Journal of Archaeological Science 39 (2012) 205–214

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

# Experimental study of cut marks made with rocks unmodified by human flaking and its bearing on claims of ∼3.4-million-year-old butchery evidence from Dikika, Ethiopia

Manuel Domínguez-Rodrigo[a,b,*], Travis Rayne Pickering[c,d], Henry T. Bunn[c]

[a] Department of Prehistory, Complutense University of Madrid Ciudad Universitaria, s/n 28040, Madrid, Spain
[b] IDEA (Instituto de Evolución en África), Museo de los Orígenes, Plaza de San Andrés 2, 28005 Madrid, Spain
[c] Department of Anthropology, University of Wisconsin–Madison, Madison, WI 53706, USA
[d] Institute for Human Evolution, University of the Witwatersrand, WITS 2050, Johannesburg, South Africa

ARTICLE INFO

*Article history:*
Received 27 February 2011
Received in revised form
7 March 2011
Accepted 8 March 2011

*Keywords:*
Early hominin carcass foraging
Taphonomy
Cut marks
Random striae
Trampling

ABSTRACT

In order to assess further the recent claims of ∼3.4 Ma butchery marks on two fossil bones from the site of Dikika (Ethiopia), we broadened the actualistic-interpretive zooarchaeological framework by conducting butchery experiments that utilized naïve butchers and rocks unmodified by human flaking to deflesh chicken and sheep long limb bones. It is claimed that the purported Dikika cut marks present their unexpectedly atypical morphologies because they were produced by early hominins utilizing just such rocks. The composition of the cut mark sample produced in our experiments is quite dissimilar to the sample of linear bone surface modifications preserved on the Dikika fossils. This finding substantiates and expands our earlier conclusion that—considering the morphologies and patterns of the Dikika bone surface modifications and the inferred coarse-grained depositional context of the fossils on which they occur—the Dikika bone damage was caused incidentally by the movement of the fossils on and/or within their depositional substrate(s), and not by early hominin butchery. Thus, contrary to initial claims, the Dikika evidence does not warrant a major shift in our understanding of early hominin behavioral evolution with regard to carcass foraging and meat-eating.



## 1. Introduction

If correct, the recent interpretation of two ∼3.4 Ma, surficially marked, ungulate fossils from the site of Dikika (Ethiopia) as evidence of Pliocene (presumably pre-*Homo*) hominin butchery (McPherron et al., 2010) would have a major impact on our understanding of human evolution. The finding would: (1) demonstrate meat-eating almost one million years earlier than previously inferred; (2) imply that large carcass foraging and meat-eating was unrelated to the invention of flaked stone tool technology; (3) imply that the behavior(s) responsible for concentrating certain hominin activities in discrete spots in the landscape was not related to meat-eating; (4) imply that development of skills necessary to acquire large animal carcass resources was not linked to encephalization; (5) and imply that dietary and adaptive reconstructions based on morphological and wear analyses of hominin dentition are incomplete and/or inaccurate. Given the potential of the Dikika claims for instituting these important conceptual shifts in our understanding of human evolution, the data underpinning them deserve very close scrutiny.

We did subject the Dikika data to this close scrutiny and produced a critique of the claims for hominin butchery, concluding that the published evidence did not, in fact, support the identification of bone surface marks on the two published Dikika fossils as unequivocal stone tool butchery damage (Domínguez-Rodrigo et al., 2010). We further asserted that any equivocation surrounding butchery claims of this great antiquity (i.e., ∼800 ka older than oldest known butchery marks from Gona [Ethiopia], where marked animal bones are derived from fine-grained sediments and in spatial association with hominin-flaked stone tools [Semaw et al., 2003; Domínguez-Rodrigo et al., 2005]) should lead to rejection of such claims. This assertion is not equivalent to contending the impossibility of >2.6 Ma butchery by hominins. We simply raised two straightforward contentions: (1) that the Dikika fossils derived from a potentially abrasive sedimentary context, and (2) that the Dikika fossils show surface damage that is indistinguishable from that imparted on bone surfaces randomly (by trampling and/or other incidental movement) in such deposits.

* Corresponding author. Department of Prehistory, Complutense University of Madrid Ciudad Universitaria, s/n 28040, Madrid, Spain. Tel./fax: +34 91 394 60 08.
*E-mail address:* manueldr@ghis.ucm.es (M. Domínguez-Rodrigo).

0305-4403/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jas.2011.03.010


**Fig. 1.** Some examples of the rocks used for butchery (left) and the morphology of their edges (right). Every piece was selected taking into account that no flaking feature (platform, impact point, overlapping flaking scars, bulb, sinuous ventral profile, concave scars) was present. Frequently other features, such as irregular ragged ventral surface, acute stepping on ventral and dorsal surfaces, and flat ventral profiles, suggestive of natural breakage when present together, were used to document a natural non-anthropogenic origin of the pieces.

A response from McPherron et al. in the popular press (Vergano, 2010) alluded to as-yet unpublished results of experiments that purport to corroborate their inference that the Dikika bone surface marks were imparted by hominins butchering with humanly unmodified rocks (HURs). In that publication, it was argued that "Domínguez-Rodrigo et al. based their assessment on standard flaked stone tools, which are not present at Dikika. As we suggested, the most likely tool was unflaked stone, which has broader less sharp edges than flaked stone." The actual scientific article to which Domínguez-Rodrigo et al. (2010) responded was considerably less definitive: "It is not possible to demonstrate from the [Dikika] modified bones whether the stone tools were knapped for this purpose or whether naturally-occurring sharp-edged stones were collected and used" (McPherron et al., 2010: 858). No HURs were reported from the Dikika locality. And, no planned and/or in-progress experimentation with HURs (i.e., "naturally-occurring sharp-edged stones") was mentioned in that report.

Interpretations outside a referential framework are invalid scientifically. In order to support claims for an anthropogenic (i.e., butchery) origin of the Dikika bone surface marks, McPherron et al. (2010) should have either successfully matched those marks to butchery marks of known origins, produced by various types of stone tools, or, alternatively, provided analogical links, based on mark morphologies and other attributes, between the Dikika marks and those produced experimentally with HURs.

Either way, hypothesizing an anthropogenic origin for the Dikika bone surface marks should have followed first from a serious, informed consideration of other natural processes as potential causes of the marks, and a rejection of all of those. We argued that such a crucial inferential step was not taken by McPherron et al., and that the Dikika bone surface marks actually were indistinguishable from trampling marks (Domínguez-Rodrigo et al., 2010). Further, we stressed that an analogical framework for understanding bone surface modifications created by HURs did not exist. Thus, we initiated our own experimental program using HURs to butcher fleshed animal bones, on which we report here. Our results, in contrast to those alluded by McPherron et al. in Vergano (2010), which result from their own experimentation, do *not* "show that the Dikika marks are a tight fit to marks produced by unflaked stone."

