Estimating the Magnitude of Data Asymmetry in Palaeozoological Biogeography

R. L. LYMAN*
Department of Anthropology, 107 Swallow Hall, University of Missouri-Columbia, Columbia, MO 65211 USA

ABSTRACT Palaeozoological data are asymmetrical because they indicate the presence of a species in the area in which its remains are recovered, but the absence of remains is not necessarily evidence for the absence of a species. It is impossible to measure the magnitude of data asymmetry with respect to biogeography, but the magnitude can be estimated. The proportion of sites in an area where a species is known to occur that have not produced remains of that species is an estimate of counter-ubiquity. The proportion of identified faunal remains in an area where a species is known to occur that do not represent that species is another estimate of data asymmetry. Bivariate plots of the number of sites in an area that have produced remains of a taxon against the total number of faunal remains identified in the area indicate that data asymmetry can be estimated as the inverse of sample size. More sites and more identified faunal remains tend to produce more occurrences of a species, so more sites and more identified faunal remains will tend to provide a database that is not only more accurate but also less asymmetrical with respect to the geographical range of a species.

Key words: biogeography; data asymmetry; palaeozoology; presence/absence data

Introduction

Twenty-five years ago, Donald Grayson (1981) argued that the palaeozoological record provided asymmetrical data on the presence/absence of taxa. In his view palaeozoological data are asymmetrical because they only provide evidence regarding the presence of taxa, evidence in the form of preserved, usually skeletal, remains such as bones, teeth, shells and the like. Data on the absence of taxa are not provided by the palaeozoological record for several reasons, including the possibilities that the remains of a taxon were not preserved, or they were not collected as a result of a poor sampling strategy, or they were not recovered because of inappropriate collection techniques such as use of a screen mesh too coarse to retain remains of a taxon (see Pluskowski, 2006 for other possibilities).

It is difficult to fault Grayson’s (1981) argument, and probably for this reason it has been repeated several times over the years (Lyman, 1995a; Ervynck, 1999; Etnier, 2002). The potential magnitude of data asymmetry has, however, not been measured or estimated. As indicated in the preceding paragraph, the variables creating the asymmetry are several. It is probably because those variables are for the most part not measurable in palaeozoological contexts that the magnitude of asymmetry has not been measured. By measurement I mean a quantitative value that is assigned based on observation and rules of assignment; if a measurement cannot be taken (as a result of,
say, inability to observe), a phenomenon is not measurable. The magnitude of asymmetry might be estimated, however, if it is granted that an estimate is a measurement (description) assigned to a phenomenon based on incomplete observation or incomplete data. In this paper I discuss methods for estimating the magnitude of data asymmetry.

One might argue that sampling intensity could be measured in terms of the fraction of a site that was excavated, but that requires intimate knowledge of the horizontal and vertical boundaries of the sediments being sampled, knowledge that often is imperfect at best. In addition, recovery and preservation are rather more difficult to quantify, and require knowledge that we simply do not have access to as palaeozoologists. Knowing the amount of faunal remains that was not recovered requires intimate knowledge of what was available to be recovered; this is a manifestation of the sampling paradox, where the reason we are recovering materials is to find out what kinds of materials and how much of each kind are there. Efforts to measure the amount of preservational skewing require detailed knowledge of what was accumulated and deposited but what did not preserve: a rather tricky analytical endeavour that palaeozoologists and taphonomists have not yet perfected (Lyman, 1994a). Preservation can, however, sometimes be estimated.

Given that the asymmetry of biogeographical data is difficult (if not impossible) to measure, the question of whether or not asymmetry can be estimated arises. I suspect we can get within an order of magnitude or so of that degree of asymmetry via estimation. The process of estimation incorporates aspects of sampling intensity and general preservation and recovery. In this paper I use a small, osteologically distinctive mammal with a relatively limited modern geographical distribution and its palaeozoological record to illustrate techniques of estimating the degree or magnitude of the asymmetry of palaeobiogeographical data. The resulting estimates allow assessment of just how asymmetrical our knowledge might be with respect to the prehistoric distribution of the exemplary species.

