Toolstone constraints on knapping skill: Levallois reduction with two different raw materials

Metin I. Eren a,b,*, Stephen J. Lycett b, Christopher I. Roos a, C. Garth Sampson c

a Department of Anthropology, Southern Methodist University, Heroy Building (ISEM), Dallas, Texas 75275 0336, USA
b Department of Anthropology, University of Kent, Canterbury, UK
c Department of Anthropology, Texas State University, USA

ABSTRACT

Lithic raw material constraints are widely assumed to be a determining factor of flaked stone tool morphology, but this assumption remains largely untested. We conducted a controlled experiment to determine whether a knapper’s growing replication skills would be hindered if the toolstone used was switched from large flakes of an easily worked chert to nodules of less tractable one. Two batches of Preferential Levallois cores were knapped, an earlier series made from standardised large flakes of sediments dominated by chalcedonic quartz and a more challenging one using variably-shaped, cortical nodules of microcrystalline quartz that varies in the completeness of quartz replacement of calcite and dolomite. Skill level markers were designed to measure the knapper’s ability to achieve a series of set goals. These were quantified and subjected to statistical testing. In all but one test, significant increases in skill could be detected from the earlier to the later batch of reductions, despite the drop in toolstone quality. Significant improvements in the consistency of the knapper’s output could also be detected. However, the switch to a more challenging, nodular chert did require extra shaping, which resulted in more waste. This masked visible progress towards producing a less costly core. Overall, our results do not support the assumed primacy of toolstone constraints over other factors in influencing the morphology of flaked stone tools.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

It has long been recognized that flaking stone tools is a reductive process in which slivers of stone are knapped from a larger rock (e.g. Evans, 1872: 13; Frison, 1968; Shott, 2010: 276–277). It is equally well known that isotropic rocks, i.e. those free of cleavage planes or other inclusions that inhibit the free passage of energy, are the ones most suitable for successful initiation and control of conchoidal fracture. The best exhibit properties that combine elasticity, brittleness, hardness, and homogeneity (Goodman, 1944; Inizan et al., 1999; Whittaker, 1994). The exact combination of these properties varies from one potentially suitable raw material (toolstone) to another, and it is reasonable to suspect that the exact mix of such properties in any particular toolstone could affect the form of Palaeolithic artefacts made from it (Archer and Braun, 2010; Dibble, 1985; Goodman, 1944; Isaac, 1972; Jones, 1979). Untested assumptions about toolstone constraints reached their apogee at the so-called Movius Line – a putative subcontinental pattern of artefact variation where bifacial and prepared core reduction is present in India and to the west, but is all but absent farther east. The purported lack of better-quality toolstone east of the Line, although once a popular explanation, has now fallen out of favour (Bar-Yosef et al., In press; Brantingham et al., 2000; Lycett and Bae, 2010; Norton et al., 2006; Schick, 1994). While there have been repeated calls to consider carefully the effects of raw material parameters and their influences on artefact form and assemblage composition (Clegg, 1977; Isaac, 1972; Kuhn, 1992), an understanding of how these factors play out at the assemblage level remains limited.

Two different sets of toolstone properties are thought to influence lithic artefact form (e.g. Goodman, 1944; Jones, 1979; Clark, 2001: 1). The first set is internal, i.e. the mechanical flaking properties already listed. The second set is external, namely the form (size, shape, surface regularity, and presence of cortex) of the initial nodule, block, or blank from which the flakes are struck. External parameters have been claimed to be a primary cause of differences between whole assemblages made of chert.
that may exhibit a similar mix of internal properties (Ashton and McNabb, 1994).

Despite its seemingly logical basis, the case for toolstone constraints on artefact form is hardly supported by experiment, and there are “few studies that actually document the impact of the properties of raw materials on the way in which particular tool types were produced” (Holdway and Stern, 2004: 55). Indeed, several recent case studies suggest that raw material may not be the dominant factor in regulating the final form of knapped artefacts. Archer and Braun (2010), for instance, undertook three-dimensional geometric morphometric analyses of Acheulean handaxes from Elandsfontein (South Africa) and compared them to experimentally-produced replicas. Their comparisons reveal that much of the shape variation between handaxes could not be accounted for solely by raw material. A more emphatic refutation of the argument may be found at Castel di Guido (Italy) where handaxes were knapped in both stone and bone. Here, Costa’s (2010) morphometric analysis of outline form failed to detect any statistical difference in plan shape between stone and bone handaxes despite the fundamental differences in flaking properties of the two materials used. In a similar manner, Clarkson’s (2010) multivariate morphometric comparison of cores from the Howiesons Poort techno-complex (South Africa) revealed that classification scores based on geographic region were more accurate than those based on raw material (72.8 versus 46.0% accuracy, respectively). While these results do not deny a role for raw material in causing differences in cores between regions, they do indicate that “raw material differences would appear to be subservient to other causes of variation” Clarkson (2010: 53).

