



A probabilistic model for distinguishing between sheep and goat postcranial remains



Jesse Wolfhagen^{a,*}, Max D. Price^b

^a Interdepartmental Doctoral Program in Anthropological Sciences, Stony Brook University, Stony Brook, United States

^b Department of Anthropology, Harvard University, Cambridge, United States

ARTICLE INFO

Article history:

Received 21 November 2016

Received in revised form 11 February 2017

Accepted 14 February 2017

ABSTRACT

Distinguishing morphologically-similar taxa is a perennial problem for zooarchaeologists. As an example, zooarchaeological methods to distinguish sheep and goat bones are limited by the fact that diagnostic traits are not exclusive to each species. For this reason, we argue that sheep/goat bones should be considered probabilistically, not absolutely. That is, each bone should be considered as falling along a spectrum from sheep to goat. To that end, we present a probabilistic method for assigning taxonomic status to sheep/goat postcranial specimens. Suites of diagnostic traits present on each specimen are translated into probabilities, which we then use in simulations to estimate the number of sheep and goat bones. We apply our model to two assemblages: 'Ain Dara (Syria) and Tell Sakhariya (Iraq). We evaluate the effectiveness of standard diagnostic traits, finding that distal tibias and second phalanges particularly problematic to identify to species. In addition, our probabilistic method improves the reliability and replicability of sheep/goat identifications in zooarchaeological assemblages by making explicit the diagnostic criteria used and providing clear standards that other projects can follow and thus be directly compared.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Replicable taxonomic assignments are central to the practice of zooarchaeology. Archaeologists rarely have the opportunity to verify published species assignments, forcing them to treat such assignments as absolute facts once they have been published. This can be problematic, though, because experiments on the replicability of zooarchaeological identifications show variability between analysts (e.g., Atici et al., 2013; Gobalet, 2001; Morin et al., 2016; Twiss et al., 2016; Driver, 2011; Fish, 1978). This variability makes it difficult to assess the replicability of taxonomic identifications.

Sheep (*Ovis aries*) and goats (*Capra hircus*) have historically been difficult for zooarchaeologists to distinguish. Although numerous publications guide researchers in differentiating between sheep and goats (e.g., Boessneck, 1969; Buitenhuis, 1995; Payne, 1969; Prummel and Frisch, 1986; Zeder and Lapham, 2010), separation using skeletal morphological criteria, or “traits,” remains difficult. Even traits that are considered diagnostic are not exclusive to either sheep or goat skeletons (Zeder and Lapham, 2010). Although zooarchaeologists recognize this uncertainty, standard recording methods and publication norms

produce false senses of security in taxonomic assignments. Thus once identifications of sheep and goat have been made and published, they are treated as immutable facts despite their shaky foundations.

To address this problem, we use data published by Zeder and Lapham (2010) to calculate posterior probabilities of taxonomic status for each included skeletal element using discriminant function analysis. Zeder and Lapham's publication is an ideal starting point for this paper. Although it does not present a complete list of the various diagnostic traits or elements used by zooarchaeologists (for example, they do not distinguish horn cores or vertebrae), it does provide a list of the more commonly-preserved traits. These traits are concentrated on the limb bones, which are useful for additional research into survivorship analysis and biometrics – analyses that assume reliable species-specific identifications. Finally, their work stands out as the only “test” of morphological traits against a large sample of known sheep and goats.

Using Zeder and Lapham's study as a starting point, we present a method based on probabilistic modeling for capturing the uncertainty in sheep/goat identifications. This *probabilistic model* allows us to expand our understanding of sheep and goat management in the ancient past by 1) including previously-taxonomically indeterminate material and 2) improving the replicability of metrics such as sheep:goat ratios and comparisons of body part representation. We apply this approach to two zooarchaeological assemblages to test its applicability to fragmentary remains: 'Ain Dara (Syria) and Tell Sakhariya (Iraq).

* Corresponding author at: Department of Anthropology, SBS S-501, Stony Brook University, Stony Brook, NY 11790-4364, United States.

E-mail address: jesse.wolfhagen@stonybrook.edu (J. Wolfhagen).

1.1. Absolutist and probabilistic paradigms in taxonomic assignments

The paradigm under which current zooarchaeological assignments are published can be considered “absolutist” because uncertainty is rarely acknowledged. When a specimen is publicly identified as a sheep or a goat, any uncertainty about that identification becomes invisible. Yet specimens vary in their identifiability due to state of preservation, state of epiphyseal fusion, elements present, access to and quality of comparative collection, analyst optimism/confidence, and whether second opinions were sought (Driver, 2011; Wolverson, 2012). Publications mask these inconsistencies.

Meanwhile, specimens that lack key traits due to bone breakage or have contradictory trait morphologies (e.g., trait 1 = sheep; trait 2 = goat) may not be assigned to either taxon. Zooarchaeologists often code these specimens as “sheep/goat.” The decision to assign a specimen to “sheep/goat” as opposed to a specific taxon varies from analyst to analyst and even throughout the recording of an assemblage.

Greater transparency is achievable. It requires reporting individual trait states (which can be investigated for their inter-analyst reliability, e.g., Zeder and Lapham, 2010), and using a *probabilistic* model to make taxonomic assignments. Unlike an absolutist paradigm, a probabilistic paradigm takes into account the uncertainties inherent to morphological trait identifications. Advances in computational power have empowered archaeologists to examine the process of identification more thoroughly and improve the rigors of standard research protocols. Analogous debates have occurred in other sub-disciplines, such as the essentialist/materialist paradigms in lithic analysis (Shott, 2010).