Grayson (1981) was concerned with the asymmetry of data at the scale of a collection or assemblage of faunal remains recovered from a single site or from a single stratum in a site. This is the least inclusive geographical scale. A scale of greater inclusiveness concerns the geographical range of a taxon where a range consists of a bounded area, the taxon is present within that bounded area but absent from areas outside that area (Gaston, 1991). Determination of ranges, or the history of stability and fluctuation in a taxon’s range (e.g. Vojta, 2005) could become an increasingly important but slippery analytical task that requires having a more developed conception of the asymmetry of palaeobiogeographical data. This is so because palaeozoologists and others increasingly use those data in the service of conservation biology and wildlife management (e.g. Lyman, 1998, 2006; Barnosky et al., 2005; Grayson & Delpech, 2005).

Methods and materials

I chose the mountain beaver (Aplodontia rufa) as the taxon with which to illustrate techniques for estimating the magnitude of the asymmetry of palaeobiogeographical data. I chose it in part because of a recent report on a biogeographically unique, prehistoric extralimital record (Wake, 2006) which helps illustrate the point I wish to make here. I also chose this taxon in part because of its relatively limited modern geographical range and the amount of palaeozoological research that has been done within that range.

Mountain beaver belong to the monotypic family Aplodontidae. The taxonomic family name derives from the Greek words for ‘simple tooth’. The common name often used, mountain beaver, reflects the fact that they occasionally gnaw bark and small limbs from trees. Otherwise, this rodent is thought not to be related to the common beaver (Castoridae, Castor canadensis), but rather to be an evolutionarily primitive rodent most closely related to squirrels (Sciuridae) (Maser, 1998). The mountain beaver today is found only in southwestern British Columbia, western Washington state, western Oregon state, northwestern California state and a portion of eastern California (Carraway & Verts, 1993) (Figure 1). Within its modern range, local residents sometimes refer to the animal as a
'boomer' because of one of its vocalisations (Maser, 1998: 127).

Mountain beavers are similar in size and appearance to a medium-sized muskrat (*Ondatra zibethicus*) except that the tail of a mountain beaver is very short (20–40 mm) and well furred. Mean body mass of adults is 800 g and body length is 300–450 mm (Carraway & Verts, 1993). Their bones and teeth are of sufficient size to be regularly caught in the archaeological industry standard of quarter-inch (0.625 cm) mesh screens (Shaffer, 1992). Intersite variance in recovery should not create intersite variance in data asymmetry. Mountain beavers generally live in coniferous forests and riparian habitats. They are fossorial, and their burrow systems produce entrances approximately 20 cm in diameter. Burrow entrances are surrounded by loose sediment, rocks and other debris and are quite noticeable. Burrows are often just beneath the ground surface. Mountain beavers are herbivorous and although primarily active nocturnally, they are often active diurnally as well (Maser, 1998; Verts & Carraway, 1998).

Various groups of American Indians that lived in the modern range of mountain beavers exploited them (Suckley, 1860; Ray, 1938; Elmendorf, 1960). Archaeological evidence indicates that this small mammal was exploited by at least some human groups (Lyman & Zehr, 2003). Thus, there is some potential that remains of mountain beaver were anthropogenically accumulated and deposited in prehistoric sites. Major non-human predators of mountain beaver include bobcat (*Felis rufus*), coyote (*Canis latrans*) and eagles (Carraway & Verts, 1993). This means that some clusters of remains of mountain beaver may have been accumulated and deposited by non-human agents.