It would appear, then, that the jury is still out on whether or not toolstone quality exercises any significant constraint on resulting tool form. We doubt that the question will be resolved conclusively by ever-more sophisticated analyses of prehistoric materials, because there always remain too many unknowns. To fill these gaps requires so large a body of refitted material as to be beyond our present means and abilities. To circumvent the problem we designed a replication experiment which documented how the improving knapping skills of an individual knapper (Eren) might be interrupted when the toolstone on which he was honing his technique was switched to a material that was less tractable in both its internal and external properties.

The experiment focused on Eren’s ability to execute a particularly challenging knapping routine — that of Levallois core production. Levallois reduction was common during the Middle Palaeolithic and has been recorded at sites in Africa, Europe, and western Asia (Dibble and Bar-Yosef, 1995 and chapters therein). Being part of a group of ‘prepared core’ reduction strategies, there is disagreement on how to define Levallois (Chazan, 1997; Dibble and Bar-Yosef, 1995; Sandgathe, 2005; Schlanger, 1996; Van Peer, 1992). Traditional definitions are rooted in the concepts of preparation or predetermination of the flake products, but more recent definitions see only a reduction sequence that results in a core with specific ‘volumetric’ properties (Boëda, 1995; Chazan, 1997; Brantingham and Kuhn, 2001). Here, Levallois reduction results in a bifacial core on which both faces are hierarchically related while displaying distinctive flaking patterns. One (lower) face sets up a series of dihedral striking platforms (Bradley, 1977), while the other (upper) surface is for the removal of flakes initiated from those dihedral platforms to render that surface suitably convex. The ultimate ‘Levallois flake(s)’ run parallel to a plane of intersection that divides the two faces. The upper surface exhibits distal and lateral convexities, which facilitate control over the size and shape of the Levallois flake(s) removed just prior to reparation or discard (Boëda, 1995; Chazan, 1997; Eren and Bradley, 2006; Van Peer, 1992). Archaeological examples of Levallois may vary widely in size, but tend to conform to a restricted set of shape properties in line with this ‘volumetric’ concept of Levallois (Lycett et al., 2010).

Having learnt the rudiments of Levallois reduction, Eren produced a series of 100 Levallois cores from a pre-selected set of blanks (flakes) of high quality (i.e. with good mechanical flaking properties) and of uniformly unchallenging size and shape (Eren et al., 2011). A series of attributes designed to capture exhibited skill level and regularity of product were recorded on these reductions. Having mastered the skills after 20 months of practise, a second set of Levallois reductions was then recorded, using a chert of lower quality flaking properties and of more unwieldy, nodular morphology. Following the orthodoxy of the toolstone constraints argument, we expected that the switch from the higher to lower quality chert would stifle any measurable improvement in Eren’s skill levels and/or lead to greater irregularity of product. Accordingly, we tested these predictions via a series of statistical analyses.

2. Materials and methods

2.1. Knapping the Levallois cores

Eren was taught to prepare the sub-circular Levallois tortoise core (Fig. 1a) and to detach from its upper surface the first ‘preferential Levallois flake’ (Bradley, 1977), hereafter called the ‘Levallois flake’ (Fig. 1b). We chose this reduction because it is widespread in the Middle Palaeolithic (MP) record, and requires skill to execute. MP knappers are known to have sometimes reshaped the upper surface of the core after the first Levallois flake removal and struck off subsequent flakes (Schlanger, 1996). This

![Image](image_url)
could be repeated 2–3 more times before the much-reduced core was abandoned (Van Peer, 1992). We elected to halt the reduction after the first detachment to avoid having to grapple with the unknown effects of shrinking core size on skill level—a topic for future study.

Having learnt the rudiments of Levallois tortoise core reduction in December 2007 from his tutor (Bruce Bradley), Eren then repeated the exercise 100 times, starting with blanks of the same raw material and form (Eren et al., 2011). This was done in batches over three months in 2008, with pauses of days or weeks between work-sessions. All flakes were removed using solely hard stone hammer percussion techniques. Each set of flaking debris was bagged and labelled along with the parent tortoise core and the Levallois flake. After a further 20 months of almost daily knapping practise, a second batch of 27 reductions was undertaken in 2010 for comparison with the first set.