A probabilistic paradigm uses reference data to translate the combination of “sheep” and “goat” traits for each specimen into the probability that a specimen came from a goat. In statistics, this is expressed as $P(\text{goat}|\text{traits})$ – this expression should be read as “the probability of a specimen being a goat given its observed traits.” This is the *posterior probability*. The traits of known-taxon specimens provide logically distinct information, the probability of a combination of traits coming from a goat: $P(\text{traits}|\text{goat})$. Meanwhile, *prior probabilities* – $P(\text{goat})$ and $P(\text{sheep})$ – express beliefs about the proportion of goats or sheep in an assemblage. One can calculate the posterior probability in terms of these other probabilities using Bayes’ Theorem:

$$P(\text{goat}|\text{traits}) = \frac{P(\text{traits}|\text{goat})P(\text{goat})}{P(\text{traits}|\text{goat})P(\text{goat}) + P(\text{traits}|\text{sheep})P(\text{sheep})}$$

$P(\text{goat}|\text{traits})$: probability that a specimen is a goat given its combination of traits. This is known as the *posterior probability*.

$P(\text{goat})$: probability that any specimen is a goat. This is known as the *prior probability*.

$P(\text{traits}|\text{goat})$: probability that a specimen would have a specific combination of traits given that it’s a goat. This is known as the *likelihood*.

$P(\text{traits}|\text{sheep})$: *likelihood* that a specimen would have a specific combination of traits given that it’s a sheep.

$P(\text{sheep})$: *prior probability* that a specimen is a sheep = $1 - P(\text{goat})$.

2. Materials

2.1. Reference populations: Zeder and Lapham’s (2010) study of known sheep and goats

A probabilistic approach to sheep/goat determination requires actualistic studies on known sheep and goat skeletons to identify the distribution of traits across both taxa. Such studies enable us to estimate how frequently a certain trait (e.g., a strongly expressed articular ridge on the medial facet of the astragalus) occurs in sheep as opposed to in goats. Zeder and Lapham’s (2010) paper is an excellent example of such a study (Table 1). They recorded the postcranial traits of 84 sheep and 49 goats in several museums (primarily collected from

Table 1

Element parts and number of traits from Zeder and Lapham (2010).

Element	# of traits
Distal humerus	4
Proximal radius	4
Distal radius	2
Distal tibia	3
Distal metacarpal	3
Distal metatarsal	4
Astragalus	4
Calcaneus	4
First phalanx	4
Second phalanx	2

Iran), identifying the accuracy of 31 traits on 8 skeletal elements (Zeder and Lapham, 2010: Tables 3, 13). Importantly, Zeder and Lapham (2010) provided their raw data in a supplemental table, allowing researchers to investigate the distributions of different traits across both sheep and goat populations.

2.2. Archaeological test cases: ‘Ain Dara and Tell Sakhariya

To show how the probabilistic method can be used on archaeological samples, we used sheep and goat remains recovered from two Middle Eastern archaeological sites: ‘Ain Dara (Syria) and Tell Sakhariya (Iraq). Both of these collections are curated by the Zooarchaeology Laboratory at Stony Brook University. These assemblages are ideal for this study because they each contain a robust sample of mixed sheep and goat postcranial remains with preserved morphological traits.

‘Ain Dara is an Iron Age (early 1st millennium BC) site located in northwestern Syria. It was excavated in the 1980s by a team led by Elizabeth Stone and Paul Zimansky (Stone and Zimansky, 1999). Zooarchaeological analysis of the ‘Ain Dara’s fauna (Table 2) was conducted by Carol Frey and Curtis Marean (Frey and Marean 1999), who differentiated between sheep and goat using Boessneck (1969), Payne (1985), and then-unpublished criteria provided by Melinda Zeder (Frey and Marean, 1999). Our sample included 86 postcranial bones identified as sheep, sheep/goat, goat, and (in a small number of cases) bovid/cervid.

Tell Sakhariya is a Middle Bronze Age site located in southern Iraq. It was excavated in a single season in 2011/2012 by Elizabeth Stone and Paul Zimansky (Zimansky and Stone, 2016). The bulk of the faunal material originated from the Sealand-Dynasty (mid-2nd millennium BC). The faunal remains were studied by Katheryn Twiss (Twiss, in press), who differentiated between sheep and goat using Zeder and Lapham (2010) (Katheryn Twiss, personal communication). Our sample included 55 postcranial bones identified as sheep, sheep/goat, and goat.

Table 2

Summary data for major taxa at ‘Ain Dara and Tell Sakhariya.

