29 What lies behind the quantification debate?

Clive Orton

29.1 AIM

The aim of this paper is to try to look behind the arguments that surround the choice of a measure of quantity of pottery. I suspect that these apparently technical differences mask different approaches to archaeological data and to archaeological argument. If so, it is only by bringing them into the open and acknowledging them that we will be able to communicate properly with each other.

29.2 BACKGROUND

29.2.1 Definitions
In this paper I shall use the term "quantification" as a general term for the various ways in which archaeologists attempt to answer the question «how much pottery is there?» (e.g. in a deposit; on a site). This is usually followed by the corollary «how much is of each type?», "type" being defined according to the needs and inclinations of the individual researcher. The term "measure" is used to denote a way of measuring the quantity of a group of pottery. In other words, "quantification" is a process and "measures" are different means of expressing its outcome.

29.2.2 History
The question began to be asked seriously, and for purposes of comparison, after the introduction of seriation to the study of collections of sherds (Kroeber 1916). For over 40 years, the only measure was the sherd count (Orton *et al.* 1993). From some of the literature of the time, it is hard to recall that sherds have anything to do with pots. The sherd–based approach was supplemented by other approaches, either vessel–based or material–based, between about 1960 and 1980 (see Table 29.1). Of the later measures, two (nos. 2, 4) are vessel–based and three (nos. 3, 5, 6) are material–based.

Once there was more than one measure, a further question arose: «which is the best?», and archaeologists began to compare various measures. The main attempts at comparison are listed in Table 29.2: uses of two or more measures to throw light on (for example) site formation processes are not included.

29.3 THE DEBATE

The debate consists of these (and probably other) written comparisons, plus much that has not been committed to writing but which is clearly strongly felt. The written part of it centres on the criteria that are used to judge between one measure and another; the unwritten part consists of more general views and feelings.

29.3.1 Criteria
The most important criteria seem to be: speed, practicality, "accuracy" and the formal properties of measures as estimators.

29.3.1.1 Speed
Speed is clearly important, especially in today's climate of tight funding and strict deadlines. However, it is secondary to considerations of "accuracy", since it would be unwise to prefer a fast but inaccurate measure to a slower but accurate one, unless it could be shown that the difference in accuracy was small and could be compensated by a greater through–put in a given time. Opinions have varied as to whether measure 1 or 3 (count or weight) is the fastest of all the measures.

29.3.1.2 Practicality
It is clearly important that a measure should be easy to carry out, without the need for either special equipment or advanced training. This criterion has been used to rule out measure 6 (displacement volume), and sophisticated forms of

Clive Orton

measure	date of introduction	author	measures compared	author	date
1. sherd count	1916	Kroeber	1, 3	Solheim	1960
1a. rim sherd count					

and is really outside the scope of this paper, which is about different approaches to quantification, given a need for it. But we could say in passing that this criterion seems to confuse means with ends, while we leave full rebuttal for another occasion. It must be acknowledged that many useful studies of pottery can be made without the use of quantification as defined here.

2) «the proper use of quantification is the reconstruction of vessels from pieces scattered across a site» (e.g. Moorhouse 1986: 118). While accepting the value of this approach for answering some questions about (for example) the use and development of a site, I do not think it can be called quantification. «C'est magnifique, mais ce n'est pas la guerre».

3) «what we have are sherds, so we count them». This view is characteristic of the typological phase of ceramic studies, c 1916–60 (Orton et al. 1993). Anything more subtle is dismissed as "purist" (e.g. Ford 1962) and too complicated. Although dated, this view persists today, especially among those who consider themselves to be primarily excavators. Another potential group of culprits consists of museum curators, who insist that collections consist of discrete objects so that they can be audited (is anything missing?). In pottery terms, "objects" tend to mean sherds, so a house–keeping need becomes an analytical tool. This view can be countered by standard arguments about (for example) differential breakage rates, but people who do not have the imagination to realise that sherds come from pots are not likely to be impressed.

4) «we are dealing with pots, not sherds, so we must know how many pots we have». This is a much more sophisticated and well–thought–out view, and carries emotional weight as well as some reason. It also has the attraction of being analogous to the mni (minimum number of individuals) statistics used by animal bone specialists, although that statistic has its critics. It tends to gloss over the practical difficulties of sorting sherds to individual vessels, and the statistical implications of the different strategies used to overcome such difficulties. There is also a hint of machismo in the claimed ability to relate non–joining sherds to the same vessel. Theoretical arguments against it concentrate on its role as a descriptor of excavated assemblages rather than an estimator of the parameters of "life" assemblages.

5) «what we have is a sample of what we want to know about. Our basis for interpretation is therefore sampling theory and its associated formal criteria (bias, etc.). Measures should be chosen which best reflect the structure of "life" assemblages». This, of course, is my position. Nevertheless, it has its problems. The two main ones, which are related, are (i) if the excavated assemblage is a sample, what is the population? (ii) how representative is an assemblage of its parent population?

There is a trade–off between the answers to the two questions: the more narrowly we define the parent population, the more likely it is that the sample will be representative, but the less archaeologically useful the information about that population is likely to be — a sort of archaeological analogy of the Uncertainty Principle of physics.