2.2. Raw materials

At present, the relationship between the chemical, petrological, and microstructural properties of toolstones and their suitability for knapping is only coarsely known. While it may be possible to quantify differences between various types of chert in terms of texture, lustre, translucency, and colour (e.g. Luedtke, 1992: 62–66), lithic analysts are currently able to do little more than qualitatively assign raw materials to ‘high’ versus ‘low’ knapping potential (but see Callahan, 1979 for a semi-quantitative effort). Additionally, the combination of several interacting variables (e.g. elasticity, brittleness, hardness, and homogeneity) may be more important than the isolated expression any single variable (Whittaker, 1994), and unique combinations of such variables may differ across different rock types or even within a single type.

Stone toolmakers rely (and relied) on macroscopic properties and strike-testing to identify differences between raw materials. We primarily employed this same simple strategy to distinguish ‘highly workable’ from ‘less workable’ toolstone. This strategy is justifiable for two reasons. First, visual ‘in hand’ tests of toolstone properties we have applied are precisely those that would have been accessible to hominins when making decisions regarding raw material exploitation. Second, our study is designed to determine if hypotheses of raw material constraints—based, as they frequently are, on the macroscopic determinations of raw material quality (e.g. Morgan and Renne, 2008; Noll and Pettaglia, 2003)—are valid. Our study is designed, in part, to determine if such assessments merit reconsideration, at least as incontrovertible rules that may be applied in all archaeological situations.

The external properties of the raw material (morphology and cortex) appear to present fewer difficulties. There is general agreement among lithic specialists that regular nodules present fewer difficulties for the knapper than irregular ones, and non-cortical flake blanks are considered to influence the final forms of artefacts in different ways to cortical nodules (Ashton and McNabb, 1994; Jones, 1994; White, 1998). Hence, in regard to the external (form) variation we provide quantitative descriptions of raw material variation and statistical evidence of differences between the two blank forms employed (Table 1). We also originally discriminated between ‘higher’ and ‘lower’ quality materials based on both their macroscopic internal qualities and their macroscopic external properties. Petrological observations of thin sections made from flakes produced during the experiment support the ‘in hand’ observations (described below).

The first (2008) batch of cores was made on fine-grained silicious sedimentary rock from the Cretaceous-aged Upper Greensand, located in the chalk cliffs of the South Devon coast, U.K., hereinafter referred to as Greensand silicate. Chert and other silicious rocks occur in a variety of sedimentary contexts within the Upper Greensand, resulting in a variety of textures, chert morphologies, and content of clastic and residual materials (Williams, 1986). From the knapper’s perspective, the Greensand silicate used in this study is a homogenous stone without voids or cleavage lines, and is thus of very good flaking quality. Microscopically, (Fig. 2, column 3) our samples of Greensand silicate are primarily made up of chalcedonic quartz (pseudo-fibres of stacked quartz minerals <1 micron in width) supporting 2–15% fine-to very fine sand sized (<250 microns, Wentworth Scale), dispersed clastics that include rounded quartz grains, carbonate fragments, mica, and terrigenous materials (e.g., fossil charcoal) with very rare (<1%) <100 micron-sized voids with thin clay coatings. Chaledonies (i.e., cherts made up of chalcedonic quartz) are among the most workable stone tool-making raw materials (Callahan, 1979; Luedtke, 1992:85–88), thus, the chalcedonic nature of the Greensand silicate corroborates the ‘in hand’ observations of the knapper that this raw material is of high quality. The blanks used were large, pre-selected flakes of uniform size and shape, struck from outcrop or beach boulders (Fig. 3a; Table 1) and were also, thus, of high quality. The flake’s bulb surface was shaped into the convex, upper surface of the core.

For the second (2010) batch, a raw material with more challenging external and internal properties was chosen. We selected a chert from the Cretaceous-aged Fredericksburg Group (Fredericksburg variety of Texas chert) that occurred as irregular, cortical nodules as replacement features within the Edwards limestone (Pittman, 1959) and varied tremendously in texture and internal cleavages (Fig. 3b; Table 1), ‘in hand’ assessments of the Texas chert indicated that it has good flaking quality in places but it grades into rougher textures bordering on limestone, thus making it much more difficult to fracture (Pittman, 1959:130). Observations of thin sections of both ‘higher quality’ and ‘lower quality’ Texas chert corroborate these observations (Fig. 2, columns 1 and 2). The ‘higher quality’ Texas chert is made up almost exclusively of very fine (<10 microns) microcrystalline quartz with rare domains of chalcedonic or macrocrystalline quartz. ‘Lower quality’ Texas chert has a coarser microcrystalline quartz matrix with up to 10–15% unrecalculated calcite crystals throughout the specimen. Petrographically, the Texas chert does grade from ‘higher quality’ nearly pure microcrystalline quartz chert to a ‘lower quality’ calcitic, coarser grained microcrystalline facies. Additionally, there are small but troublesome cleavage planes and concretion-like inclusions (not represented in the thin sections) that hamper flaking. One nodule had so many cleavages it fell apart and had to be abandoned.

