Journal of Archaeological Science 45 (2014) 270–283

Contents lists available at ScienceDirect

Journal of Archaeological Science

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

The effect of raw material on inter-analyst variation and analyst accuracy for lithic analysis: a case study from Olduvai Gorge

Tomos Proffitt*, Ignacio de la Torre

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

A R T I C L E I N F O

Article history:
Received 21 November 2013
Received in revised form
29 January 2014
Accepted 25 February 2014
Available online 12 March 2014

Keywords:
Lithic technology
Olduvai Gorge
Blind tests
Inter-analyst variability
Analyst accuracy
Refit analysis

A B S T R A C T

This study aims to understand what effect, in terms of inter-analysis variation and analyst accuracy, different raw material types have on modern technological analyses of lithic assemblages. This is done through a series of blind analysis tests undertaken on experimentally derived assemblages of cores and flakes. Novelties of our approach include the introduction of refit studies as a method to assess analyst's accuracy, and the use of statistical tests specifically designed to address inter-analyst variability, common in other disciplines but rarely used in Archaeology. The experimental assemblages were produced from raw materials collected at Olduvai Gorge, an archaeological sequence that has been a source for studies of early human technology for several decades, and where re-analyses of the same assemblages have usually offered different interpretations. The results of the blind analyses are compared to the true technological values obtained through full refit analysis of the experimental material, and suggest that there is a significant difference in terms of inter-analyst variability as well as accuracy related to different raw materials. Our paper highlights the interpretative problems posed by difficult-to-analyse raw materials such as quartzite, and stresses subjectivity present in stone-tool technological studies, which may contribute to explain differences in the interpretation of Early Stone Age lithic assemblages.

1. Introduction

Since Mary Leakey's (1971) seminal work, the importance of the Olduvai assemblages as a continued source of archaeological data and interpretations regarding technological skill (de la Torre and Mora, 2005), manual dexterity, cognitive evolution (Ludwig, 1999), wider landscape use (Potts, 1988; Blumenschine et al., 2008) and raw material procurement (Kimura, 1997, 1999, 2002; Kyara, 1999) has been made clear. It is noticeable, however, that from the same assemblages different results and varying (sometimes contradictory) interpretations are produced by different researchers (see review by de la Torre and Mora, 2009).

The implementation of lithic technological studies requires the correct identification of numerous common technological characteristics and markers located on lithic material. Nonetheless, to date surprisingly few studies have attempted to identify the level of inter-analyst accuracy associated with the identification of technological characteristics of lithic assemblages. At a general level, studies investigating inter-analyst variability have concentrated on the identification of variability caused by random and systematic error of the analyst (Gnaden and Holdaway, 2000), of typological identification between analysts and by individual analysts over time (Beck and Jones, 1989; Fish, 1978; Whittaker et al., 1998), the correctness of lithic measurements (Fish, 1978) and cortex coverage (Fish, 1978).

In terms of identifying analyst accuracy on assemblages made up of a single raw material type, Driscoll (2011), Jeske and Lurie (1993), and Perpère (1986) conducted blind tests based on a technological approach. Driscoll (2011) investigated how the correctness of a number of analysts of varying skill level was affected by quartz. Jeske and Lurie (1993) conducted blind tests in order to identify attributes which would distinguish bipolar knapping from freehand knapping, and Perpère (1986) used three lithic analysts to identify Levallois and non-Levallois flakes from an archaeological assemblage.

More generally, blind analyses have been implemented to assess the level of inter-analyst correctness in the identification of bone modification on faunal remains (e.g. Blumenschine et al., 1996; Gobalet, 2001), and the identification and quantification of use wear patterns within lithic residue and microwear analysis (Newcomer et al., 1986; Wadley et al., 2004; Rots et al., 2006; Crowther and Haslam, 2007). Nonetheless, to date no studies

* Corresponding author. Tel.: +44 7889 369 826.
E-mail address: t.proffitt@ucl.ac.uk (T. Proffitt).

http://dx.doi.org/10.1016/j.jas.2014.02.028
0305-4403/

have investigated what effect multiple raw material types within an assemblage has on the accuracy of the identification of lithic characteristics commonly used in technological lithic analyses.

This paper reports on an investigation designed to identify how different raw materials available at Olduvai Gorge may affect an analyst's ability to correctly undertake a technological lithic analysis. This is achieved through an experimental programme incorporating both experimental lithic material and detailed blind analysis of that material, with the primary aim of assessing how different raw materials affects lithic analysis, specifically that of Oldowan material from Olduvai Gorge, for which a variety of often-contradictory interpretations have been made of the same assemblages (e.g. Leakey, 1971; Kimura, 1999; Ludwig, 1999; de la Torre and Mora, 2005).

2. Materials and methods

2.1. Raw materials

Three of the primary raw materials documented in the Olduvai Beds I and II assemblages are quartzite, basalt and chert (although chert was available during specific periods only, particularly in Bed II) (Hay, 1976). For this reason, these rock types were selected for the blind tests, and locally sourced from Olduvai Gorge for the experimental programme.