Taxa	‘Ain Dara	Sakhariya
<i>Bos</i>	208	160
<i>Ovis/Capra</i>	413	303
(<i>Ovis</i>)	109	80
(<i>Capra</i>)	89	2
<i>Gazella</i>	0	19
<i>Equus</i>	61	93
<i>Sus</i>	329	75
<i>Canis</i>	23	50
Reference	Frey and Marean, 1999: 126	Twiss, in press: Table 2

3. Methods

3.1. Identification of sheep and goat traits at 'Ain Dar and Tell Sakhariya

- (1) We discussed the Zeder and Lapham (2010) criteria, using modern sheep and goat bones as a reference, to ensure comparability between our assignments. Zeder and Lapham (2010) coded their trait values as “C”, “O”, “O/C”, “C/O”, or not-preserved (“-”, “x”); we re-coded these variables as “Capra”, “Ovis”, “O/C”, and “NA.” With a maximum of four traits per element, each specimen had up to 16 possible trait combinations.
- (2) We each independently coded sheep and goat postcranial remains from 'Ain Dara and Tell Sakhariya according to Zeder and Lapham's (2010) criteria. To eliminate trait identification interdependency (i.e., the bias of recording trait 2 as “Ovis” because trait 1 was “Ovis”), we identified traits one at a time on all elements from both sites.
- (3) We combined our evaluations and resolved discrepancies. If one analyst chose “Capra” and the other “Ovis”, the consensus identification was “O/C”; if one analyst chose “O/C” and the other chose a specific taxon, the consensus identification of that trait was the specific taxon.

Leave-one-out cross-validation evaluated how well the Zeder and Lapham (2010) trait data separated the two taxa (Kovarovic et al., 2011). Cross-validation used the discriminant function analysis method described in Section 3.2 on the reference data multiple times, leaving out a single specimen each time to calculate its group membership probabilities (Evin et al., 2013). 1000 simulated taxonomic estimates derived from the membership probabilities of every specimen were used to calculate ‘chance-corrected’ classification accuracies (TAU) for each element (Kovarovic et al., 2011).

3.2. Calculating the likelihood: discriminant function analysis (DFA)

We calculated $P(\text{traits}|\text{goat})$ and $P(\text{traits}|\text{sheep})$ likelihood values with discriminant function analysis (DFA). DFA is a multivariate statistical method that uses a series of *predictor variables* to predict group membership (an *explanatory variable*) based on a set of training data of known group membership (Baxter, 2003). DFA first calculates a series of “discriminant functions” (one less than the number of groups) that, when applied to the reference data, maximizes the distance between groups (Kovarovic et al., 2011). We used these multivariate distances to calculate the probability that each specimen was a goat or sheep using the Zeder and Lapham actualistic data as a reference (Zeder and Lapham, 2010: Table S2).

We calculated likelihood values via DFA for each element using the following steps:

- (1) The archaeological and reference trait presence/absence data for each element were converted into a matrix of Jaccard distances, which reflect how different two specimens are to one another based on how many shared traits they have. Because suites of diagnostic traits were coded as presence/absence of a species-specific trait-state (i.e., “Capra-like Trait 1”), only shared presences of a trait-state conveyed meaningful similarities between specimens; mutual lack of a trait-state was not meaningful.
- (2) This distance matrix was rescaled with Principal Coordinate Analysis (PCoA) to approximate the Euclidean distances between specimens (Legendre and Legendre, 1998; Quinn and Keough, 2002). Rescaling distance to Euclidean distances simplifies the application of the DFA. An added bonus is that measurement data (properly centered and rescaled) could also be incorporated into the trait information.
- (3) PCoA values of the reference data were analyzed via a linear discriminant analysis (a form of DFA) to provide a discriminant

function that is applied to the archaeological specimens, resulting in probabilities of group membership. The LDA uses all of the PCoA eigenvalues that sum up to 95% of the total variation in the sample. This procedure is similar to that developed by Evin et al. (2013) for geometric morphometrics.

- (4) $P(\text{traits}|\text{goat})$ and $P(\text{traits}|\text{sheep})$ likelihood values were calculated for each unknown specimen using a normal approximation of the discriminant function scores of the known specimens.

3.3. Estimating posterior probabilities and other parameters

$P(\text{goat}|\text{traits})$ posterior probabilities for each specimen were calculated with Bayes' Theorem, based on the likelihood values calculated via DFA and a sample of prior probability estimates for $P(\text{goat})$ that reflect our estimate of the proportion of goats at the site prior to the analysis. As we were interested in testing the validity of the method, we wanted to provide a wide range of $P(\text{goat})$ values and therefore used a vague prior for $P(\text{goat})$: a Beta distribution with the parameters α and $\beta = 0.5$ (Lunn et al., 2013). For each site, we randomly drew from this distribution 10,000 times to provide prior estimates of $P(\text{goat})$, which resulted in 10,000 posterior estimates of $P(\text{goat}|\text{traits})$ for each specimen (Smith and Gelfand, 1992). $P(\text{goat}|\text{traits})$ values for each specimen were summarized by their medians and 89% Highest Posterior Density Intervals (HPDI), which present the range of the 89% likeliest values of the sample. HPDI values are similar in approach to confidence intervals, but provide the range of likeliest values rather than values at certain percentiles of the data (McElreath, 2016a). To move from $P(\text{goat}|\text{traits})$ values for each specimen to actual “Capra” or “Ovis” identities, we used random Binomial draws of each $P(\text{goat}|\text{traits})$, which is essentially flipping a weighted coin with the odds of “Capra” equal to the $P(\text{goat}|\text{traits})$ value. Combining these individual identities in each iteration allowed us to calculate site-level parameters: total numbers of sheep and goats, the proportion of goats in the assemblage, and element-specific counts of sheep and goats (Bréhard et al., 2014; Crema, 2011).

