Brief Comment

Assessing Criticisms of Faunal Analyses and Environmental Reconstructions in the Tehuacán Valley Project

Christopher C. Fennell¹

INTRODUCTION

This comment examines a portion of the statistical data compiled and analyzed by Flannery (1967) in his archaeological work at Coxcatlán Cave and other caves near Tehuacán in the early 1960s. The Tehuacán Valley project was undertaken by MacNeish, Flannery, and others to establish a chronological, cultural phase, and climatic sequence for prehistoric periods in this region of Mexico. This large-scale research project covered a time period spanning approximately 30,000 years to the present, including phases referred to as the Ajuereado (starting at 30,000 B.C. ± 10,000), El Riego (8650 B.C. ± 2000), Coxcatlán (5705 B.C. ± 500), and Abejas (3825 B.C. ± 300, and ending 2600 B.C. ± 200) (MacNeish, 1997, pp. 666–670).

The Tehuacán Valley project received pointed criticism from Hardy (1999, 1996), who focused her doctoral dissertation on reevaluating the extensive data and the related chronological analyses formulated by MacNeish, Flannery, and others. Hardy (1996, p. 700) emphasized the prominence of the Tehuacán Valley project's chronological sequences within the field of Mesoamerican archaeology, and this provided a primary impetus for her critical review of the data and related interpretations. This comment evaluates Hardy's criticisms, and concludes that she applied an inappropriate statistical test in her evaluation of Flannery's analysis of rodent remains and related inferences of environmental changes over time. In contrast, Flannery (1967) applied statistical tests appropriate to the types of data used in his analysis, if those data in fact comprised a random sample. The most intriguing issues in

¹Department of Anthropology, University of Virginia, Charlottesville, Virginia,

this debate concern Flannery's propositions that certain natural processes can serve as a proxy for creating statistically random samples, and those samples can then be used in formulating chronologies of past environmental changes.

BACKGROUND

Among a variety of research efforts in this multiyear project, Flannery (1967) analyzed the remains of rodents excavated from a series of strata in various caves near Coxcatlán and Tehuacán, and related the varying frequencies of the presence of different rodent species to changes in climatic conditions over time. These rodent remains were deposited in the caves by owls, whose diet was deemed to "constitut[e] a random sample of the local rodent population" (Flannery, 1967, p. 140). After swallowing the rodent whole, the owls regurgitated "pellets" of the bones and skin particularly amenable to statistical analysis:

Owls occupied the Tehuacan caves during periods between human occupations and left behind a stratified series of disintegrating pellets which were eventually incorporated into the human refuse. These constitute a series of random samples of the available rodent population during all periods of prehistory, an almost too-perfect situation for statistical analysis. (Flannery, 1967, pp. 140–141)

He based this position on his observation that "[t]here is no evidence that owls discriminate against certain species of rodents, and hence their diet constitutes a random sample of the local rodent population" (1967, p. 140). His team excavated approximately 538 rodent bones, which they interpreted as comprising approximately 188 identifiable individuals, from 48 stratigraphic and site locations, which they categorized by 21 prehistoric time zones (referred to as preceramic zones) reflected in the stratigraphy. These 21 zones were grouped into nine broader temporal phases, such as the Early Ajuereado phase.

Applying a Difference of Proportions statistical test to data that indicated differences in the presence of various rodent species over time, Flannery (1967, p. 144) concluded that the "rodent population, and presumably the environment as well, were significantly different" during certain phases, and contended the statistical data on differential frequencies of rat species over time were evidence of a corresponding difference in climatic conditions in the region over time. Using these and other faunal data, Flannery inferred that the climate in the Tehuacán Valley was more arid in an early temporal phase, referred to as the Early Ajuereado, than previously posited. Such a difference in climatic conditions provided indicators of the past habitat in which human populations subsisted. For example, Flannery (1967, p. 144) believed the absence of cotton rats in the Early Ajuereado

indicated that at that time "[t]he valley floor would have been open steppe, grazed by horse and antelope, but to judge by the lack of cotton rats, it could not have had a very rich cover of tall weeds or grass."