The quartzite was collected from the main quartzite source, the Naibor Soit inselberg, located north of the confluence of the Main Gorge and Side Gorge. Hay (1976) referred to the artefacts derived from Naibor Soit as quartzite, a rock classification that was maintained in subsequent cornerstone studies of the Olduvai assemblages (e.g. Potts, 1988; Jones, 1994; Leakey and Roe, 1994). This quartzite is of metamorphic origin, coarse-grained, and possesses micaceous layers which are foliated and lineated (Hay, 1976). Naibor Soit is easily reached today as it stands prominent within the landscape, with variable-sized unmodified blocks of quartzite eroding from the inselberg.

The Olduvai Gorge chert is chemically produced through the precipitation of sodium silicate minerals from the saline, alkaline Olduvai Lake during Lower Bed II times, over a relatively short period of time (less than 10,000 years) (Hay, 1976; Stiles et al., 1974). These fine-grained nodules have an irregular shape, vary greatly in dimensions, and possess a clearly visible chalk cortex. A large outcrop of chert known as the Main Chert Unit (Stiles et al., 1974) is located on the southern side of the Side Gorge. This source is currently exposed in a layer ranging in thickness from 5 cm to 30 cm (Stiles et al., 1974), from which the nodules for the experimental programme were collected.

Basalt was available as boulders, blocks and cobbles within seasonal rivers and streams in Beds I and II, originating from the surrounding volcanic outcrops (Kyara, 1999; Hay, 1976). The lava types available to hominins in this region included basalt, andesite, trachyte, phonolite and nephelinite (Hay, 1976). In this study, basalt was selected to conduct the experimental analysis, and collected from ravines within the Main Gorge.

2.2. Production of experimental assemblages

In total, six separate experimental assemblages were produced, two of each raw material (blocks of quartzite, nodules of chert and basalt cobbles). The target of each knapping sequence was to produce core and flake assemblages which were comparable to classic Oldowan assemblages. Knapping was undertaken by two individuals (Tomos Proffitt and Adrian Arroyo) with between 3 and 4 years' experience in flaking raw materials from Olduvai Gorge. Each nodule was knapped using free-hand hard hammer percussion technique and with a quartzite hammerstone. Rather than obtaining a specific core form, each knapper's primary aim was to produce the highest number of useable sharp-edged flakes as possible from each nodule, as has been suggested for Oldowan use (Toth, 1985).

The knappers adhered to a number of flaking protocols in order to maintain the highest level of repeatability as possible, and to maintain the experimental assemblages' consistency with known Oldowan archaeological assemblages. Firstly, no intentional platform preparation of the core was undertaken, with only naturally-occurring platforms being utilised; secondly, once an insufficient angle between the knapping platform and the flaking surface was produced, no rejuvenation of the knapping platform was conducted. Similarly, when a flaking surface exhibited a loss of convexity, no rejuvenation of knapping platforms or flaking was followed; at the point where rejuvenation was necessary, the knapper was instructed to locate a new naturally-occurring knapping platform and flaking surface and continue reduction from this location. Thirdly, knappers were informed that the core could be rotated in any direction required in order to facilitate the production of flakes. Furthermore, no intentional retouch of flakes was undertaken. Following previous classifications of Olduvai Gorge assemblages (e.g. de la Torre and Mora, 2005), only pieces larger than 20 mm ($n = 427$; quartzite $n = 208$, chert $n = 130$, basalt $n = 89$) were subjected to full refit and technological analysis.

2.3. Blind analysis

Four lithic analysts (BAs) assessed the entire experimental assemblages; these analysts possessed varying levels of experience in lithic analysis, ranging from over 10 years' experience (BA1 and BA3) to between 3 and 10 years' experience (BA2 and BA4). Each of the six experimental assemblages was presented to each analyst in no set order. Each piece was analysed initially in terms of technological category, followed by the relevant observations as set in Fig. 1.

Similarly to Driscoll (2011), each analyst was given a definition of all possible attributes to be identified. The first stage of analysis consisted of assigning each experimental artefact to a technological category (complete flakes, broken flakes, angular chunks, retouched material, cores, and small waste debitage). Further attributes were recorded on complete flakes (de la Torre and Mora, 2005; Toth, 1985). For fragmented flakes, only the identification of dorsal surface cortex and the presence of knapping accidents were encoded. All attributes used for this blind analysis have been used to some extent in a number of technological analyses of Oldowan material (Kimura, 1997, 1999, 2002; Ludwig, 1999; de la Torre and Mora, 2005; de la Torre, 2011; Stout et al., 2010). Furthermore, from each analyst's identification of striking platform and dorsal surface cortex coverage, Toth's (1985) flake categories were calculated, providing a further attribute that is widely used in Early Stone Age lithic studies.

Once all assemblages were studied by each analyst according to the attributes detailed above, results had to be compared to the "real values". In previous work based on blind tests, it has been common to utilise a "Gold Standard", i.e. a set of true values, usually set by one experienced analyst or Reference Observer (Gnaden and Holdaway, 2000), to which all blind analysis results can be compared (see also Driscoll, 2011). It has been noted, however, that this method of obtaining the "true values" can be biased, for an experienced analyst may also be incorrect in their analysis of the material (Gnaden and Holdaway, 2000).

In order to minimise subjectivity of the results, rather than using the observations of the most experienced of the four analysts as the true values, following the blind tests the experimental assemblages