| Table 1 |
| Size and shape of the large flake blanks knapped by M.I.E. in 2008 (Eren et al., 2011), and the corticated nodules knapped in 2010 (present study). |

<table>
<thead>
<tr>
<th></th>
<th>1st Batch – Large Flakes (n = 100)</th>
<th>2nd Batch – Chert Nodules (n = 27)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>Mean = 1718 g, Std. Dev. = 783 g</td>
<td>Mean = 2655 g, Std. Dev. = 761 g</td>
<td>t = 3.5621, p &lt; 0.0001, df = 42.0870</td>
</tr>
<tr>
<td>Maximum length</td>
<td>Mean = 198 mm, Std. Dev. = 32 mm</td>
<td>Mean = 199 mm, Std. Dev. = 32 mm</td>
<td>t = 0.0164, p = 0.8670, df = 40.4447</td>
</tr>
<tr>
<td>Width</td>
<td>Mean = 144 mm, Std. Dev. = 23 mm</td>
<td>Mean = 132 mm, Std. Dev. = 18 mm</td>
<td>t = 2.8589, p = 0.0057, df = 51.3550</td>
</tr>
<tr>
<td>Thickness</td>
<td>Mean = 61 mm, Std. Dev. = 17 mm</td>
<td>Mean = 70 mm, Std. Dev. = 10 mm</td>
<td>t = 3.5049, p &lt; 0.0001, df = 71.0456</td>
</tr>
</tbody>
</table>
Although the nodules were of similar length to the flake blanks of the first batch, they were narrower, thicker, and heavier than the first batch of blanks (Table 1).

Both raw materials are almost entirely made of microscopic quartz crystals and are, thus, cherts (Luedtke, 1992:5–10; Tucker, 2001:212). Relative to the Texas chert (knapped in 2010), the Greensand silicate (knapped in 2008) is a ‘higher quality’ chert than the Fredericksburg variety of Texas chert in terms of both its internal and external properties. In general, cherts are more workable than some raw materials used by toolmakers in prehistory (e.g., quartzite or andesite) but less workable than others (e.g., obsidian). Future experiments should continue to explore the issue.
of raw material constraints using a range of materials from ‘very low quality’ to ‘high quality’.

### 2.3. Measuring skill level

A series of attributes was measured that aimed to capture the skill level of the knapper in removing most (but not all) of the tortoise core’s upper surface, thereby creating a Levallois flake that was symmetrical and conformed to the subcircular outline of the core’s upper surface (Fig. 1). ‘Skill’ was therefore quantified according to a specific set of goals to be targeted in each Levallois reduction. Explicitly, therefore, the motivation of the knapper was directed towards achieving the goals used to measure ‘skill’, so defined. These points are methodologically important since typically the use of a single experimental participant with *a priori* knowledge of such goals might be considered a confounding factor (e.g. Orne, 1962). Here, however, we turn this to experimental advantage since skill is ultimately defined (and deliberately motivated) according to his ability to achieve set, quantifiable goals (Eren et al., 2011). In other words, this sort of experiment is only possible with the knapper’s full awareness of production goals.

We have no way of knowing if MP knappers set themselves the same goals, indeed we do not even know why they went to so much trouble to make Levallois flakes. The only certainty is that our chosen attributes require considerable technical skill to produce – it is the nature of isotropic rock that dictates this, no matter who the knapper is, or was. We also developed procedures for measuring how much stone was wasted during core preparation, particularly from the underside of the core. In addition, we determined whether the knapper’s output had become more consistent.

Skill level markers recorded as follows:

#### 2.3.1. Percentage of core surface area removed by Levallois flake

To ascertain how much of the upper core surface was removed by the first Levallois flake, we used a computer imaging method similar to that developed by Eren et al. (2008, see also www.thinkcomputer.com/research), but here we used a plug-in for Adobe Illustrator CS2 called “path area filter” (created by Telegraphics and downloaded from http://www.telegraphics.com.au/sw/#filterfoundry). Areas (mm²) of both the Levallois flake and the prepared core’s total upper surface were measured. The percentage of the total core surface area removed by the Levallois flake is a fair measure of knapping skill. Higher percentage values mark superior knapping judgement.

