Journal of Archaeological Science (1996) **23,** 493–507

Blind Tests of Inter-analyst Correspondence and Accuracy in the Identification of Cut Marks, Percussion Marks, and Carnivore Tooth Marks on Bone Surfaces

Robert J. Blumenschine

Department of Anthropology, Rutgers University, P.O. Box 270, New Brunswick, NJ 08903, U.S.A.

Curtis W. Marean

Department of Anthropology, SUNY at Stony Brook, Stony Brook, NY 11794-4364, U.S.A.

Salvatore D. Capaldo

Department of Anthropology, Rutgers University, P.O. Box 270, New Brunswick, NJ 08903, U.S.A.

(*Received 6 January 1995, revised manuscript accepted 9 May 1995*)

We show through blind tests that marks inflicted on bone surfaces by carnivore teeth, hammerstone percussion, and metal knife cutting and scraping can be distinguished with near perfect reliability without scanning electron microscopy or consideration of only conspicuous marks. Using low-cost and high-volume hand lens and low-power light microscope techniques, we determined the presence or absence of conspicuous and inconspicuous marks with 97% three-way correspondence, and diagnosed marks of known origin to actor and effector with 99% accuracy. Novices with less than 3 h training on control collections correctly diagnosed 86% of classic but mainly inconspicuous marks. Novices spending several more hours studying control specimens elevated their diagnostic accuracy on morphologically representative marks to near-expert levels of 95%.

Our results show that published cautions about mimicry among cut marks, percussion marks, and carnivore tooth marks are overstated. All types of marks examined can be identified reliably, regardless of conspicuousness. As such, fully standardized comparisons of mark frequencies can be drawn among assemblages, even those documented by different analysts. However, such robust interpretations can be attained only if analysts base diagnoses on (a) a firm familiarity with bones marked under strictly controlled conditions, (b) the systematic application of published morphological and contextual criteria, and (c) the use of prescribed low-power magnification techniques.

Keywords: BONE MODIFICATION, TOOTH MARKS, CUT MARKS, PERCUSSION MARKS, TAPHONOMY, ZOOARCHAEOLOGY, BLIND TESTS.

Introduction

Marks inflicted on bone surfaces by feeding carnivores and hominids can be informative about many details of the accumulation, modification, and dispersal of archaeological and paleontological assemblages of larger mammal fossils. Most notably for paleoanthropology, surface modification studies are becoming increasingly useful in reconstructing the behavioural and ecological contexts in which hominid activities occurred (e.g. Guilday, Parmalee & Tanner, 1962; Walker & Long, 1977; Potts & Shipman, 1981; Binford, 1981, 1988; Bunn, 1981, 1991; Shipman & Rose, 1983; Marshall, 1986; Bunn & Kroll, 1986, 1988; Blumenschine & Selvaggio, 1988; Olsen & Shipman, 1988; Noe-Nygaard, 1989; White, 1992; Blumenschine & Marean, 1993; Selvaggio, 1994; Cruz-Uribe & Klein, 1994; Blumenschine, Cavallo & Capaldo, 1994; Capaldo, 1995; see also Lyman, 1987, 1994 for reviews, and various contributions to Bonnichsen & Sorg, 1989).

Paleoanthropologists can only realize the great interpretive potential of data on bone surface modification if two requirements are satisfied. First, marks must be shown to be accurately identifiable to actor (e.g. carnivore or hominid) and effector (e.g. tooth, stone flake cutting edge, or hammerstone) (terminology following Gifford-Gonzalez, 1991). The identifiability requirement encompasses the need for identifications to enjoy high intra-analyst reproducibility and consistently high

0305–4403/96/040493+15 $18.00/0

inter-analyst agreement. We are concerned here not only with the elementary issue of the correct identification of individual marks, but, of greater interpretive importance, with the reliable estimation of the frequency and anatomical distribution of various types of marks in a bone assemblage. Second, behavioural and ecological interpretations of surface mark frequencies and distributions in fossil bone assemblages must be based on high resolution actualistic and experimental models designed specifically to these ends (see Blumenschine & Marean, 1993; Blumenschine, 1995; and Capaldo, 1995 for discussions).

In this paper we explore the identification requirement of robust interpretive models based on bone surface marks. Specifically, we conduct a series of blind tests that are used to evaluate the degree of correspondence in mark identifications between analysts, and the accuracy of the identifications of several types of known marks to actor and effector. This type of testing has not been reported previously for studies of bone surface modification, though Shipman (1988) reported the results of a blind test of eight marks encompassing three states (fresh marks on fresh bone, old marks on old bone, and fresh marks on old bone). As advocated recently by Ringrose (1993), blind testing is needed to validate interpretations that hinge on the analyst's ability to discriminate between marks produced by carnivores, hominids, and other relevant agents of bone modification. The initial round of blind tests we report on allows us to meet Ringrose's challenge for several types of surface marks.

The tests are warranted by a substantial body of literature focusing on potential mimicry of marks produced by different actors and effectors. Hence, Potts and Shipman (1981) and Shipman and Rose (1983) contend that very fine carnivore tooth scratches and stone tool cut marks are so similar macroscopically as to be easily mistaken for one other. Blumenschine & Selvaggio (1988) suggest that hammerstone percussion pits are superficially similar to carnivore tooth pits. Behrensmeyer, Gordon & Yanagi (1986), Olsen & Shipman (1988), and Fiorillo (1989) emphasize the potential mimicry of trampling marks and stone tool cut marks, and Andrews & Cook (1985) showed that natural dispersal processes can produce mimics of stone tool cut marks. Other researchers also note the potential confusion of preparator's marks with stone tool cut marks (Shipman, 1981; Shipman & Rose 1983; see Lyman, 1987: table 5.4).

This diversity of studies focusing on mark mimicry highlight the potential ambiguity in identifying surface modifications to actor and effector, where the ambiguity can detract severely from the credibility of archaeological interpretations. For example, Haynes makes the following remarks about surface modifications on bones from the Lubbock Lake site illustrated in the Lubbock Lake monograph (Johnson, 1987): "Illustrated 'cutmarks' on muskrat elements look like rodent gnawmarks; a fractured mammoth bone looks trampled; 'skinning' cuts on a camelid phalange could be trample-marks" (Haynes, 1991: 85). Such disagreement between analysts undermines the credibility of surface modification studies in general. It also underscores the need for studies that quantitatively evaluate the magnitude of the alleged ambiguity in surface modification identifications. The blind tests we report here allow us to make this evaluation.

Our blind test results also permit us to evaluate two key and long-standing methodological debates upon which rest the usefulness of surface modification studies to paleoanthropological interpretation. One debate concerns the type of instrumentation needed to identify surface modifications. This issue has existed ever since the first systematic reports of the morphology and anatomical distribution of stone tool cutmarks were published jointly by Bunn (1981) and Potts & Shipman (1981). Bunn (1981, 1991) and Bunn and Kroll, (1986) advocate a low power "macroscopic" method as adequate to reliably detect the distinguishing micromorphological characteristics of stone tool cut marks and carnivore tooth marks. Bunn is not explicit on the range of magnification, but most of his illustrations are at less than 15 power magnification. Shipman (1981) and Shipman and Rose (1983) have maintained that scanning electron microscopy (SEM) is required to reliably distinguish marks made by different actors, fostering many studies that base final diagnoses on SEM analysis (e.g. Potts & Shipman, 1981; Shipman & Rose, 1983; Behrensmeyer, Gordon & Yanagi, 1986; Olsen & Shipman, 1988; Fiorillo, 1989). Blumenschine (1995; Blumenschine & Marean, 1993; Blumenschine & Selvaggio, 1988, 1991) and Capaldo (1995) state that a 10–16 power hand lens used in conjunction with strong lighting is sufficient to reliably detect all diagnostic micromorphological features of a wide variety of surface modifications. They emphasize several disadvantages of a reliance on the SEM, including its often prohibitive time and financial costs, the consequent limitations these costs impose on assemblage-wide analyses of surface marks, and an over-reliance during diagnoses on the micromorphology of individual marks at the expense of the mark's context. Diagnostically valuable contextual clues include the orientation of the mark with respect to a specimen's long axis, the number of marks present on a specimen, and the mark's location on a specimen in relation to anatomical landmarks, fracture features, and other marks (e.g. Bunn, 1982, 1991; Shipman & Rose, 1983; Olsen & Shipman, 1988; Gifford-Gonzalez, 1991; White, 1992; Blumenschine & Selvaggio, 1991; Selvaggio, 1994; Capaldo, 1995). These assertions about the dispensibility of SEM diagnosis must be supported by a test of the reliability of the low-cost, high-volume hand lens and light microscopic techniques. This paper reports on such a test.