I examined much of the literature concerning palaeozoological research in areas within or near the modern range of mountain beaver. The latter was included because two prehistoric records of this taxon are in locations that are outside of, or extralimital to, the modern range of mountain beaver. I did not perform an exhaustive search of the palaeozoology literature, particularly for California, as that is unnecessary to illustrate the techniques for assessing the asymmetry of palaeozoological data. The number of identified specimens (NISP) of mammals identified to at least taxonomic family for each site or locale was recorded. The NISP of mountain beaver, if any, per site or, in a few instances, per several sites in the same locale was recorded (Table 1). Sites that had been sampled but from which no faunal remains were recovered or the recovered faunal remains had not been studied and reported were excluded. Including such sites would have markedly increased the apparent magnitude of data asymmetry. The location of each site (or

Figure 1. Modern range (shaded) of mountain beaver (*Aplodontia rufa*) and locations of archaeological sites (numbers) that have produced identified faunal remains. See Table 1 for site-specific data.
locale) that had produced identified faunal remains was marked on a map (Figure 1). Most mammalian remains discussed below date to the last 4000–5000 years; many of them date to the last 3000 years, reducing the probability that shifts in the biogeographical range of mountain beaver are represented in the studied materials.

Grayson’s (1981) key point was that the absence of remains of a taxon was not necessarily evidence that the taxon was indeed absent from the geographical area of concern, usually a limited catchment area around the site that produced the faunal remains. To determine the potential degree of asymmetry, we must assume that if one excavates and recovers remains of animals from a site within the modern range of a taxon, remains of that taxon will have a very good chance of being recovered. This in turn assumes

