Blind Testing of Faunal Identification Protocols: A Case Study with North American Artiodactyl Stylohyoids

Patrick M. Lubinski, R. Lee Lyman, and Matthew P. Johnson

Taxonomic identification of archaeofauna relies on techniques and anatomical traits that should be valid, reliable, and usable, but which are rarely tested. Identification protocols (techniques and anatomical traits), particularly those used to distinguish taxa of similar size and morphology, should be rigorously tested to ensure a solid interpretive foundation. Blind testing of a protocol for identifying stylohyoid bones of North American artiodactyls was performed by three analysts who independently employed the protocol to identify 77 anatomically complete specimens of known taxonomic identity, representing 54 individuals and 11 species. Identifications were identical in 89% of cases and in conflict in 3% of cases. The remainder involved differences in resolution; two analysts identified specimens to species, whereas the third identified specimens to more general taxonomic groups. Inter-analyst variability in identification was a result of differences in protocol application. Identifications were consistent with known taxon in 92%–96% of cases. Results indicate that the protocol is valid, reliable, and usable, and it can be applied to archaeological specimens with confidence. Testing of other identification criteria employed by zooarchaeologists is encouraged.

Keywords: artiodactyls, method testing, stylohyoid, taxonomic identification, zooarchaeology

La identificación taxonómica de la arqueofauna se basa en técnicas y rasgos anatómicos que deberían ser válidos, confiables y utilizables, pero que rara vez se prueban en la práctica. Sostenemos que los protocolos de identificación (técnicas y rasgos anatómicos), particularmente aquellos utilizados para distinguir taxones de tamaño y morfología similares, deben estar sujetos a pruebas rigurosas para garantizar una base interpretativa sólida. Tres analistas realizaron pruebas a ciegas de un protocolo propuesto recientemente para identificar huesos estilohioides de artiodactilos de América del Norte que emplearon de forma independiente el protocolo para identificar 77 especímenes anatómicamente completos de identidad taxonómica conocida y que representan 54 individuos y 11 especies. Las identificaciones fueron idénticas en el 89% de los casos y en conflicto en el 3% de los casos. El resto involucraba diferencias en la resolución; dos de nosotros identificamos especímenes de especies, mientras que el tercero identificó especímenes de grupos taxonómicos más generales. La variabilidad entre analistas en la identificación fue el resultado de diferencias en la aplicación del protocolo. Las identificaciones fueron consistentes con el taxón conocido en el 92%–96% de los casos. Los resultados indican que el protocolo es válido, confiable y utilizable, y puede aplicarse a especímenes arqueológicos con confianza. Se alienta la prueba de otros criterios de identificación empleados por los zooarqueólogos.

Palabras clave: artiodactilos, método de prueba, stylohioide, identificación taxonómica, zooarqueología

Taxonomic identification of faunal remains recovered from archaeological excavations is a requisite first step to attaining many of zooarchaeology's analytical goals (e.g., Gifford-Gonzalez 2018; LeFebvre and Sharpe 2018). The identification process requires first ascertaining which anatomical traits of skeletal parts are the result of variation between

Patrick M. Lubinski (Pat.Lubinski@cwu.edu, corresponding author) ■ Department of Anthropology, Central Washington University, 400 East University Way, Ellensburg, WA 98926-7544, USA
R. Lee Lyman ■ Department of Anthropology, University of Missouri, 112 Swallow Hall, Columbia, MO 65211, USA
Matthew P. Johnson ■ Cultural & Environmental Resource Management Program, Central Washington University, 400 East University Way, Ellensburg, WA 98926-7544, USA

American Antiquity 85(4), 2020, pp. 781–794

doi:10.1017/aaq.2020.45

species (inter-taxonomic variation) and which reflect variation within a species (intra-taxonomic variation). The former traits allow zooarchaeologists to assert that a particular distal left tibia represents a yellow-bellied marmot (*Marmota flaviventris*) as opposed to some other species of similar body (and skeletal) size. Zooarchaeologists have spent countless hours scrutinizing skeletons of known taxonomic identity to determine which skeletal traits are taxonomically diagnostic and which ones are not. In North American zooarchaeology, this process began in the late nineteenth century (Eaton 1898) and has extended continuously up to the present (e.g., Adams and Crabtree 2012; Balkwill and Cumbaa 1992; Brainerd 1939; Brown and Gustafson 1979; Gilbert 1973; Jacobson 2003; Lawrence 1951; Lubinski and Hale 2018; Olsen 1964).

Twenty years ago, Gobalet (2001) asked four experts to identify the same zooarchaeological sample of fish bones. He found considerable variability across investigators in both the number of specimens identified and the taxonomic identifications made. These results were not surprising to us given the history of (not) reporting taxonomic identification protocols in zooarchaeology. Gobalet's (2001) results suggest that different analysts employ different skeletal traits or traits of varied diagnostic value to make identifications. Although rarely cited, we believe, as Driver does, that Gobalet's study "should have been a wake-up call to the discipline" (2011:37).

Beyond the well-acknowledged problem that limited access to sufficient comparative materials can significantly reduce identification accuracy (see references in Lyman 2019), Driver described the key issue as an assumption that the methods and anatomical traits used "for identification are sufficiently well tested that one does not need to justify most identifications, except in relatively rare circumstances" (Driver 1992:39). Importantly, he noted that "we do not systematically test the quality of our identifications using 'blind' tests. . . . Consequently we have no idea of the accuracy of our methods" (Driver 1992:40–41). He added, "While most zooarchaeologists would probably agree that most identifications are probably accurate, they have no empirical or theoretical basis for this claim" (Driver 1992:41).

This argument could as easily be made today as in 1992 (e.g., Lau and Kansa 2018; Wolverton 2013). Given that anatomical traits used for identifications are rarely published, there can be little peer review or testing of them, although some may occur at conferences and workshops and on listservs. In our combined 60+ years of experience, however, no one has questioned our taxonomic identification techniques or requested that we describe the anatomical traits used to make identifications. We worry that this is typical of the discipline. Since accurate identifications provide the basis for much of zooarchaeology, and insofar as paleozoological identifications of species are having increasing influence on conservation biology decisions (e.g., Barnosky et al. 2017; Dietl and Flessa 2017; Lyman and Cannon 2004; Wolverton et al. 2016), an inaccurate identification could have implications beyond archaeological inquiry. Three examples illustrating uncertainty about identification protocols and the changing status of some skeletal traits from taxonomically diagnostic to nondiagnostic are provided in Supplemental Text 1. It is in large part because of such cases that we, Driver (1992), and Gobalet (2001) believe that skeletal traits thought to be diagnostic among taxa of similar size and morphology should be tested.

