifacts across the board, with the lowest representation for very elongated (2.4%) and compact (5.4%) lithics (see Figure 4B). Selection of the assemblages with a higher number of elements (Figure 4A) shows the almost identical shape distribution of artifacts in TKLF and TKUF and a broad similarity for FLK Zinj and FLKN1-2, apart from the higher frequency of very bladed artifacts in FLK Zinj. Interestingly, these four assemblages (and particularly TKLF and TKUF) do not conform to the average high frequency of platy shapes, whilst containing a higher than average representation of bladed artifacts (see Figure 4A).

Table 8 and Figure 4C-D show artifact shape distributions following Tarrío's (2015) classes. The following types dominate: class #52 (i.e., tabular, thick shapes) and class #53 (tabular, carinate), with 29.4% and 26.3% respectively, followed by class #62 (16.8%: bladed, short, thick lithics). As expected, the four main assemblages (TKUF, TKLF, FLKN1-2 and FLK Zinj; Figure 4C) show a very similar distribution of artifact shapes when compared to the total assemblage sample (Figure 4D).

A more detailed analysis of artifact length, width and thickness is shown in histograms from Figure 5, which indicate two clusters of archaeological assemblages. The first cluster includes the TKUF, TKLF and FLK Zinj assemblages, where artifact length displays a unimodal right-tailed asymmetric distribution (Figure 5A1, Supplementary Material 1), defined by positive and high skewness and kurtosis values (Table 9). TKUF and TKLF length distributions are very similar and show higher values of kurtosis and skewness than FLK Zinj. The main peaks

are at 18-20mm in TKUF, 14-16mm in TKLF and at 10-12mm in FLK Zinj (Figure 5A1, Table 9).

The second cluster of assemblages shows a polymodal distribution, with a general bimodal trend. In the case of FLKN1-2, FLKN3, FLKN5, FLKN6 and FLKNN3, length distribution shows a main mode between 18-30mm, and a secondary mode around 66-86mm (Figure 5B1, Supplementary Material 1). A similar main bimodal trend is also noted in FLKN SC and HWKEL1 length distributions. The sample from FLK Deino, FLKN4 and FLKNN1 is too low, although peaks at 26-28mm and 58-80mm are observed (Supplementary Material 1).

These two patterns are also distinguishable, and more clearly defined, when the distribution of width and thickness values is also considered (Supplementary Material 1). Width and thickness histogram curves of Figure 5A2 and Figure 5A3 show the same unimodal right-tailed asymmetric distribution in artifact length measurements for TKUF, TKLF and FLK Zinj (Table 9), while the rest of the assemblages have a polymodal distribution (Supplementary Material 1), driven by a generally bimodal trend (Figure 5B2 and Figure 5B3).

Experimental assemblages

Analysis of experimental collections focused on length and width patterns. Results show a unimodal right tailed distribution in both dimensions, although significant differences exist by raw material (Figure 6). Quartzite experimental artifact lengths peak at 4-6mm (Table 10), with relative frequencies ranging between 30-40% (Exp #15, Exp #39 and Exp #40), and up to 50% in Exp #13 (Figure 6A). Chert length peaks at a lower range (class 2-4mm), which contains 50-60% of experimental artifacts. The three basalt experiments produced largely homogeneous distributions, with around 60% of artifacts also clustering in the 2-4mm length class. Mean and median values are similar in chert and basalt length distribution, while they are higher in the quartzite experiments (Table 10).

The decreasing left tails of basalt and chert artifacts are missing in the width distributions, since most of the debitage (50-60%) are clustered within the 0-2mm width class. In contrast, the peak in quartzite experiments is within the 2-4mm width class, therefore producing a descending left tail (Figure 6B).