### Table 1. Palaeozoological data in or near the modern range of mountain beavers

<table>
<thead>
<tr>
<th>Site</th>
<th>Mammalian NISP</th>
<th>Aplodontia NISP</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glenrose (1)</td>
<td>327</td>
<td>0</td>
<td>Imamoto (1976)</td>
</tr>
<tr>
<td>Belcarra Park, Noons Creek (2)</td>
<td>1692</td>
<td>0</td>
<td>Galdikas-Brindamour (1972)</td>
</tr>
<tr>
<td>Lillooet (3)</td>
<td>1339</td>
<td>0</td>
<td>Langemann (1987)</td>
</tr>
<tr>
<td>45WH34 (4)</td>
<td>743</td>
<td>0</td>
<td>Nokes (2004)</td>
</tr>
<tr>
<td>45KP2 (5)</td>
<td>121</td>
<td>2</td>
<td>Snyder (1956)</td>
</tr>
<tr>
<td>45CA227 (6)</td>
<td>4580</td>
<td>0</td>
<td>Blukis Onat &amp; Larson (1984)</td>
</tr>
<tr>
<td>45CA426 (7)</td>
<td>495</td>
<td>0</td>
<td>Lyman (1999)</td>
</tr>
<tr>
<td>45CA21 (8)</td>
<td>3266</td>
<td>0</td>
<td>Wigen &amp; Stucki (1988)</td>
</tr>
<tr>
<td>45CA24 (9)</td>
<td>52,938</td>
<td>0</td>
<td>Huelsbeck (1994)</td>
</tr>
<tr>
<td>45DO372 (12)</td>
<td>538</td>
<td>2</td>
<td>Moura (1986)</td>
</tr>
<tr>
<td>45KI428 (13)</td>
<td>667</td>
<td>90</td>
<td>Lyman (1995b)</td>
</tr>
<tr>
<td>45KI23 (14)</td>
<td>721</td>
<td>1 skeleton</td>
<td>Lyman (1981); Livingston (1987)</td>
</tr>
<tr>
<td>45LE266 (15)</td>
<td>9</td>
<td>2</td>
<td>Miss (1989)</td>
</tr>
<tr>
<td>45LE410 (16)</td>
<td>39</td>
<td>22</td>
<td>Lyman (1994c)</td>
</tr>
<tr>
<td>45LE222 (17)</td>
<td>802</td>
<td>8</td>
<td>Gustafson (1987a)</td>
</tr>
<tr>
<td>45LE223 (18)</td>
<td>214</td>
<td>13</td>
<td>Gustafson (1987b)</td>
</tr>
<tr>
<td>45LE263 (19)</td>
<td>348</td>
<td>9</td>
<td>Lyman (1994b)</td>
</tr>
<tr>
<td>45LE209 (20)</td>
<td>171</td>
<td>96</td>
<td>Canaday (1991)</td>
</tr>
<tr>
<td>Seaside (21)</td>
<td>2013</td>
<td>0</td>
<td>Colten (2002); Connolly (1992)</td>
</tr>
<tr>
<td>35CLT21 (22)</td>
<td>39</td>
<td>1</td>
<td>Minor (1991)</td>
</tr>
<tr>
<td>35LNC60 (24)</td>
<td>452</td>
<td>0</td>
<td>Lyman (1991)</td>
</tr>
<tr>
<td>35LNC14 (25)</td>
<td>1826</td>
<td>0</td>
<td>Lyman (1991)</td>
</tr>
<tr>
<td>35DO83 (26)</td>
<td>2983</td>
<td>5</td>
<td>Lyman (1991)</td>
</tr>
<tr>
<td>35DO147 (27)</td>
<td>144</td>
<td>0</td>
<td>O’Neill (1989)</td>
</tr>
<tr>
<td>35CS43 (28)</td>
<td>261</td>
<td>2</td>
<td>Lindsay &amp; Keith (1986)</td>
</tr>
<tr>
<td>35LIN11 (29)</td>
<td>262</td>
<td>0</td>
<td>Murray (1966)</td>
</tr>
<tr>
<td>35LIN36 (30)</td>
<td>52</td>
<td>0</td>
<td>Sanford (1975)</td>
</tr>
<tr>
<td>35LA53 (31)</td>
<td>27</td>
<td>0</td>
<td>Olsen (1975)</td>
</tr>
<tr>
<td>35LA39 (32)</td>
<td>151</td>
<td>0</td>
<td>Baxter et al. (1983)</td>
</tr>
<tr>
<td>Elk Creek (33)</td>
<td>741</td>
<td>0</td>
<td>Budy et al. (1986); Lyman (1987); Macdougall (1991)</td>
</tr>
<tr>
<td>35JA42 (34)</td>
<td>151</td>
<td>0</td>
<td>Schmitt (1986)</td>
</tr>
<tr>
<td>CA-SK4 (35)</td>
<td>1383</td>
<td>0</td>
<td>Grayson (1973)</td>
</tr>
<tr>
<td>Iron Gate (36)</td>
<td>200+</td>
<td>0</td>
<td>Leonhardy (1967)</td>
</tr>
<tr>
<td>Surprise Valley (37)</td>
<td>1457</td>
<td>0</td>
<td>James (1983)</td>
</tr>
<tr>
<td>Potter Creek Cave (38)</td>
<td>?</td>
<td>Present</td>
<td>Sinclair (1904)</td>
</tr>
<tr>
<td>Samwell Cave (39)</td>
<td>?</td>
<td>Present</td>
<td>Furlong (1906)</td>
</tr>
<tr>
<td>CA-SON348/H (40)</td>
<td>1309</td>
<td>9</td>
<td>Wake (2006)</td>
</tr>
<tr>
<td>CA-ALA309 (41)</td>
<td>11,605</td>
<td>0</td>
<td>Broughton (1999)</td>
</tr>
<tr>
<td>CA-ALP212/223 (42)</td>
<td>109</td>
<td>0</td>
<td>Schmitt (1992)</td>
</tr>
</tbody>
</table>

Number in parenthesis in ‘Site’ column corresponds to location in Figure 1. British Columbia, 1–3; Washington, 4–20; Oregon, 21–34; California, 35–42.
that there was no major prehistoric change in the range of the taxon. Available evidence suggests this is the case for the mountain beaver (see below).

Given how most faunal remains described below were recovered, recovery does not seem to be an issue as screen mesh was typically quarter-inch (0.625 cm) mesh. Preservation is more difficult to ascertain, but overall preservation was good at virtually all sites included in the data set. Therefore, the magnitude of asymmetry is here estimated in two ways. Firstly, the proportion of sites located within the modern range of a taxon, and which did not produce remains of the taxon of concern, serves as one estimate. This is the complement or counterpart of a ubiquity measure, within the range of a taxon, one would expect ubiquity to be very close to 1.0, say > 0.9, but outside the range the estimate should be very low, say < 0.1, rather than exactly 0.0, to allow for prehistoric shifts in range of the taxon. The estimate comprising the proportion of sites without remains of the taxon can be considered a counter-ubiquity estimate.