## 2. Materials and methods

### 2.1. Experimental design

Previous experimental research on cut marks indicated that a researcher may introduce bias into the process when he/she is also the experimental subject, performing the butchery himself/herself; for instance, experimenter-subjects inflict cut marks in frequencies and locations that differ from those reported in experiments in which the butcher is a different person than the experimenter and is also unaware of the hypothesis/hypotheses being tested (Domínguez-Rodrigo, 1997). The same phenomenon may occur if the experimenter and experimental subject are one-in-the-same and the topic of experimentation is cut mark morphology rather than mark frequency. For this reason, we recruited 28 individual, naïve butchers (i.e., they had no previous butchery experience and no knowledge of the hypothesis being tested) from MD-R's 2010–2011 Archaeology Grade course on the Paleolithic (Complutense University, Madrid, Spain).

To sample a wide range of naturally produced edge forms on the HURs utilized, each butcher conducted two separate butchery episodes with two separate HURs. Most HURs were obtained from terraces of the Manzanares and Jarama Rivers and from the footslope of the Sierra de Madrid mountain system. The raw materials used include quartzite (used for 35 butcheries), granite (nine butcheries), chert (six butcheries), limestone (four butcheries), and metamorphic (two butcheries) rocks. Most of the utilized HURs are fragments detached from their sources by thermal contrast (e.g., gelifraction) or other natural processes. They are unknapped by humans and their surfaces present rugged microtopographies, with most edges showing variable trajectories, instead of the typical straight outlines of intentionally knapped, unretouched flakes (Fig. 1). In sum, the utilized HURs are naturally sharp pieces, readily available in fluvial and/or colluvial landscapes, as is hypothesized as the general type of source rock(s) that effected the purported butchery marks on the Dikika fossils (see, McPherron et al., 2010). The HURs used range between 39 and 98 mm in length (mean = 56 mm; median = 59 mm). For each HUR, the edge length spans most of the specimen's length (at least, on one side), although this applies only if defining as edge the portion of the outline of the specimen that does not present a blunt flat surface thicker than 2 mm, given the irregular and variable thickness and outline of the edges of several of the tools used.

Seventy-two chicken (*Gallus gallus domesticus*) and 20 sheep (*Ovis aries*) bones were butchered. Each butchery episode included either the defleshing of two raw (and de-feathered) chicken hindlimbs or a single raw sheep hindlimb (only meat-bearing bones were butchered; meatless chicken tarsometatarsus and sheep metatarsal bones were excluded from butchery). Chicken and sheep were selected for this experiment because they were affordable enough that a large experimental sample could be produced. U Mann–Whitney tests, with $p$ values $>0.05$, confimed that, for all the bone surface mark variables employed in this study (see below), bone type (chicken versus sheep) did not influence bone surface mark morphology or frequency.

Each butchery episode was performed without any assistance or instruction from the authors, other than that the butcher should deflesh each bone by cutting with the HUR placed perpendicular to the long axis of bone being butchered.

### 2.2. Analyses

After butchery, bones were cleaned by boiling them in a solution of water and neutral detergent. Butchery mark analysis was carried

**Table 1**
Definitions and distributions of studied cut mark variables (see also, Domínguez-Rodrigo et al., 2009).

| | |
|---|---|
| Groove trajectory | Straight (a), curvy (b) or sinuous (c). This variable is applied to the outline of the mark, without taking into account the presence of barb (when it is present) at the end of the mark. |
| Groove cross-sectional shape | Narrow V-shaped (a) and wide V-shaped (_/) (b). Narrow V-shaped grooves are deeper than they are wide. Wide V-shaped grooves have bases that are horizontal and are wider than the grooves are deep. |
| Shoulder effect on the groove | Presence (a), or absence (b) |
| Flaking on the shoulders of the groove | Presence (a), or absence (b). In some cases flaking inside the groove of the mark is present. |
| Groove that is a forked | Presence (a), or absence (b). This type of mark is produced by a single cutting stroke, but branches into a forked shape at one end (Fig. 3). |
| Multiple-clustered mark | This type of mark is actually a set of multiple, non-contacting marks produced by a single cutting stroke (Fig. 4). |
| Number of forked marks | Number of marks in contact with the main groove of the cut mark. |
| Number of multiple-clustered marks | Number of marks occurring near the main groove of the cut mark. |

out by first using the naked eye to identify each cut mark. Each mark was then numbered and examined using hand lenses and a Motic™ binocular microscope, at magnifications of 20×–40×. Marks were photographed with a digital camera (MC V3) incorporated into the microscope, which transfers high-resolution images in .mix, .bmp and .jpeg formats into a computer. The images were thus downloaded directly to a computer and processed with Motic Image Plus 2.0™ software. Marks were also observed under a hand-held digital microscope (Dinolite AM413FVT™) with magnifications of 10×–200× and analyzed in the computer with Dino Capture 2.0™ software.

A comparative butchery mark sample consisting of: (1) 246 cut marks created by simple (i.e., unretouched), intentionally produced flakes; (2) 105 cut marks created by intentionally produced, retouched flakes; and (3) 212 cut marks created by intentionally produced handaxes was utilized. The simple and retouched flake cut mark samples are drawn from that described in Domínguez-Rodrigo et al. (2009), and that sample created by handaxes from de Juana et al. (2010). A sample of 251 trampling marks (from Domínguez-Rodrigo et al., 2009) was also used in one of our comparative analyses.

Previous experimental research on protocols to differentiate trampling marks and cut marks made with simple and retouched flakes showed that four variables (mark cross-sectional shape, mark trajectory, incidence of shoulder effects and incidence of flaking on the mark shoulder), out of fourteen, explained most of the variance between samples, with a discrimination power that enabled differentiating trampling and cut marks samples in >90% of cases (Domínguez-Rodrigo et al., 2009). Additional experiments comparing cut marks made with retouched flakes with those created by handaxes revealed four discriminant variables (presence of multiple-clustered marks, presence of forked marks, number of multiple-clustered marks and number of forked marks) that separated the two types of butchery marks (retouched flake versus handaxe cut marks) in 80% of cases (de Juana et al., 2010). Because the present study, by providing new results on the morphology of