Identification Protocols: Best Practices and Examining Success

We begin by characterizing the identification process and defining concepts we use to evaluate an identification protocol. The terms "reference" and "comparative" specimens designate bones of known taxonomy wherein the taxonomy was determined by standard biological criteria among living organisms. Zooarchaeologists have long used such collections to develop means to identify ancient faunal remains. There are today no recommended best practices or generally accepted standards for reporting identifications—that is, which skeletal parts were identified and how those specimens were identified (Clason 1972; Driver 1982; Lawrence 1973; Wolverton 2013). Furthermore, there tends to be minimal description of the identification protocol used in a particular study (e.g., Baker and Shaffer

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

1999; Driver 1992, 2011; Emery 2004; Lupo 2011; Lyman 2005; Stewart 2005). Authors of recent zooarchaeology textbooks and articles state identification "methods" should be described and published, particularly listing the reference collections and published skeletal guides consulted (e.g., Beisaw 2013; Driver 1992; Gobalet 2001; Hesse and Wapnish 1985; LeFebvre and Sharpe 2018; Nims and Butler 2017; O'Connor 2000; Rea 1986; Reitz and Wing 2008), but the suggestion is not consistently made, even in textbooks (e.g., Davis 1987; Gifford-Gonzalez 2018; Hillson 1992; Rackham 1994). Some authors have urged analysts to publish the skeletal traits used to make identifications (e.g., Butler 2011; Butler and Lyman 1996; Driver 1982; Gobalet 2001; Graham and Semken 1987; Klein and Cruz-Uribe 1984; Lawrence 1973; Lyman 2005, 2011a; Wolverton 2013; Wolverton and Nagaoka 2017). This suggestion, however, is only followed when an identification is particularly interpretively significant or difficult (Bochenski 2008), such as the distinction of remains of domestic sheep (*Ovis aries*) from those of domestic goat (*Capra hircus*) (e.g., Balaase and Ambrose 2005; Boessneck 1969; Halstead et al. 2002; Haruda 2017; Payne 1985; Salvagno and Albarella 2017; Wolfhagen and Price 2017; Zeder and Pilaar 2010).

However, reporting that the University of California Berkeley Museum of Vertebrate Zoology skeletal collection was consulted and that Gilbert's (1973) guide was examined reveals little about the actual identification *protocol* followed—the techniques and anatomical traits used. Anatomical traits can be morphological (shape), metric (size), or meristic (frequency, as in the dental formula). Techniques can involve visual comparison of a zooarchaeological specimen with some (typically unreported) number of reference skeletons of similar size and morphology (which species were examined and why) or other approaches, such as measuring skeletal dimensions for statistical comparison (which dimensions were measured and why, and how many reference specimens of each of which taxa were measured). Best practice should be "transparency in identification methodology" (Lau and Kansa 2018:34), including everything just mentioned.

We use general science concepts to evaluate the identification process. "Validity" concerns the question, Are we measuring what we think we are measuring, or does the measurement technique provide an empirical indication of the concept of interest (Carmines and Zeller 1979; Lastrucci 1967; VanPool and Leonard 2011)? "Reliability" (replicability, repeatability) concerns the question, If we measure something twice, do we get the same answer, or, on repeated trials, do we obtain the same results (Carmines and Zeller 1979; Lastrucci 1967; Zar 1996)? Two aspects of reliability and validity (Golafshani 2003; Heale and Twycross 2015; Whittemore et al. 2001) are important with respect to testing taxonomic identification protocols. The first is the "stability" of results (Heale and Twycross 2015) across repeated testing, known as "test–retest" (Roberts et al. 2006). A test–retest method has been suggested by zooarchaeologists (Nims and Butler 2017; Wolverton and Nagaoka 2017), but reidentifying zooarchaeological specimens identified a few weeks ago does not test validity or accuracy, because the true identities are unknown. The second aspect of reliability and validity is "equivalence," or consistency of results across multiple respondents (Heale and Twycross 2015). We think of the equivalence of results across multiple investigators as "usability." By this, we mean, Is an identification protocol readily applied by different investigators, especially researchers who were not involved in development of the protocol? And does use of a protocol by multiple investigators influence the validity and/or reliability of the identifications?

"Precision" is defined as the closeness of repeated measurements of the same specimen (Sokal and Rohlf 1969; VanPool and Leonard 2011; Zar 1996). In this sense, a precise identification would be one made by several investigators of the same specimens as, say, *Lepus americanus* (snowshoe hare). That does not necessarily mean the identification is correct or accurate; "accuracy" means how close a measurement is to the true value (Sokal and Rohlf 1969; VanPool and Leonard 2011; Zar 1996).

Finally, "resolution" concerns the grain of a measurement—think dots per square inch in a photograph, where a greater number of dots

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

means finer grain and greater resolution (a "sharper" or "higher-definition" image). A measurement of 0.01 mm is of greater resolution than a measurement of 0.1 mm, and a species-level identification is of greater resolution than a genus-level identification (Findley 1964; Gifford-Gonzalez 2018). In both cases, a fine-grained or high-resolution measurement (identification) contains all of the information of a coarse-grained or low-resolution measurement, but the latter only includes some of the information of the former.

Mindful of these concepts, we suggest criteria for the most rewarding testing of identification protocols. Testing would most simply employ protocols with explicitly described techniques and anatomical traits. If it has not been used much, it will not have entered the standard operating procedures of zooarchaeologists who might find it of value, and it will not have provided invalid identifications at a worrisomely high rate. Furthermore, it will be easy to correct any ambiguities or inaccuracies in the protocol discovered during the test. The researcher who developed the protocol should be involved in the testing, along with researchers of varied levels of experience with identifying ancient faunal remains. The developer will have intimate knowledge of the protocol and be able to correct problems detected by those researchers without such knowledge. Including researchers with varied levels of experience ensures that both novices and experienced researchers can use the protocol. The protocol should be tested with reference specimens of known taxonomy so that its validity can be evaluated. The test reference specimens should be of individual organisms not used to develop the protocol to maintain independence of the test from development of the protocol. An additional benefit would attend such independent testing because a greater range of intraspecific variability would likely be involved, particularly if the test specimens originate in populations different from the specimens used to develop the protocol (Polly and Head 2004).

We do not expect all such tests to be published, but we think it is appropriate to describe one such test here in the hopes that the testing method will be mimicked for many allegedly taxonomically diagnostic traits. We also hope the results of such tests will be incorporated into a freely accessible online database. This suggestion follows one made by Lawrence (1973) to create and maintain a hard copy of such a thing—a suggestion not followed due to, as Lawrence recognized, its impracticality. The suggestion of an online database was made more recently by Barr (2008), and it should now be both possible and practical (see also Lyman 2019).

Materials and Methods

We implemented a test of the protocol proposed by Lubinski and Hale (2018) for identifying stylohyoids of North American artiodactyls, hereafter referred to as the LH protocol. This protocol provides an ideal test case. First, its recent development means that it has not yet been widely applied, and consequently, that potentially incorrect identifications have not been produced. Second, its recency means weaknesses have not yet been detected. Third, the LH protocol involves an explicit dichotomous key made up of detailed descriptions of anatomical traits—both taxonomically diagnostic and not diagnostic—and a particular order for considering those traits, which lends itself to rigorous testing. Fourth, testers included one of the protocol originators (PML) and users who had nothing to do with developing the protocol and who have not previously used it, both an experienced zooarchaeologist (RLL) and a novice (MPJ). Our test exemplifies an ideal model for testing any identification protocol.