A related methodological debate influencing the veracity of surface mark studies concerns the extent to which interpretations should rely on only conspicuous marks or additionally on inconspicuous marks. The latter comprise the vast majority of marks on bone assemblages (e.g. Blumenschine & Marean, 1993). Although the term "conspicuous" is highly subjective, such marks are those easily detectable with the naked eye without the type of systematic, dedicated analysis we advocate for identifying the location and character of all marks (see below). Conspicuous marks would include furrows and punctures [following Binford's (1981) definitions] associated with gross gnawing of articular ends by carnivores, deeply incised cut marks and chop marks, and percussion battering. Examples of such marks have been well illustrated by Binford (1981), Brain (1981), Haynes (1983*a*) and Blumenschine and Selvaggio (1991). Inconspicuous marks include isolated, small tooth pits and scores, fainter cut marks, and the vast majority of percussion marks. The latter type of mark has only recently been described systematically and distinguished from other marks by Blumenschine & Selvaggio (1988, 1991); [see Turner (1983) and Freeman (1983) for early references to percussion marks, and Blumenschine (1995), Marean (1991, 1992), Oliver (1994), Turner and Turner (1992), Delpech and Villa (1993) and White (1992) for systematic identifications of percussion marks on archaeological assemblages]. The diagnostic feature of percussion marks, transversely oriented patches of microstriations, is very inconspicuous. The reliable detection of microstriations requires, at the minimum, examination with a hand lens under strong light at the correct angle of incidence. Otherwise, the general shape of percussion marks can macroscopically resemble carnivore toothmarks in more conspicuous examples.

The issue of mark conspicuousness is therefore important for lower-order interpretive goals such as identifying the range of actors and effectors responsible for the formation of bone assemblages. However, the issue of mark conspicuousness is most critical for interpretive goals that rely on accurate estimates of the anatomical distribution and frequency of bone specimens with particular marks.

In various publications, we and others have argued that identification must include a search for all marks, and that even inconspicuous marks typically preserve published, diagnostic, morphological and contextual criteria allowing accurate identification of actor and effector if three conditions are met.

(1) The analyst has experience examining control collections of specimens known to have been marked by a single, empirically observed actor and effector. The study of bones from modern or fossil hyaena dens, or any other assemblage whose entire formational history was not observed, is an unsatisfactory way to establish a reliable search image for each type of mark.
(2) The published diagnostic criteria are applied consistently.
(3) The search for marks is conducted using a hand lens or light microscope under strong light, systematically examining all parts of the surface at different angles with respect to the incoming light for conspicuous *and* inconspicuous marks.

We suspect that many studies that report on surface modifications, no matter how incidentally, do not meet all of these conditions. We are familiar with very few studies that are explicit about such analytical limitations. One is provided by Cruz-Uribe & Klein (1994), who conduct a comparative study of butchery patterns using the frequency of conspicuous tooth marks and cut marks, as defined above. They note that many more marks either visible at low magnification (8–20 ×) or obscured by post-depositional processes were present but excluded from their analysis. They justify their methods with the argument that an analysis of the inconspicuous or ambiguous marks would require hundreds or thousands of hours (Cruz-Uribe & Klein 1994: 42). While they do not contend that their results are replicable, we infer that their pursuit of a comparative analysis based on the frequency of conspicuous marks is based on a belief that standardized inter-assemblage comparisions can be accomplished without investing the large amount of time needed to identify inconspicuous marks. We suspect that such analytical standards and justifications thereof characterize most reports of surface modifications.

We also suspect that the exclusion of inconspicuous marks from assemblage-wide estimates of mark frequencies is especially prevalent in studies that conclude that carnivore involvement in bone accumulation or post-butchery modification of an archaeological assemblage was minor in relation to hominid involvement. Oliver's (1994) recent analysis of the Plio-Pleistocene larger mammal bone assemblage from FLK *Zinjanthropus*, Olduvai Gorge, provides a clear example. While Oliver includes all cut marks, percussion marks, and other hammerstone damage in tallies of hominid-induced modification, he excludes from his tooth mark tallies specimens bearing "faint or isolated lineations or pits" because of their alleged greater mimicry with marks produced by various geological processes and trampling (Oliver, 1994: 271). Such differential quantification of subtle surface marking inflicted by hominids and carnivores skews interpretations of the nature and sequence of assemblage formation by these actors. The magnitude of the resulting bias can be great, as seen in the different intensity of modification and orders of access assigned to hominids and carnivore in the formation of FLK *Zinjanthropus* by Oliver (1994) and Blumenschine (1995), the latter of whom based results on conspicuous and inconspicuous hammerstone percussion marks and carnivore tooth marks.

When interpreting tooth mark frequencies, many analysts seem to make an implicit assumption that the frequency of conspicuously tooth-marked (=heavily

Table 1. Classification of marks and criteria applied in diagnosing actor and effector in the blind tests

Actor	Effector	Mark(s) produced	Criteria
Spotted hyena and lion	Teeth	Pit, score (furrows and punctures excluded from tests)	*Morphological*: high breadth:depth ratio, with shallow U-shaped cross-section. Internal surface shows crushing. Microstriations rare, occurring in low-density patches. *Contextual*: often multiple, on cortical and medullary surfaces, and/or on thickness
Human	Metal knife	Cut mark Scrape mark	*Morphological*: low breadth:depth ratio for individual striae, with deep, V-shaped cross section. Internal surface with longitudinal microstriations; lacks crushing. *Contextual*: cut marks often in subparallel groups. Scrape marks broad shallow fields oriented parallel to long axis of bone, often with dimpling.
Human	Hammerstone and anvil	percussion pit and groove, isolated percussion microstriations	*Morphological*: high breadth:depth ratio for pits and grooves but internal surface typically lacks crushing. Very shallow microstriations in and/or emanating from pits and grooves, oriented transverse to the long axis and occurring in dense superficial patches. *Contextual*: usually within 5mm of fracture edge and restricted to cortical surface. Commonly found at or opposite point of percussion impact

Criteria are drawn from numerous sources, including Binford (1981); Blumenschine & Selvaggio (1988, 1991); Brain (1981); Bunn (1981); Haynes, 1983b; Horton & Wright (1981); Maguire, Pemberton & Collett (1980); Noe-Nygaard (1989); Potts & Shipman (1981); Shipman & Rose (1983); and White (1992).

gnawed) bone is directly proportionate to the intensity of carnivore contribution to, and modification of, bones in an assemblage (Cruz-Uribe, 1991; Oliver, 1994; Stiner, 1991). The premise is generally true if frequency estimates are based on both inconspicuous and conspicuous tooth marks, as shown experimentally (see Blumenschine, 1988, 1995; Blumenschine & Marean, 1993). However, these experiments suggest that the frequency of heavily gnawed, conspicuously tooth-marked bone may be inversely proportionate to the extent of carnivore ravaging of a bone assemblage. Intense carnivore ravaging of bone discarded by humans will probably leave low levels of conspicuous marking, because most or all of the articular ends, areas of muscle, tendon, and ligament attachment, and cylinders that would bear the conspicuous marks have been destroyed or deleted (Blumenschine, 1988; Marean & Spencer, 1991). What dominates the remains of such heavily ravaged assemblages are long bone shaft fragments with often isolated and less conspicuous marks. Frequencies of tooth-marking on heavily ravaged assemblages would often be further underestimated because most researchers typically exclude such "non-identifiable" shafts from detailed analysis. Clearly, blind tests are needed to demonstrate that inconspicuous and isolated marks can be located and identified to actor and effector as accurately as conspicuous marks associated with gross gnawing and vigorous butchery.

General Methods

We restrict this first round of testing to marks on modern long bone specimens inflicted by one of two sets of actors, humans or carnivores (Table 1). The butchery marks derive from experimental assemblages created by either CWM or RJB [variously referred to as "hammerstone only" or "hominid only" by Blumenschine (1988, 1995), Blumenschine and Marean (1993), Blumenschine and Selvaggio (1988, 1991), Marean and Spencer (1991) and Marean *et al*. (1992)] using metal knives to deflesh and disarticulate, and stone hammers and anvils to breach marrow cavities. The tooth marks in RJB's sample were produced during observed episodes of carcass consumption of hunted or scavenged wild bovid prey by spotted hyaenas (*Crocuta crocuta*), and, in a few instances, lions (*Panthera leo*) from the Serengeti and Ngorongoro ecosystems in Tanzania (referred to as "carnivore only" in above publications). Tooth marks in CWM's sample derive from observed episodes of consumption of sheep long bones by captive spotted hyaenas in a colony in Berkeley, California.