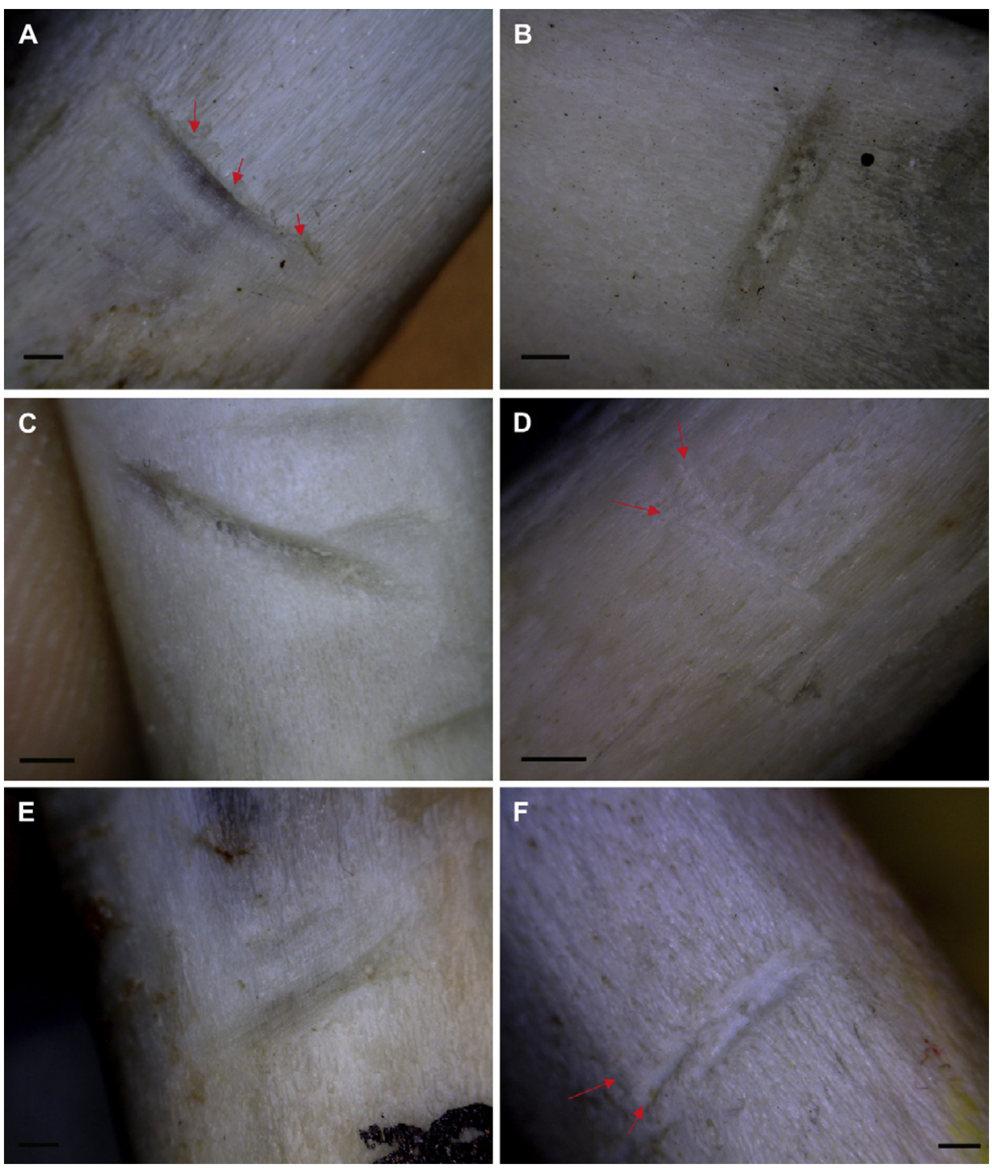


**Fig. 2.** Examples of marks commonly created by butchery with humanly unmodified rocks (HURs): (A) typical mark that is composed of deep, wide groove with internal microstriations and flaking on the mark shoulder (arrows); (B) a mark similar to A, but shallower; (C) a mark with a "classic" cut mark V-shaped cross-section; (D) forked mark with converging striae (arrows); (E) multiple-clustered marks, consisting of broad grooves with microstriations; (F) combined multiple-clustered and forked mark (arrows). As apparent with all the figured marks, the most common trajectory of HUR-produced marks is straight. Bar scale = 1 mm.

**Table 2**
Confidence interval (of null-hypothesis acceptance) from the robust Wilcoxon–Mann–Whitney (Mee's) test, comparing the cut mark sample created with natural rocks (this study) to cut mark sample made with flaked tools. Null-hypothesis of equality of proportions is accepted when 0.5 can be found within the interval. Bold intervals indicate non-significant differences. In the absence of *, $\chi^2$ tests also confirm significant differences between tool sets for each variable.

| | Cut mark sample created with simple (unretouched) flakes | Cut mark sample created with retouched flakes | Cut mark sample created with handaxes | Frequency of occurrence of each variable in cut mark sample created with humanly unmodified rocks |
|---|---|---|---|---|
| Trajectory | 0.014–0.090 | 0.001–0.054 | 0.780–0.912 | 91% = straight |
| Groove shape | 0.132–0.261 | 0.208–0.376 | 0.182–0.335 | 69% = open |
| Presence of shoulder effects | 0.110–0.222 | **0.338–0.541*** | 0.282–0.446** | 47% |
| Presence of flaking | 0.087–0.195 | **0.276–0.537**** | 0.259–0.417** | 37% |
| Forked marks | 0.048–0.149 | **0.469–0.648*** | 0.012–0.078 | 21% |
| Multiple-clustered marks | 0.052–0.153 | 0.170–0.376 | 0.598–0.771 | 19% |
| # forked marks | 0.012–0.210 | 0.034–0.119 | 0.624–0.798 | 0–1 |
| # multiple-clustered marks | 0.102–0.311 | 0.245–0.404 | 0.710–0.854 | 0–4 |

* $\chi^2$ values are significant when the robust Wilcoxon–Mann–Whitney test shows no significant differences.
** $\chi^2$ values are not significant.

cut marks produced by butchery with HURs (previously undocumented), is aimed at broadening the referential framework for interpreting the surficial marks on the Dikika fossils, we used these eight variables as potential discriminatory variables. These variables are defined in Table 1 (see also, Domínguez-Rodrigo et al., 2009; de Juana et al., 2010).

Quantitative data were analyzed statistically with R. R is a software that allows a high degree of plasticity in the design of the statistical tools and parameters to apply in the analysis (www.r-project.org). All graphs derived from these data were also created in R.

The first series of analyses was conducted to discriminate the characteristics of butchery marks created with HURs, when considering the *entire* sample of marks. Given the categorical nature of several variables, their distribution was not normal. Thus, instead of simply applying only logarithmic and square-rooted transformations to compare mean values of different variables, as is common procedure with continuous numeric variables, we used a robust approach to the sample in order to avoid the bias introduced by the use of non-normal samples. A robust method allows for an estimation of the matrix of variances-covariances of variables using robust estimators, which overcomes the biases introduced by skewed distributions (Wilcox, 2005). For instance, percentage bend correlations, Winsorized correlations and biweight midcovariances use M-estimators, which increase the reliability of statistical inference (Wilcox, 2005).

We used three complementary approaches for the analysis of median values and equality of variances when comparing each of the eight variables according to the four groups of marks: i.e., those created by (1) simple (unretouched) humanly produced flakes; (2) retouched flakes; (3) handaxes; and (4) HURs. The first approach was a robust Rust & Fligner test, which is a robust version of the Kruskal–Wallis test applicable to heterocedastic samples (García-Pérez, 2008). This test was applied to detect differences in each variable when comparing all four mark groups simultaneously. The second approach was a pair-wise comparison of the marks created by HURs and those resulting from the other three types of knapped stone tools per variable. It was performed with a robust Wilcoxon–Mann–Whitney analysis, through Mee's test. Last, as a comparative control, we also applied $\chi^2$ tests. However, this latter non-parametric test is usually less powerful, since it frequently identifies as different samples which are not when using a more robust estimator (García-Pérez, 1996; Hair et al., 1998; Wilcox, 2005). To prevent this shortcoming of the standard $\chi^2$ test, the probability values of this test were estimated using a Monte Carlo simulation based on 2000 replicates. All samples were bootstrapped 1000 times prior to submitting them to any of the three types of tests. The functions in $R^{mo}$ used were "rfanova" for Rust & Fligner/Kruskal–Wallis test and "mee" for Mee's test (García-Pérez, 2008). The standard "chisq.test" function for R was used for the $\chi^2$ test, with the argument "simulate.p.value = TRUE" for the Monte Carlo *p* simulation.