Overall, frequency distributions are very similar when each raw material is considered separately, although they are more clustered in the basalt experiments (Figure 6A); chert and quartzite experiments are slightly more scattered, and show higher variation in their peak distributions (30-50% in quartzite and 40-60% in chert, while all basalt experiments are constrained between 59-65%). In the case of the quartzite experiments, this is accompanied by minor irregularities in the right tail of size distributions. Nevertheless, the position of the main size class peak is consistent across all quartzite experiments, and peaks in the size distributions are consistent by raw material (i.e., quartzite, basalt and chert), irrespective of the experiment (Figure 6A, Table 10). In basalt and chert distributions, mean length values are situated around 4.8-6.1mm and median between 3.0-3.6mm (Table 11). The quartzite size distribution consists of higher values (mean=9.0mm, median=5.9mm), showing the main peak at 4-6mm (40-55%) for length, and at 2-4mm (40-50%) for width.

These trends are also observed in the cumulative histograms that consider all experiments by raw material (Figure 6C). Quartzite distributions tend to be slightly displaced to the right, towards higher length and width values. Conversely, basalt and chert distributions share very similar patterns, defined by a larger proportion of smaller debitage.

Comparison of archaeological and experimental patterns

Our experimental results show that artifact size production during *in situ* knapping activities follows a unimodal, leptokurtic and positive skewed frequency distribution (Figure 6, Table 10). Our results also indicate that knapping using basalt and chert from Olduvai tends to produce a more left-skewed and more leptokurtic main peak than in Olduvai quartzite (Figure 6A, Table 10), which is due to a higher production of smaller debitage during chert and basalt reduction. This pattern explains why the left tail of basalt and chert width distributions is absent, considering the minimum class (0.1- 2mm) used in this paper.

This unimodal frequency distribution of flaking experiments is different from what is observed in the Olduvai archaeological assemblages. Most of the Olduvai sites analyzed here (FLKN SC, HWKEL1, FLKN1-2, FLKN3, FLKN5, FLKN6, FLKNN3; Figure 5; Supplementary Material 1) show polymodal distribution driven by a bimodal trend, and only three assemblages (TKUF, TKLF and FLK Zinj) present main unimodal positive skewed distribution (Figure 5). Such patterns are observed not only in length distribution, but also in width and thickness values, whose distribution curves are actually neater than that of length.

All experimental materials (Table 11) are compared in Figure 7 to the main archaeological assemblages in one single size distribution. The histogram curve of the experimental assemblages (ALL EXP in Figure 7) continues to follow a unimodal positive skewed distribution, very different from that of the archaeological sites. Even the size distributions of TKUF, TKLF and FLK Zinj, which also display unimodal right tailed shapes, vary significantly from the ALL EXP size curve. Such differences are pronounced in both the histogram and cumulative curves and concern both length and width values. Differences are defined by: 1) a more pronounced peak in the ALL EXP curve, and 2) a more positive skewed position of the ALL EXP mean peak. Both differences are due to a much higher presence of smaller pieces (<5mm length) in ALL EXP than in the archaeological assemblages.

In order to refine the comparison, and given the absence of <5mm length and <3mm width objects in the archaeological assemblages, experimental size distributions were recalculated only for length >5mm, and width >3mm. Experimental distribution curves are shown in Figure 7 as ALL EXP Length >5mm (Figure 7A) and ALL EXP Width >3mm (Figure 7B). These graphs show that even when size ranges undetected or uncollected during excavations are excluded, experimental curves are considerably different from the archaeological assemblages; archaeological artifacts in the 5-10mm length range are still underrepresented when compared to the experimental curve, while those >10mm are overrepresented (Figure 7). Differences are also highlighted in the lower mean and median values of ALL EXP length distributions (Table 11), when compared to the values of the most complete archaeological assemblages (TKUF, TKLF and FLK Zinj) (Table 9). Results thus confirm that even when considering collection bias, the most complete archaeological lithic assemblages clearly show sorting with respect to experimental size distributions. This sorting is more accentuated in HWKEL1 and FLKN1-2 than in TKUF, TKLF and FLK Zinj, but the latter are also different from experimental distributions.