In her reevaluation of the prehistoric sequence formulated in the Tehuacán Valley project. Hardy criticized the analysis of stone tools conducted by MacNeish and others, the faunal and related environmental analysis by Flannery, and the precision and consistency of the stratigraphic and chronological sequences they formulated. She tested the statistical significance of Flannery's rodent population data by running a "Kolmogorov-Smirnoff (sic) test on all five microfaunal species present in the preceramic zones at Coxcatlán Cave" (Hardy, 1996, p. 705). Hardy's results indicated that there was not sufficient statistical significance in rodent population differences to warrant Flannery's conclusions. He responded that his data and conclusions remain sound, and that Hardy's reevaluation should be rejected (Flannery, 1997).

Hardy criticized Flannery's analysis on three main grounds: the data on rodent species involved too small a sample to be useful; there were no statistically significant differences in these rodent populations even using Flannery's data; and the rodent species on which he based his analysis were not "diagnostic" of past environmental conditions (Hardy, 1996, pp. 703–705). This comment addresses the following issues: First, was Flannery correct in his assumption that the manner of deposition of rodent remains by owls constituted a random sample that was well-suited to statistical analysis? Second, was Flannery's use of a Difference in Proportions test appropriate for these data? Third, was Hardy's use of the Kolmogorov–Smirnov test appropriate for these data? Finally, can one use a statistically significant difference in rodent populations as a basis for making inferences on past climatic conditions in the manner undertaken by Flannery?

SUMMARY OF RELEVANT DATA

The data on rodent remains from the Coxcatlán and Tehuacán area caves were set out in detail in Flannery's publication of his findings (1967, pp. 142–143, Table 15). The data consisted of 48 cases, each case representing the location of owl pellets containing rodent remains in a specific stratigraphic zone within a specific site. Seven types of rodent species were located in these sites: deer mouse, wood rat, spiny mouse, kangaroo rat, cotton rat, harvest mouse, and pygmy mouse. In summary, a total of 188 individuals uncovered in these 48 locations consisted of 28 deer mice, 40 wood rats, 36 spiny mice, 27 kangaroo rats, 55 cotton rats. I harvest mouse, and I pygmy mouse. Flannery (1967, pp. 142–143, Table 15) divided these 48 specific stratigraphic zones containing rodent remains into nine temporal phases.

ASSESSMENT

Sample Size and Randomness

Flannery (1967, pp. 140–141) contended in his analysis that the remains of these individual rodents deposited by owls in caves provided a sample of adequate size and statistical randomness. He thus utilized the entire data set of rodent remains uncovered in these caves as a sample from which to extrapolate an estimate of the relative proportions of the entire populations of each rodent species in past periods in that region. He was not seeking to estimate with precision the exact population counts for each rodent species, but rather to draw inferences as to their relative presence or absence in the environment. From this, he sought to infer characteristics, such as aridity, of the environment at different periods.

One of the more deficient methods for creating a sample is for the analyst purposefully and subjectively to select those variates she or he wishes to include in the sample, based on some working sense of observable patterns. This approach introduces the bias of the analyst directly into the composition of the sample, and is not replicable due to its idiosyncratic character. Flannery avoided such a subjective approach for creating a sample by utilizing the entire set of rodent remains uncovered, which were deposited by owls (particularly Tyto alba, the common Barn Owl), as the equivalent of a random sample. He proceeded on the assumption that the behavior of past owl populations in targeting, capturing, and consuming different species of rodents was a direct proxy for a recognized device of selecting a random sample from a population. He stated that "[t]here is no evidence that owls discriminate against certain species of rodents" (1967, p. 140). In his response to Hardy's criticisms, he emphasized this point again: "Regardless of the routes flown by individual owls, such pellet samples are archaeologically valuable because owls eat all small rodents in the region, not just the ones humans like" (Flannery, 1997, p. 661, emphasis in original). However, Flannery did not cite any sources or data to substantiate these statements on the behavior of owls and rodents that existed several thousand years ago.

Difficulties in the character of Flannery's sample directly impact the types of tests that can be applied and the inferences one can derive. The statistical definition of a random sample is not simply a collection of cases selected through some agency other than the analyst's personal biases: "[s]pecifically, each element in the population must have had an *equal* and *independent* chance for selection" (Thomas, 1986, p. 340, emphasis in original). In the absence of a valid random sample, numerous types of statistical tests cannot be applied to the data in a valid or meaningful way. Such statistical

methods include the Difference of Proportions test applied by Flannery and the Kolmogorov-Smirnov test applied by Hardy.