3.4. Statistical analyses and reproducibility

All of these analyses were undertaken in the R Statistical Environment, version 3.3.1 (R Core Team, 2016). The Jaccard distance matrix was calculated using the `vegdist()` function in the “vegan” package, version 2.4-2 (Oksanen et al., 2016). Principal Component Analysis was calculated using the `pcoa()` function in the “ape” package, version 3.5 (Paradis et al., 2004). Linear discriminant analysis was calculated with the `lda()` function in the “MASS” package (Venables and Ripley, 2002). Calculation of HPDI was calculated using the `HPDI()` function in the “rethinking” package, version 1.59 (McElreath, 2016b). All plots were drawn using the “ggplot2” package (Wickham, 2009).

Following the transparency and reproducibility principles in Marwick (2016), we include the R script used to structure the Zeder and Lapham (2010) reference data (after saving it as a CSV format) and run all the analyses and visualizations in this paper in our supplemental online material (SOM) attached to this publication. We also include the raw data from 'Ain Dara and Tell Sakhariya, necessary to independently reproduce the analyses presented here, stored in an associated Mendeley dataset with, accessible at <http://dx.doi.org/10.17632/sz7jjfg6fc.1>. Our R script is released under the MIT license, our data as CC-0, and our figures as CC-BY, to provide maximal reuse capabilities.

4. Results

4.1. Evaluation of the Zeder and Lapham (2010) traits

The ‘chance-corrected’ classification accuracy rates for Zeder and Lapham's (2010) sheep/goat traits are shown in Fig. 1. In general, the cross-validated success rates are high (over 80% for all elements, over

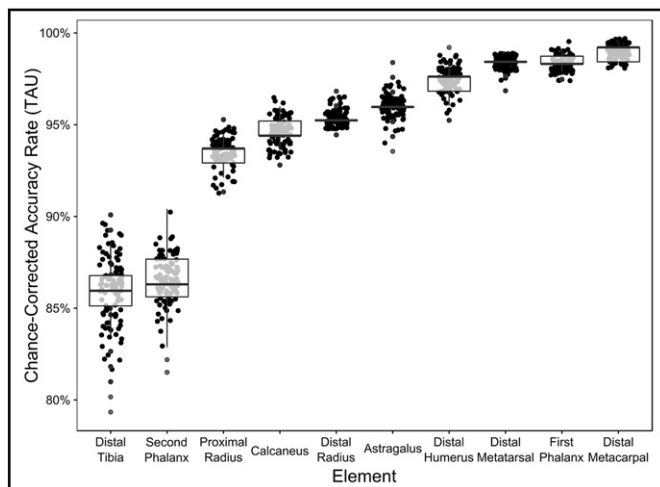


Fig. 1. Accuracy rates of cross-validated linear discriminant analysis of Zeder and Lapham (2010) reference data.

90% for elements excluding the distal tibia and second phalanx), indicating that these traits are useful for distinguishing between sheep and goat postcranial remains. The lower average success rates for distal tibias and second phalanges may mean that these specimens are less diagnostic or that current diagnostic traits are insufficient; Zeder and Lapham (2010) noted that their analysts performed poorly on the distal tibia. Table 3 compares the average classification inaccuracy (1-TAU) to the “Projected % Misidentified” rates from the experienced analysts in Zeder and Lapham’s (2010) analysis (Table 14). Overall, the cross-validation rates suggest that the LDA method described in this paper appears to be as accurate as identifications made by experts.

4.2. Archaeological applications

Incorporating the uncertainty inherent in sheep/goat taxonomic assignments increased the amount of information that can be gleaned from bone assemblages. The original identifications by Twiss, Frey, and Marean were either “Ovis”, “Capra”, or various indeterminate categories (“Ovis/Capra” or “Bovid/Cervid”). We can use a 90% threshold advocated by Evin et al. (2013) to interpret $P(\text{goat}|\text{traits})$ posterior probabilities into those same categories. Doing so decreases the amount of indeterminate specimens in both assemblages. 23% of the ‘Ain Dara specimens (20/86) were originally identified as “Bovid/Cervid” or “Ovis/Capra”; our method halved the number of unidentified specimens to 10% (9/86) (Supplementary Table 1). For Tell Sakhariya, 31% of the sheep/goat specimens (17/55) were originally “Ovis/Capra”; the probabilistic method reduced the rate to 9% (5/55) based on median posterior probabilities (Supplementary Table 2). The probabilistic method also

identified 4 specimens as likely goats (median $P(\text{goat}|\text{traits}) > 90\%$; none of these was originally assigned to *Capra* (Twiss, in press: Table 2).

By increasing the frequency of reliable sheep vs. goat identifications, we gain interpretive power for derived analyses. Simulated assemblages derived from the posterior probabilities $P(\text{goat}|\text{traits})$ makes this possible. Table 4 shows sheep and goat parameters for both ‘Ain Dara and Tell Sakhariya assemblages compared to the results of the original analyses of these specimens. The probabilistic model suggests that there was a somewhat higher proportion of sheep than originally estimated at ‘Ain Dara and goats at Tell Sakhariya. Furthermore, our approach allows us to make probabilistic statements about the balance of sheep and goats at these sites: 85% of the simulated ‘Ain Dara assemblages had more sheep than goats, whereas >99% of the Tell Sakhariya simulations had more sheep than goats.