Given the differential size distributions per raw material observed in the experiments, we also analyzed the archaeological assemblages according to size curves of quartzite (Figure 8) and basalt (Figure 9) artifacts (chert was excluded from the analysis due to its low frequency in the archaeological collection). Results of quartzite size distributions are consistent with those presented above for the individual experimental sets. Thus, size distribution of the entire quartzite experimental assemblage (ALL EXP Qz) is characterized by more peaked and more

positive skewed curves than in the archaeological quartzite artifacts (Figure 8). As previously, we also based the comparison on the size ranges shared by both archaeological and experimental samples, i.e., considering only artifacts available in the archaeological assemblages. The experimental curves display similar patterns as presented earlier, and are conspicuously different from archaeological size distributions. The presence of larger artifacts in archaeological assemblages (Figure 7 and Table 9) when compared to experimental ranges (Table 10 and Table 11) could introduce a bias in distribution shapes and statistical results. To mitigate the impact of such disparity, Figure 8C limited comparisons to artifacts with <125mm length, which is the maximum size in the quartzite experimental sets. Figure 8C shows, however, that no significant changes to the patterns of Figure 8A and 8B occur, and therefore results on the differential size distributions of quartzite archaeological and experimental curves are confirmed.

Differences between experimental and archaeological size distributions are even more noticeable on the basalts. As shown in Figure 9, basalt distributions are scattered, and are characterized by irregular polymodal size distribution patterns where small pieces are almost entirely missing (Figure 9). This is particularly relevant when taking into account that basalt experiments produced higher quantities of small pieces than in the case of quartzite. According to our experimental results, archaeological assemblages are expected to contain high numbers of small basalt artifacts, should sites correspond to *in situ* flaking activities. However, size distributions of basalt pieces in the archaeological contexts studied here show the contrary, as the smallest lava archaeological artifacts show length values of 11mm (TKLF) or 13mm (TKUF, FLK Zinj), and their frequency is <3%. Differences between archaeological and experimental distributions of lava artifacts are also pronounced when considering the minimum size documented in the archaeological assemblages (length >5mm, width >3mm).

DISCUSSION

Interpretation of results

Experimental results indicate that quartzite produces fewer smaller lithics than basalt and chert, at least with regards to those quantifiable in the > 0.1mm class. As shown in Figure 2B1, quartzite also produces 'dust' that would be invisible in an archaeological context. This is contrary to what is often claimed for the Olduvai quartzite (see Byrne et al, 2016), and is likely related to the coarse texture of this raw material (i.e., smallest size of debitage is dictated by the minimum dimensions of quartzite individual grains), as opposed to the fine texture of most Olduvai lavas and all chert.

Based on size classes, all experiments show unimodal curves, but size distribution patterns indicate pronounced differences between the archaeological and experimental assemblages. Such differences include: 1) a bimodal trend in most archaeological assemblages that is opposed to a dominantly unimodal experimental distribution, and 2) in the only archaeological assemblages where distribution is mainly unimodal (TFLF, TKUF and FLK Zinj), there is a significant underrepresentation of pieces <10mm in length and an overabundance of artifacts >10mm, when compared to the experimental distributions. This difference persists when only quartzite assemblages are considered, and it is still more pronounced in the case of lavas, where distributions are polymodal and irregular in all sites.

The predominance of polymodal size patterns across most archaeological assemblages indicates that artifact accumulation processes differ notably from an *in situ*, pristine knapping activity, thus suggesting the presence of biasing factors. Such factors may include excavation

bias, natural sorting agents (e.g., water runoff), raw material properties, original size of blanks, intensity of reduction, and technological and functional constraints.

The absence of <5mm materials in the Leakey collection can be explained, at least partially, by limitations in the identification and collection of smaller pieces during archaeological excavation. Although we are not aware of the specific recovery methods employed by Mary Leakey (1971), we assume that she routinely sieved sediments, given the large amount of small artifacts preserved in the collection, and the fact that in the case of FLKN SC she stated that in this instance the deposits were not screened (Leakey, 1971: 111). In archaeological excavations, sieves used during dry screening often have a mesh opening of 5mm. The minimum length present in the archaeological assemblages under study is precisely 5mm (TKLF and FLK Zinj). Therefore, it may well be the case that artifacts <5mm were present in the archaeological record, but were not identified or collected.