The artiodactyl stylohyoid (1) is large enough for consistent archaeological recovery (measuring 4–19 cm in length for adult specimens); (2) is commonly found at North American archaeological sites (up to half as common as the most abundant skeletal element at some sites); (3) often exhibits evidence of butchery and is sometimes manufactured into pendants; and (4) is not well known by many zooarchaeologists (no North American artiodactyl skeletal key includes this element), so documenting it will contribute to zooarchaeological knowledge (Lubinski and Hale 2018). The stylohyoid is the largest bony element in the hyoid complex of artiodactyls in

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Figure 1. Example stylohyoid shown in (a) dorsal view, and (b) lateral view. This is a left bighorn sheep (*Ovis canadensis*) specimen (University of Washington Burke Museum 81695). It measures 69 mm in total length (anterior end to dorsal process). (Color online)

the Antilocapridae, Bovidae, and Cervidae. It is a paired bone in the throat below the mandibles. The element (Figure 1) is anchored at either end with cartilaginous rods, from the posterior, dorsal process to the temporal bone of the cranium (Saber and Hoffman 1985), and from the anterior end to remaining hyoid-complex elements, which are small and indistinct in these taxa.

Lubinski and Hale (2018) recorded metric and morphological traits of nearly 500 reference hyoids of 13 species (10 genera, three families) of North American artiodactyls. They determined which traits seemed to distinguish taxa and estimated the rate of accurate identifications using the traits. Some traits allowed only distinguishing large from small artiodactyls; other traits permitted Bovidae to be distinguished from Cervidae; still other traits were diagnostic of genera; and some traits allowed identification to a particular species. Some congeneric species (e.g., *Ovis aries*, *O. canadensis*, *O. dalli*) could not be distinguished, but it seemed that most anatomically complete and skeletally mature stylohyoids could be identified at least to taxonomic family. No independent test of the validity and reliability of the stylohyoid traits was performed. Furthermore, no researcher who had not been involved in developing the protocol employed it to identify taxonomically known specimens, so there was no test of its usability.

To test the validity, reliability, and usability of the LH protocol, we obtained stylohyoids of known taxonomy from four zooarchaeology reference collections: California State University, Chico (CSUC); Eastern Tennessee State University (ETSU); Iowa State University (ISU); and Indiana University (IU). None of these test specimens had been used by Lubinski and Hale (2018) to develop this protocol. Lubinski received collection loans and removed specimens with significant breakage or adhering connective tissue that would complicate measurement, yielding a test sample of 77 stylohyoids representing 11 species of North American artiodactyls (Table 1). This sample derives from 54 different animals and includes 23 bilateral pairs, allowing evaluation of left-right identification consistency. It also includes at least one animal from each of the eight taxonomic groups that can be identified with the protocol: *Antilocapra americana*, small bovid, small cervid (*Odocoileus* sp.), large bovid, *Alces alces*, *Bos taurus*, *Cervus elaphus*, and *Rangifer tarandus*.

The three test analysts represent a range of experience in identification of mammalian skeletal elements and with stylohoid identification. Lubinski has approximately 25 years of experience and no previous training or collaboration with Lyman, and he developed the stylohyoid identification protocol. Lyman has approximately 40 years of experience and had never before attempted to identify artiodactyl stylohyoids to species, in part due to the few specimens of this element in comparative collections he has used, and because no taxonomically

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Table 1. Specimens in Study Sample.

Size Group	Family	Species	Stylohyoids	Animals
Small	Antilocapridae	*Antilocapra americana*	5	4
	Bovidae	*Capra hircus*, *Ovis aries*, *Ovis canadensis*[a]	18	13
	Cervidae	*Odocoileus hemionus*, *Odocoileus virginianus*[b]	27	18
Large	Bovidae	*Bison bison*	12	9
		Bos taurus	5	4
	Cervidae	*Alces alces*	2	1
		Cervus elaphus	5	3
		Rangifer tarandus	3	2
TOTAL:			77	54

[a]The Lubinski and Hale (2018) identification protocol does not allow for distinction of these three small bovid species, so they are combined here.
[b]The Lubinski and Hale (2018) identification protocol does not allow for distinction of these two small cervid species, so they are combined here.

diagnostic criteria were known to him. Johnson has approximately three years of experience but none in stylohyoid identification. Neither Lyman nor Johnson helped develop the LH protocol, which is why they could evaluate its usability. None of us was provided with the taxonomic identities of the test specimens until after we each independently identified them using the LH protocol.

Lubinski recorded the two morphological attributes and 10 osteometrics called for in Lubinski and Hale (2018:Table 5), recording metrics to 0.01 mm using digital calipers. He then followed the eight steps of the key in their Table 5 to make identifications, which involve two morphological traits and five of the osteometrics. After identification was complete, Lubinski sent the 77 test specimens first to Lyman and then to Johnson, with no guidance other than illustrations from Hale (2016:Figure 11) regarding how to complete the measurements.

Upon receipt of the 77 test specimens from Lubinski, Lyman practiced the identification protocol using 12 reference stylohyoids representing four species and curated in the University of Missouri Zooarchaeology Comparative Collection (five *Antilocapra americana*, two *Cervus elaphus*, four *Odocoileus* sp., and one *Ovis aries*). In every case, an accurate identification was determined. Feeling confident in his having learned the protocol correctly (and therefore believing the usability of the protocol to be high), he applied the protocol to the 77 test specimens. All metrics Lyman recorded were measured to the nearest 0.02 mm using dial calipers.

Johnson first practiced the protocol using seven specimens representing six species in the Central Washington University Zooarchaeology collection (one *Antilocapra americana*, two *Bos taurus*, one *Capra hircus*, one *Cervus elaphus*, one *Odocoileus* sp., one *Ovis aries*). Having success with these, he applied the protocol to the 77 test specimens. He recorded measurements to the nearest 0.01 mm using digital calipers.

Although all analysts used the LH protocol, we did not apply it identically. Lubinski used all criteria (except dorsal curvature, discussed below) for the eight steps of the key in Lubinski and Hale's Table 5 (2018) to make identifications. This means he faced the possibility of conflicting results in each step. When there was a conflict, he made a conservative identification as suggested by Lubinski and Hale (2018:373). For example, if in Step 7 Lubinski found 7a to key to *Alces/Cervus*, whereas 7b keyed to *Rangifer*, he would exit the key and list the result as unknown large cervid. Johnson also used all criteria except dorsal curvature and employed Lubinski and Hale's Table 5 (2018). When a conflict was noted, he was not certain how to proceed, so the criterion from Table 4 was employed, and the other criterion was ignored. Lyman used only one criterion for each step of the key, following Lubinski and Hale's Table 4 (2018), and had no conflicting results in a step to consider.