#### 2.3.2. Levallois flake symmetry

The symmetry (eccentricity) of each Levallois flake was also determined via use of computer imaging and Adobe Illustrator.
tools (see Eren et al., 2011). In each case, the flake outline was fitted to a standard graphic jig with 16 axes passing through the midpoint of the flake’s axial length. Midpoints of each axis are plotted and joined to form a complex polygon. The area of the polygon reflects total divergence of the flake margin from its perfectly symmetrical equivalent. Polygon area reduces as a flake approaches perfect symmetry. Symmetry was measured in this manner for each successful Levallois flake.

2.3.3. Number of overshots

Overshots occur when almost the entire upper surface of the core is detached, along with the distal sinuous edge of the core itself (Fig. 4a). Although we do not know if MP knappers regarded overshots as failures/errors, we are persuaded that they probably did because an overshot Levallois core may be permanently ruined and is more difficult to refurbish (Fig. 4b). That many Levallois cores show evidence of flakes that are not overshot (when to overshoot is easier than not) is also suggestive. Regardless of the actual intention of MP knappers, the ability to not overshoot the Levallois flake requires considerable skill and judgement on the part of the knapper, and the goal set here was to avoid such incidents. Since the avoidance of overshots requires controlling several different elements on the part of the knapper (such as delicate Levallois flake platform facetting, upper core surface flake invasiveness, and trimming of the core’s distal and lateral edges) the execution of which may all covary with toolstone related properties, the risk of overshoot flaking can reasonably be expected to increase as raw material properties become more challenging.

2.3.4. Economy of core reduction

One measure of skill is to determine how swiftly (in terms of mass removed) a knapper can ready a core for its first Levallois flake removal. To measure rock wastage during core preparation, the mass of the finished core was calculated as a percentage of the weight of the original rock blank. Percentage values rise as flaking economy improves.

2.3.5. Relative economy of underside surface preparation (Ratio of upper to lower surface flake mass)

Much care is needed to prepare the convex upper surface of the Levallois core, while the underside requires less attention. It follows that flaking economy can be achieved more effectively by minimal shaping of the underside. Accordingly, we computed the mass of flakes struck from the curved top side of the core and divided this by the mass struck from the core’s deeper underside (Fig. 1). This ratio (top:underside) increases with improving allocation of effort to the top surface.

We acknowledge that knapping skill could be measured in other ways not considered here, such as the maximising of cutting edge, or the production of usable blanks during core preparation (Eren et al., 2008; Jennings et al., 2010; Prasciunas, 2007). However, neither can be measured using the flaked output from this experiment because they conflict with, and were excluded from, the knapper’s chosen goals. For example, to maximise total cutting edge it becomes necessary to produce more smaller flakes during core preparation (Eren et al., 2008: 959), and this conflicts directly with our two stated goals of improving economy of core production. Likewise, the goal of increasing usable banks calls for an increase in preparation flakes, not fewer. In other words, distinct production and performance efficiency outputs require a knapper to approach a particular reduction sequence in different ways (Eren et al., 2011; see also Surovell, 2009).

2.4. Expectations and tests

Three sets of analyses were undertaken. The first set involved the percentage of core surface removed, Levallois flake symmetry, economy of overall core reduction, and economy of underside core preparation. The expectation tested in this analysis was that if a switch to inferior toolstone does not inhibit the knapper’s rising skill levels, then the mean values of all four attributes in the second (2010) batch of cores will be significantly higher than those for the first (2008) batch. Conversely, if a switch to the more challenging raw material does interfere in some way with rising skill level, then mean values for all (or some) attributes will be indistinguishable between batches. Indeed, mean values for some attributes in the second batch could even significantly decrease if toolstone effects are preeminent. Both batches of cores were compared for statistical differences using t-tests ($\alpha = 0.05$). Since the factors we are examining (i.e. increasing skill level versus a switch to inferior toolstone) do not inhibit the knapper’s chosen goals. For example, to maximise total cutting edge it becomes necessary to produce more smaller flakes during core preparation (Eren et al., 2008: 959), and this conflicts directly with our two stated goals of improving economy of core production. Likewise, the goal of increasing usable banks calls for an increase in preparation flakes, not fewer. In other words, distinct production and performance efficiency outputs require a knapper to approach a particular reduction sequence in different ways (Eren et al., 2011; see also Surovell, 2009).