We distinguished between several types of surface marks on bones. These include cut marks inflicted with a metal knife during defleshing and disarticulation, and scrape marks inflicted with a metal knife while removing periosteum in a scraping motion parallel to the long axis of the bone. Tool marks also included percussion pits, grooves, and isolated patches of microstriations (collectively called percussion marks) inflicted through hammerstone-on-stone-anvil breakage of marrow cavities (see Blumenschine & Selvaggio, 1988 and White, 1992). CWM's sample included detailed notes on the location and identity of butchery marks taken at the time the marks were inflicted, thus allowing us to unambiguously separate cutmarks from percussion marks. All carnivore tooth marks tested for

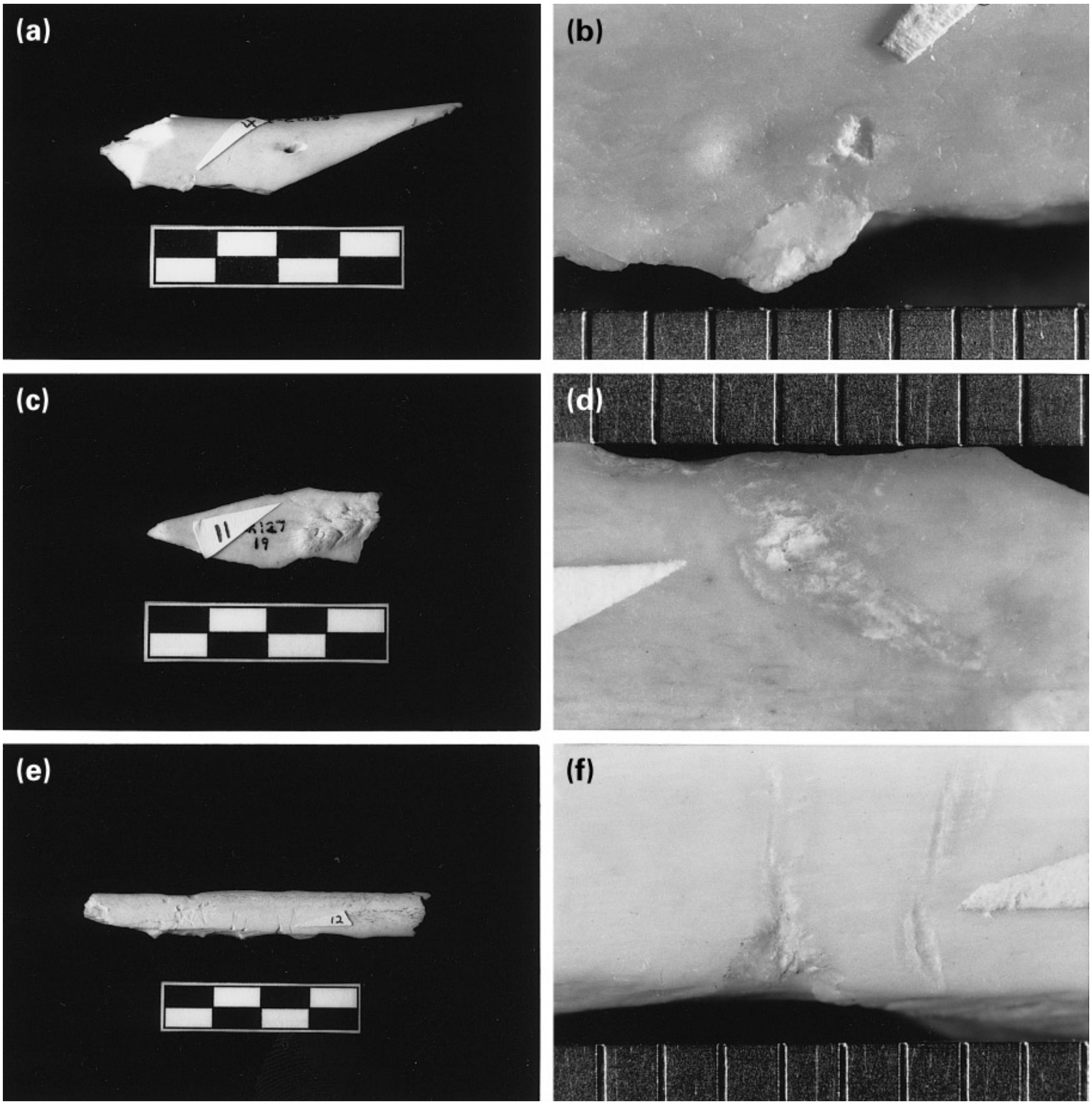

Figure 1. Representative series of three of the carnivore tooth-marked specimens used in the blind tests. Macroscopic, whole-specimen images of SER127-3 (a), SER127-19 (c), and HC17-18 (e) are shown on the left, with the tested mark indicated by a press-on arrow. Corresponding microscopic images (b), (d), (f) of the tested mark photographed under approximately 16 power are shown to the right of each macroscopic image. The tested mark on HC17-18 ((e) and (f)) can be considered to be conspicuous and classic examples of carnivore tooth-scoring. The isolated tooth pit on SER127-3 ((a) and (b)) and the isolated and shallow tooth score on SER127-19 ((c) and (d)) are less conspicuous, but still class examples of carnivore tooth-marking. For macroscopic images, each scale increment is 1 cm. For microscopic images, each scale increment is 1 mm.

include pits and scores, following Binford's (1981) classification. We did not include the far more conspicuous and easily diagnosed tooth furrows or punctures on cancellous epiphyses.

Sample sizes for individual blind tests were chosen so that the level of error could be measured to 5% or less. This was achieved by administering a minimum of 20 specimens (and in some cases 30 or 40 specimens) to each tester for any single test run.

The morphological and contextual criteria employed to distinguish these marks during the blind tests are described in Table 1. These are the criteria we have specified and employed in prior studies of mark frequencies (Blumenschine & Selvaggio, 1988, 1991; Blumenschine & Marean, 1993; Capaldo, 1995), which are based on our assessments and those of a number of researchers (see Table 1 for references). The criteria are not exhaustive (see, for example, Lyman, 1987), but they are readily applied during low magnification analysis of surface marks.

Macroscopic and low-power microscopic photographs of representative tooth marks, percussion

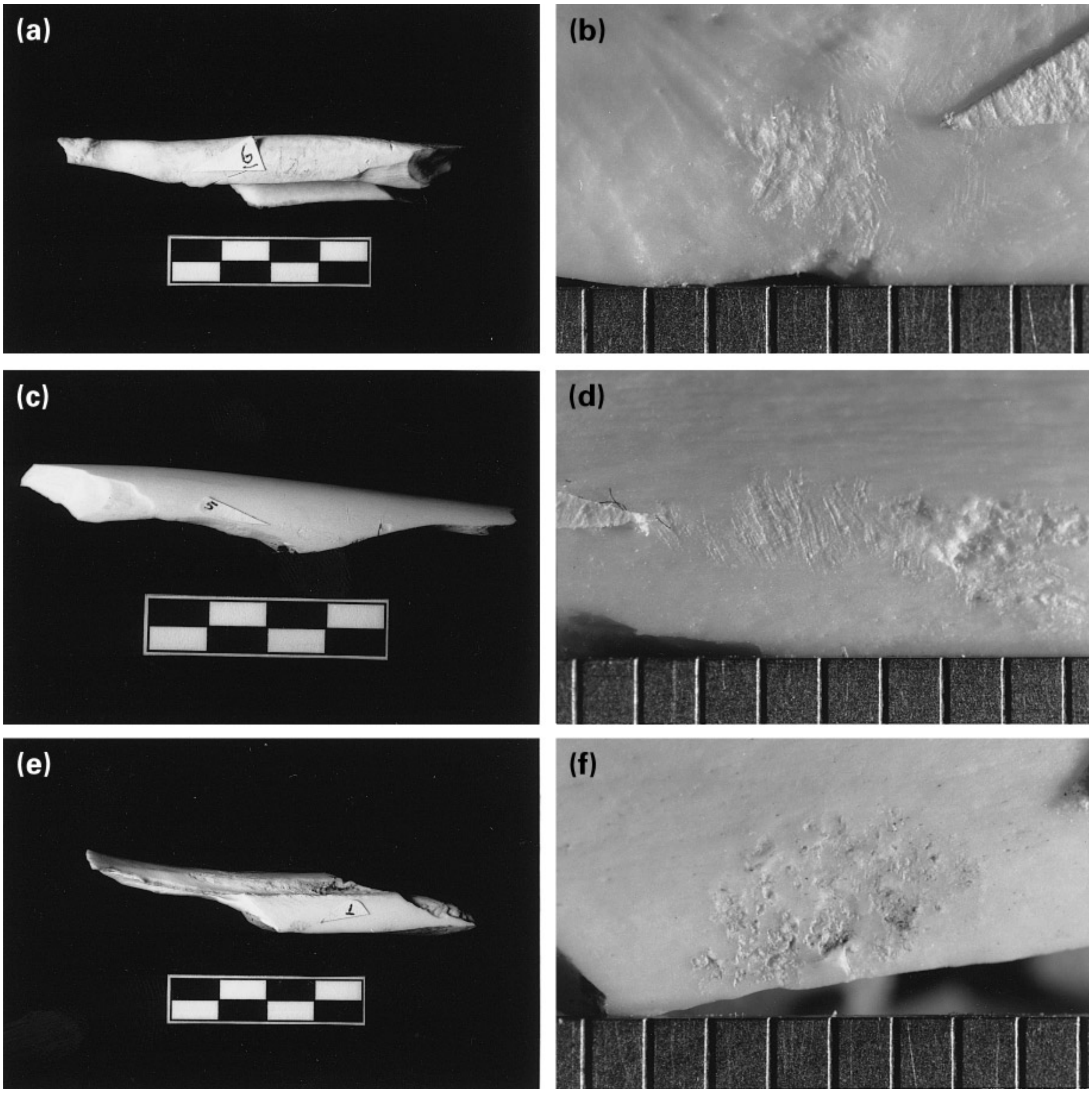

Figure 2. Representative series of three of the hammerstone-on-anvil percussion-marked specimens used in the blind tests. Macroscopic, whole-specimen images of IKO11-12 (a), 3-5-1380 (c), and 3-5-1357 (e) are shown on the left, with the tested mark indicated by a press-on arrow. Corresponding microscopic images (b), (d), (f) of the tested mark photographed under approximately 16 power are shown to the right of each macroscopic image. The tested marks on IKO11-12 ((a) and (b)) and 3-5-1380 ((c) and (d)) are classic examples of percussion-marking, particularly in being associated with microstriations clearly visible in the magnified images. Both, however, are typically inconspicuous macroscopically. The percussion mark on 3-5-1357 ((e) and (f)) is both inconspicuous and not classic in that it lacks obvious microstriations when viewed under low magnification. This is the one specimen misattributed by RJB in Run 2 of Test 2. For macroscopic images, each scale increment is 1 cm. For microscopic images, each scale increment is 1 mm.