Once each variable was individually analyzed for all samples jointly and pair-wise, a multiple correspondence analysis (MCA) was carried out with the purpose of analyzing the differences of marks created by the five different processes studied (i.e., butchery

**Table 3**
Distribution of the main variables used to differentiate among marks created by various tool types and trampling (see data in Domínguez-Rodrigo et al., 2009 and de Juana et al., 2010). Data are in percentages.

| Tool type | Trajectory | Groove shape | Shoulder | Shoulder flaking | Fork-shaped marks | Multiple marks | No. of fork-shaped marks | No. of multiple marks |
|---|---|---|---|---|---|---|---|---|
| Handaxe | Straight = 90<br>Curvy = 5.1<br>Sinuous = 4.9 | V = 8.4<br>_/ = 91.6 | Present = 77<br>Absent = 23 | Present = 74.1<br>Absent = 25.9 | Present = 83.5<br>Absent = 16.5 | Present = 93<br>Absent = 7 | Range = 2–9 | Range = 2–10 |
| Retouched flake | Straight = 97<br>Curvy = 0<br>Sinuous = 2.9 | V = 5.7<br>_/ = 94.3 | Present = 74.3<br>Absent = 25.6 | Present = 51.4<br>Absent = 48.6 | Present = 41.9<br>Absent = 58.1 | Present = 61<br>Absent = 39 | Range = 1–2 | Range = 1–3 |
| Simple flake | Straight = 93.5<br>Curvy = 6.5<br>Sinuous = 0 | V = 96.7<br>_/ = 3.3 | Present = 32.9<br>Absent = 67.1 | Present = 14.6<br>Absent = 85.4 | Present = 1<br>Absent = 99 | Present = 0<br>Absent = 100 | 0 | 0 |
| Trampling | Straight = 29.8<br>Curvy = 16.7<br>Sinuous = 53.4 | V = 4<br>_/ = 96 | Present = 5.9<br>Absent = 94.1 | Present = 2.7<br>Absent = 97.3 | Present = 0<br>Absent = 100 | Present = 0<br>Absent = 100 | 0 | 0 |
| HUR | Straight = 91<br>Curvy = 7.9<br>Sinuous = 1.1 | V = 31<br>_/ = 69 | Present = 47<br>Absent = 53 | Present = 37<br>Absent = 63 | Present = 21<br>Absent = 79 | Present = 19<br>Absent = 81 | 0–1 | 0–4 |

with simple [unretouched] flakes, retouched flakes, handaxes and HURs, and trampling bones). This was also done in R, using the "ade4" library and the "dudi.acm" function. Correspondence analyses distribute variables spatially according to their inertia (that is, the amount of variance explained), treating each variable as a factor variable. Bidimensional plots were used to display the results of the scores of each individual mark. Points in these plots do not just show overlapping individual case values but each point also shows a type of mark, defined by the configuration of the different factors of each variable when all variables are combined. Biplots were also used to display the proportional inertia of each variable simultaneously with the scores of each mark type within each sample. Areas displaying 95% confidence intervals per sample type are also shown.

## 3. Results

### 3.1. Gross mark morphology

The experimental butchery sample includes 113 individual cut marks made by HURs. The marks are of variable morphologies, but overall very similar morphologically to those made by retouched flakes. This is unsurprising, as both tool types usually have cutting edges that are irregular and undulating in outline (i.e., not simple and straight, like a typical unretouched, humanly produced flake) (Fig. 2). A Rust & Fligner/Kruskal–Wallis test detected no statistically significant difference in mark morphology according to raw material type in any of the eight variables used for the present study ($p$ values >0.05).

The most common mark produced by HURs is a single broad groove with a straight trajectory. The broadness of a typical HUR-produced groove is similar to the broadness of grooves created by retouched flakes, as is the varying depths of both types of grooves (Fig. 2a,b), and the rareness of V-shaped cross-sections (the "classic" cross-sectional shape of cut mark main-grooves produced by simple [unretouched], humanly produced flakes) for grooves created by both types of effectors (Fig. 2c). Multiple-clustered and forked grooves are also present in low frequencies in the complete sample of HUR-produced marks (Fig. 2d,e). The HUR-created marks also evince microstriations in the troughs and walls of their main grooves; the microstriations are continuous along the lengths of groove troughs and walls. Marginally, some shallow marks with winding trajectory were documented, but in these rare cases, the winding showed a single S-shaped trajectory, which contrasts with the multiple S-shaped trajectories of other biostratinomic processes (e.g., trampling).