One difficulty in assessing the randomness of the owl deposits as a sample of the overall rodent populations is that we do not have sufficient information on either the feeding habits of the owls or the presence of different types of rodent species within the owls' range. The sample used consists of only those rodents captured by owls and then deposited in the caves. This raises a number of questions concerning the sample. Were there any patterns in the locations in which owls captured and consumed rodents and later regurgitated rodent pellets? Did any differential foraging habits of the different species of rodents tend to make them more vulnerable to being captured and consumed by owls? Were there any cyclical aspects to the behavior of the owls or particular rodent species that would create nonrandom patterns in the sample? Any of these factors could significantly lessen one's ability to assume there is an independence of events in each selection of a representative of a particular rodent species by a hunting owl from the general populations that existed in that area (see, e.g., Bernard, 1995, pp. 81-83; Thomas, 1986, p. 340).

A number of studies of owl predation patterns support the proposition that the Barn Owl (*Tyto alba*) is typically nondiscriminatory in its hunting practices, and that the contents of its pellets should therefore accurately reflect the relative proportions of the populations of varying small mammal species that were present in its local habitat (e.g., Bunn *et al.*, 1982, p. 82; Mikkola, 1983, p. 47; Ticehurst, 1935). Other studies found that the Barn Owl's pellets do not accurately reflect the proportions of those populations of prey species (e.g., Bellocq, 1998; Wallick and Barrett, 1976; Yom-Tov and Wool, 1997). In particular, the Barn Owl's predation patterns will be skewed when species of voles (*Microtus*) and shrews (*Sorex*) are present, with those prey being overrepresented in owl pellets and other species of mice and rats being underrepresented relative to actual populations (Andrews, 1990, pp. 178–180; de la Torre, 1990, p. 160; Derting and Cranford, 1989; Fast and Ambrose, 1976; Glue, 1974; Herrera and Jaksic, 1980; Johnsgard, 1988, pp. 100–101; Taylor, 1994, pp. 77–79; Voous, 1988, pp. 16, 18).

The reports of Flannery (1967) and others in the Tehuacán Valley project do not indicate the frequencies, if any, with which vole and shrew remains were uncovered in different time periods and localities included in their investigations. General studies of the zoological and geographic characteristics of vole species over time indicate that their populations have been present in varying ecological habitats in North and Central America from the Pleistocene period to the present day (Elton, 1942, p. 104; Hoffman and Koeppl, 1985, pp. 105–113). However, more specific studies of modern populations indicate that the presence and range of *Microtus* species have been

much more limited in the areas of Central and South America than in North America and Europe (Bellocq, 1998; Bunn et al., 1982, p. 82; Herrera and Jaksic, 1980). A modern sample of over 460 individual rodents contained in owl pellets collected in the Cueva de los Afligidos cave in the Oaxaca Valley region of Mexico yielded no Microtus or Sorex specimens (Flannery and Wheeler, 1986). It is possible that populations of voles and shrews may have been very low or nonexistent in the specific habitat zones and time periods investigated by Flannery and his colleagues within the Tehuacán Valley region. In the absence of significant populations of voles or shrews, Barn Owls have a greater tendency to prey on the diversity of other rodent species in a nondiscriminatory manner, and to produce pellets that would reflect the relative proportions of those other species' populations in the local habitat (see, e.g., Taylor, 1994, pp. 36–37, 80–81).

Flannery has been sensitive to issues concerning possible skewing in deposition behavior in other studies. For example, he observed elsewhere a differential presence of domesticated or wild food products that people brought to caves and consumed at those locations. In such instances, the samples of proportions from each type will be distorted and inferences must be undertaken with great caution, if at all (Flannery, 1976, p. 116). Thus, without additional supporting data, the unknown capture, consumption, and regurgitation behavior of past owl populations cannot serve as a proxy for rigorous methods of selecting random samples.