marks, and cut marks used in this study are shown in Figures 1–3. The whole-specimen, macroscopic photographs accurately convey the range of conspicuousness of marks included. While some marks are readily visible macroscopically (e.g. Figures 1(a), (e), 2(e), 3(c), others (e.g. Figures 2(a), (c), 3(e)) can only be detected, and their diagnostic features examined, with the aid of low magnification, as seen in the associated microscopic views.

Blind Test 1: Inter-analyst Correspondence in Locating Marks

In this basic test, we evaluate the level of inter-analyst agreement on the presence or absence of extrinsic marks on individual bone specimens. We are for the moment unconcerned with the actor or effector responsible for producing the mark, beyond the fact that the mark was inflicted by an extraneous agent after the

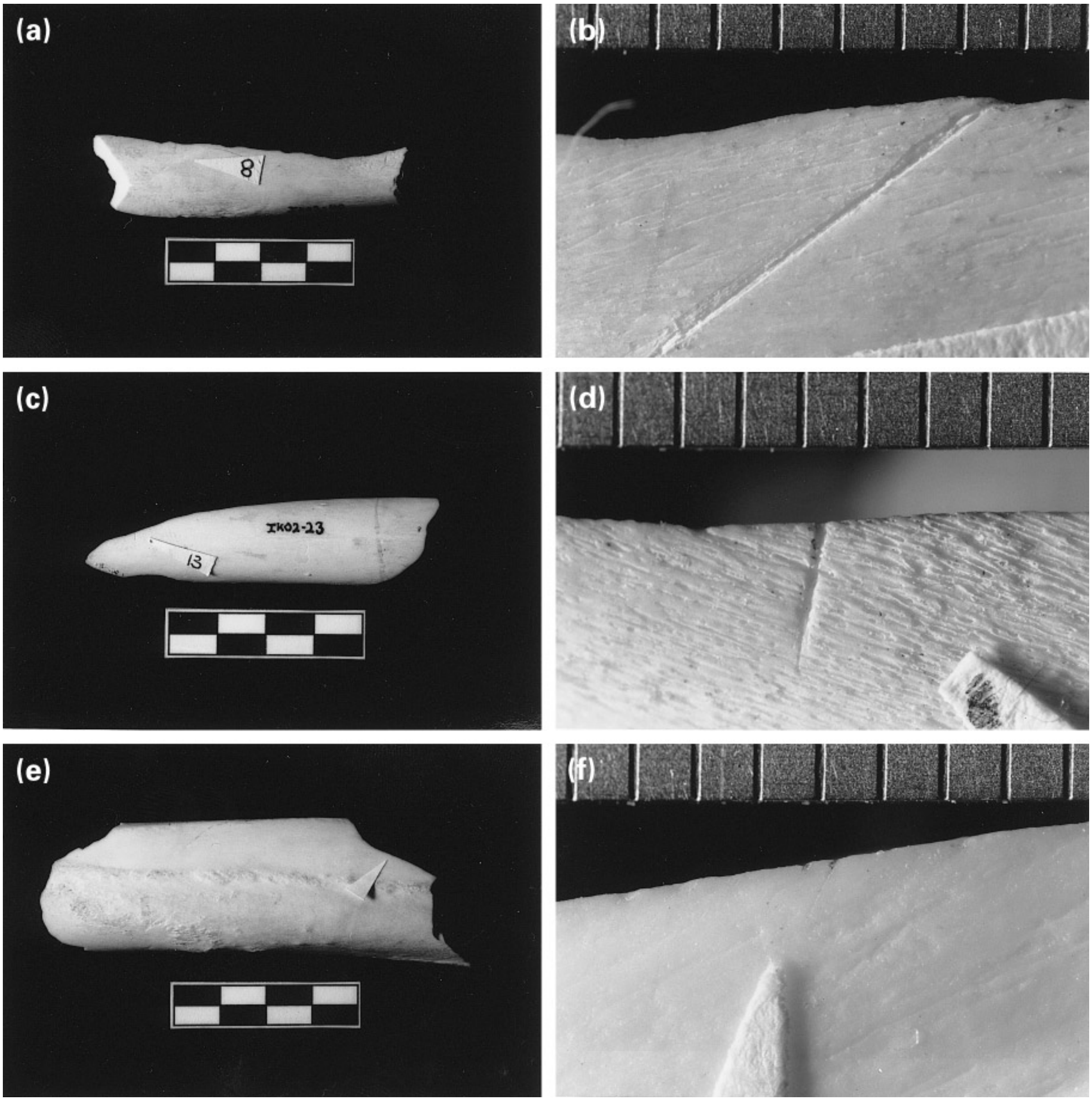

Figure 3. Representative series of three of the metal knife cut-marked specimens used in the blind tests. Macroscopic, whole-specimen images of IKO2-40 (a), IKO2-23 (c), and IKO3-24 (e) are shown on the left, with the tested mark indicated by a press-on arrow. Corresponding microscopic images (b), (d), (f) of the tested mark photographed under approximately 16 power are shown to the right of each macroscopic image. The tested mark IKO2-23 ((c) and (d)) can be considered to be a classic and fairly conspicuous examples of metal knife cut-marking. The tested mark on IKO2-40 ((a) and (b)) is less conspicuous but still classic. The extremely short and macroscopically faint mark on IKO23-24 ((e) and (f)) is identifiable as a cut mark by its V-shaped cross-section and low breadth:depth ratio. The latter mark is the one that SDC failed to locate in Test 1. For macroscopic images, each scale increment is 1 cm. For microscopic images, each scale increment is 1 mm.

death of the animal. This exercise therefore tests the reliability of two components of surface mark analysis.

(1) The degree of correspondence among three experienced analysts in distinguishing intrinsic surface features (e.g. vascular grooves) from extrinsic ones inflicted after the death of the animal.
(2) The degree of correspondence in locating extrinsic marks, regardless of their conspicuousness. While conspicuous marks, by definition, should be consistently located on an inter-analyst basis, we do not know the reliability and repeatability of identifying inconspicuous marks. Clearly, this is crucial for evaluating the reproducibility of zooarchaeological results that rely on the frequency of marked specimens.

Methods and samples

RJB and CWM selected 10 long bone shaft fragments from their respective collections. Another 10 specimens

Table 2. Inter-analyst correspondence in identifying presence or absence of marks on 30 modern long bone midshaft fragments (Test 1). Two-way correspondence refers to agreement (number and proportion of same responses) within each of three pairs of analysts, while three-way correspondence is agreement among all three analysts

	Naked-Eye		Instrument-Assisted			
	Empirical and recording		Empirical and recording		Empirical only	
	n	%	*n*	%	*n*	%
Two-way correspondence						
RJB & SDC	26	86.7	28	93.3	**29**	96.7
SDC & CWM	26	86.7	26	86.7	**29**	96.7
CWM & RJB	24	80.0	25	83.3	**30**	100.0
Three-way correspondence						
RJB, CWM & SDC	22	73.3	24	80.0	**29**	96.7

were selected from RJB's collections by SDC. Most fragments were between 2–8 cm long, and preserved from 1/5 to 1/2 of the original shaft's circumference. Selection was random with respect to the presence or absence of marks, and their conspicuousness. Since marks are common on even small shaft fragments from our collections, a short, 1–3 cm length of each specimen was designated for study by strapping a rubber band around the shaft by the individual selecting the shaft. With this random approach we were trying to ensure that designated sections of at least some specimens would bear no extraneous marks.

Each of us analysed all 30 specimens, first with the naked eye under strong light, and, in a subsequent rotation, with either a 16-power hand lens (RJB and SDC, as is consistent with their typical methodology), or with a stereo zoom microscope set at 16-power (CWM, as is consistent with his typical methodology). Naked-eye examination required approximately 5 min for each set of 10 specimens, while approximately twice the time was taken in the instrument-assisted examination.