2.4. Expectations and tests

Three sets of analyses were undertaken. The first set involved the percentage of core surface removed, Levallois flake symmetry, economy of overall core reduction, and economy of underside core preparation. The expectation tested in this analysis was that if a switch to inferior toolstone does not inhibit the knapper’s rising skill levels, then the mean values of all four attributes in the second (2010) batch of cores will be significantly higher than those for the first (2008) batch. Conversely, if a switch to the more challenging raw material does interfere in some way with rising skill level, then mean values for all (or some) attributes will be indistinguishable between batches. Indeed, mean values for some attributes in the second batch could even significantly decrease if toolstone effects are preeminent. Both batches of cores were compared for statistical differences using t-tests ($\alpha = 0.05$). Since the factors we are examining (i.e. increasing skill level versus a switch to a poorer quality raw material) could potentially affect statistical parameters in opposing directions, two-tailed tests were undertaken.

The second analysis involved a single attribute, namely number of overshots. The expectation tested here was that if the switch from Greensand silicate to Texas chert does not exert any influence on the knapper’s skill in avoiding overshots, then the frequency of overshots in the second (2010) batch should be less than that for the first (2008) batch. If, on the other hand, the switch to a more challenging toolstone does in some way promote overshooting, then overshoot frequencies should be similar in both batches. It is
even conceivable that the frequency could increase in the second (2010) batch. The proportion of overshots in each batch was assessed for statistically significant difference ($\alpha = 0.05$) using a Chi-squared test.

The third set of analyses examined variability within the measures of percentage of core surface area removed, Levallois flake symmetry, economy of core reduction, and economy core underside preparation. In the larger first (2008) batch there are clear signs of improving standardization of output, as reflected in all these attributes, particularly in those last two involving economy of core reduction. Hence, if the switch to a more challenging toolstone has no effect on this trend towards greater standardisation (i.e. less variability from one core to the next), then the trend will continue unabated in the second (2010) batch. Conversely, if the introduction of inferior raw material does cause more variable output in all (or some) of these attributes, then the trend will plateau in the second (2010) batch or may even decline. To assess this, $F$-tests were undertaken to determine whether standard deviations were significantly different ($\alpha = 0.05$) in each of these variables across core sets in the first and second batches.

3. Results

In the first set of analyses, the tests for percentage of core surface removed (by the Levallois flake), and for economy of core underside preparation both show statistically significant improvements over the course of the two core batches (Table 2). Hence, skill levels continued to rise uninterrupted, despite the insertion of inferior raw material into the experiment.

However, the test for differences in Levallois flake symmetry between batches failed to detect any significant increase in symmetry in the second batch. This is because there are two conspicuous outliers in the dataset. One resulted in a broken flake due to simple knapper error, namely a mis-hit (bounce) as the Levallois flake was struck. Raw material played no part in its failure. The other outlier was very much affected by raw material quality. This particular Texas chert nodule blank had a tough, flat cortex much of which was retained on the core's upper surface and was carried away on the Levallois flake's (badly skewed) dorsal surface. When these two outliers were removed from the dataset, a strong and statistically significant ($p = 0.0146$) difference between flake batches emerges, despite the reduction in overall sample size. The two special cases were masking an overall increase in skill level. Evidently the switch to less tractable raw material only rarely and sporadically gets the better of skill level for flake symmetry.

By contrast, the switch to Texas chert overwhelmed skill level when it came to economy of core preparation. There is no statistically significant difference between measures of the first and second batches (Table 2). Enhanced skill level in core preparation has been completely masked by the demands of the more challenging material and nodular form both of which called for more blows to reach the desired shape. Repeating each of these tests with a non-parametric Mann—Whitney $U$-test, which makes less strict assumptions of the data than the $t$-tests, results in a matching statistical pattern (see Supplementary information).

The second set of analyses concerned the frequency of overshots. About 20% of Eren’s first (2008) Levallois batch resulted in overshots. Over the course of 100 reductions, the rate of overshooting did not decline, but occurred in bursts of overshots separated by runs of properly constrained strikes (Eren et al., 2011). By contrast, his second (2010) batch contains only one overshot (3.7% of the total). The difference in the proportion of overshots in these two Levallois batches is statistically significant ($\chi^2 = 4.3049$; $df = 1$; $p = 0.038$). The introduction of the more difficult Texas chert evidently had no effect on his progressive ability to avoid overshooting the Levallois detachment.

The third set of analyses addressed consistency (i.e. variance) of output. The standard deviations of three measured variables (percentage of core surface area removed, economy of core reduction, economy of core underside preparation) were wider in the first (2008) batch but narrower in the second (2010) batch, pointing to greater consistency of output with time and increasing experience. Differences between standard deviations were statistically significant for each of the variables (Table 3).