The second estimate is the proportion of the total number of identified specimens (\( \sum \) NISP) that does not represent the taxon of interest. This estimate is a bit more difficult to interpret than the counter-ubiquity estimate because it is at least in part a function of the availability of the taxon on the landscape, and that availability may vary tremendously across the range of the taxon. For example, the easternmost site in California (site 42), while within the modern range of mountain beaver, is in a low-elevation xerix habitat in which mountain beavers are not found. This small mammal has probably never been locally available near site 42. Nevertheless, the proportion of \( \sum \) NISP not representing the taxon provides another measure of the general magnitude of asymmetry. It will provide, for example, an estimate of the sampling intensity (measured by NISP) requisite to finding remains of the taxon of concern (Lyman, 1995a).

Finally, the two estimates can be combined in a single bivariate graph to illustrate the magnitude of ubiquity (Styles, 1981: 43), and thus by implication the degree of counter-ubiquity. To produce the graph and to avoid multiple zero points being plotted, all collections from each of the four politically defined geographical units included here are summed. That is, all materials from British Columbia are summed, all materials from Washington state are summed, and so on. The total NISP per political province plotted against the number of sites per political province that have produced remains of mountain beaver provides a graph that reveals much about ubiquity (and counter-ubiquity) and its relationship to sample size or sampling intensity.

Results

Data for 42 sites and locales were collected (Table 1, Figure 1). Of these, 33 are located within what is generally taken to be modern mountain beaver range and nine are outside of that range (Carraway & Verts, 1993). Nineteen sites and locales produced remains of mountain beaver; two of the sites that have produced remains of mountain beaver are outside of the modern range of that species (Figure 1).

Available data indicate that the range of the mountain beaver has been relatively stable, but obviously not completely static, for much of the past several thousand years; that range is apparently smaller now than previously. Considering only those collections from within the modern mountain beaver range, 17 of 33 sites produced remains of mountain beaver. This means that the counter-ubiquity index is 16/33 or 0.48; just less than half of the sites within the modern range of mountain beaver have not produced remains of mountain beaver.

Considering only those collections from within the modern mountain beaver range, a total NISP of 87,556 has been identified. Of those, only 418 (0.5%) have been identified as representing mountain beaver. At 0.995, the proportion of \( \sum \) NISP not representing mountain beaver is very high. Not only did many collections that probably should have produced remains of mountain beaver not produce such remains, but many remains that perhaps should have represented mountain beaver do not. The proportion of NISP not representing mountain beaver is, as noted earlier, difficult to interpret because it depends on a host of variables, the values of which are unknown and unknowable, in
particular the availability of mountain beaver on the landscape and the behaviours of faunal-remains accumulating agents. That mountain beaver could be relatively abundant in a collection is indicated by assemblages such as those from 45KI428 (13.5% of all NISP) and 45LE209 (56.1% of all NISP). That mountain beaver is rarely represented overall suggests that it was seldom accumulated, or that it was generally rare on the landscape, or both.

The total NISP per political province was log-transformed to graph the bivariate relationship between \( \sum \text{NISP} \) and the number of sites per political province that produced remains of mountain beaver. That graph (Figure 2) suggests a long-known fact (Grayson, 1984). As sample size (measured as NISP) increases, the number of sites that produce remains of a taxon also increases (Pearson's \( r = 0.954, p = 0.046 \)). Thus, counter-ubiquity as reflected by Figure 2 reduces to a function of sample size, a fact that seems intuitively obvious in hindsight. As more sites are excavated, NISP increases; as NISP increases, the probability of finding remains of mountain beaver increases.