Flannery proposed to use all of the rodent remains uncovered in the Tehuacán Valley project as a sample of the extrapolated, inferred populations of rodents existing in the region in these past time periods rather than take an even smaller sample from this sample. However, an inference of which populations of species were larger or smaller relative to one another depends upon the ability to infer the size of each population to estimate those proportions. Therefore, one cannot dependably make an inference of overall population sizes based on his chosen sample.

If Flannery's sample was not random and probabilistic, what was it? The closest category is that of "convenience or haphazard sampling," in which the analyst uses as a sample those cases that present themselves for recording in some convenient manner. However, such a convenience sampling method should be used as a preliminary step only, and not as a sample from which one would attempt to draw inferences. Basing strong inferences on such a convenience sample is "plain hazardous" in terms of statistical rigor (Bernard, 1995, p. 94). In the absence of a probabilistic sampling, principles of statistics dictate that one should not generalize beyond the sample (Bernard, 1995, p. 96).

Hardy (1996) focused her criticisms on the small size of the sample, a characteristic that also affects the likelihood that the sample was random.

The equal probability of any one individual rodent being selected by an owl from the general population would be lessened and distorted if the overall population was small, and thus the resulting sample was small. Neither Flannery nor Hardy discuss any measures that may have been taken to correct for such possible distortions in the randomness of the sample due to a small population and associated sample. This would likely be difficult in any event, since Flannery did not have data indicating the sizes of the overall populations. He could only extrapolate those population sizes based on the limited sample exhibited in the archaeological record uncovered in the caves.

Applying the Difference in Proportions Test

Flannery's application of a Difference of Proportions test to the data on the frequency with which different species of rodents appeared in the various strata and time periods was inappropriate. Although this test may be conducted on nominal data, one must utilize a random sample to conduct a Difference of Proportions test. For the reasons discussed earlier, the sample utilized by Flannery should not be assumed to be the equivalent of a random sample. However, for the sake of argument and further analysis, this comment will assume that this sample was random. The Difference of Proportions test can then be conducted, using the different percentages of each species' presence in each temporal phase. This was the test Flannery conducted, using an alpha (" α ") value of .01 (1967, p. 144).

Specifically, Flannery compared the percentages of the presence of cotton rat, kangaroo rat, and deer mouse in the Early Ajuereado time period (30,000 B.C. \pm 10,000) to the presence of each of those species in the post-Pleistocene period. He used the term "post-Pleistocene" to be equivalent to "Present Climatic Conditions" (1967, pp. 142–144). Flannery (1967) did not set forth the exact calculations he applied. Assuming he applied the standard formula for the Difference of Proportions test, his calculations should have been based on a null hypothesis ("H₀") that the proportions of a given rodent species present in the Early Ajuereado phase are equal to the proportion of that same species present in the post-Pleistocene (or " $\mu_1 = \mu_2$ "). The alternative hypothesis H_n is that these proportions are not equal (or " $\mu_1 \neq \mu_2$ "). The formula is as follows:

$$Z = \frac{p_{s1} - p_{s2}}{\delta p_{s1} - p_{s2}}$$

where

$$\delta p_{\rm N1} - p_{\rm S2} = \sqrt{p_{\rm u} q_{\rm u}} \times \frac{N_1 + N_2}{N_1 N_2},$$

where

$$p_{\rm u} = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2}$$
 and $q_{\rm u} = 1 - p_{\rm u}$.

We will reject the null hypothesis H_0 if $Z > Z\alpha$, or if $\alpha >$ the actual "p" value, where α equals .01, and the actual p value equals .5 $-Z\alpha$.

Applying the Difference of Proportions test to each case shows that we can reject the null hypothesis as to the cotton rat (Z=2.9566) and the deer mouse (Z=5.5526). We can therefore conclude that the proportions of each of these species present in the Early Ajuereado phase were not equal to those proportions present in the post-Pleistocene period. These test results support Flannery's conclusion that the cotton rat was notably absent, and the deer mouse notably present, in the Early Ajuereado phase as compared to the post-Pleistocene (1967, p. 141). However, the null hypothesis cannot be rejected for the kangaroo rat (Z=1.8826), wood rat (Z=1.7485), spiny mouse (Z=1.6225), harvest mouse (Z=.3364), or pygmy mouse (Z=.3364).