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Results

We each predicted taxon for 72–74 of the 77 test specimens (Table 2). All of us omitted two specimens with unfused angles, two of us omitted two specimens as unfused or slightly broken, and two of us omitted one pathological specimen. This provided a final sample of 72–74 individual bones and 21–22 paired sets. Identification accuracy varied due to differences described above in how we applied the protocol. Given that some of Lubinski's identifications were to lower-resolution (higher) taxonomic groups than distinguishable by the LH protocol, we consider both instances in which there is a perfect match in the three investigators' identifications, and consistent matches in which the three investigators' identifications are not identical but are taxonomically consistent (e.g., two of us identified a specimen as *Cervus elaphus,* and the third identified it as large cervid).

If the LH protocol is reliable, taxonomic identifications made of left and right sides of paired elements will match, and taxonomic identifications made by each analyst of individual specimens will match. If the LH protocol is valid, taxonomic identifications should, regardless of investigator, match known taxonomy (if not resolution). If the LH protocol is usable, Lyman's and Johnson's identifications should closely match Lubinski's identifications as well as the known taxonomy.

Results indicate that the LH protocol is reliable; identifications were consistent across analysts for nearly all test specimens (Table 3). There were no mismatches in identifications between left and right sides across the three analysts. All identifications are consistent between Lubinski and Lyman, and they differ from Johnson for just two specimens (3%). Results also indicate that the LH protocol is valid but not 100% of the time. All of us misidentified a few specimens (see below), and Johnson misidentified two specimens correctly identified by the other two analysts (Table 3). Finally, Johnson and Lyman's results suggest that the LH protocol is quite usable; both successfully applied it to small pilot sets of specimens to learn the protocol, and then accurately identified 92%–94% of the test specimens.

Our identifications matched known species in the majority of cases. Counting differences in resolution that are consistent (e.g., large cervid vs. *Cervus elaphus*) as correct, there were 68–70 correct identifications—a 92%–96% success rate. One case was a partial mismatch. Lubinski identified specimen # 70 as large cervid and Lyman and Johnson identified it as *Cervus elaphus*; IU records list it as *Alces alces.*

In three cases, all analysts' identifications failed to match the "known" species. We identified specimen # 6 as *Antilocapra americana*, whereas CSUC records list it as *Odocoileus hemionus*. We identified # 15 as *Cervus elaphus*, but ETSU records list it as *Bos taurus*, and we identified # 71 as *Cervus elaphus*, whereas IU records list it as *Alces alces*. These may be failures of the LH protocol's validity, or they may be errors in the records associated with the comparative collections. Such problems can occur as a result of labeling errors or incomplete removal of bones from dermestid beetle colonies used to skeletonize carcasses (e.g., Grayson and Maser 1978), so the stylohyoids are collected with bones of a different animal. Lubinski has noted indisputable taxonomic recording errors in several collections, such as four stylohyoids labeled 21715 from the University of Wisconsin Zoological Museum and four stylohyoids labeled 28721 from the Slater Museum of Natural History. That the potential exists for misidentified reference specimens has been noted before (e.g., LeFebvre and Sharpe 2018; Lyman 2019), but it has not been sufficiently emphasized. Accuracy in identification of archaeological specimens will depend on the accuracy with which the consulted reference specimens are identified.

Discussion

Our test of the LH protocol is thorough because it employs a large sample of stylohyoids from more than 50 animals (although there are few large artiodactyl specimens), and the test specimens were *not* included in the sample used to design the LH protocol. Results demonstrate that the protocol is reliable—identifications were replicable between anatomical sides and analysts. The LH protocol is also valid and accurate:

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Table 2. Results of Blind Test Identifications.

No.	Specimen[a]	Lubinski ID	Lyman ID	Johnson ID	Known ID[b]
1	CSUC: 352	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
2	CSUC: 395	*Antilocapra*	*Antilocapra*	*Antilocapra*	*Antilocapra americana*
3	CSUC: 568	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
4	CSUC: 574-R	UNFUSED	UNFUSED	UNFUSED	*Odocoileus hemionus*
5	CSUC: 574-L	UNFUSED	UNFUSED	UNFUSED	*Odocoileus hemionus*
6	CSUC: 944	*Antilocapra*	*Antilocapra*	*Antilocapra*	*Odocoileus hemionus*
7	CSUC: 1103	*Antilocapra*	*Antilocapra*	*Antilocapra*	*Antilocapra americana*
8	CSUC: 1420-R	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
9	CSUC: 1420-L	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
10	ETSU: ETVP 5171	Small bovid	Small bovid	Small bovid	*Ovis aries*
11	ETSU: ETVP 7109	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
12	ETSU: ETVP 7259	Small bovid	Small bovid	Small bovid	*Ovis candensis*
13	ETSU: ETVP 9210	Large bovid	Large bovid	Large bovid	*Bison bison*
14	ETSU: NAU QSP 13474	Large cervid	*Cervus*	*Cervus*	*Cervus elaphus*
15	ETSU: NAU QSP 2791	*Cervus*	*Cervus*	*Cervus*	*Bos taurus*
16	ETSU: NAU QSP 7535	Large bovid	Large bovid	Large bovid	*Bison bison*
17	ETSU: NAU QSP 7674	*Bos taurus*	*Bos taurus*	*Bos taurus*	*Bos taurus*
18	ETSU: Z174	Small bovid	Small bovid	Small bovid	*Ovis canadensis*
19	ISU: H0002-L	BROKEN	UNFUSED	Small cervid	*Odocoileus virginianus*
20	ISU: H0002-R	BROKEN	UNFUSED	Small cervid	*Odocoileus virginianus*
21	ISU: H0021-L	Small artio.	*Antilocapra*	Small cervid	*Antilocapra americana*
22	ISU: H0021-R	*Antilocapra*	*Antilocapra*	Small cervid	*Antilocapra americana*
23	ISU: H0023	*Antilocapra*	*Antilocapra*	*Antilocapra*	*Antilocapra americana*
24	ISU: H0025-L	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
25	ISU: H0025-R	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
26	ISU: H0032-R	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
27	ISU: H0032-L	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
28	ISU: H0052	Small cervid	Small cervid	Small cervid	*Odocoileus hemionus*
29	ISU: H0105-R	Small bovid	Small bovid	Small bovid	*Capra hircus*
30	ISU: H0105-L	Small bovid	Small bovid	Small bovid	*Capra hircus*
31	IU: 10085	Small bovid	Small bovid	Small bovid	*Ovis aries*
32	IU: 10270-R	Small bovid	Small bovid	Small bovid	*Ovis aries*
33	IU: 10270-L	Small bovid	Small bovid	Small bovid	*Ovis aries*
34	IU: 1010047	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
35	IU: 131002-L	Large bovid	Large bovid	Large bovid	*Bison bison*
36	IU: 131002-R	Large bovid	Large bovid	Large bovid	*Bison bison*
37	IU: 131003-L	Large bovid	Large bovid	Large bovid	*Bison bison*
38	IU: 131003-R	Large bovid	Large bovid	Large bovid	*Bison bison*
39	IU: 1710002	Small bovid	Small bovid	Small bovid	*Capra hircus*
40	IU: 9210356-R	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
41	IU: 9210356-L	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
42	IU: 9310764-L	*Cervus*	*Cervus*	*Cervus*	*Cervus elaphus*
43	IU: 9310764-R	*Cervus*	*Cervus*	*Cervus*	*Cervus elaphus*
44	IU: 9410350	Small bovid	Small bovid	Small bovid	*Capra hircus*
45	IU: 940138-R	Small bovid	Small bovid	Small bovid	*Ovis aries*
46	IU: 940138-L	Small bovid	Small bovid	Small bovid	*Ovis aries*
47	IU: 9510207	Small bovid	Small bovid	Small bovid	*Capra hircus*
48	IU: 9510211-L	Large cervid	*Cervus*	*Cervus*	*Cervus elaphus*
49	IU: 9510211-R	*Cervus*	*Cervus*	*Cervus*	*Cervus elaphus*
50	IU: 9610045-L	Large bovid	Large bovid	Large bovid	*Bison bison*
51	IU: 9610045-R	Large bovid	Large bovid	Large bovid	*Bison bison*
52	IU: 9610352-R	Unknown	*Rangifer*	*Rangifer*	*Rangifer tarandus*
53	IU: 9610352-L	Unknown	*Rangifer*	*Rangifer*	*Rangifer tarandus*
54	IU: 9710078-L	Small bovid	Small bovid	Small bovid	*Capra hircus*