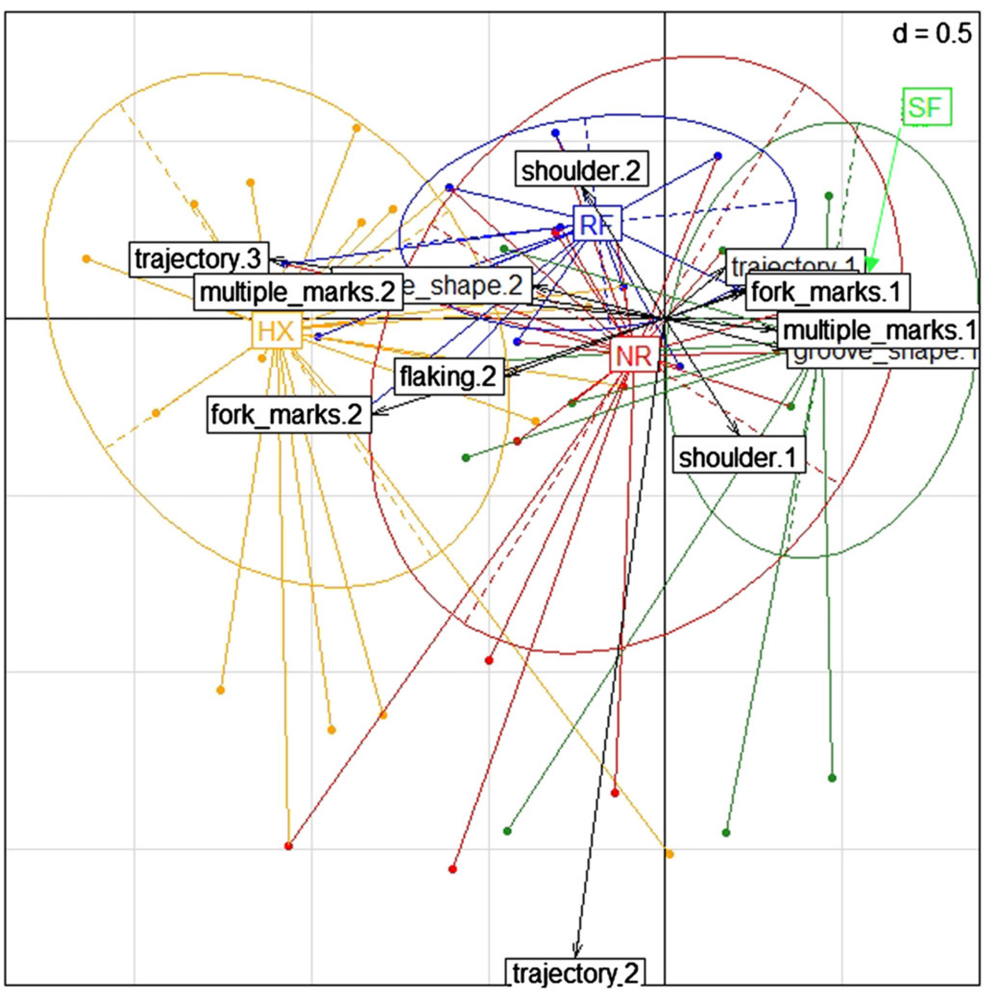


**Fig. 3.** Biplot of the scores of the MCA, showing the distribution of the cases and variables (black arrows and labels). Each group of marks (NR = those created by naturally-occurring, humanly unmodified rocks; SF = those created by simple [unretouched] flakes; RF = those created by retouched flakes; HX = those created by handaxes) is shown in different color with their 95% confidence intervals containing the centroid of each group. Each point is a cluster of individual marks consisting on a combination of factors in each variable. Therefore, each point can also be considered as a type of mark shaped by the interplay of the set of variables: mark trajectory, 1 = straight, 2 = curved, 3 = sinuous; groove cross-sectional shape, 1 = V-shaped, 2 = broad V-shaped; shoulder effect, 1 = absent, 2 = present; flaking, 1 = absent, 2 = present; forked marks, 1 = absent, 2 = present; multiple-clustered marks, 1 = absent, 2 = present. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.2. Statistical comparative results

The Rust & Fligner test detects significant differences (with alpha values <0.05) among the four types of marks (i.e., those produced by: [1] simple (unretouched) humanly produced flakes; [2] retouched flakes; [3] handaxes; [4] HURs) in all the eight variables analyzed. This result indicates that the four lithological effectors utilized do not create the same proportions of different mark-types at the assemblage (or total mark-sample) level. This, in turn, means that distributions of mark-types hold the potential to differentiate mark-samples created by each effector.

However, since the Rust & Fligner test cannot be used to detect which assemblage(s) is/are significantly different from the others, we applied pair-wise comparisons (provided by a robust Wilcoxon–Mann–Whitney test) of individual variables between mark-samples to detect which samples are significantly different when compared directly. Test results are summarized in Table 2, which shows that the sharpest contrast in samples is between that sample created by HURs and that created by simple (unretouched) flakes. These two samples are significantly different in their proportions of (1) various mark trajectories, (2) types of main-groove shape, (3) incidence of shoulder effects, (4) incidence of flaking on mark shoulders, (4) the presence or absence of forked and multiple-clustered marks, and (5) in the raw number of forked and multiple-clustered marks. These significant differences, detected by Wilcoxon–Mann–Whitney tests, are confirmed by $\chi^2$ tests.

The Rust & Fligner test also shows that the mark-sample created by HURs differs in proportion of varying mark-types from that produced with handaxes in all variables, but to a lesser extent than that made with simple (unretouched) flakes (the latter show confidence intervals with smaller values). The $\chi^2$ test partially confirms this result, but also indicates that in incidences of shoulder effects and flaking on the mark shoulder, the two mark-samples do not differ significantly.

Finally, the Rust & Fligner test indicates that the greatest resemblance between mark-samples is between that created by HURs and that created by retouched flakes, whose proportions of marks with shoulder effects, flaking on the mark shoulder and incidence of forked marks is indistinguishable statistically. This result is partially confirmed by the $\chi^2$ test, although for the incidences of shoulder effects and forked marks this test does not provide a significant value (Table 2).

In summary, the mean values for each variable within each mark-sample are statistically different (Table 3). However, it is important to determine how reflective this result is of the proportional representation of various mark-types created by each effector in order that one or more of these variables might be employed to discriminate mark-samples of unknown origins, like

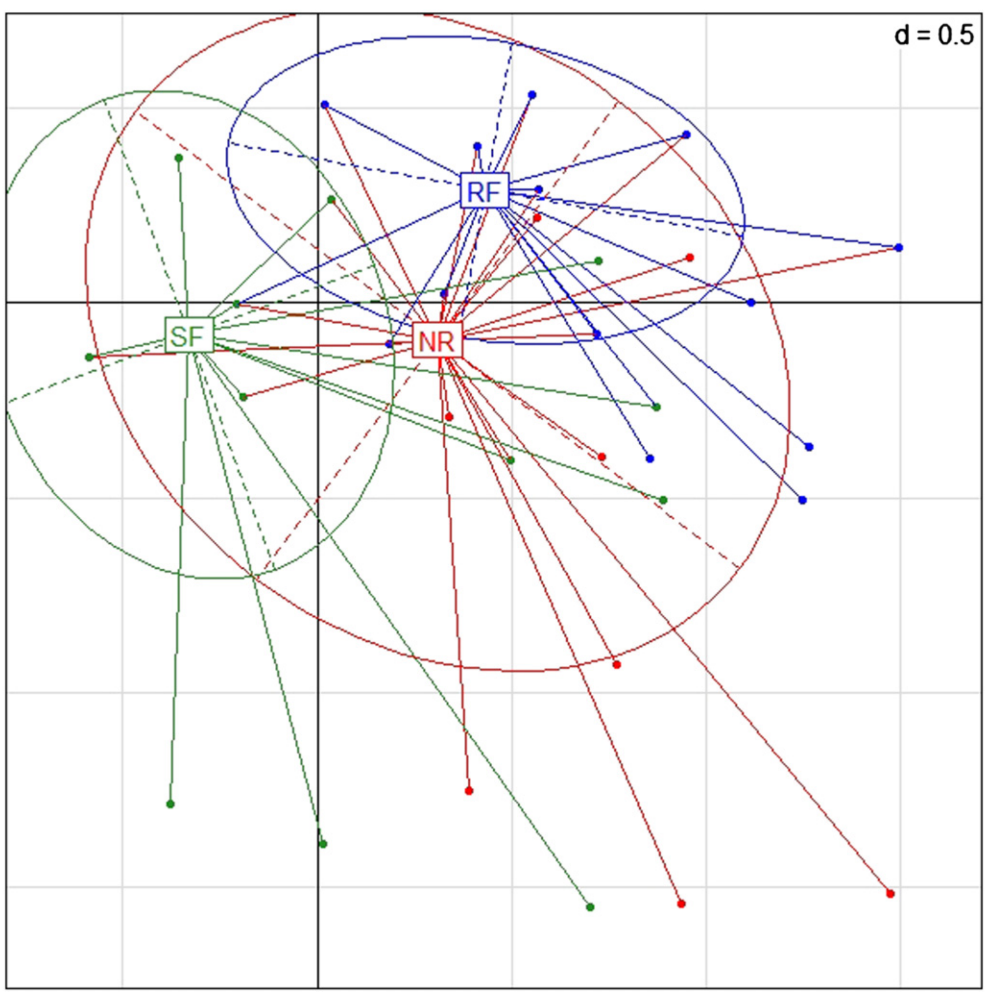


**Fig. 4.** Plot of the MCA scores showing the distribution of cases and their confidence intervals per group, excluding the handaxe-created mark-sample. Each group centroid is displayed marks (NR = those created by naturally-occurring, humanly unmodified rocks; SF = those created by simple [unretouched] flakes; RF = those created by retouched flakes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that sample from Dikika. The results of the MCA we conducted provided results relevant to this goal.