Results of Test 1

Two-way correspondence on instrument-assisted identifications of the presence or absence of a mark ranges from 96·7% to 100% (Table 2). CWM and RJB achieved perfect correspondence on all 30 specimens, while both pairs involving SDC disagreed on only one of the 30 specimens. The single contentious specimen (Figures 3(e), (f)) bore a faint cut mark that, upon closer examination after the test results were tabulated, was acknowledged to be an identifiable mark by SDC. Three-way correspondence in locating an extraneous mark is 96·7%.

Correspondence is lower when differences in recording conventions are considered. Marks without appreciable depth were detected by SDC and RJB using a hand lens, but were not coded as identifiable marks following a long-standing convention between them. CWM, on the other hand, coded such marks as being present, but, following his conventions, as a "non-id." mark that would not contribute analytically to mark frequencies. These different recording conventions are reflected in the higher empirical and recording correspondence attained between RJB and SDC (93·3%) than either analyst enjoyed with CWM (86·7% and 83·3%; Table 2). Regardless, the difference in recording would have had no effect on analytical results, as in neither case would the marks in question have contributed to assemblage-wide frequency estimates.

Inter-analyst correspondence in determining presence/absence of marks is lowest with naked eye analysis. Two-way correspondence is as low as 80%, and three way correspondence is only 73·3% (Table 2).

Blind Test 2: Accuracy of Diagnosing the Agent of Mark Production

The primary purpose of this test is to evaluate the accuracy by which analysts can identify a mark whose agent of production (actor and effector) is known to the test administrator but not to the analysts. The marks tested for include percussion marks (pits, grooves, isolated patches of microstriations), metal knife cut marks and scrape marks, and carnivore tooth marks (pits and scores only) (see Figures 1–3 for examples). Cut marks produced by stone knives are not examined because relatively few examples of these are available in our experimental collections.

This test also examines inter-analyst correspondence in identifying and applying diagnostic criteria used to distinguish marks of various types. In making mark identifications, the analyst typically applies a combination of micromorphological criteria combined with somewhat more subjective contextual criteria (Table 1). To the extent that mark identifications in this way become something of an "art" renders the possibility that divergent and erroneous interpretations of these criteria will come to be applied either idiosyncratically by an individual analyst, or as a "tradition" between analysts who work together commonly.

Several variables affecting accuracy of mark identifications were explored preliminarily in this test. These include the following.

(1) The effect of the use of micromorphological versus contextual criteria on identification accuracy. This was accomplished by using both classic examples of each type of mark, as well as marks considered by the administrator to be confidently attributable to the (known) agent, but which did not necessarily display all of the micromorphological criteria of the classic marks (compare specimens in Figures 1–3). In the latter case, greater reliance must be placed on contextual criteria.
(2) The effect of the experience of the analyst (novice versus expert).
(3) The effect of different instrumentation (naked eye versus 16 power hand lens versus 16 power stereo microscope).
(4) The effect of the analyst's familiarity with the type of bone surface on which marks occur (smooth surfaces of adult bovid long bone diaphyses versus the flakey surface of juveniles' diaphyses).

Methods and samples

Two separate runs of this test were conducted. In both runs, a single mark, flagged by a stick-on arrow, was selected by the test administrator on each of 20 long bone fragments from assemblages known by the administrator to have been defleshed and fragmented by either carnivores (spotted hyaenas and/or lions) or by metal knife butchery followed by hammerstone percussion. All test specimens were administered to analysts who had no prior experience with the specimens. Analysts were asked to make a distinction between carnivore tooth marks (regardless of whether it was a pit or score), cut marks, scrape marks, and percussion marks (regardless of whether it was a pit, groove, or isolated patch of microstriations). In both runs of Test 2, a response was scored as correct only if it was correctly attributed to one of the four kinds of marks.

In Run 1, RJB selected 20 specimens bearing a "classic" example of one of the three types of marks from his "carnivore only" and "hammerstone only" collections. The sample contained eight carnivore tooth marks, six percussion marks, and four metal knife cut marks and two metal knife scrape marks. Each mark was classic in that it bore all published, diagnostic micromorphological features. Conspicuousness was not a criterion in making the selection, such that diagnostic features of some marks (e.g. microstriations of percussion marks) could only be detected reliably using the techniques and aids we recommend (light and low magnification).

The 20 specimens for Run 1 were submitted to CWM and SDC, and, on a different day, to 15 undergraduate and two graduate students enrolled in a practical zooarchaeology course taught by RJB. CWM and SDC identified the marks independently of one another, using techniques and aids well familiar to each. CWM required about 19 min to complete the test with the aid of a 10–40 × stereo microscope with low incidence light set at 16 power. SDC required about half that time with a 16 power hand lens under low incident light.

The 17 students given the same test specimens were novices in the identification of marks on bone surfaces, and in zooarchaeology in general. Each had completed 10 weeks of course work emphasizing element and species identification of mammalian skeletal material. One week prior to the test, each student had also attended a 1·5 h lecture on surface marks illustrated with SEM micrographs of individual marks and macro photographs of marked bones. The lecture was followed by a demonstration of the hand-lens-under-light technique of identifying marks, and a 45 min period during which students could examine, under supervision, a teaching collection of control specimens bearing flagged marks of known and stated origin.

The test was administered 1 week later by circulating the 20 specimens among the students. Students were given 3 min to independently identify each mark with the aid of a 10 power hand lens. Students were permitted to consult a guide distributed and explained the previous week that contained a systematic and detailed description of the shared and distinguishing features of marks, as in Table 1. Prior to the test, students were asked to indicate the amount of time, to the nearest hour, they had spent examining the teaching collection of control specimens outside of class hours during the preceding week. These estimates are not likely to be biased because RJB told the class that the test and work-hour estimates would not influence their grade in the course.

A second version of Test 2 was run using a selection of marks of known origin that were morphologically more representative (i.e. more difficult to identify) than the series of classic marks used in Run 1. The two test administrators (RJB & CWM) each selected 20 test marks from a sample that each would have coded as a confident attribution to a particular actor and effector. RJB selected from his assemblages, which are composed of long bone fragments from adult, wild bovids that possess hard and smooth bone surfaces. CWM selected from his collections, composed of long bones from subadult domestic bovids with "flakey" bone surfaces. As in Run 1, this run included inconspicuous marks. However, unlike Run 1, this sample contained marks that lacked some of the micromorphological and contextual criteria defined above. The marks are therefore representative of the typical variation found in an assemblage, where the potential for ambiguity in diagnoses of actor and effector is greatly increased over that presented by classic marks.

Run 2 of Test 2 was completed first by the authors. The two samples were subdivided into sets of 10 and examined independently by each of us. RJB examined all of CWM's specimens, and vice versa, while SDC examined 10 from each in two sessions that each required a maximum of 13 min to complete. RJB and SDC again utilized a 16 power hand lens under light, while CWM used the stereo microscope set at 16 magnification.

Run 2 of Test 2 was also completed by four of the undergraduates in RJB's zooarchaeology class who had developed sufficient interest in bone modification to continue with the study after completing Run 1. By the time Run 2 was administered, each student had completed a total of 5–6 h of study of the original teaching collection of control marks, and of a larger sample of specimens from RJB's single-agent "carnivore only" and "hammerstone only" assemblages. All of these assemblages are composed of long bone fragments from adult, wild bovids, whose bone surfaces are hard and smooth. Prior to Run 2, the students were not exposed to any specimens from CWM's collections. In contrast to RJB's collections, these are composed of long bone fragments from subadult domestic animals displaying "flakey" surfaces. During the pre-test period, the students studied these materials in pairs, so as to be able to discuss ambiguities, and to mutually reinforce identification and systematic application of diagnostic criteria. Students did not switch between these "learning pairs," creating an opportunity to test whether two divergent interpretive traditions might arise in the 3–4 h during which members of each pair exchanged opinions.

Each student examined all 40 specimens in Run 2 independently of one another, in sets of 10 over four rotations. Thirty specimens were examined using a 10 power hand lens under strong light, while the remaining 10 specimens were analysed without the benefit of the hand lens. Within each learning pair, a hand lens was used to identify marks on the same specimens.

Table 3. Results on accuracy of diagnosing 20 known marks, using classic examples from Blumenschine's experiments (Test 2, Run 1)

	No. of identifications	% Correct identifications
Novice students		
All combined (n=17 students)	340	85.9
Course Performance		
Excellent (n=6)	120	90.8
Good (n=5)	100	85.0
Fair (n=6)	120	81.7
Time spent studying controls		
<2 h (n=14)	280	85.0
2–3 h (n=3)	60	90.0
Students later participating in Test 2, Run 2 (n=4)	80	92.5
Experienced analysts		
CWM & SDC (n=2)	40	100.0

"Time spent studying controls" includes an initial 45 min of supervised examination of known marks flagged by stick-on arrows on control specimens plus the amount of time spent outside of class hours, as reported by each student. Course performance is an ordinal ranking of students' performance on a series of skeletal element and species identification quizzes administered earlier in the course. "Students later participating in Test 2, Run 2" include four undergraduates who had developed a sufficient interest in mark identifications as to volunteer for continued study and blind testing.