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Table 2. Continued.

No.	Specimen[a]	Lubinski ID	Lyman ID	Johnson ID	Known ID[b]
55	IU: 9710078-R	Small bovid	Small bovid	Small bovid	*Capra hircus*
56	IU: 9810010-L	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
57	IU: 9810010-R	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
58	IU: 9810035	*Rangifer*	*Rangifer*	*Rangifer*	*Rangifer tarandus*
59	IU: 9810423	Large bovid	Large bovid	Large bovid	*Bison bison*
60	IU: 9910001	Unknown	PATHOL.[c]	PATHOL.[c]	*Odocoileus virginianus*
61	IU: 9910213A-L	Small bovid	Small bovid	Small bovid	*Capra hircus*
62	IU: 9910213A-R	Small bovid	Small bovid	Small bovid	*Capra hircus*
63	IU: A4-L	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
64	IU: A4-R	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
65	IU: BICA 06-01	Small bovid	Small bovid	Small bovid	*Ovis candensis*
66	IU: D82	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
67	IU: G77	Large bovid	Large bovid	Large bovid	*Bos taurus*
68	IU: GEARS 10-1	Large bovid	Large bovid	Large bovid	*Bison bison*
69	IU: GEARS 10-3	Large bovid	Large bovid	Large bovid	*Bison bison*
70	IU: J1-R	Large cervid	*Cervus*	*Cervus*	*Alces alces*
71	IU: :J1-L	*Cervus*	*Cervus*	*Cervus*	*Alces alces*
72	IU: J24	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
73	IU: O1-R	*Bos*	*Bos*	*Bos*	*Bos taurus*
74	IU: O1-L	*Bos*	*Bos*	*Bos*	*Bos taurus*
75	IU: R27	Large bovid	Large bovid	Large bovid	*Bison bison*
76	IU: U71-L	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*
77	IU: U71-R	Small cervid	Small cervid	Small cervid	*Odocoileus virginianus*

[a]Listed as Facility: Specimen Number - Side (if both sides included).
[b]As reported by loaning curation facility.
[c]This specimen was keyed as small bovid but was noted as pathological, and no identification would have been made if it had been an archaeological specimen.

92%–96% of the taxonomic identifications were correct per investigator. The high level of agreement in results based on metric attributes indicates that any inter-analyst differences in measurement are not significant (e.g., Breslawski and Byers 2015). And the usability of the LH protocol is good—if not perfect (see below) because its major aspects are spelled out explicitly. Consequently, both a highly experienced analyst and a novice produced a majority of correct identifications.

There were significant differences between analysts in identification resolution, a component of validity and a subject of some debate. Put simply, is it better to have low-resolution (e.g., genus-level) identifications with low rates of misidentification or high-resolution (e.g., species-level) identifications with higher rates of misidentification? Some advocate the former (e.g., Chaplin 1971; Gilbert 1973; Lawrence 1957, 1973; Peres 2010). Others do not necessarily advocate the latter but seem less concerned about the rate of inaccurate identification (e.g., Ziegler 1973). Gobalet (2001:380) seems to imply that accurate species-level identification is of less importance for a zooarchaeologist interested in human-behavioral questions than for a biologist interested in biogeography. To us, the implied acceptance of inaccurate identifications is ill-advised, regardless of the research question.

Lubinski's application of the LH protocol led to accurate low-resolution identifications, whereas Lyman and Johnson's led to inaccurate high-resolution identifications. In our view, low-resolution identification is preferable to misidentification because there is less potential for errors. It is undeniable that differences in taxonomic level of identification affect comparability of results between analysts, a component of usability. But it is also true that incorrect taxonomic identifications can range from having a minor influence on interpretations (e.g., Lyman 2012) to a fairly major influence on what we think happened in the past (e.g., Lyman 2011b). An

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Table 3. Evaluating Blind Test Identification Success.

Expectation	Analyst	Identical Match[a]	Consistent Match[b]	No Match
Left-right paired identifications match	Lubinski	18/21 (86%)	21/21 (100%)	0
	Lyman	21/21 (100%)	21/21 (100%)	0
	Johnson	22/22 (100%)	22/22 (100%)	0
Agreement: Analyst identifications match each other	All three	65/73 (89%)	71/73 (97%)	2/73 (3%)
Accuracy: Analyst identifications match known species	Lubinski	63/73 (86%)	70/73 (96%)	3/73 (4%)
	Lyman	68/72 (94%)	68/72 (94%)	4/72 (6%)
	Johnson	68/74 (92%)	68/74 (92%)	6/74 (8%)

[a]Perfect match in identifications—that is, at the same taxonomic resolution (e.g., all identifications were for *Cervus elaphas*).
[b]Identical match or agreement at a different taxonomic resolution (e.g., one identification was for *Cervus elaphas*, and the other for unknown large cervid).

example of the latter is that the paleoenvironmental implications of bird bones inaccurately identified as turkey (*Meleagris gallopavo*) from the Great Basin of western North America were negated once the identifications were corrected to sage grouse (*Centrocercus urophasianus*; Grayson 1977).

The high degree of reliability and validity suggests that the LH protocol is usable, although differences in our applications led to different degrees of resolution of identification and slightly different misidentification rates. In performing this blind test, Lubinski also realized that Step 4a (S- or C-shaped curvature in dorsal view) of the LH protocol was not easy to use. Based on the test sample of stylohyoids, this trait appeared to vary in a more continuous way than in an unambiguously distinct way, and it should be deleted from the key. None of the three analysts used this trait in this study. Other improvements to usability are to note typographical errors for Step 6a and 6b in Table 5 of Lubinski and Hale (2018), which should have no additional steps for large bovids and direct the user to Step 7 for large cervids, and to provide the photographs of how to take metrics from Hale's (2016:23–24) thesis.