Flannery (1967) applied the Difference of Proportions test to the data for the cotton rat, kangaroo rat, and deer mouse, but did not apply it to the data for the wood rat, spiny mouse, harvest mouse, and pygmy mouse. Having applied the same test to these latter species, it is notable that the null hypothesis of equal proportions over time cannot be rejected for five of the seven species of rodents on which his analysis focused. This does not necessarily weaken the findings for the cotton rat and deer mouse under the Difference of Proportions test. However, it is an additional indicator that there may be flaws in the assumption that owl deposition of rodent remains was the equivalent of a random sample. The differential shifts, or lack of shifts, in the proportions of each species' presence may reflect some form of discriminatory consumption by owls over time, rather than changes in the population sizes of each species.

Hardy's Use of the Kolmogorov-Smirnov Test

Hardy's application of the Kolmogorov-Smirnov test to Flannery's data was inappropriate. The Kolmogorov test was designed for application to ordinal data and random samples (see, e.g., Shennan, 1997, pp. 55, 57, 65; Siegel, 1956, pp. 48–51, 135–136; Thomas, 1986, p. 336). Flannery's data were nominal in nature, consisting of counts of individuals, fitting each exclusively into one of seven classifications of rodent species at each site and stratum. There was no ordering of the data to provide an ordinal scale, nor was there any equal unit of measure between variates to provide an interval scale (see, e.g., Thomas, 1986, pp. 19–26). Flannery's rodent data did not contain ordinal

characteristics to which the Kolmogorov-Smirnov test could be applied with any valid statistical meaning.

It is curious that Hardy applied the Kolmogorov test to Flannery's data after contending that the sample was too small to be useful or random. She also provided no indication that she considered attempting to convert (or in fact converted) the original nominal data into an ordinal data form for the purpose of conducting this type of test. Hardy instead applied the test to the wrong type of data, and to what she believed to be the wrong type of sample. Could the rodent count data be converted into a useful form of ordinal data? No: the data presented do not include any characteristic that can be used for establishing a meaningful property of asymmetry (see Siegel, 1956, pp. 23–26; Thomas, 1986, p. 22). Flannery's data on the presence or absence of each of these different rodent species do not yield to a meaningful ordinal seriation of the separate species. While one might speculate as to a particular, asymmetrical relation between two of these species (e.g., C > F), one cannot line them all up in a meaningful continuation of ranked gradation so that A > B > C > D > E > F > G.

Extrapolations for Climatic Conditions

It is hazardous to make robust extrapolations from nonrobust, nonparametric statistical tests. An additional step for Flannery's analysis was the ability to treat the relative differences of rodent populations as diagnostic of different environmental conditions. This is an issue of anthropological and environmental science, and not of statistical theory. Flannery provided little or no discussion of sources for his confidence that such rodent population differences were diagnostic. Hardy criticized his assumption by offering comparably sweeping statements of rodent population tendencies without supporting sources. Absent solid support for this assumption of diagnostic significance, Flannery's anthropological conclusions are substantially weakened.

However, Flannery's use of inferences from the rodent remains was not critical to his overall analysis. His inferences on past climatic conditions were based on a variety of faunal remains uncovered in these extensive excavations, including deer, jack rabbit, Mexican cottontail, and iguana (1967, p. 144). The data on rodent remains thus provided additional support to his overall analysis, and were not used in isolation.

CONCLUSIONS

Hardy's criticisms of the value and validity of Flannery's rodent populations analyses fail. She brought no new excavation findings to bear on

this debate. Rather, she focused solely on undertaking a critical review of the data compiled and analyzed by those whose conclusions she sought to disprove. Yet she applied incorrect tests and failed adequately to consider and question all of the assumptions concerning the nature of the data under examination.

Flannery's proposed proxy for a random sample remains intriguing: Can a naturally-occurring process, such as owl deposition of rodent remains, be used as the equivalent of a random sample? Thus far, it appears to be a questionable proposition. However, to date there is no definitive empirical evidence offered by proponents or detractors by which to conclude if the typical behavior of rats, mice, and owls could have resulted in such convenient randomness.

ACKNOWLEDGMENTS

The author thanks Kent Flannery, Karen Hardy, and Rachel Most for their helpful comments and critiques of earlier versions of this commentary. Any shortcomings remain the author's responsibility.