These simulations can also be used to compare derived traits of sheep and goats at each site. Fig. 2 shows body part representation of sheep and goats for both sites with 89% HPDI values. While the same procedure could be used to analyze other aspects of sheep and goat husbandry, such as biometrics or survivorship, this was outside the purview of this study. For ‘Ain Dara, most elements are better-represented by sheep, only the first phalanx is better represented by goats. Species-specific counts of both the distal tibia and second phalanx are too uncertain to tell whether sheep or goats are better represented, complementing the results of the cross-validation analysis. In the Tell Sakhariya assemblage, sheep are more numerous for every element; for every element except the first phalanx, the 89% HPDI intervals include 0, which suggests that there may be in fact no goat specimens for these elements.

4.3. Testing hypotheses

The probabilistic approach provides a straightforward way to test hypotheses about sheep/goat management. Richard Redding (1981, 1984) calculated expected models of sheep:goat ratios for flocks with different herding strategies and goals. Redding (1984) posited that sheep:goat ratios of 5:1 (~17% goat) are expected when herders aim to maximize meat production, whereas herders focused on risk minimization are expected to have sheep:goat ratios of 1.7:1 (~37% goat). Fig. 3 shows posterior distributions of the percentage of goats at both sites; the proportion of the distribution that exceeds the values posited by Redding (1984) has been filled in and translated into probabilities that the site’s sheep:goat ratio exceeds these limits. For ‘Ain Dara, 62% of the simulations had sheep:goat ratios ≤ 1.7 and 1% of simulations had a sheep:goat ratio $\geq 5:1$, strongly suggesting that the herds supplying ‘Ain Dara were managed for herd security. In contrast, for Tell Sakhariya, 71% of the simulations had sheep:goat ratios $\geq 5:1$, whereas sheep:goat ratios $\leq 1.7:1$ were only found in 3% of the simulations. These calculations, in the context of Redding’s models, suggest that it is highly likely that the Tell Sakhariya herds were managed to maximize meat production. While these initial comparisons are simplistic and speculative, they demonstrate the ability to directly relate aspects of herd management to parameters derived from this method.

5. Discussion and conclusion

This paper demonstrates the utility of probabilistically distinguishing between sheep and goat bones. Shifting the focus of zooarchaeological identifications from an absolutist paradigm that condenses all diagnostic information into a single assignment towards a probabilistic paradigm emphasizes the reporting of more replicable data (individual traits). Moving forward, then, it is imperative to investigate the potentially biasing role of inter-analyst variability on the identification of specific traits. Despite both of us having several years’ experience using the Zeder and Lapham (2010) traits and reviewing them together on reference specimens, one author listed a trait as “goat-like” and the other listed it as “sheep-like” 15% (64/434) of the time. Differences in analysts’ skill-levels/experience are a well-known

Table 3
Misidentification rates for cross-validated linear discriminant analysis (this study) and expert identifications (Zeder and Lapham (2010): Table 14).

Element	Average TAU Inaccuracy %	Average expert Projected % misidentified
Distal humerus	2	2
Proximal radius	6	5
Distal radius	5	5
Distal tibia	14	24
Distal metacarpal	1	5
Distal metatarsal	2	2
Astragalus	4	5
Calcaneus	5	5
First phalanx	2	5
Second phalanx	14	5
Total	6	6
Total (minus distal tibia)	5	4

Table 4

Summary data from probabilistic sheep/goat identifications for 'Ain Dara and Tell Sakhariya. Original identifications from Frey and Marean (1999) and Twiss (in press), specimen-level data in Supplementary Tables 1 and 2.

Site	Sheep NISP			Goat NISP			Proportion goat		
	Original	Median	89% HPDI	Original	Median	89% HPDI	Original	Median	89% HPDI
'Ain Dara (N = 86)	38	52	38–63	28	34	21–46	42%	40%	24–53%
Tell Sakhariya (N = 55)	38	47	42–54	0	8	1–13	0%	15%	2–24%

source of variation in zooarchaeological identification (e.g., Zeder and Lapham, 2010: 2898); there is a real need for more object means of identifying traits remains that go beyond subjective determinations of traits as being more “goat-like” vs. “sheep-like.” A geometric morphometrics approach, such as that recently proposed by Haruda (in press), is one possible solution. Importantly, geometric morphometric or traditional metric information can be included in the probabilistic framework presented here. This method can thus be applied in field identifications, unlike geometric morphometric analyses (Haruda, in press: 3).

The probabilistic approach helps highlight biases in taxonomic assignments. For example, our analyses relied on a vague prior – i.e., we made few assumptions about trait identifications – in order to be maximally conservative (Lambert et al., 2005). Use of a vague prior may

have resulted in overestimating the goats at Tell Sakhariya. Using a more precise prior distribution based on other information might result in the identification of Tell Sakhariya's assemblage as containing sheep exclusively. The benefit of the probabilistic approach adopted here, and Bayesian statistics more generally, is that it forces researchers to state their assumptions in order to derive results (McElreath, 2016a). Such a methodology improves the overall transparency of zooarchaeological identification.

While the approach does require analysts to include additional steps in the field, the work presented here can easily be incorporated into existing workflows. Zooarchaeologists generally have relevant reference manuals available when recording remains. To incorporate a probabilistic approach, all an analyst must do is record the traits of a specimen when they examine it; these traits are typically noted by the analyst to make a traditional decision of “sheep” vs “goat” for the specimen anyway. For previously-analyzed assemblages like the two sites analyzed here, we collected the relevant specimens and recorded their traits over the course of two days. Since zooarchaeologists are already trained to examine these traits, we feel that the benefits of probabilistic identifications across a wide range of projects can be achieved by the minor change to workflows: writing down and presenting traits for each specimen.