This test has demonstrated inter-analyst variability in the use of an explicit key-based identification protocol. Despite our goals to each follow the directions in Lubinski and Hale (2018), we did not employ the protocol identically. This reflects two things. First, as a developer of the LH protocol, Lubinski was familiar with it and used it in the manner he believed to be the most correct. Second, as non-developers, Lyman and Johnson were less familiar with the protocol, so they tested their understanding of it with a pilot study. Finding success, they adopted it to identify the test specimens. In hindsight, they learned a lesson worth emphasizing: study closely any identification protocol with which you are not intimately familiar and, if possible, consult the developer of that protocol as to the correctness of your understanding of it. We differed in how we operationalized the protocol, and our resulting identifications varied accordingly. If viewed in terms of identification resolution, the differences in results are notable, with 92%–94% identical matches between Lyman and Johnson's identifications and known taxonomy, and 86% with Lubinski's identifications. On the other hand, if viewed in terms of mismatches between identifications and known taxonomy, regardless of resolution, the variability in accuracy is quite modest—6%–8% compared to 4% inaccurate. In our view, this constitutes an acceptable level of agreement and underscores the overall validity, reliability, and usability of the LH protocol.

Earlier we expressed our preference for what are probably accurate low-resolution identifications (taxonomic genus or family) over potentially inaccurate high-resolution identifications (species). Others may prefer the opposite. Whichever is preferred, our test underscores the need for explicit description of protocols, including the anatomical traits thought to be taxonomically diagnostic, and whether high- or low-resolution identifications were attempted.

https://doi.org/10.1017/aaq.2020.45 Published online by Cambridge University Press

Conclusions

Gobalet recommended that "identifications should be accepted only when there has been a discussion of the comparative methods used and the criteria used in the discriminations" (2001:385). We have expanded Gobalet's recommendation by outlining a specific procedure for testing identification protocols—both their analytical techniques and their anatomical traits. Following the wisdom of Lawrence (1973), we see the critical nature of taxonomic identification as a warrant for explicit description in reports of the reference specimens and illustrated skeletal keys consulted, as well as the analytical techniques and skeletal traits used to make identifications. We also urge the creation of an open-access database of skeletal traits believed to be taxonomically diagnostic. It is a huge task to state identification criteria for every specimen, and editors are not likely to accept such criteria in published works, even for a limited set of the most problematic and interpretively significant identifications. We therefore suggest starting small, by focusing on more difficult identifications, explicitly testing anatomical traits, and describing those traits used to make identifications. The creation of an open-access database of skeletal traits would facilitate this process, providing a place for publication, peer review, and discussion. The database should allow editing when a trait is found to be difficult to employ (has low usability) or when it does not always result in accurate identifications (has low validity; e.g., Barr 2008; Lyman 2019). Comments regarding the resolution provided by an anatomical trait should be part of the discussion of each one proposed to be taxonomically diagnostic. A trait naming system would also be helpful so that future zooarchaeologists might be able to list identifiers such as "Trait MA-106." A nomenclatural system like the one developed by Driesch (1976) for osteometry, with which zooarchaeologists the world over are familiar, suggests that such a system is not implausible (its practicality is unclear).

It is our hope that the sort of test described here will be undertaken and reported going forward so that problems such as those described by Gobalet (2001) can be avoided in the future. Like Wolverton (2013) before us, we hope to see a marked improvement in zooarchaeological quality control. Nims and Butler's (2017) recommendation for continuous evaluation of the identification protocol being used over the duration of a project is a superb additional suggestion to enhance the probability of correct taxonomic identifications and reduce inter-observer variation. Much of modern zooarchaeology rests on high-resolution taxonomic identifications. Insofar as those identifications are weakly substantiated and poorly documented, so too are any interpretations based on them.

Acknowledgments. Loans of comparative stylohyoid bones were provided by Laura Scheiber (Indiana University), Matt G. Hill (Iowa State University), Chris Widga (Eastern Tennessee State University), and Matt O'Brien (California State University, Chico). This manuscript was improved in light of comments provided by the editor and four anonymous reviewers.

Data Availability Statement. A Microsoft Excel spreadsheet containing all of the original measurements generated in this study is available from Pat Lubinski at Central Washington University.

Supplemental Materials. For supplemental material accompanying this report, visit https://doi.org/10.1017/aaq.2020.45. For supplemental figures showing how to take stylohyoid measurements, see Figure 10 from Hale (2016:23–24), available at https://digitalcommons.cwu.edu/etd/344/.

Supplemental Text 1. On Taxonomic Identification and the Changing Status of "Diagnostic" Skeletal Traits.

References Cited

Adams, Bradley, and Pam Crabtree
2012 *Comparative Osteology*. Elsevier, Amsterdam.

Baker, Barry W., and Brian S. Shaffer
1999 Assumptions about Species: A Case Study of Tortoise Bones from SE Texas. *Journal of Field Archaeology* 26:69–74.

Balaase, Marie, and Stanley H. Ambrose
2005 Distinguishing Sheep and Goats Using Dental Morphology and Stable Carbon Isotopes in C_4 Grassland Environments. *Journal of Archaeological Science* 32:691–702.

Balkwill, Darlene, and Stephen L. Cumbaa
1992 *A Guide to the Identification of Postcranial Bones of* Bos taurus *and* Bison bison. Canadian Museum of Nature, Ottawa, Ontario.

Barnosky, Anthony D, Elizabeth A. Hadly, Patrick Gonzalez, Jason Head, David Polly, A. Michelle Lawing, Jussi T. Eronen, David D. Ackerly, Ken Alex, Eric Biber, Jessica Blois, Justin Brashares, Gerardo Ceballos, Edward Davis, Gregory P. Dietl, Rodolfo Dirzo, Holly Doremus, Mikael Fortelius, Harry W. Greene, Jessica Hellmann,

Thomas Hickler, Stephen T. Jackson, Melissa Kemp, Paul L. Koch, Claire Kremen, Emily L. Lindsey, Cindy Looy, Charles R. Marshall, Chase Mendenhall, Andreas Mulch, Alexis M. Mychajliw, Carsten Nowak, Uma Ramakrishnan, Jan Schnitzler, Kashish Das Shrestha, Katherine Solari, Lynn Stegner, M. Allison Stegner, Nils Chr. Stenseth, Marvalee H. Wake, Zhibin Zhang
2017 Merging Paleobiology with Conservation Biology to Guide the Future of Terrestrial Ecosystems. *Science* 355. DOI:10.1126/science.aah4787.

Barr, William A.
2008 Taxonomic Ambiguity in Faunal Analysis: Theory, Practical Considerations, and Implications. Master's thesis, University of Texas, Austin .

Beisaw, April M.
2013 *Identifying and Interpreting Animal Bones*. Texas A&M University Press, College Station.

Bochenski, Zbigniew M.
2008 Identification of Skeletal Remains of Closely Related Species: The Pitfalls and Solutions. *Journal of Archaeological Science* 35:1247–1250.