REFERENCES

- Andrews, P. (1990). Owls, Caves and Fossils: Predation, Preservation and Accumulation of Small Mammal Bones in Caves, With an Analysis of the Pleistocene Cave Faunas From Westburysub-Mendip, Somerset, UK, University of Chicago Press, Chicago.
- Bellocq, M. I. (1998). Prey selection by breeding and nonbreeding Barn Owls in Argentina. The Auk 115(1): 224–229.
- Bernard, R. (1995). Research Methods in Anthropology, AltaMira Press, London.
- Bunn, D. S., Warburton, A. B., and Wilson, R. D. (1982). The Barn Owl. Buteo Books, Vermillion, SD.
- de la Torre, J. (1990). Owls: Their Life and Behavior, Crown, New York.
- Derting, T. L., and Cranford, J. A. (1989). Physical and behavioral correlates of prey vulnerability to Barn Owl (Tyto alba) predation. American Midland Naturalist 121(1): 11–20.
- Elton, C. (1942). Voles, Mice and Lemmings; Problems in Population Dynamics, Oxford University Press, Oxford.
- Fast, S. J., and Ambrose, H. W., III (1976). Prey preference and hunting habitat selection in the Barn Owl. American Midland Naturalist 96(2): 503–507.
- Flannery, K. V. (1967). The vertebrate fauna and hunting patterns. In Byers, D. S. (ed.), The Prehistory of the Tehuacán Valley, Vol. 1, Environment and Subsistence, University of Texas Press, Austin, TX, pp. 132–177.
- Flannery, K. V. (1976). Empirical determination of site catchments in Oaxaca and Tehuacán. In Flannery, K. V. (ed.), The Early Mesoamerican Village, Academic Press, New York, pp. 103–116.
- Flannery, K. V. (1997). In defense of the Tehuacán Project. Current Anthropology 38(4): 660–662.
- Flannery, K. V., and Wheeler, J. C. (1986). Comparing the preceramic and modern microfauna. In Flannery, K. V. (ed.), Guilá Naquitz: Archaic Foraging and Early Agriculture in Oaxaca, Mexico, Academic Press, New York, pp. 239–246.
- Glue, D. E. (1974). Food of the Barn Owl in Britain and Ireland. Bird Study 21(3): 200-210.

Hardy, K. (1996). The preceramic sequence from the Tehuacán Valley: A reevaluation. Current Anthropology 37(4); 700–716.

Hardy, K. (1999). On the Tehuacán Project: Reply to Flannery and MacNeish. Current Anthropology 40(1): 63–69.

Herrera, C. M., and Jaksic, F. M. (1980). Feeding ecology of the Barn Owl in Central Chile and Southern Spain: A comparative study. The Auk 97(4): 760–767.

Hoffman, R. S., and Koeppl, J. W. (1985). Zoogeography. In Tamarin, R. H. (ed.). Biology of New World Microtus. The American Society of Mammalogists, Shippensburg, PA, pp. 84–115. Special Publication No. 8.

Johnsgard, P. A. (1988). North American Owls: Biology and Natural History, Smithsonian Institution Press, Washington, DC.

MacNeish, R. S. (1997). In defense of the Tehuacán Project. Current Anthropology 38(4): 663–672.

Mikkola, H. (1983). Owls of Europe, Buteo Books, Vermillion, SD.

Shennan, S. (1997). Quantifying Archaeology, University of Iowa Press, Iowa City.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York. Taylor, 1, (1994). Barn Owls: Predator-Prey Relationships and Conservation, Cambridge University Press, Cambridge.

Thomas, D. H. (1986). Refiguring Anthropology, Waveland Press, Prospect Heights, IL.

Ticehurst, C. B. (1935). On the food of the Barn Owl and its bearing on Barn Owl population. The lbis 5(2 Serial no. 13): 329–335.

Voous, K. H. (1988). Owls of the Northern Hemisphere, MIT Press, Cambridge.

Wallick, L. G., and Barrett, G. W. (1976). Bioenergetics and prey selection of captive Barn Owls. The Condor 78(1): 139–141.

Yom-Tov, Y., and Wool, D. (1997). Do the contents of Barn Owl pellets accurately represent the proportion of prey species in the field? The Condor 99(4): 972–976.