Results of Test 2, Run 1

Success rates in identifying classic examples of carnivore tooth marks and three types of tool marks were high (Table 3); 100% accuracy was achieved by the two experienced analysts (CWM, SDC). Using different visual aids, and possessing largely independent histories of practice in mark identifications, the perfect scores also show that the two researchers achieved 100% correspondence between themselves.

The 17 novice zooarchaeology students collectively scored an accuracy rate of 86% in identifying the same classic marks as the two experts (Table 3). Accuracy in identifications shows minor covariation with course performance, a general measure of student motivation, and perhaps also of aptitude for zooarchaeology. Hence, the six students, including two graduate students, whose performance on prior taxonomic and element identification tests had been scored as excellent, achieved 91% accuracy in identifying the marks (Table 3). Those students who had performed less admirably on the general zooarchaeology examinations scored 85% ("good") and 82% (fair) (Table 3). Time spent independently with the teaching collections of known marks on control specimens also shows the expected covariation with accuracy of mark identifications: students who had studied the controls for less than 2 h scored lower (85%) than those claiming to have devoted more time (2–3 h) to acquiring a distinct "search image" of each type of mark (90% accuracy). Finally, the highest accuracy rate (93%) was achieved by that subgroup of novice students who would later volunteer to pursue further study and testing on surface modifications (Table 3).

Results of Test 2, Run 2

Similarly high accuracy rates were achieved on identifying marks more representative of the range of micromorphological and contextual variation than seen on "classic" marks. Among the authors, a 98·3% accuracy was attained, with only one misattribution to actor or effector among the 60 identifications made collectively (Table 4). That error was committed by RJB on a percussion mark that lacked clearly defined microstriations when viewed under a hand lens, and which, very uncharacteristically, had closely associated marking on the thickness of the bone, as is often the case with carnivore tooth-marking (Figures 2(e) and (f)). The four undergraduates who had by the time of

Table 4. Results on accuracy of diagnosing 20 known marks, chosen as identifiable but not necessarily classic (Test 2, Run 2)

	I.D.s of RJB's specimens		I.D.s of CWM's specimens		Total	
	n	% Correct	*n*	% Correct	*n*	% Correct
Novice undergraduates with 5–6 h experience (*n*=4)						
Naked eye	20	75.0	20	80.0	40	77.5
With hand lens	60	95.0	60	83.3	120	89.2
Experienced analysts (*n*=3)						
RJB (with hand lens)	0	—	20	95.0	20	95.0
CWM (with microscope)	20	100.0	0	—	20	100.0
SDC (with hand lens)	10	100.0	10	100.0	20	100.0
Total	30	100.0	30	96.7	60	98.3

Run 2 acquired 5–6 h of experience also scored well. They achieved an overall accuracy of 89% for those identifications made with the hand lens (Table 4).

While the accuracy of identifications made by expert analysts is uniformly high, two factors contributed to substantial variability in the success rate of less experienced analysts. Overall accuracy of naked-eye identifications is 77·5%, almost 12 percentage points lower than that achieved using the 10 power hand lens. The instrumentation discrepancy, however, is evident mainly on RJB's specimens (20 percentage points, versus 3% for CWM's specimens). Using a hand lens, the students attained accuracy rates for marks on RJB's specimens that closely approach those achieved by the authors, collectively making only three misattributions to actor or effector among the total of 60 identifications (95% accuracy). Conversely, hand-lens accuracy rates on CWM's specimens, which had surface textures unfamiliar to the students, were almost 12 percentage points lower (83·3%), and do not display the marked improvement over naked eye identifications seen on RJB's specimens (Table 4).

High inter-analyst correspondence on instrument-assisted identifications is not dependent on shared learning. Among the authors, two-way and three-way correspondence was 100% (SDC did not analyse the subset of 10 specimens provided by CWM that contained the specimen on which RJB committed the single error). Among the students, two-way correspondence ranges from 75% to 95% (Table 5). However, overall correspondence within learning pairs (85% agreement on 60 identifications in common within each pair) is identical to that between learning pairs (85% on 80 such identifications; Table 5).

Discussion

The results of the two blind tests demonstrate a high level of inter-analyst correspondence and great accuracy in the identification of known tooth marks and tool marks. Near perfect results were attained by the authors, who achieved a 96·7% three-way correspondence in locating marks (disagreeing on 1 in 30; Table 2), and a 99% accuracy rate for identifying known marks (one error among a total of 100 attributions from the two runs of Test 2; Tables 3 & 4). Even novices with less than 3 h training on control collections correctly identified 86% of classic, but not necessarily conspicuous, examples of tooth marks, cut marks, scrape marks, and percussion marks (Table 3). An arguably more motivated or talented subgroup of these novices attained an accuracy of 90% or greater (Table 3). When these same novice analysts had acquired several more hours of experience with controls, their accuracy in identifying *representative* samples of the four types of marks occurring on specimens with familiar surface textures reached near-expert levels of 95% (Table 4).

Table 5. Two-way inter-analyst correspondence among four undergraduates of Table 4 in identifying the representative marks used for Test 2, Run 2

	Total no. of specimens	Correspondence	
		n	%
Two-way correspondence			
Within learning pairs			
SD & GA	30	25	83.3
SR & RC	30	26	86.7
Total	60	51	85.0
Between learning pairs			
SD & SR	20	15	75.0
SD & RC	20	17	85.0
GA & SR	20	17	85.0
GA & RC	20	19	95.0
Total	80	68	85.0

Data are reported only for those specimens in Run 2 analysed using a hand lens by all four students. Correspondence refers to agreement (number and proportion of same responses) within each pair of analysts. Undergraduates GA and SD studied teaching collections of control specimens as a pair, while SR and RC formed a second learning pair.

Several factors account for variability in inter-analyst correspondence and accuracy. Experience is one factor with two components revealed by our tests. Above, we summarized the expected effects of general experience in studying marks, specifically time spent with control collections of marks inflicted by known actors and effectors. A second component of experience concerns familiarity with the range of bone surface textures on which marks occur. The lower accuracy rate achieved by students on CWM's specimens compared to RJB's (Table 4) is directly attributable to their complete lack of experience with flakey surface textures of bones from subadult, domestic animals. A similar phenomenon has been encountered by RJB and CWM when beginning analysis of a fossil assemblage, despite the fact that both of us have conducted detailed surface modification studies on thousands of specimens from a number of fossil assemblages. These assemblages bear a variety of extraneous marks in addition to tooth and tool marks. Additionally, each new assemblage contains bones with a novel combination of colours, patinas, and surface textures. These factors combine to require us to study variation in surface mark morphology on a sample of several hundred specimens before feeling that we can reliably and consistently apply criteria for diagnosing actor and effector.

Instrumentation has an effect on surface mark identification. Novice analysts with 5–6 h of training scored markedly higher accuracy rates when using a hand lens than the naked eye alone on specimens with familiar surface textures (Table 4). There was a less dramatic improvement with the flakey-surface bones (CWM's sample). Likewise, correspondence among the authors in locating marks is well below 90% using naked-eye search, while high, two-way correspondence is achieved regardless of whether search is assisted with a hand lens or low power light microscopy (Table 2). Here, the artificial magnification is less essential for locating even minute marks, than for determining if the mark has appreciable depth.

A combination of greater depth of field, higher magnification, and a lessening of eye fatigue gives the stereo light microscope a slight advantage in mark identifications over the hand lens. The microscope's greater depth of field makes assessment of whether a mark has appreciable depth easier, though with negligible effect on correspondence. Likewise, accuracy in identifying known marks with the hand lens was essentially as high as that utilizing the stereo microscope (Tables 3 & 4). Here, however, the single misidentification made among the 100 marks analysed by the authors was a percussion mark whose microstriations became clearer to RJB under the microscope set at 40 power than they were under a hand lens during the test. The slight advantages of the microscope are offset somewhat by the greater amount of time required to analyse a specimen in this way, due to the greater difficulty of shifting specimen orientation to the light source while maintaining focus.

It is important to note that both the hand lens and the light microscope techniques result in comparable improvements over naked-eye identifications. Our near-perfect correspondence and accuracy in mark identifications renders the expense, time, and volume restriction of the SEM unnecessary.