Boessneck, J.
1969 Osteological Difference between Sheep (*Ovis aries* Linné) and Goat (*Capra hircus* Linné). In *Science in Archaeology*, rev. ed, edited by Don Brothwell and Eric S. Higgs, pp. 331–358. Prager, New York.

Brainerd, George W.
1939 An Illustrated Field Key for the Identification of Mammal Bones. *Ohio State Archaeological and Historical Quarterly* 48:324–328.

Breslawski, Ryan P., and David A. Byers
2015 Assessing Measurement Error in Paleozoological Osteometrics with Bison Remains. *Journal of Archaeological Science* 53:235–242.

Brown, Christopher L., and Carl E. Gustafson
1979 *A Key to Postcranial Skeletal Remains of Cattle/Bison, Elk, and Horse*. Reports of Investigations No. 57. Laboratory of Anthropology, Washington State University, Pullman.

Butler, Virginia L.
2011 Comments on "Identification, Classification, and Zooarchaeology." *Ethnobiology Letters* 2:31.

Butler, Virginia L., and R. Lee Lyman
1996 Taxonomic Identifications and Faunal Summaries: What Should We Be Including in Our Faunal Reports. *Society for American Archaeology Bulletin* 14(1):22.

Carmines, Edward G., and Richard A. Zeller
1979 *Reliability and Validity Assessment*. Sage, Beverly Hills, California.

Chaplin, Raymond E.
1971 *The Study of Animal Bones from Archaeological Sites*. Seminar Press, London.

Clason, Antje Trientje
1972 Some Remarks on the Use and Presentation of Archaeozoological Data. *Helinium* 12:139–153.

Davis, Simon J. M.
1987 *The Archaeology of Animals*. Yale University Press, New Haven, Connecticut.

Dietl, Gregory P., and Karl W. Flessa (editors)
2017 *Conservation Paleobiology*. University of Chicago Press, Chicago.

Driesch, Angela von den
1976 *A Guide to the Measurement of Animal Bones from Archaeological Sites*. Bulletin 1. Peabody Museum of Archaeology and Ethnology, Harvard University, Cambridge, Massachusetts.

Driver, Jonathan C.
1982 Minimum Standards for Reporting of Animal Bones in Salvage Archaeology: Southern Alberta as a Case Study. In *Directions in Archaeology*, edited by Peter D. Francis and Eric C. Poplin, pp. 199–209. University of Calgary Archaeological Association, Calgary, Alberta.
1992 Identification, Classification, and Zooarchaeology. *Circaea* 9:35–47.
2011 Twenty Years After "Identification, Classification, and Zooarchaeology." *Ethnobiology Letters* 2:36–39.

Eaton, George F.
1898 The Prehistoric Fauna of Block Island, as Indicated by Its Ancient Shell-Heaps. *American Journal of Science* 6:137–159.

Emery, Kitty F.
2004 In Search of Assemblage Comparability: Methods in Maya Zooarchaeology. In *Maya Zooarchaeology*, edited by Kitty F. Emery, pp. 15–33. Costen Institute of Archaeology Monograph 51. University of California, Los Angeles.

Findley, James S.
1964 Paleoecologic Reconstruction: Vertebrate Limitations. In *The Reconstruction of Past Environments*, assembled by James J. Hester and James Schoenwetter, pp. 23–25. Publication No. 3. Fort Burgwin Research Center, Taos, New Mexico.

Gifford-Gonzalez, Diane
2018 *An Introduction to Zooarchaeology*. Springer, Cham, Switzerland.

Gilbert, B. Miles
1973 *Mammalian Osteo-Archaeology: North America*. Missouri Archaeological Society, Columbia, Missouri.

Gobalet, Kenneth W.
2001 A Critique of Faunal Analysis; Inconsistency among Experts in Blind Tests. *Journal of Archaeological Science* 28:377–386.

Golafshani, Nahid
2003 Understanding Reliability and Validity in Qualitative Research. *Qualitative Report* 8:597–607.

Graham, Russell W., and Holmes A. Semken Jr.
1987 Philosophy and Procedures for Paleoenvironmental Studies of Quaternary Mammalian Faunas. In *Late Quaternary Mammalian Biogeography of the Great Plains and Prairies*, edited by Russell W. Graham, Holmes A. Semken Jr., and Mary Ann Graham, pp. 1–17. Scientific Papers Vol. 22. Illinois State Museum, Springfield.

Grayson, Donald K.
1977 A Review of the Evidence for Early Holocene Turkeys in the Northern Great Basin. *American Antiquity* 42:110–114.

Grayson, Donald K., and Chris Maser
1978 An Efficient, Inexpensive Dermestid Colony for Skeleton Preparation. *Journal of Field Archaeology* 5:246–247.

Hale, Thomas A.
2016 Species Identification of the Stylohyoid Bone for North American Artiodactyls. Master's thesis, Cultural and Environmental Resource Management Program, Central Washington University, Ellensburg.

Halstead, Paul, Patricia Collins, and Valasia Isaakidou
2002 Sorting the Sheep from the Goats: Morphological Distinctions between the Mandibles and Mandibular