Comparing all the eight variables together, the MCA yielded a two-dimension solution with an inertia that explains 59% of the sample variance, with the number of forked marks and number of multiple-clustered marks as mostly relevant to differentiating samples created by handaxes from all other samples. Excluding those two variables, a new MCA yielded a more powerful result, with an inertia that accounts for 66% of the variance of the complete comparative sample set (Fig. 3). The first dimension explains as much as 46% of the sample variance and is determined (using factor variables with scores >0.5) by sinuous mark trajectory, the presence of flaking on the mark shoulder, incidence of forked marks and multiple-clustered marks and a main groove shape that is open. This set of factor variables discriminates most efficiently mark-samples created by handaxes from the other mark-samples. Curved mark trajectory and the presence of mark shoulder effects explain most of the second dimension. Most of the mark-types made with simple (unretouched) and retouched flakes are commonly produced by HURs, probably because HURs, collectively, possess a wide range of cutting edge morphologies, from uncomplicated, quite sharp borders to very irregular, duller ones (Fig. 1).

To better appreciate the degree of overlap in mark-samples created by simple (unretouched) flakes, retouched flakes, and HURs, we conducted a separate MCA that excluded the mark-sample created by handaxes. However, the degree of overlap among the three remaining groups was very similar to the results of the more inclusive MCA summarized above (Fig. 4).

Finally, employing the same complete set of variables listed in Table 1, we used another MCA to compare all four mark-samples created by butchery actions to the sample of marks created by the trampling experiments of Domínguez-Rodrigo et al. (2009). This results in some clear separations between samples. The MCA yielded a two-dimension solution where the inertia explains 61% of the sample variance. If using those factor variables with scores >0.5, presence of shoulder flaking, presence of shoulder effects, the number of forked and the number of multiple-clustered marks explain most of the first component (inertia = 0.35); proportion of trajectory types and groove shapes explains a large part of the second component (inertia = 0.26). Once again, these results appear to be due to analytical inclusion of handaxe-created mark-sample, more easily discriminated than the other samples from each other. The confidence interval of the trample-mark sample overlaps just barely with cut mark samples created by simple,

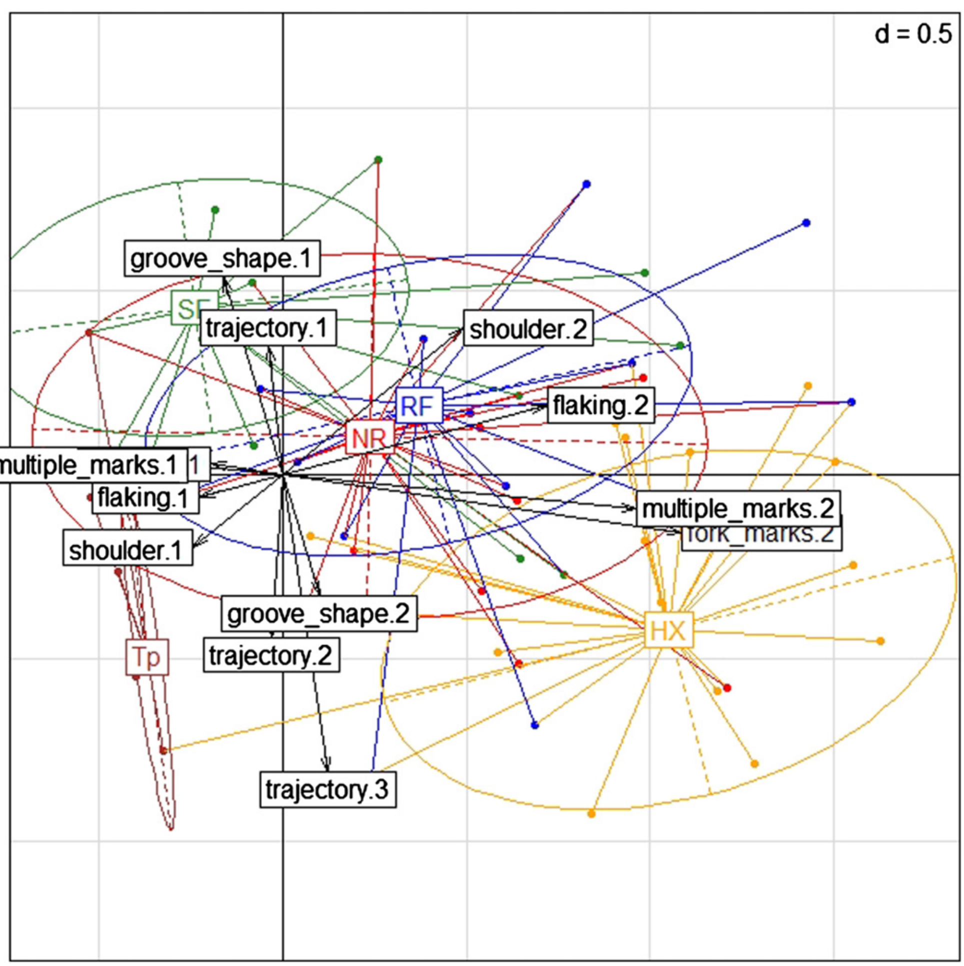


**Fig. 5.** Biplot of the scores of the MCA, showing the distribution of the cases and variables (black arrows and labels) for all the cut-marked samples and trampling. Each group of marks (NR = those created by naturally-occurring, humanly unmodified rocks; SF = those created by simple [unretouched] flakes; RF = those created by retouched flakes; HX = those created by handaxes; TP = those created by trampling) is shown in different color with their 95% confidence intervals containing the centroid of each group. Each point is a cluster of individual marks consisting on a combination of factors in each variable. Therefore, each point can also be considered as a type of mark shaped by the interplay of the set of variables: mark trajectory, 1 = straight, 2 = curved, 3 = sinuous; groove cross-sectional shape, 1 = V-shaped, 2 = broad V-shaped; shoulder effect, 1 = absent, 2 = present; flaking, 1 = absent, 2 = present; forked marks, 1 = absent, 2 = present; multiple-clustered marks, 1 = absent, 2 = present. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

retouched and handaxe tools, *but almost half of it overlaps* with the mark-sample created by HURs (Fig. 5).

## 4. Discussion and conclusion

Collectively, our results show the great variability in shape and other properties of cut marks created by butchery utilizing HURs. Cut marks inflicted by HURs overlap extensively in their morphologies with those linear marks created by simple (unretouched), retouched and handaxe tools used for butchery *and* with linear trampling marks. These results counsel extreme caution in diagnosing linear marks on fossil bone surfaces of unknown origin as cut marks created by hominin butchery with HURs.