One source of potential error that we were unable to assess was variation in diagnostic traits across geographic, temporal, and environmental contexts. Our model assumes that our archaeological assemblages have the same distribution of diagnostic traits as our reference population. This problem, known as “reference mimicry,” exists in faunal analysis more broadly (Millard, 2006). Do traits or combinations of traits vary across populations in time and space? Further research on the distribution of diagnostic traits is necessary to establish better reference populations (e.g., Haruda, in press; Fernandez, 2001). Studies of other diagnostic criteria for sheep and goats – particularly those not included in Zeder and Lapham's (2010) study – should include statements on the inter- and intra-observer reliability of these traits along with their ability to differentiate the two taxa (Zedda et al., 2016); ideally, these studies should also include specimen-level suites of traits to act as reference populations. Specialized chemical techniques, like Zooarchaeology by Mass Spectrometry (ZooMS) or bulk bone DNA metabarcoding, may also be useful for establishing the distribution of diagnostic criteria in zooarchaeological assemblages (Buckley et al., 2010; Murray et al., 2013).

A probabilistic approach allows for more replicable and comparable datasets in zooarchaeology. As the discipline grows in scope and ambition, it is vital that we adjust our methodological standards in order to capture the uncertainty that is inherent in specimen identification. The recent increased interest in regional-scale “meta-data” analyses highlights this need (e.g., Arbuckle et al., 2014; Conolly et al., 2011; Manning et al., 2015). These analyses assume that analyses performed by different analysts have reproducible results; that is, it does not matter which analysts recorded a site because all analysts would have reached the same conclusion. However, even if primary data is shared (Atici et al., 2013), identifications are still based on implicit decisions by different analysts to prioritize certain diagnostic criteria. To improve the reliability of these types of research agendas, it is paramount to develop methods that encourage valid taxonomic assignments across multiple analysts and projects.

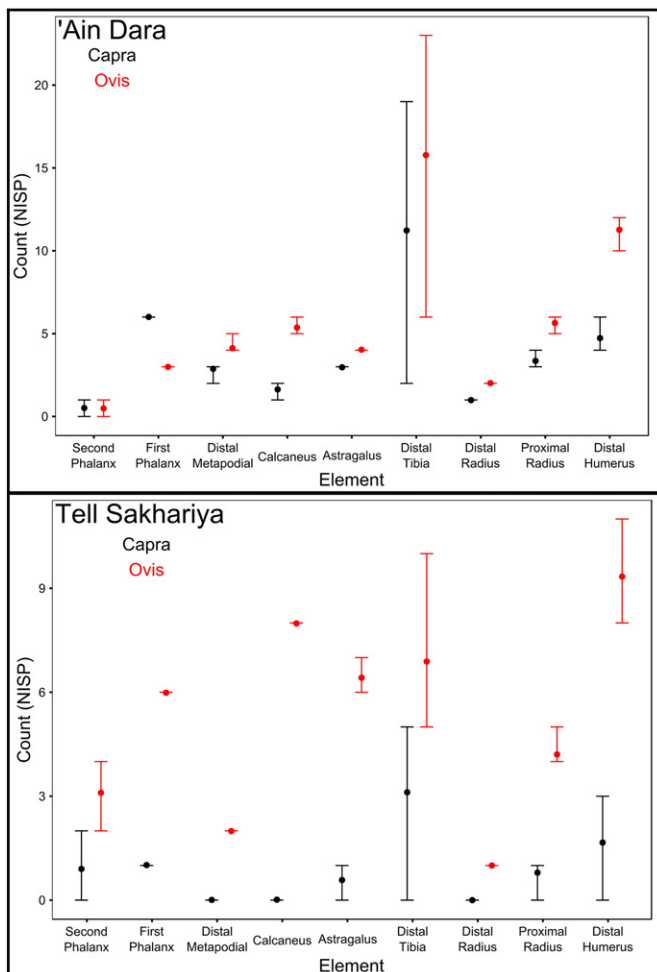


Fig. 2. Probabilistic body part representation at 'Ain Dara (top) and Tell Sakhariya (bottom). Points show the mean counts of different elements, error bars reflect 89% HPDI (McElreath, 2016b). Sheep are the right-hand value for each element.