Length class underrepresentation, however, affects artifacts that were clearly recovered through sieving (e.g., pieces <10mm), suggesting that excavation bias alone does not explain the overall deficit of small pieces in the archaeological assemblages, and thus alternative scenarios should be sought. Compartmentation of the landscape by hominins (Blumenschine et al, 2012), functional variability (e.g., Leakey, 1971; Isaac and Crader, 1981; Mora and de la Torre, 2005) as well as the fragmentation of Oldowan and Acheulean *chaînes opératoires* (de la Torre and Mora, 2005) have all been widely discussed for Olduvai. Technological differences across Bed I and II assemblages (Leakey, 1971; de la Torre and Mora, 2014) may have also contributed to the increased artifact size variability; in this regard, it is worth noting the remarkably similar shape patterns in both levels of TK (see Figure 4A), which are the only two Acheulean assemblages studied here and whose size distributions differ from most of the other sites. All these factors certainly obscure a direct comparison with our experimental results, which cannot replicate all naturalistic scenarios that would account for the formation of Olduvai assemblages, and are limited to sequences of small debitage reduction.

Despite these hindrances, it is parsimonious to consider that at least some knapping took place in most of the classic Olduvai assemblages –although see Mora and de la Torre (2005) regarding alternative activities for some sites. If that was the case, it is also very likely that additional sorting agents removed part of the assemblage, as all our experimental results consistently show size classes that do not exist (or are considerably underrepresented) in the archaeological assemblages. The most probable natural agent for this phenomenon is water action.

Water disturbance in some of the classic Olduvai sites was already considered by Leakey (1971), and subsequent work (e.g., Potts, 1982; Petraglia and Potts, 1994; Benito and de la Torre, 2011; de la Torre and Benito, 2013) has dwelled further on this scenario. Interestingly, polymodal patterns produced by hydraulic action has already been proposed for the Olduvai assemblages based on the study of stone tool and fossil orientations (Benito-Calvo and de la Torre, 2011; de la Torre and Benito-Calvo, 2013). These authors have argued that Bed I sites show two oblique modes indicative of two flow directions (Benito-Calvo and de la Torre, 2011), whereas Bed II assemblages present more heterogeneous (albeit still non-random) orientation patterns (de la Torre and Benito-Calvo, 2013). Orientation of items is a taphonomic variable independent from size classes, and the fact that the present study also indicates polymodal patterns reinforces the notion that natural agents have played a relevant role in site formation at Olduvai Beds I and II. Transport sedimentary processes often generate preferred orientated patterns of sedimentary particles and sort the sediments, producing in many occasions polymodal particle size distributions, especially in hydraulic and aeolian continental

contexts (Ashley, 1978; Sun et al., 2002). The total absence of the smallest lithic fraction, the general paucity of artifacts <10mm, and the polymodal nature of archaeological assemblages, contrast starkly with patterns derived from our experimental framework, and strongly suggest that all sites under study were affected by water action.

Degree of disturbance in the Olduvai assemblages

If all sites under study were disturbed by water action, the extent of such disturbance in each assemblage must be discussed. Detailed actualistic studies of water action over experimental assemblages, such as those made in Koobi Fora (Schick, 1984), are not yet available for the Olduvai contexts, and thus no specific framework exists against which the archaeological sample can be compared. Instead, we can assess the relative degree of disturbance of each site with regards to the rest of archaeological assemblages studied in this paper.

All weight, shape and size variables analyzed here yielded consistent results. Flaked:detached ratios (Figure 3A), mean weight (Figure 3B), frequency of artifacts < 100g (Figure 3 C) and 1-10g (Figure 3D), percentages of pieces smaller than 20mm (Figure 3F), and overall dimensions of stone tools (Figure 5), all suggest the same pattern; in FLK Zinj, TKLF and TKUF there is a predominance of small and light artifacts, with a higher than average percentage of bladed objects (Figure 4A). At the other end, small debitage percentages at FLKN1, FLKN Deino, FLKN4 and FLKN SC are low. Other assemblages such as FLKN1-2, FLKN3, FLKN5, FLKN6 and HWKL1 show variable percentages, but in general they seem to have undergone an intermediate stage of small debitage depletion.