**Discussion and conclusion**

Were a palaeozoologist trying to generate a prehistoric range map for mountain beaver including the last several thousand years, that map would be incomplete relative to the known modern range of the taxon. That map would largely reflect where within that range excavations have taken place and which excavations had produced faunal remains that had been identified and reported (Figure 1); the problem would be how to interpolate between geographical points of occurrence, a problem common to modern efforts to establish ranges of taxa (Graham & Semken, 1987). This fact, too, seems intuitively obvious yet it has, to my knowledge, seldom been explicitly acknowledged (see Lyman, 2004, for an example). It is in a way a detailed restatement of Grayson's (1981) original observation regarding the asymmetry of palaeozoological data with respect to biogeography. But Grayson's observation was at the geographical scale of a site or collection location; the observation above concerns the geographical scale of a taxon's biogeographical range. That is, the combination of techniques for estimating the asymmetry of palaeobiogeographical data represented in Figure 2 is at the scale of a biogeographical range. The lesson there, however, is pertinent both to the scale of a single site or collection location, and to the scale of a range or region. Larger samples will tend to be less asymmetrical.

The asymmetry of the palaeozoological data presented here on mountain beaver is readily apparent. That asymmetry may well be greater (or less) in light of more faunal data from California. Indeed, the northwestern coast of California has produced much palaeozoological data, but those data are thus far limited to large mammals because small mammals were not analysed (Hildebrandt, 1984). This highlights yet another cause of asymmetry in biogeographical data. Unless the research question requires it, various categories of faunal remains may not be identified. Accuracy of identification has been noted before as a potential cause of problems in palaeoenvironmental research (Findley, 1964; Graham & Semken, 1987). Simple lack of study
and identification of available specimens has not been, so far as I know, explicitly listed as a cause of asymmetry in palaeobiogeographical data (but see Lyman, 1998).

It is likely that counter-ubiquity, or absence, of mountain beaver is low because most sites included in analyses here are located within the modern range of mountain beaver (Figure 1). Including the sites that are outside of the modern range of mountain beaver, the counter-ubiquity index increases from 0.48 (16 of 33 sites produced no remains of mountain beaver) to 0.55 (23 of 42 sites produced no remains of mountain beaver). But why, one might reasonably ask, is the counter-ubiquity index 0.48? Should not that index be, say, much closer to 0.1, given that 33 of the sites are within the modern range of mountain beaver? In other words, should not many more than 17 of those 33 sites have produced remains of mountain beaver? Perhaps, but recall that at least site 42 (Figure 1) is not in a microhabitat expected to include mountain beavers as part of the local fauna. This is as much of a problem with how the modern range map was generated as with the palaeozoological record. In short, how does one interpolate ranges from a series of point-location data, whether prehistoric or modern (Graham & Semken, 1987)? The counter-ubiquity index may be unexpectedly large because of difficulties with identification of fragmentary remains; identification procedures and definitive criteria used are seldom reported in the literature examined, so this possibility must remain just that – a possibility. Understanding why the counter-ubiquity index is greater than it perhaps should be may reveal something about taphonomic history, collection history or analysis history with respect to these collections.

A critical variable that is revealed here becomes clear in light of Figure 2. That variable is sample size measured as NISP. The graph in Figure 2 should give one pause with respect to interpreting any measure of ubiquity (or counter-ubiquity), whether that measure is aimed at detecting biogeography or some other variable. With respect to the focus of this analysis, the graph in Figure 2 suggests that an estimate of the biogeographical asymmetry of palaeozoological data can be found in sample size. The larger the \( \sum NISP \) value for a site, the less asymmetric palaeozoological data will be with respect to reflecting accurately the presence and absence of a taxon in that location (all else being equal, such as preservation and recovery of remains of taxa in different size classes). Similarly, the larger the number of sites examined in an area, the less asymmetrical the data will be with respect to reflecting the range of a taxon in that area.

Acknowledgements

I thank Dave Schmitt for a helpful phone call, and an anonymous reviewer for identifying two ambiguously worded sentences.

References


Copyright © 2007 John Wiley & Sons, Ltd.