More particularly, our results—showing such a high degree of morphological overlap in linear bone surface marks created by four distinct lithological effectors—demonstrate that diagnostically useful patterns of mark form can only be detected at the level of the total mark-sample. This conclusion negates the supposition that the morphology of any individual bone surface mark drawn from a sample of unknown origin is sufficient to accurately diagnose the lithological effector that produced that mark—effectively corroborating our earlier assertion (Domínguez-Rodrigo et al., 2010), and that of previous researchers (e.g., Binford, 1981; Bunn, 1991; White, 1992), that a configurational approach is the only appropriate means to evaluate the veracity of potential evidence of early hominin butchery.

By all available measures, the claim of butchery marks on two fossil bone specimens from Dikika fails the configurational litmus test. It might be possible to experimentally impart, using an HUR, butchery marks that are similar morphologically to some of those preserved on the Dikika fossils (McPherron et al. claim to have produced such marks using HURs in their popular press, online response to Domínguez-Rodrigo et al., 2010). But, that potentiality in no way nullifies the fact, as documented in detail by Domínguez-Rodrigo et al. (2010), that randomly imparted striae can also mimic the Dikika marks, nor does it change the abrasive sedimentary context (i.e., a depositional context in which there is a high predictive likelihood for the production random striae on bone surfaces) of the fossils. Simply on these bases, the null hypothesis of non-anthropogenic origin(s) for the Dikika marks remains unfalsified. The experimental work reported here—by broadening the actualistic framework on lithologically derived linear bone surface mark morphologies— only amplifies this conclusion.

Specimen DIK-55-3, one of the two Dikika fossil bone specimens with purported butchery damage (McPherron et al., 2010), exhibits eight individual mark-types: (1) striae fields (mark DIK-55-3-D); (2) wide shallow grooves with winding trajectories (DIK-55-3-G1); (3) multiple parallel microstriations, also with winding trajectories (DIK-55-3-H2); (4) broad and deep grooves with marginal microstriations at the base of mark walls, with substantial flaking of the mark bases (DIK-55-3–I); (5) tick (or checkmark)-shaped marks (DIK-55-3-B and C); (6) sub-parallel, intersecting marks with V-shaped cross-sections (DIK-55-3-A); (7) multiple sub-parallel and partially curved marks that lack microstriations (DIK-55-3–F); (8) semi-lunate-shaped marks (DIK-55-3-J). Specimen DIK-55-2, on the other hand, shows deep linear marks, with V-shaped cross-sections, in addition to other marks alleged by McPherron et al. (2010) to be “high-confidence stone-tool cut marks and hammerstone percussion marks”, but which are actually virtually identical to known trample marks (Domínguez-Rodrigo et al., 2010).

Relevant to these basic observations on the published Dikika bone surface damage are two points from our current experimental results. First, in the extremely large, aggregated sample of *all butchered bones from all experiments combined*, no single specimen evinces such a wide diversity of marks as is documented on DIK-55-3—a single fossil bone specimen from Dikika. Second and more specifically, our large HUR mark-sample *lacks completely* most of the mark-types listed above that occur on DIK-55-3. For instance, marks with V-shaped cross-sections and broad grooves with microstriations in our HUR mark-sample have straight or, less frequently, curved (Fig. 1) but never multiple sinuous or winding trajectories, as is, in contrast, documented in the Dikika mark-sample (McPherron et al., 2010). No tick-shaped marks with the same morphology as DIK-55-3-B and C are found in our HUR mark-sample; neither are marks with the morphologies of DIK-55-3-A, DIK-55-3-F, DIK-55-3-G1, DIK-55-3-H2, DIK-55-3-I and DIK-55-3-J.

Further, such a diversity of mark types on a single bone specimen is not documented in Domínguez-Rodrigo et al.’s (2009) large collection of experimentally butchered bones. However, *all* the mark-types most prone to be mistaken with (and argued by McPherron et al., (2010) to be) cut marks, such as DIK-55-2-A, DIK-55-2-B, DIK-55-3-A, DIK-55-3-D, DIK-55-3-G1, DIK-55-3-G2 and DIK-55-3-H2, occur in experimentally trampled bone assemblages (Domínguez-Rodrigo et al., 2010; see also Behrensmeyer et al., 1986, 1989; Fiorillo, 1989). These congruities in mark types between the Dikika and modern trample-mark samples, along with the presence of microabrasion on the Dikika fossil specimens, led us to suggest that trampling and/or some other process(es) of incidental on/within-substrate movement of the Dikika fossils was the major factor in the production of the Dikika bone surface modifications (Domínguez-Rodrigo et al., 2010).

That does not mean, however, that other non-anthropogenic processes can be completely excluded as other producers of *some* of the Dikika bone surface marks. For instance, DIK-55-3-I is too broad a linear feature to have been created by any experimentally studied lithological effector, including a sedimentary particle(s) during incidental bone movement on and/or within its depositional substrate. The exfoliated base of the mark suggests it was created by a biochemical process, an inference supported by its contrasting color compared to the rest of the specimen’s bone surface and by its overall ragged and flaked microtopography. The survival of lamellar “islands” of original cortical bone surface within the mark’s base (see McPherron et al., 2010, supplementary information), and its lack of parallel microstriations along most of its main groove also point to bone modification accomplished by a process other than one employing a lithological effector. The few microstriations the mark does contain occur along the wall and margin of one of the sides of the groove and could have been caused by postdepositional trampling of the mark (as documented in Domínguez-Rodrigo et al., 2010) or by sediment particles trapped between the surface of the groove and a growing plant root, either process of which could have occurred *subsequent* to the mark’s initial production through biochemical erosion.

Dikika marks DIK-55-3-B and C are tick-shaped. Tick-shaped marks on bone surfaces can be created by Nile crocodiles (*Crocodylus niloticus*) (Njau and Blumenschine, 2006), and also by Griffon vultures (*Gyps fulvus*) (Domínguez-Solera and Domínguez-Rodrigo, in prep). However, neither crocodiles nor vultures produce tick-shaped marks with the same microscopic characteristics observed on DIK-55-3-B. Close examination of the published image of this mark (McPherron et al., 2010, supplementary information: Figure 15) reveals that this mark might not even be taphonomic in origin, but instead a vascular feature. It certainly lacks the microstriations typical of most linear damage caused by various lithological effectors.

On balance, the replication of all mark-types preserved on DIK-55-2 and DIK-55-3 in modern bone trampling experiments (Domínguez-Rodrigo et al., 2010), and their lack of replication in our HUR butchery experiments (reported above) reinforces our initial diagnosis that the Dikika fossil marks were probably caused primarily by incidental movement of the marked bone specimens on and/or within their depositional substrate(s) (Domínguez-Rodrigo et al., 2010). Further, whether or not McPherron et al. can use HURs to produce bone surface marks that mimic those on the Dikika fossils (as-yet unreported in a peer-reviewed scientific journal, but alluded to in the popular press) is almost immaterial to the ultimate relevance of the fossil marks to increasing our understanding of the evolution of early hominin butchery. The Dikika marks are still on fossils derived from a sedimentary context that includes a coarse-grained facies, one with high potential to impart random striae on bone surfaces with any movement of those bones on and/or within that depositional substrate. The marked Dikika fossils are still a singular occurrence of purported butchery evidence from >2.6 Ma. The marked Dikika fossils are still surface finds and there is still no associated, larger bone assemblage reported that could provide a site-level context for their taphonomic interpretation. There are still no flaked stone tools associated with the marked Dikika fossils, nor, for that matter, are there HURs associated with the marked Dikika fossils that are claimed to be the actual implements of butchery. Behrensmeyer et al. (1989) and Oliver (1989) report on morphologically convincing butchery mark mimics in unambiguous non-anthropogenic contexts (see also, Fiorillo, 1989). Together, this accumulated knowledge demonstrates unequivocally that bone surface mark morphological equifinality is a pervasive concern in the case of the Dikika bone modifications. That is why we continue to stress just how important depositional and site contexts are for evaluating the veracity of the Dikika team's claims of >2.6 Ma butchery. Those contexts are insufficient to support a hypothesis that the Dikika marked fossils are bones butchered by early hominins.