Figure 10 shows associations between several of the variables discussed here, and clearly relates the site cluster of FLKNN1, FLKN4, FLKN SC and FLKN Deino with a predominance of heavier artifacts, flaked (rather than detached) pieces, and compact shapes (which can be assimilated to core forms). On the opposite end, FLK Zinj, TKUF and TKLF are associated with higher percentages of smaller artifacts and elongate shapes. FLKN5, HWKEL1, FLKN3, FLKNN3, FLKN6 and FLKN1-2 are positioned in the middle, with the latter two assemblages closer to the cluster of TK and Zinj. Given the consistency of all variables in the classing of Olduvai sites, this figure shows the heuristic potential of combining multiple size and shape proxies to assess assemblage composition.

Leakey (1971) originally considered FLKNN1, FLKNN3, FLK Zinj, HWKEL1, TKLF and TKUF as living floors, as opposed to butchery sites (FLKN6 and FLKN Deino) and assemblages with dispersed materials –FLKN5, FLKN4, FLKN3 (she did not provide an assessment for FLKN SC). This interpretation has been nuanced and/or challenged in subsequent work (e.g., Isaac and Crader, 1981; Blumenschine and Masao, 1991; de la Torre and Mora, 2005; Dominguez-Rodrigo et al, 2007; Benito-Calvo and de la Torre, 2011). Although site functionality, sedimentary contexts and other taphonomic aspects other than size and shape variables are beyond the scope of this paper, our results clearly rank the assemblages into two main clusters, with FLK Zinj, TKUF and TKLF on one side, opposed to a heavily depleted group (FLKNN1, FLKN4, FLKN SC and FLKN Deino), with the rest of the assemblages in an intermediate position.

The fact that both Zinj and the two TK assemblages were recovered from clay contexts (Leakey, 1971) might account for lower energy processes, preserving the archaeological record better than in sandy contexts (e.g., FLKN SC). Nonetheless, it is important to note that a clay context does not automatically result in pristine conservation of the assemblage, as several of the sites (e.g., FLKNN, most of FLK North units) were located in claystone deposits (Leakey, 1971), and have been depleted of substantial parts of the smaller lithic fraction. In fact, running

water probably affected most (if not all) of the Bed I and II assemblages (Benito-Calvo and de la Torre, 2011; de la Torre and Benito-Calvo, 2013), highlighting the importance of accounting for abiotic agents in the formation of Olduvai sites. The experimental framework developed in this paper confirms this assertion, as even the most complete archaeological assemblages (e.g., TKLF, TKUF and FLK Zinj) present substantial deviations from the size distributions produced during experimental flaking. In summary, our results strongly suggest that water action played an important role in the configuration of the Olduvai sites, and shaped to a lesser or greater extent the composition of all assemblages.

CONCLUSIONS

Artifact size distribution is known to be a key variable for the assessment of hydraulic sorting in archaeological sites (Isaac, 1967; Harris, 1978; Schick, 1984; Petraglia and Nash, 1987; Bertran et al, 2006, 2012; Mallinsky-Buller et al, 2011; Byers et al, 2015; Lotter et al, 2016). Small- sizeddebitage is particularly sensitive to post-depositional disturbance (Fladmark, 1982; Stein and Telsen, 1989), and its relative frequency is considered as a proxy to address water winnowing (e.g., Harris, 1978; Schick, 1984; Sitzia et al, 2012).

In this paper, we have applied these principles to the study of 14 assemblages excavated by Mary Leakey (1971) in Olduvai Beds I and II. Rather than sampling the assemblages, we decided to analyze all materials; although significantly time consuming (a total of 12,083 objects were measured), this strategy provided a robust statistical ground to the study. Analysis of archaeological material was accompanied by production of an experimental dataset, for which a statistically significant sample (n=12,388 individual pieces) was studied. Our methods have drawn on previous studies on artifact dimensions (e.g., Schick, 1984; Petraglia and Potts, 1994; Bertran et al, 2006), and combined those with less conventional (in Archaeology) approaches to particle sorting (Sneed and Folk, 1958; Tarriño, 2015) and automatized grain size counting (Tanabata et al, 2012).