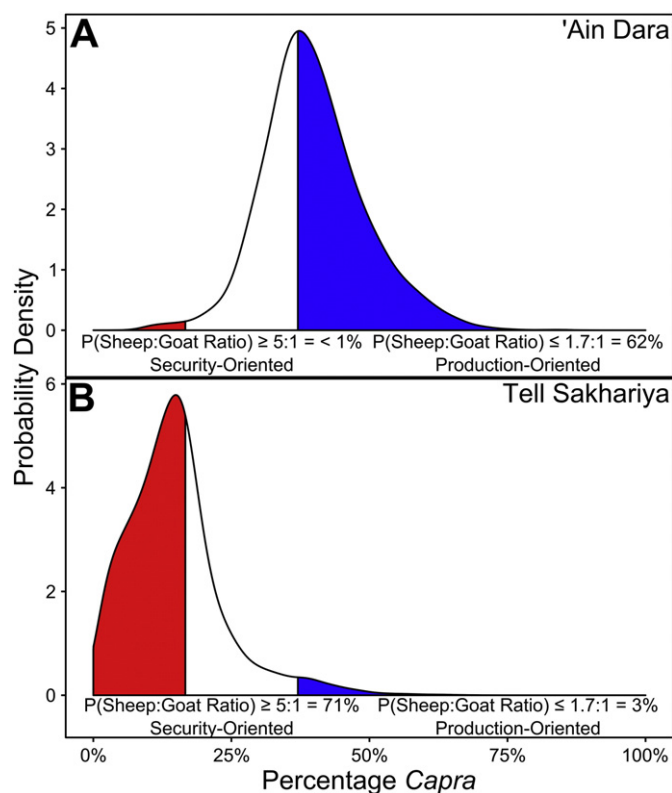


Fig. 3. Distributions of percentage *Capra* at (A) 'Ain Dara and (B) Tell Sakhariya. Shaded regions reflect probabilities that sheep:goat ratios exceed limits recognized by Redding (1984).

Acknowledgements

We wish to thank Drs. Elizabeth Stone and Paul Zimansky for access to the 'Ain Dara and Tell Sakhariya faunal material. Thank you as well to Dr. Katheryn Twiss for laboratory space, original data files for both assemblages, access to her in press faunal report, and critical advice on the manuscript. Thank you to Erik Otárola-Castillo for methodological advice early in this project's conception. Comments by two anonymous reviewers greatly improved the flow of the article's arguments and the utility of the R script. Any errors and misrepresentations are our own.

Appendix A. Supplementary data

CSV data files for 'Ain Dara and Tell Sakhariya sheep-goat trait IDs and the R script for structuring the Zeder and Lapham (2010) reference data (in CSV format) and analyses and visualizations are available in linked Mendeley Data files, at <http://dx.doi.org/10.17632/sz7jifg6fc.1>. Supplementary results tables for both 'Ain Dara and Tell Sakhariya can be found in the online version, at <http://dx.doi.org/10.1016/j.jasrep.2017.02.022>.

References

Arbuckle, B.S., Kansa, S.W., Kansa, E.C., Orton, D., Cakirlar, C., Gourichon, L., Atici, A.L., Galik, A., Marciniak, A., Mulville, J., Buitenhuis, H., Carruthers, D., De Cupere, B., Demiregi, A., Frame, S., Helmer, D., Martin, L., Peters, J., Pöllath, N., Pawlowska, K., Russell, N., Twiss, K.C., Württemberg, D., 2014. Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS One* 9, 1–11.

Atici, L., Kansa, S.W., Lev-Tov, J., Kansa, E.C., 2013. Other people's data: a demonstration of the imperative of publishing primary data. *J. Archaeol. Method Theory* 20, 663–681.

Baxter, M., 2003. *Statistics in Archaeology*. Oxford University Press, New York.

Boessneck, J., 1969. Osteological differences between sheep (*Ovis aries* Linné) and goat (*Capra hircus* Linné). In: Brothwell, D.R., Higgs, E.S. (Eds.), *Science in Archaeology: a Survey of Progress and Research*. Thames and Hudson, London, pp. 331–358.

Bréhard, S., Radu, V., Martin, A., Hanot, P., Popovici, D., Bălăşescu, A., 2014. Food supply strategies in the Romanian eneolithic: sheep/goat husbandry and fishing activities from Hârşova tell and Borduşani-Popină (5th millennium BC). *Eur. J. Archaeol.* 17, 407–433.

Buckley, M., Whitcher Kansa, S., Howard, S., Campbell, S., Thomas-Oates, J., Collins, M., 2010. Distinguishing between archaeological sheep and goat bones using a single collagen peptide. *J. Archaeol. Sci.* 37, 13–20.

Buitenhuis, H., 1995. A quantitative approach to species determination in ovicapridae. In: Buitenhuis, H., Uerpmann, H.-P. (Eds.), *Archaeozoology of the Near East II: Proceedings of the Second Symposium on the Archaeozoology of Southwestern Asia and Adjacent Areas*. Backhuys Publishers, Leiden, pp. 140–155.

Conolly, J., Colledge, S., Dobney, K., Vigne, J.-D., Peters, J., Stopp, B., Manning, K., Shennan, S., 2011. Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. *J. Archaeol. Sci.* 38, 538–545.

Core Team, R., 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Crema, E.R., 2011. Modelling temporal uncertainty in archaeological analysis. *J. Archaeol. Method Theory* 19, 440–461.

Driver, J.C., 2011. *Identification, Classification and Zooarchaeology*. Society of Ethnobiology 2, 19–39.

Evin, A., Cucchi, T., Cardini, A., Strand Vidarsdottir, U., Larson, G., Dobney, K., 2013. The long and winding road: identifying pig domestication through molar size and shape. *J. Archaeol. Sci.* 40, 735–743.

Fernandez, H., 2001. *Ostéologie comparée des petits ruminants eurasiatiques sauvages et domestiques (genres *Rupicapra*, *Ovis*, *Capra* et *Capreolus*): diagnose différentielle du squelette appendiculaire*. Zoologie et Biologie Animale, Université de Genève, Geneva, p. 441.

Fish, P.R., 1978. Consistency in archaeological measurement and classification: a pilot study. *Am. Antiq.* 43, 86–89.

Frey, C.J., Marean, C.W., 1999. Mammal remains. In: Stone, E.C., Zimansky, P.E. (Eds.), *The Iron Age Settlement of 'Ain Dara, Syria: Survey and Surroundings*. Archaeopress, Oxford, pp. 123–137.