Rejection of the Dikika butchery claims is not equivalent to rejection of the possibility that hominins acquired and consumed vertebrate carcass resources >2.6 Ma. Based on available fossil evidence and on theoretical grounds, we think it very likely that hominins engaged in some sort(s) of regular meat-eating prior to the invention of flaked stone technology (e.g., Pickering, 2010; Pickering and Egeland, 2009; Pickering and Domínguez-Rodrigo, 2010, 2011; Pickering and Bunn, 2011). But, the evidentiary basis of the claim that Dikika is the archaeological validation of this belief does not withstand configurational testing.

## Acknowledgements

We thank the students of the 2010–2011 Paleolithic course of the Archaeology Grade, Complutense University, for conducting the butchery experiments reported on here. We are extremely thankful to D. Chessel, A. B. Dufour and S. Dray for their design of the "ade4" package for R. We are also thankful to the members of the R Project Foundation for Statistical Computing for keeping this software available (http://www.r-project.org). Pickering thanks his family for support and patience. Thanks to Richard Klein and three anonymous reviewers for helpful comments that improved this paper.

## References

Behrensmeyer, A.K., Gordon, K.D., Yanagi, G.T., 1986. Trampling as a cause of bone surface damage and pseudo-cutmarks. Nature 319, 768–771.

Behrensmeyer, A.K., Gordon, K.D., Yanagi, G.T., 1989. Nonhuman bone modification in Miocene fossils from Pakistan. In: Bonnichsen, R., Sorg, M. (Eds.), Bone Modification. Center for the Study of the First Americans, Orono (ME), pp. 99–120.

Binford, L.R., 1981. Bones: Ancient Men and Modern Myths. Academic Press, New York.

Bunn, H.T., 1991. A taphonomic perspective on the archaeology of human origins. Annual Review of Anthropology 20, 433–467.

Domínguez-Rodrigo, M., 1997. Meat-eating by early hominids at the FLK 22 Zinjanthropus site, Olduvai Gorge, Tanzania: an experimental approach using cut mark data. Journal of Human Evolution 33, 669–690.

Domínguez-Rodrigo, M., Pickering, T.R., Semaw, S., Rogers, M.J., 2005. Cutmarked bones from Pliocene archaeological sites at Gona, Afar, Ethiopia: implications for the function of the world's oldest stone tools. Journal of Human Evolution 48, 109–121.

Domínguez-Rodrigo, M., de Juana, S., Galán, A., Rodríguez, M., 2009. A new protocol to differentiate trampling marks from butchery cut marks. Journal of Archaeological Science 36, 2643–2654.

Domínguez-Rodrigo, M., Pickering, T.R., Bunn, H.T., 2010. Configurational approach to identifying the earliest hominin butchers. Proceedings of the National Academy of Sciences (USA) 107, 20929–20934.

Fiorillo, A.R., 1989. An experimental study of trampling: implications for the fossil record. In: Bonnichsen, R., Sorg, M. (Eds.), Bone Modification. Center for the Study of the First Americans, Orono (ME), pp. 61–71.

García-Pérez, A., 1996. Behaviour of sign test and one sample median test against changes in the model. Kybernetika 32, 159–173.

García-Pérez, A., 2008. Métodos avanzados de estadística aplicada. Métodos robustos y de remuestreo. UNED, Madrid.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C., 1998. Multivariate Data Analyses. Prentice Hall, Englewood Cliffs.

de Juana, S., Galán, A.B., Domínguez-Rodrigo, M., 2010. Taphonomic identification of cut marks made with lithic handaxes: an experimental study. Journal of Archaeological Science 37, 1841–1850.

McPherron, S.P., Alemseged, Z., Marean, C.W., Wynn, J.G., Reed, D., Geraads, D., Bobe, R., Béarat, H.A., 2010. Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. Nature 466, 857–860.

McPherron, S., quoted in Vergano, D., 2010. Cut Mark Controversy Commences over 3.4 Million-year-old Bones. USAToday.com. http://content.usatoday.com/communities/sciencefair/post/2010/11/cut-mark-controversey-commences/1.

Njau, J.K., Blumenschine, R.J., 2006. A diagnosis of crocodile feeding traces on larger mammal bone, with fossil examples from the Plio-Pleistocene Olduvai Basin, Tanzania. Journal of Human Evolution 50, 142–162.

Oliver, J.S., 1989. Analogues and site context: bone damages from Shield Trap Cave (24CB91), Carbon County, Montana, USA. In: Bonnichsen, R., Sorg, M. (Eds.), Bone Modification. Center for the Study of the First Americans, Orono (ME), pp. 73–98.

Pickering, T.R., 2010. Invited comment on Shipman "The animal connection and human evolution". Current Anthropology 51, 529–530.

Pickering, T.R., Egeland, C.P., 2009. Experimental zooarchaeology and its role in defining the investigative parameters of the behavior of Early Stone Age hominids. In: Schick, K., Toth, N. (Eds.), The Cutting Edge: New Approaches to the Archaeology of Human Origins. Stone Age Institute Press, Bloomington (IN), pp. 171–177.

Pickering, T.R., Bunn, H.T., 2011. Meat-foraging by Pleistocene African hominins: tracking behavioral evolution beyond baseline inferences of early access to carcasses. In: Domínguez-Rodrigo, M. (Ed.), Theoretical Debates in the Archaeology of Human Origins. Cambridge University Press (submitted), Cambridge.

Pickering, T.R., Domínguez-Rodrigo, M., 2010. Chimpanzee referents and the emergence of human hunting. The Open Anthropology Journal 3, 107–113.

Pickering, T.R., Domínguez-Rodrigo, M., 2011. Can we use chimpanzee behavior to model early hominin hunting? In: Domínguez-Rodrigo, M. (Ed.), Theoretical Debates in the Archaeology of Human Origins. Cambridge, Cambridge University Press (submitted).

Semaw, S., Rogers, M.J., Quade, J., Renne, P.R., Butler, R.F., Domínguez-Rodrigo, M., Stout, D., Hart, W.S., Pickering, T.R., Simpson, S.W., 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.

White, T.D., 1992. Prehistoric Cannibalism at Mancos 5MTUMR-2346. Princeton University Press, Princeton (NJ).

Wilcox, R.R., 2005. Introduction to Robust Estimation and Hypothesis Testing. Elsevier Academic Press, London.