Our particle size perspective has enabled us to create a strong referential framework for interpreting the Olduvai assemblages, and the results provide a solid ground to assess hydraulic disturbance. Our analysis has shown clear differences between assemblages based on size and shape distributions, with sites that may have been severely depleted of the smaller lithic fraction, and others which preserve a better contextual integrity. Nonetheless, comparison of the Olduvai archaeological assemblages with our experimental results clearly indicate that even the best-preserved sites underwent postdepositional disturbance. In reality, this is unsurprising for, as long highlighted by Isaac (1984), most if not all Early Stone Age assemblages have been subjected to some degree of disturbance. Creation of behavioral models should take into consideration that Early Stone Age sites in pristine conditions are an oxymoron, and the analytical tools to disentangle the role of abiotic agents on site formation should continue to be refined.

ACKNOWLEDGEMENTS

We thank Rafael Mora (Universidad Autonoma de Barcelona) and John Mwangi (National Museums of Kenya) for their assistance in the analysis of the Leakey collections, and Adrian Arroyo (UCL) for his help in the experiments. We are also grateful to A. Tarriño (CENIEH) for access to Bi-plot(GEO)_v1-4 spreadsheet to calculate shape percentages, and to two anonymous reviewers on an earlier draft for their very helpful comments. Fieldwork at Olduvai to collect raw materials was authorized by the Ngorongoro Conservation Area Authority, the

Commission for Science and Technology (COSTECH) and the Department of Antiquities, Tanzania. Funding from the European Research Council-Starting Grants (283366) is gratefully acknowledged.

REFERENCES

- Ashley, G. M. (1978). Interpretation of polymodal sediments. *The Journal of Geology* 86, 411–421.
- Bertran, P., Claud, É., Detrain, L., Lenoble, A., Masson, B. & Vallin, L. (2006). Composition granulométrique des assemblages lithiques, application à l'étude taphonomique des sites paléolithiques. *Paléo* 18, 7-36.
- Bertran, P., Lenoble, A., Todisco, D., Desrosiers, P. M. & Sørensen, M. (2012). Particle size distribution of lithic assemblages and taphonomy of Palaeolithic sites. *Journal of Archaeological Science* 39, 3148-3166.
- Blumenschine, R. J. & Masao, F. T. (1991). Living sites at Olduvai Gorge, Tanzania? Preliminary landscape archaeology results in the basal Bed II lake margin zone. *Journal of Human Evolution* 21, 451-462.
- Blumenschine, R. J., Masao, F. T., Tactikos, J. C. & Ebert, J. I. (2008). Effects of distance from stone source on landscape-scale variation in Oldowan artifact assemblages in the Paleo-Olduvai Basin, Tanzania. *Journal of Archaeological Science* 35, 76-86.
- Blumenschine, R. J., Stanistreet, I. G., Njau, J. K., Bamford, M. K., Masao, F. T., Albert, R. M., Stollhofen, H., Andrews, P., Prassack, K. A., McHenry, L. J., Fernández-Jalvo, Y., Camilli, E. L. & Ebert, J. I. (2012). Environments and hominin activities across the FLK Peninsula during *Zinjanthropus* times (1.84 Ma), Olduvai Gorge, Tanzania. *Journal of Human Evolution* 63, 364-383.
- Boggs, S. (2011). *Principles of Sedimentology and Stratigraphy*, Fifth Edition. New Jersey: Pearson Prentice Hall.
- Bunn, H. T. (1982). *Meat-Eating and Human Evolution: Studies on the Diet and Subsistence Patterns of Plio-Pleistocene Hominids in East Africa*. Unpublished Ph.D. Berkeley: University of California.
- Byers, D. A., Hargiss, E. & Finley, J. B. (2015). Flake Morphology, Fluvial Dynamics, and Debitage Transport Potential. *Geoarchaeology* 30, 379-392.
- Byrne, F., Proffitt, T., Arroyo, A. & de la Torre, I. (2016) A comparative analysis of bipolar and freehand experimental knapping products from Olduvai Gorge, Tanzania. *Quaternary International*. 424: 58-68
- Domínguez-Rodrigo, M., Barba, R. & Egeland, C. P. (2007). *Deconstructing Olduvai. A Taphonomic Study of the Bed I Sites*. Dordrecht: Springer.
- Fladmark, K. R. (1982). Microdebitage analysis: Initial considerations. *Journal of Archaeological Science* 9, 205-220.
- Graham, D. J. & Midgley, N. G. (2000). Graphical representation of particle shape using triangular diagrams: an Excel spreadsheet method. *Earth Surface Processes and Landforms* 25, 1473-1477.
- Harris, J. W. K. (1978). *The Karari Industry. Its Place in East African Prehistory*. Berkeley: Unpublished PhD, University of California, Berkeley.