Gobalet, K.W., 2001. A critique of faunal analysis; inconsistency among experts in blind tests. *J. Archaeol. Sci.* 28, 377–386.

Haruda, A.F., 2016. 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. *Int. J. Osteoarchaeol.* <http://dx.doi.org/10.1002/oa.2576> (in press).

Kovarovic, K., Aiello, L.C., Cardini, A., Lockwood, C.A., 2011. Discriminant function analyses in archaeology: are classification rates too good to be true? *J. Archaeol. Sci.* 38, 3006–3018.

Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., Jones, D.R., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat. Med.* 24, 2401–2428.

Legendre, P., Legendre, L., 1998. *Numerical Ecology*. Second English ed. Elsevier, New York.

Lunn, D.J., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2013. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC.

Manning, K., Timpson, A., Shennan, S., Crema, E., 2015. Size reduction in early European domestic cattle relates to intensification of Neolithic herding strategies. *PLoS One* 10, e0141873.

Marwick, B., 2016. Computational reproducibility in archaeological research: basic principles and a case study of their implementation. *J. Archaeol. Method Theory*.

McElreath, Richard L., 2016a. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC, London.

McElreath, Richard L., 2016b. *Rethinking: Statistical Rethinking Book Package*, R Package Version 1.59.

Millard, A.R., 2006. A Bayesian approach to ageing sheep/goats from toothwear. In: Ruscillo, D. (Ed.), *Recent Advances in Ageing and Sexing Animal Bones*. Oxbow Books, Oxford, pp. 145–154.

Morin, E., Ready, E., Boileau, A., Beauval, C., Coumont, M.-P., 2016. Problems of identification and quantification in archaeozoological analysis, part I: insights from a blind test. *J. Archaeol. Method Theory*.

Murray, D.C., Haile, J., Dortch, J., White, N.E., Haouchar, D., Bellgard, M.I., Allcock, R.J., Prideaux, G.J., Bunce, M., 2013. Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Sci. Report* 3, 1–8.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2016. *vegan: Community Ecology Package*, R Package Version 2.4-1.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.

Payne, S., 1969. A metrical distinction between sheep and goat metacarpals. In: Ucko, P.J., Dingleby, G.W. (Eds.), *The Domestication and Exploitation of Plants and Animals*. Aldine Publishing Company, Chicago, pp. 295–305.

Payne, S., 1985. Morphological distinctions between the mandibular teeth of young sheep, *Ovis*, and goats, *Capra*. *J. Archaeol. Sci.* 12, 139–147.

Prummel, W., Frisch, H.-J., 1986. A guide for the distinction of species, sex and body side in bones of sheep and goat. *J. Archaeol. Sci.* 13, 567–577.

Quinn, G.P., Keough, M.J., 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.

Redding, R.W., 1981. *Decision Making in Subsistence Herding of Sheep and Goats in the Middle East*. Anthropology. The University of Michigan, Ann Arbor, MI, p. 422.

Redding, R.W., 1984. Theoretical determinants of a herder's decisions: modeling variation in the sheep/goat ratio. In: Clutton-Brock, J., Grigson, C. (Eds.), *Animals and Archaeology*. British Archaeological Reports, Oxford, pp. 223–241.

- Shott, M.J., 2010. Size dependence in assemblage measures: essentialism, materialism, and “SHE” analysis in archaeology. *Am. Antiq.* 75, 886–906.
- Smith, A.F.M., Gelfand, A.E., 1992. Bayesian statistics without tears: a sampling-resampling perspective. *Am. Stat.* 46, 84–88.
- Stone, E.C., Zimansky, P.E., 1999. The Iron Age Settlement at ‘Ain Dara, Syria: Survey and Soundings. British Archaeological Reports, Oxford.
- Twiss, K.C., 2017. Animals of the Sealands: Ceremonial Activities in the Southern Mesopotamian “Dark Age”, Iraq (in press).
- Twiss, K.C., Wolfhagen, J., Madgwick, R., Foster, H., Demiregi, G.A., Russell, N., Everhart, J.L., Pearson, J., Mulville, J., 2016. Horses, hemiones, hydruntines? Assessing the reliability of dental criteria for assigning species to southwest Asian equid remains. *Int. J. Osteoarchaeol.* (n/a–n/a).
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics With S*. fourth ed. Springer, New York.
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Wolverton, S., 2012. Data quality in zooarchaeological faunal identification. *J. Archaeol. Method Theory*.
- Zedda, M., Palombo, M.R., Brits, D., Carcupino, M., Sathé, V., Cacchioli, A., Farina, V., 2016. Differences in femoral morphology between sheep (*Ovis aries*) and goat (*Capra hircus*): macroscopic and microscopic observations. *Zoomorphology*.
- Zeder, M.A., Lapham, H.A., 2010. Assessing the reliability of criteria used to identify postcranial bones in sheep, Ovis, and goats, Capra. *J. Archaeol. Sci.* 37, 2887–2905.
- Zimansky, P., Stone, E.C., 2016. Tell Sakhariya and Gaes. In: Kaelin, O., Stucky, R.A., Mathys, H.-P. (Eds.), *Proceedings of the 9th International Congress on the Archaeology of the Ancient Near East: June 9–13, 2014*. University of Basel, Harrassowitz, Basel, pp. 57–66.