Teeth of Adult *Ovis* and *Capra*. *Journal of Archaeological Science* 29:545–553.
Haruda, Ashleigh F.
2017 Separating Sheep (*Ovis aries* L.) and Goats (*Capra hircus* L.) Using Geometric Morphometric Methods: An Investigation of Astragalus Morphology from Late and Final Bronze Age Central Asian Contexts. *International Journal of Osteoarchaeology* 27:551–562.
Heale, Roberta, and Alison Twycross
2015 Validity and Reliability in Quantitative Studies. *Evidence-Based Nursing* 18:66–67.
Hesse, Brian, and Paula Wapnish
1985 *Animal Bone Archaeology*. Manuals on Archaeology No. 5. Taraxacum, Washington, DC.
Hillson, Simon
1992 *Mammal Bones and Teeth*. Institute of Archaeology, University College London, London.
Jacobson, Jodi A.
2003 Identification of Mule Deer (*Odocoileus hemionus*) and White-Tailed Deer (*Odocoileus virginianus*) Postcranial Remains as a Means of Determining Human Subsistence Strategies. *Plains Anthropologist* 48:287–297.
Klein, Richard G., and Kathryn Cruz-Uribe
1984 *The Analysis of Animal Bones from Archaeological Sites*. University of Chicago Press, Chicago.
Lastrucci, Carlo L.
1967 *The Scientific Approach: Basic Principles of the Scientific Method*. Schenkman, Cambridge, Massachusetts.
Lau, Hannah, and Sarah W. Kansa
2018 Zooarchaeology in the Era of Big Data: Contending with Interanalyst Variation and Best Practices for Contextualizing Data for Informed Reuse. *Journal of Archaeological Science* 95:33–39.
Lawrence, Barbara
1951 Post-Cranial Skeletal Characters of Deer, Pronghorn, and Sheep–Goat with Notes on *Bos* and *Bison*. *Papers of the Peabody Museum of American Archaeology and Ethnology* 35(3):7–40.
1957 Zoology. In *The Identification of Non-Artifactual Archaeological Materials*, edited by Walter W. Taylor, pp. 41–42. National Research Council Publication 565. National Academy of Sciences, Washington, DC.
1973 Problems in the Inter-Site Comparisons of Faunal Remains. In *Domestikationsforschung und Geschichte der Haustiere*, edited by Janos Matolcsi, pp. 397–402. Ckademiai Kiado, Budapest.
LeFebvre, Michelle J., and Ashley E. Sharpe
2018 Contemporary Challenges in Zooarchaeological Specimen Identification. In *Zooarchaeology in Practice*, edited by Christina M. Giovas and Michelle J. LeFebvre, pp. 35–57. Springer, Dordrecht, Netherlands.
Lubinski, Patrick M., and Thomas A. Hale
2018 Identifying and Siding the Stylohyoid Bone for North American Artiodactyls. *International Journal of Osteoarchaeology* 28:364–374.
Lupo, Karen D.
2011 Comments on "Identification, Classification, and Zooarchaeology." *Ethnobiology Letters* 2:32.
Lyman, R. Lee
2005 Zooarchaeology. In *Handbook of Archaeological Methods*, Vol. II, edited by Herbert D. G. Maschner and Christopher Chippindale, pp. 835–870. AltaMira Press, Lanham, Maryland.
2011a Comments on Jon Driver's "Identification, Classification, and Zooarchaeology." *Ethnobiology Letters* 2:33–34.
2011b Paleoecological and Biogeographical Implications of Late Pleistocene Noble Marten (*Martes americana nobilis*) in Eastern Washington State, USA. *Quaternary Research* 75:176–182.
2012 Human-Behavioral and Paleoecological Implications of Terminal Pleistocene Fox Remains at the Marmes Site (45FR50), Eastern Washington State, USA. *Quaternary Science Reviews* 41:39–48.
2019 Assumptions and Protocol of the Taxonomic Identification of Faunal Remains in Zooarchaeology: A North American Perspective. *Journal of Archaeological Method and Theory* 26:1376–1438.
Lyman, R. Lee, and Kenneth D. Cannon (editors)
2004 *Zooarchaeology and Conservation Biology*. University of Utah Press, Salt Lake City.
Nims, Reno, and Virginia L. Butler
2017 Assessing Reproducibility in Faunal Analysis Blind Tests: A Case Study from Northwestern North America. *Journal of Archaeological Science: Reports* 11:750–761.
O'Connor, Terry
2000 *The Archaeology of Animal Bones*. Texas A&M University Press, College Station.
Olsen, Stanley J.
1964 *Mammalian Remains from Archaeological Sites, Part I: Southeastern and Southwestern United States*. Papers of the Peabody Museum of Archaeology and Ethnology 56. Harvard University, Cambridge, Massachusetts.
Payne, Sebastian
1985 Morphological Distinctions between the Mandibular Teeth of Young Sheep, *Ovis*, and Goats, *Capra*. *Journal of Archaeological Science* 12:139–147.
Peres, Tanya M.
2010 Methodological Issues in Zooarchaeology. In *Integrating Zooarchaeology and Paleoethnobotany*, edited by Amber M. VanDerwarker and Tanya M. Peres, pp. 15–36. Springer, New York.
Polly, P. David, and Jason J. Head
2004 Maximum-Likelihood Identification of Fossils: Taxonomic Identification of Quaternary Marmots (Rodentia, Mammalia) and Identification of Vertebral Position in the Pipesnake *Cylindrophis* (Serpentes, Reptilia). In *Morphometrics*, edited by Ashraf M. T. Elewa, pp. 197–221. Springer, Berlin.
Rackham, James
1994 *Animal Bones*. University of California Press, Berkeley.
Rea, Amadeo M.
1986 Verification and Reverification: Problems in Archaeofaunal Studies. *Journal of Ethnobiology* 6:9–18.
Reitz, Elizabeth J., and Elizabeth S. Wing
2008 *Zooarchaeology*. 2nd. ed. Cambridge University Press, Cambridge.
Roberts, Paula, Helena Priest, and Michael Traynor
2006 Reliability and Validity in Research. *Nursing Standard* 20:41–45.
Saber, Ashraf S., and Reinhold R. Hoffman
1985 A Comparative Study of the Hyoid Bone (Os hyoides) of 6 Ruminant Species (Ruminantia Scopoli, 1777). *Anatomischer Anzeiger* 158:43–50.
Salvagno, Lenny, and Umberto Albarella
2017 A Morphometric System to Distinguish Sheep and Goat Postcranial Bones. *PLoS ONE* 12(6):e0178543. DOI:10.1371/journal.pone.0178543.

Sokal, Robert R., and F. James Rohlf
1969 *Biometry*. W. H. Freeman and Company, San Francisco.

Stewart, John R.
2005 The Use of Modern Geographical Ranges in the Identification of Archaeological Bird Remains. In *Feathers, Grit and Symbolisms: Birds and Humans in the Ancient Old and New Worlds*, edited by Gisela Grupe and Joris Peters, pp. 43–54. Proceedings of the ICAZ Bird Remains Working Group, Documenta Archaeobiologiae 3. Leidorf, Rahden, Westfalen, Germany.

VanPool, Todd L., and Robert D. Leonard
2011 *Quantitative Analysis in Archaeology*. Wiley-Blackwell, Chichester, West Sussex, United Kingdom.

Whittemore, Robin, Susan K. Chase, and Carol Lynn Mandle
2001 Validity in Qualitative Research. *Qualitative Health Research* 11:522–537.

Wolfhagen, Jesse, and Max D. Price
2017 A Probabilistic Model for Distinguishing between Sheep and Goat Postcranial Remains. *Journal of Archaeological Science: Reports* 12:625–631.

Wolverton, Steve
2013 Data Quality in Zooarchaeological Faunal Identification. *Journal of Archaeological Method and Theory* 20:381–396.

Wolverton, Steve, and Lisa Nagaoka
2017 Zooarchaeology: Investigating Past Interactions between Humans and Other Animals. In *Ethnozoology: Animals in Our Lives*, edited by Romulo R. N. Alves and Ulysses P. Albuquerque, pp. 25–43. Academic Press, London.

Wolverton, Steve, Lisa Nagaoka, and Torben C. Rick
2016 *Applied Zooarchaeology*. Eliot Werner Publications, Clinton Corners, New York.

Zar, Jerrold H.
1996 *Biostatistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey.

Zeder, Melinda A., and Suzanne E. Pilaar
2010 Assessing the Reliability of Criteria Used to Identify Postcranial Bones in Sheep, *Ovis*, and Goats, *Capra*: *Journal of Archaeological Science* 37:2887–2905.

Ziegler, Alan C.
1973 *Inference from Prehistoric Faunal Remains*. Addison Wesley Module in Anthropology No. 43. Addison-Wesley, Reading, Massachusetts.

Submitted September 9, 2019; Revised April 26, 2020; Accepted April 27, 2020