- Isaac, G. L. (1967). Towards the interpretation of occupation debris: some experiments and observations. *Kroeber Anthropological Society Papers* 37, 31-57.
- Isaac, G. L. (1984). The Archaeology of Human Origins: Studies of the Lower Pleistocene in East Africa 1971-1981. In (F. Wendorf & A. E. Close, Ed.) *Advances in World Archaeology*. Orlando: Academic Press, 1-87.
- Isaac, G. L. (1986). Foundation stones: early artifacts as indicators of activities and abilities. In (G. N. Bailey & P. Callow, Ed.) *Stone Age Prehistory: studies in Memory of Charles McBurney*. Cambridge: Cambridge University Press, 221-241.
- Isaac, G. L. & Crader, D. C. (1981). To What Extent Were Early Hominids Carnivorous? An Archaeological Perspective. In (R. S. O. Harding & G. Teleki, Ed.) *Omnivorous Primates: Gathering and Hunting in Human Evolution*. New York: Columbia University Press, 37-103.
- Isaac, G. L., Isaac, B. & eds. (1997). *Koobi Fora Research Project. Volume 5: Plio-Pleistocene Archaeology*. Oxford: Oxford University Press.
- Jorayev, G., Wehr, K., Benito-Calvo, A., Njau, J. & Torre, I. de la (2016). Imaging and photogrammetry models of Olduvai Gorge (Tanzania) by Unmanned Aerial Vehicles: A high-resolution digital database for research and conservation of Early Stone Age sites. *Journal of Archaeological Science*, in press.
- Leakey, M. D. (1971). *Olduvai Gorge. Vol 3. Excavations in Beds I and II, 1960-1963*. Cambridge: Cambridge University Press.
- Lotter, M. G., Gibbon, R. J., Kuman, K., Leader, G. M., Forssman, T. & Granger, D. E. (2016). A Geoarchaeological Study of the Middle and Upper Pleistocene Levels at Canteen Kopje, Northern Cape Province, South Africa. *Geoarchaeology* 31, 304-323.
- Malinsky-Buller, A., Hovers, E. & Marder, O. (2011). Making time: 'Living floors', 'palimpsests' and site formation processes – A perspective from the open-air Lower Paleolithic site of Revadim Quarry, Israel. *Journal of Anthropological Archaeology* 30, 89-101.
- Mora, R. & Torre, I. de la (2005). Percussion tools in Olduvai Beds I and II (Tanzania): Implications for early human activities. *Journal of Anthropological Archaeology* 24, 179-192.
- Pante, M. (2010). *The Larger Mammal Fossil Assemblages from Beds III and IV, Olduvai Gorge, Tanzania: Implications for the Feeding Behavior of Homo erectus*. New Brunswick: Ph D dissertation, Rutgers University.
- Pante, M. C. & Blumenshine, R. J. (2010). Fluvial transport of bovid long bones fragmented by the feeding activities of hominins and carnivores. *Journal of Archaeological Science* in press,
- Petraglia, M. D. & Nash, D. T. (1987). The Impact of Fluvial Processes on Experimental Sites. In (D. T. Nash & M. D. Petraglia, Ed.) *Natural Formation Processes and the Archaeological Record*. Oxford: BAR International Series, 108-130.
- Petraglia, M. D. & Potts, R. (1994). Water Flow and the Formation of Early Pleistocene Artifact Sites in Olduvai Gorge, Tanzania. *Journal of Anthropological Archaeology* 13, 228-254.
- Potts, R. (1982). *Lower Pleistocene site formation and hominid activities at Olduvai Gorge, Tanzania*. Harvard University: Unpublished PhD.
- Potts, R. (1988). *Early Hominid Activities at Olduvai*. New York: Aldine and Gruyter.
- Proffitt, T. & de la Torre, I. (2014). The effect of raw material on inter-analyst variation and analyst accuracy for lithic analysis: a case study from Olduvai Gorge. *Journal of Archaeological Science* 45, 270-283.
- Schick, K. D. (1984). *Processes of Palaeolithic Site Formation: An Experimental Study*. Berkeley: Unpublished Ph.D. University of California, Berkeley.