The potential for developing traditions of erroneous diagnoses does not seem to be large based on our tests of correspondence or accuracy. Correspondence in locating marks (Table 2) and identifying marks (Tables 3 & 4) was no higher between RJB and SDC, who have worked together in this area for the past 7 years, than either enjoyed with CWM. CWM worked closely on mark identifications with RJB on several occasions 4 years prior to the testing reported here, and has since been conducting surface modification studies continuously but without RJB. Despite the potential for CWM and RJB to diverge in diagnoses of surface marks, this did not occur. Distinct "traditions" among the authors were seen only in recording conventions (Table 2), which, regardless, would not have affected analytical results. Still, this difference points to the importance of establishing discipline-wide standards for not only identifying marks, but also for recording their characteristics [see Lyman (1994) for this demand on all zooarchaeological measures]. A similar lack of divergent traditions in identification was seen among the novice analysts, where the correspondence in mark identifications between learning pairs of undergraduates was the same as that achieved between members of a given pair (Table 5). This suggests that even for marks that do not display all of the classic morphological features, systematic application of published criteria reiterated here in Table 1 is sufficient to achieve high correspondence and accuracy. The result also shows that mark identifications can in this way be a highly objective and reproducible exercise, rather than a subjective "art."

Conclusions

The results of the blind tests show that even inconspicuous tool marks and carnivore tooth marks, and those that lack some diagnostic traits or classic expressions thereof, can be identified with near perfect accuracy to actor and effector. The allied results showing near-perfect correspondence between experienced analysts suggests that estimates of the frequency of specimens bearing particular types of marks are reproducible, and, therefore, that comparisons of mark frequencies between assemblages analysed by different researchers can be valid. This is a key result, because it is estimates of the frequency and anatomical distribution of marks, not simply their presence or absence on particular skeletal parts or taxa, that are most informative about hominid behavioural ecology and site formation.

The results of the blind tests show that such rigour in surface mark studies can be achieved without incurring the high expense and severe time and volume restrictions of SEM, or by limiting analysis to conspicuous marks, the definition of which can never be standardized from analyst to analyst. Nonetheless, we suggest that high accuracy and inter-analyst correspondence for studies reporting data on the anatomical distribution and incidence of tooth and butchery marks using low cost, high volume search and identification techniques can be achieved only if all of three conditions are met.

First, zooarchaeologists wishing to report reliable tooth and tool mark frequencies and distributions must have experience with control collections, each marked by only one of a series of known actors and effectors. This permits development of an unambiguous search image for each type of mark, including an appreciation both for variability in the expression of their diagnostic micromophological traits, and for their anatomical contexts and common associations. Control collections can only be considered adequate if the actor and effector were actually observed to be the agent inflicting the surface modification. This excludes many types of collections that might otherwise be considered control collections. For example, modern hyaena dens cannot be considered control collections unless the actual event linking the tooth marks to the feeding hyaena was observed directly. Typically, however, this linkage for all bones from a den is inferred on the basis of observations of hyaenas introducing and modifying only a small portion of the bone assemblage, or perhaps even by the frequency of alleged tooth marks only. Such evidence linking the trace to the effector and actor is circumstantial and thus cannot be the basis of a strong methodology. Bones in a hyaena den may include cut marks and percussion marks from bones scavenged from human garbage pits. Alternatively, hyaena dens in caves may have small components of human accumulation. For these same reasons fossil hyaena dens cannot be considered as control collections. Likewise, collections from human campsites are suspect unless strict empirical controls can exclude the influence of domestic dogs or free-ranging carnivores.

Results of blind tests administered to novice zooarchaeologists show that a high, 95% accuracy in identifying known marks can be achieved with only a modest investment (6 h) of study time with control collections. Indeed, the time-consuming aspect of this requirement is gaining access to, or, preferably, generating control collections by conducting butchery and marrow extraction experiments and controlled observations of carnivore feeding. Our success in identifying marks, combined with their great interpretive potential, easily justifies this time investment. We encourage other analysts to do the same. If this is not immediately possible, we invite analysts to begin by contacting any one of us so that a study visit or a short-term loan of a small control collection might be arranged. We also suggest that surface modification studies in general would be well served if all practitioners augmented reports of fossil mark identifications and frequencies with results of blind tests similar to those reported here. This would have the desirable effect for archaeological science of increasing the standardization of measurements, thereby maintaining comparability among measurements produced by different analysts.

A second prerequisite for achieving high accuracy and correspondence in distinguishing carnivore tooth marks and tool marks is to consistently apply the published criteria reiterated here (Table 1). These criteria are easily learned using control collections. They are also sufficiently unambiguous, despite being defined in part by qualitative terms such as "high" and "low" or "shallow" and "deep," as to produce extremely high levels of accuracy in identification. Although additional criteria have been proposed, those used here are adequate to distinguish the marks tested with a high level of accuracy. Also, archaeologists studying prehistoric bone collections are primarily concerned with cut marks produced by stone implements rather than the metal knives investigated here. Nonetheless, our experience and the more systematic observations made by Capaldo (1995) suggest that cut marks produced by both materials are sufficiently similar as to sustain the distinctions with other tool marks and carnivore tooth marks used here. Systematic application of these criteria minimizes the subjectivity of mark identifications, thus preventing the development of divergent and erroneous traditions of identification among groups or schools of analysts.

Third, systematic search for marks under strong incident light using a 10–16 power hand lens or a low power stereo microscope is required. Naked-eye searches and diagnoses are unreliable, particularly for the abundant inconspicuous marks and for diagnostic features such as percussion microstriations. Search must cover every section of the bone, including cortical and medullary surfaces, and the bone's thickness. The orientation of the bone relative to the observer and the incident light must be systematically altered for each field of view under the magnifier. We recommend that analysts conduct such an analysis of marks as a separate exercise from the recording of standard zooarchaeological attributes.

Such a careful, dedicated search for surface marks is tedious. In our experience on fossil assemblages, five to 15 specimens per hour can be analysed thoroughly, depending on bone size. In our blind tests, the hand lens seemed to afford more rapid analysis with only a marginal loss of accuracy over that achieved with the stereo microscope. Choice between these instruments therefore seems to be a matter of preference and access. For very large bone assemblages, time limitations might require that only a sample of available specimens be analysed in this way. If properly drawn, a sample permits accurate assessments of assemblage-wide frequencies and anatomical distributions of marks that is not possible using SEM or naked eye searches.

If these analytical requirements are met, the results of our initial round of blind testing show that published cautions about potential mimicry between marks produced by carnivores and hominids are overstated. As such, the erosion of confidence in interpretations based on surface mark attributions that these cautions might instill is largely unfounded.

To consolidate these findings, an expanded program of blind testing should be conducted. For example, cut marks and scraping marks produced by stone tools should be investigated to test our contention that, like metal knife marks, these can be readily distinguished from other tool marks and carnivore tooth marks using the criteria and methods of examination prescribed here. The blind tests of accuracy in identification of actor and effector can also be expanded to include experimental assemblages whose observed formational histories include an event in addition to, but exclusive of, butchery or carnivore gnawing, such as trampling. Further tests of inter-analyst correspondence in identifications and frequency estimates should be conducted on experimental assemblages that have been modified by multiple, known agencies, such as our controlled samples of experimentally butchered bone that was also ravaged by carnivores. The ultimate blind tests of inter-analyst correspondence in surface mark identifications should be conducted on bones from fossil assemblages, where the possible presence of the full range of known surface marks maximizes potential ambiguity in identifications.

The success of our first round of blind testing justifies the expanded program of blind testing outlined above. It should also encourage both skeptics and devotees of surface modification studies to apply to the zooarchaeological assemblage currently under study the growing body of well controlled actualistic models of hominid behaviour and site formation that are based on surface mark frequencies. We acknowledge that there will always be error in identifications, and less than perfect correspondence among analysts studying assemblages from which mark frequencies might be fruitfully compared. However, the small error levels (*c.* 5% or less) attained thus far in the blind tests reported here are acceptable, at least in as far as the resolution of current taphonomic models based on mark frequencies are concerned (see, for example, Blumenschine, 1995). In general, our blind test results indicate that detailed surface modification studies should and can become a regular, rigourous, and powerful tool of zooarchaeological analysis and paleoanthropological interpretation.

Acknowledgements

RJB thanks all of the Rutgers students in his Spring, 1994 zooarchaeology class who participated in the blind test, particularly Greg Archetti, Rebecca Casapulla, Sean Daley, and Scott Rufolo. CWM thanks Laurence Frank, Stephen Glickman, and the staff at the Berkeley Spotted Hyaena Colony for their help. The captive colony is supported by the National Institute of Health grant No. 5R01 MH 399917 to Stephen Glickman. All of us appreciate the thoughtful comments of two anonymous reviewers.

References

Andrews, P. & Cook, J. (1985). Natural modifications to bones in a temperate setting. *Man* **20,** 675–691.

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.

Binford, L. R. (1978). *Nunamiut Ethnoarchaeology*. New York, NY: Academic Press.

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

Binford, L. R. (1988). Fact and fiction about the *Zinjanthropus* floor. *Current Anthropology* **29,** 123–149.

Blumenschine, R. J. (1988). An experimental model of the timing of hominid and carnivore influence on archaeological bone assemblages. *Journal of Archaeological Science* **15,** 483–502.

Blumenschine, R. J. (1995). Percussion marks, tooth marks, and experimental determinations of the timing of hominid and carnivore access to long bones at FLK *Zinjanthropus*, Olduvai Gorge, Tanzania. *Journal of Human Evolution* **29,** 21–51.