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PLF area as % of core surface</td>
<td>100</td>
<td>34.31%</td>
<td>27</td>
<td>48.09%</td>
<td>$t = 4.0700, p &lt; 0.0001, df = 58.1279$</td>
</tr>
<tr>
<td>PLF symmetry</td>
<td>80 (minus 20 overshots)</td>
<td>119.3148</td>
<td>26 (minus 1 overshot)</td>
<td>98.0156</td>
<td>$t = -0.8169, p = 0.4179, df = 49.5987$</td>
</tr>
<tr>
<td>Economy of core preparation</td>
<td>100</td>
<td>38.03%</td>
<td>27</td>
<td>39.31%</td>
<td>$t = -2.4875, p = 0.0146, df = 91.5888$</td>
</tr>
<tr>
<td>Economy of dorsal surface preparation</td>
<td>100</td>
<td>0.6600</td>
<td>27</td>
<td>0.7985</td>
<td>$t = -2.6808, p = 0.0121, df = 28.5109$</td>
</tr>
</tbody>
</table>

**Table 3**

<table>
<thead>
<tr>
<th>Skill measure</th>
<th>First batch sample size</th>
<th>First batch std. dev.</th>
<th>Second batch sample size</th>
<th>Second batch std. dev.</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLF area as % of core surface</td>
<td>100</td>
<td>20.22%</td>
<td>27</td>
<td>14.11%</td>
<td>$F = 0.4870, p = 0.0381, df = 26$</td>
</tr>
<tr>
<td>PLF symmetry</td>
<td>80 (minus 20 overshots)</td>
<td>130.2857</td>
<td>26 (minus 1 overshot)</td>
<td>110.2723</td>
<td>$F = 0.7164, p = 0.3495, df = 25$</td>
</tr>
<tr>
<td>Economy of core preparation</td>
<td>100</td>
<td>11.41%</td>
<td>27</td>
<td>7.91%</td>
<td>$F = 0.4814, p = 0.0352, df = 26$</td>
</tr>
<tr>
<td>Economy of dorsal surface preparation</td>
<td>100</td>
<td>0.8753</td>
<td>27</td>
<td>0.2623</td>
<td>$F = 5.6861, p &lt; 0.0001, df = 26$</td>
</tr>
</tbody>
</table>
The initial test of differences examining Levallois flake symmetry failed to detect increased consistency in the second batch of flakes. However, when the same two outliers described above were removed from the dataset, a highly significant \( p < 0.0001 \) difference between flake sets emerged in terms of consistency. With these special exceptions, consistency of knapping output improved significantly from the first to second batch, despite the switch to a poorer quality of raw material.

4. Discussion and conclusions

We have endeavoured to respond to the complaint of Holdaway and Stern (2004: 55) that few if any studies attempt to adequately document the effect of different toolstones on the products of specific knapping events. To this end, we initiated a 20 month long lithic replication experiment to determine whether, with continuous practise, a knapper’s advancing skill in Levallois flake production would continue to grow if the high quality toolstone used throughout the experiment was switched in the closing months to a less tractable chert.

Skill levels were measured in terms of the knapper’s ability to achieve a series of set goals, and these were fully quantified and subjected to statistical testing. In most tests, significant increases in skill level could be detected from the earlier to the later batch of reductions. The switch to the more challenging raw material offered no obvious impediment to the knapper’s rising skill in producing larger, more symmetrical Levallois flakes that consumed more of the core’s upper surface, without overshooting the rim. There was also a statistically significant improvement in the consistency of his output, as reflected in the narrowing of standard deviations about the means of all the tests of skill level.

Taken together, our results show that factors associated with knapping skill, and not with raw material quality, were the main causes of success or failure in achieving each of the set knapping goals. It follows that our results lend little support the notion that the use of inferior toolstone will inevitably mask improvements in a knapper’s ability to achieve set goals. This is in line with the conclusions of the other recent analyses reviewed above, all of which approach the issue of raw material constraint from somewhat different angles, using prehistoric samples (e.g. Archer and Braun, 2010; Clarkson, 2010; Costa, 2010; Sharon, 2008). Thus our results further question the primacy of toolstone constraints over other factors in producing statistically significant differences between artefacts and assemblages.