- Schick, K.D. (1987). Modeling the formation of Early Stone Age artifact concentrations. *Journal of Human Evolution* 16: 789-807.
- Schick, K.D. (1991). On making behavioural inferences from early archaeological sites. In: Clark, J.D. (ed) *Cultural Beginnings*: 79-107. Bonn: Dr. Rudolf Habelt GMBH.
- Schick, K.D. (1997). Experimental studies of site-formation processes. In: Isaac, G.I. & Isaac, B. (eds) *Koobi Fora Research Project: Volume 5, Plio-Pleistocene Archaeology*: 244-261. Oxford: Clarendon Press.
- Sitzia, L., Bertran, P., Boulogne, S., Brenet, M., Crassard, R., Delagnes, A., Frouin, M., Hatté, C., Jaubert, J., Khalidi, L., Messenger, E., Mercier, N., Meunier, A., Peigné, S., Queffelec, A., Tribolo, C. & Macchiarelli, R. (2012). The Paleoenvironment and Lithic Taphonomy of Shi'Bat Dihya 1, a Middle Paleolithic Site in Wadi Surdud, Yemen. *Geoarchaeology* 27, 471-491.
- Sneed, E. D. & Folk, R. L. (1958). Pebbles in the Lower Colorado River, Texas a Study in Particle Morphogenesis. *The Journal of Geology* 66, 114-150.
- Stein, J. K. & Teltser, P. A. (1989). Size distributions of artifact classes: combining macro-and-micro-fractions. *Geoarchaeology* 4, 1-30.
- Sun, D., Bloemendal, J., Rea, D. K., Vandenberghe, J., Jiang, F., An, Z. & Su, R. (2002). Grain-size distribution function of polymodal sediments in hydraulic and aeolian environments, and numerical partitioning of the sedimentary components. *Sedimentary Geology* 152, 263-277.
- Syvitski, J. P. M. (1991). *Principles, methods, and applications of particle size analysis*. Cambridge: Cambridge University Press.
- Tanabata, T., Shibaya, T., Hori, K., Ebana, K. & Yano, M. (2012). SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiology* 160, 1871-80.
- Tarriño, A. (2015). A New Methodology and Classification System for Describing Three-Dimensional Particle Formats: Application to Clastic Lithic Products of Archaeological and Geological Origin. *Archaeometry* 57, 928-948.
- Torre, I. de la & Benito-Calvo, A. (2013). Application of GIS methods to retrieve orientation patterns from imagery; a case study from Beds I and II, Olduvai Gorge (Tanzania). *Journal of Archaeological Science* 40, 2446-2457.
- Torre, I. de la & Mora, R. (2005). Technological Strategies in the Lower Pleistocene at Olduvai Beds I & II. *Liege: ERAUL* 112.
- Torre, I. de la & Mora, R. (2005b). Unmodified Lithic Material at Olduvai Bed I: Manuports or Ecofacts? *Journal of Archaeological Science* 32, 273-285.
- Torre, I. de la & Mora, R. (2014). The Transition to the Acheulean in East Africa: An Assessment of Paradigms and Evidence from Olduvai Gorge (Tanzania). *Journal of Archaeological Method and Theory* 21, 781-823.