Blumenschine, R. J. & Selvaggio, M. M. (1988). Percussion marks on bone surfaces as a new diagnostic of hominid behaviour. *Nature* **333,** 763–765.

Blumenschine, R. J. & Selvaggio, M. M. (1991). On the marks of marrow bone processing by hammerstones and hyaenas: their anatomical patterning and archaeological implications. In (J.D. Clark, Ed) *Cultural Beginnings*. Bonn: Dr. R. Habelt GMBH, pp. 17–32.

Blumenschine, R. J. & Marean, C. W. (1993). A carnivore's view of archaeological bone assemblages. In (J. Hudson, Ed) *From Bones to Behavior*. Carbondale, IL: Center for Archaeological Investigations, Southern Illinois University, pp. 273–300.

Blumenschine, R. J., Cavallo, J. A. & Capaldo, S. D. (1994). Competition for carcasses and early hominid behavioral ecology: a case study and conceptual framework. *Journal of Human Evolution* **27,** 197–213.

Bonnichsen, R. & Sorg, M. H., Eds (1989). *Bone Modification*. Orono, ME: Center for the Study of the First Americans.

Brain, C. K. (1981). *The Hunters or the Hunted? An Introduction to African Cave Taphonomy*. Chicago, IL: University of Chicago Press.

Bunn, H. T. (1981). Archaeological evidence for meat-eating by Plio-Pleistocene hominids from Koobi Fora and Olduvai Gorge. *Nature* **291,** 574–577.

Bunn, H. T. (1982). *Meat-eating and human evolution: studies on the diet and subsistence patterns of Plio-Pleistocene hominids in East Africa*. Ph.D. Dissertation, University of California, Berkeley, CA, U.S.A.

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

Bunn, H. T. & Ezzo, J. A. (1993). Hunting and scavenging by Plio-Pleistocene hominids: nutritional constraints, archaeological patterns, and behavioural implications. *Journal of Archaeological Science* **20,** 365–389.

Bunn, H. T. & Kroll, E. M. (1986). Systematic butchery by Plio-Pleistocene hominids at Olduvai Gorge, Tanzania. *Current Anthropology* **27,** 431–452.

Bunn, H. T. & Kroll, E. M. (1988). Reply to Binford. *Current Anthropology* **29,** 123–149.

Capaldo, S. D. (1995). *Inferring hominid and carnivore behavior from dual-patterned archaeofaunal assemblages*. Ph.D. Dissertation. Rutgers University, New Brunswick, NJ, U.S.A. (UMI Ann Arbor, MI, U.S.A.)

Cruz-Uribe, K. (1991). Distinguishing hyaena from hominid bone accumulations. *Journal of Field Archaeology* **18,** 467–486.

Cruz-Uribe, K. & Klein, R. G. (1994). Chew marks and cut marks on animal bones from the Kasteelberg B and Dune Field Midden Later Stone Age sites, Western Cape Province, South Africa. *Journal of Archaeological Science* **21,** 35–49.

Delpech, F. & Villa, P. (1993). Activites de chasse et de boucherie dans la grotte des Eglises. In (J. Desse & F. Audoin-Rouzeau, Eds) *Exploitation des Animaux Sauvages a Travers le Temps Recontres Internationales d'Archaeologie et d'Histoire d'Antibes*. Juan-les-Pins: Editions APDCA, pp. 79–102.

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

Freeman, L.G. (1983). More on the Mousterian: flaked bone from Cueva Morin. *Current Anthropology* **24,** 366–373.

Gifford-Gonzalez, D. (1991). Bones are not enough: analogues, knowledge, and interpretive strategies in zooarchaeology. *Journal of Anthropological Archaeology* **10,** 215–254.

Guilday, J. E., Parmalee, P. W. & Tanner, D. P. (1962). Aboriginal butchering techniques at the Eschelman site (36LA12), Lancaster County, Pennsylvania. *Pennsylvania Archaeologist* **32,** 59–83.

Haynes, G. (1983*a*). Frequencies of spiral and green-bone fractures on ungulate limb bones in modern surface assemblages. *American Antiquity* **48,** 102–114.

Haynes, G. (1983*b*). A guide for differentiating mammalian carnivore taxa responsible for gnaw damage to herbivore limb bones. *Paleobiology* **9,** 164–172.

Haynes, G. (1991). *Mammoths, Mastodons, and Elephants: Biology, Behavior, and the Fossil Record*. Cambridge: Cambridge University Press.

Horton, D. R. & Wright, R. V. S. (1981). Cuts on Lancefield bones: carnivorous *Thylacoleo*, not humans, the cause. *Archaeology in Oceania* **16,** 73–80.

Johnson, E. (1987). Cultural activities and interactions. In (E. Johnson, Ed) *Lubbock Lake: Late Quaternary Studies on the Southern High Plains*. College Station, TX: Texas A & M University Press, pp. 120–158.

Lyman, R. L. (1987). Archaeofaunas and butchery studies: a taphonomic perspective. In (M.B. Schiffer, Ed) *Advances in Archaeological Methods and Theory*, Vol. 10. New York, NY: Academic Press, pp. 249–337.

Lyman, R. L. (1994). *Vertebrate Taphonomy*. Cambridge: Cambridge University Press.

Maguire, J. M., Pemberton, D. & Collett, M. H. (1980). The Makapansgat Limeworks Grey Breccia: hominids, hyaenas, hystricids or hillwash? *Palaeontologia Africana* **23,** 75–98.

Marean, C. W. (1991). Measuring the post-depositional destruction of bone in archaeological assemblages. *Journal of Archaeological Science* **18,** 677–694.

Marean, C. W. (1992). Hunter to herder: large mammal remains from the hunter–gatherer occupation at Enkapuno Ya Muto rock-shelter, Central Rift, Kenya. *African Archaeological Review* **10,** 65–127.

Marean, C. W. & Spencer, L. M. (1991). Impact of carnivore ravaging on zooarchaeological measures of element abundance. *American Antiquity* **56,** 645–658.

Marean, C. W., Spencer, L. M., Blumenschine, R. J. & Capaldo, S. D. (1992). Captive hyaena bone choice and destruction, the schlepp effect, and Olduvai archaeofaunas. *Journal of Archaeological Science* **19,** 101–121.

Marshall, F. (1986). Implications of bone modification in a Neolithic faunal assemblage for the study of early hominid butchery and subsistence practices. *Journal of Human Evolution* **15,** 661–672.

Noe-Nygaard, N. (1989). Man-made trace fossils on bone. *Human Evolution* **4,** 461–491.

Oliver, J. S. (1994). Estimates of hominid and carnivore involvement in the FLK *Zinjanthropus* fossil assemblage: some socioecological implications. *Journal of Human Evolution* **27,** 267–294.

Olsen, S. L. & Shipman, P. (1988). Surface modification on bone: trampling versus butchery. *Journal of Archaeological Science* **15,** 535–553.

Potts, R. & Shipman, P. (1981). Cutmarks made by stone tools on bones from Olduvai Gorge, Tanzania. *Nature* **291,** 577–580.

Ringrose, T. J. (1993). Bone counts and statistics: a critique. *Journal of Archaeological Science* **20,** 121–157.

Selvaggio, M. M. (1994). Carnivore tooth marks and stone tool butchery marks on scavenged bones: archaeological implications. *Journal of Human Evolution* **27,** 215–228.

Shipman, P. (1981). Applications of scanning electron microscopy to taphonomic problems. In (A.M. Cantwell, J.B. Griffin & N.A. Rothschild, Eds) *The Research Potential of Anthropological Museum Collections*. Annals of the New York Academy of Science 376, pp. 357–385.

Shipman, P. (1988). Actualistic studies of animal resources and hominid activities. In (S. L. Olsen, Ed) *Scanning Electron Microscopy and Archaeology.* Oxford: BAR International Series **452,** pp. 261–285.

Shipman, P. & Rose, J. (1983). Early hominid hunting, butchering, and carcass-processing behaviors; approaches to the fossil record. *Journal of Anthropological Archaeology* **2,** 57–98.

Stiner, M. C. (1991). Food procurement and transport by human and non-human predators. *Journal of Archaeological Science* **18,** 455–482.

Turner, C. G., II. (1983). Taphonomic reconstruction of human violence and cannibalism based on mass burials in the American Southwest. In (G.M. LeMoine & A.S. MacEachern, Eds) *Carnivores, Human Scavengers & Predators: a Question of Bone Technology*. Calgary: University of Calgary Archaeological Association, pp. 219–240.

Turner, C. G., II. & Turner, J. A. (1992). The first claim for cannibalism in the Southwest: Walter Hough's 1901 discovery at Canyon Butte Ruin 3, northeastern Arizona. *American Antiquity* **57,** 661–682.

Walker, P. L. & Long, J. C. (1977). An experimental study of the morphological characteristics of tool marks. *American Antiquity* **42,** 605–616.

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