The assumption that differing levels of knapping skill cannot be compared accurately across differing types of raw material should also be called into question. An extreme example of this line of thinking is that of Morgan and Renne (2008) who suggest that “since obsidian is among the raw materials most conducive to Levallois technology, the advanced appearance of the very early MSA at Gademotta and Kulukuleti [Ethiopia] may be more related to the simple circumstance of proximally available raw materials than to the developmental stage of the toolmakers.” Our results suggest these artefacts should be read as seen, namely a demonstration of skill by hominins to generate Levallois products — always a challenging exercise no matter what the raw material.

That said, the poorer raw material did impact Eren’s ability to economize during core production, as all of his effort was consumed by the extra demands imposed by the morphology of the Texas chert nodules. These required extra flaking to be brought into shape for the first Levallois detachment. The switch to inferior chert produced more waste and masked visible progress towards producing a less costly core. During the second core production batch his acquired skill became invested in coping with a, coarser, thicker, narrower, coticated blank to achieve scores just comparable to (but not lower than) those from the first batch. Hence, the quality and shape of the raw material strongly influenced the amount of waste produced during Levallois reductions. It is unlikely, therefore, that the amount of waste production can be used as a measure of skill level where more than one rock type was used. We note, however, that the switch to Texas chert had no impact on Eren’s ability to improve the consistency of his overall economy scores. This aspect of Levallois debitage analysis may hold some potential for evaluating skill level in prehistoric assemblages.

The study of complete reduction sequences is widely touted as the correct approach to core analysis where more than one raw material is evident in the assemblage (e.g. Flenniken, 1985). While agreeing in principle, we doubt if this will become standard practise given the extreme difficulty and expense of recovering refittable reductions. Indeed, our results suggest that distinctions between knappers of differing skill levels, or of differing cultural milieus, may be less visible at the level of the entire reduction sequence, because of masking by the multiple raw materials used. Our results further suggest that these distinctions are best made by comparing attributes of the residual cores. Thus, our results support Clarkson’s (2010) suggestion that core products retain relatively high levels of information about knapping skills and traditions of manufacture.

Our results should not be taken to imply that toolstone properties impose no influence upon artefact form. What they do show, however, is that when evaluating quantifiable differences between lithic assemblages, we should not prematurely assume that raw material quality will necessarily override skill levels, culturally-mediated preferences, and/or culturally-mediated skill levels. These are all involved in the attainment of functional or culturally motivated goals, which could all ultimately differ among different hominin populations.

Like all experiments, this one should be repeated, not just by other knappers but also with more sharply contrasting toolstones. Greensand silicate and Texas chert are both ‘chert’ in its broadest definition (Luedtke, 1992), and their inherent flaking properties contrast less prominently than do their external dimensions. An experiment with the same protocols performed on materials that contrast more greatly in their internal properties (e.g., basalt and obsidian) should eventually be conducted to examine the effects of other factors in producing statistically significant differences between artefacts and assemblages.

Despite the undoubted need for further work on raw material effects, it is noteworthy that the two toolstones used here differ markedly in both their internal flaking properties and their external dimensions and regularity of form — factors accused of strongly influencing knapping outcomes (e.g. Goodman, 1944; Jones, 1979; Clark, 2001: 1). However, our results fail to support the assumed primacy of raw material over other factors in influencing flaked stone morphology. Furthermore, they also indirectly support suggestions (e.g. Braun et al., 2008) that over and above a concern to obtain a ‘knappable rock’, factors associated with issues such as edge durability of the resultant flakes may have been more important to hominins when considering factors of raw material ‘quality’ and selectivity rather than fracture mechanics and/or external form per se. In sum, the results of our analyses suggest that it may not simply be a question of stating that raw material ‘will’ or ‘will not have’ an effect on artefactual and assemblage level parameters. Rather, a more complex interaction between raw material and various additional behavioural factors may be evident.
Acknowledgements

Funding for M.E. was provided by a National Science Foundation (NSF) Graduate Research Fellowship; the Department of Anthropology and QUEST Archaeological Research Program at Southern Methodist University; and Mustafa, Kathleen, and Nimet Eren. We wish to thank Craig Ratzat and Neolithics (www.neolithics.com) for advice in selecting the Fredericksburg Texas chert for this experiment, and Spectrum Petrographics, Inc., for preparing the rock thin sections. We are grateful to Rebecca Catto for reading over an early version of this paper. Thanks to Richard Klein and four anonymous reviewers for providing comments that helped clarify this manuscript.

Appendix. Supplementary data

Supplementary data related to this article can be found online at doi:10.1016/j.jas.2011.06.011.

References


World Archaeology 41, 536–551.

World Archaeology 41, 536–551.

World Archaeology 41, 536–551.

World Archaeology 41, 536–551.

World Archaeology 41, 536–551.

World Archaeology 41, 536–551.