This view can be anathema to those who hold views (3) and (4): the former just don't see the point and think it's an elaborate waste of time, while the latter are unhappy with the concept because it leads to entities that appear absurd in their frame of reference, such as the vessel–equivalent (ve). This measure expresses each "sherd family" (all the sherds from the same vessel in the same context) as a proportion of a complete vessel, so that a relatively complete vessel has a higher measure than a relatively incomplete one. Because it's rarely possible to measure exactly the proportion of a vessel that is represented by a sherd family, this proportion must frequently be estimated (e.g. by using rim sherds, which can be measured), giving rise to the estimated vessel–equivalent (eve) (Orton 1975, 1982). A common response to the idea of an eve that has contributions from two or more vessels is «but you can't add apples to oranges». My stock response of «two fruit» is too pat — the real problem is that the measure is too abstract for them: they could hold in the hand a sherd, a sherd family (all the sherds found from the same pot), or even a kilogram of pottery (though some have difficulty even with that), but not an eve.

29.3.3 What lies behind these different views?
We now try to look for the root causes of these divergences of view. Two areas of fundamental, if unexpressed, disagreement may contribute to them:

1) the status of archaeological data,
2) the level of abstraction/concreteness that is acceptable in an archaeological argument.

29.3.3.1 Archaeological data
Archaeological evidence is intrinsically material — physical, chemical, morphological, etc. Our

data that represent it may be textual (e.g. written descriptions), mathematical, or both. How can we handle such data?

One approach is to separate the tasks of record and interpretation. It then becomes our primary duty to *record* what is found, since if we do not then the information is lost for ever. The interpretation is a secondary and quite separate duty, which may well be done by someone other than the primary excavator/recorder, and at a later date. This approach underlies the British "preservation by record" policy with its concomitant growing mountain of undigested archives.

It has been argued for some time that archaeological data are not something "given" (the French translation *donnés* conveys this idea well), but rather they are what we choose to extract from the incoherent mass of material that I have called archaeological evidence. If this is so, then there is a feed–back from interpretation to record, in that what is recorded is (or should be) what is needed for interpretation and, I would add, what is formally interpretable. Most archaeologists would, if asked, agree with this general principle, but may have more difficulty with putting it into practice. There is a school of thought that says «record everything, because someone may need it one day» or even «record everything and you can never be wrong», but in practice recourse to the original material seems necessary whenever re–interpretation is attempted. Examples of successful interpretations of recorded information can be quoted as counter–examples (Bradley 1973), but I would argue that in such cases the original recording was highly selective but done with much insight.

At the other extreme, we could argue that we are not interested in archaeological data in their own right, but only as a trace or spoor of past activities. Put more formally, the data are a mean of estimating parameters in which we are interested, not an end in themselves. The complicating factor is the series of events that happen between "life" and discovery, e.g. differential survival, retrieval strategies and all the interferences with the evidence that are lumped together as "site formation processes". These introduce all sorts of biases, well–known and not so well–known, which have to be taken into account in interpretation. But they also have to be taken into account in the recording, so that it becomes possible to interpret the ensuing data. We are here in the realm of what David Clarke would have called epistemology, for which I prefer the term "data theory" or even just "method".

How does this general discussion apply to quantification? Suppose we decide to quantify all the pottery from an excavation. The total amount for the whole site as a single figure will tell us little, because we do not know (i) what proportion of the site (in the broad sense, including associated rubbish disposal) we have excavated, and (ii) what the life–span of the pottery was. These points mean that the total figure is formally uninterpretable. Next we look at the proportions of different "types" (for example, fabric types) in a single assemblage. These proportions may or may not reflect anything about the composition of the parent population, depending on (i) the relative sizes of pots of the different types or their relative brokenness statistics, and (ii) the relative life–spans of the different types, but we cannot tell, and the figures are again uninterpretable. So we move down a step further and compare these proportions between two assemblages by (to avoid being contentious) using weight and sherd count. Two types may have different weights per vessel, but the relativity between the two is the same for the two assemblages, so we can *compare* their proportions even if we cannot use *either* independently. But if we use sherd count the relativity may not be the same, since the relative brokenness of the two types may vary between assemblages (unlike the relative weight), so we cannot compare proportions without the risk of confounding population parameters with what has happened to the material since its original deposition. Even here, the use of weight as a measure carries a hidden assumption, that the relative life–spans are the same in the two assemblages. This seems reasonable, and without assumptions of this sort of level one can say nothing about archaeological data at all. To sum up, weight is interpretable in the context of a comparison of proportions, but count is not. Considerations such as these, including our objectives (e.g. do we want to compare proportions?) must guide our choice of measures.

29.3.3.2 Level of abstraction or concreteness
Archaeology is, as we have seen, a material discipline, and archaeologists seem to prefer their data to be so: they should be at least capable of being visualised, if not actually tangible — «a pot in the hand is worth two in the head», one might say. Archaeological theory may be abstract, but for it to be tested it must be rooted in a material reality. An apparently abstract type of data, such as the vessel–equivalent, is disturbing and even frightening; even weight may be slightly suspect because it cannot be seen or handled. The ability of a formal argument to generalise about all possible assemblages also feels alien to those who are used

to arguing from real groups of real material. I sense a great divide here, across which it is very difficult, if not impossible, to communicate. It is the old chestnut, the arts–science divide, or is it even deeper than our respective educations?

29.4. CONCLUSIONS

Where does all this lead us? We can first make some practical points:

1) the case is well made that the choice of measure must be based on the needs of the problem in hand, and must be interpretable in terms of the questions being asked.
2) the real question of quantification is not the measure, but the level. To what extent do we aggregate contexts to form useful assemblages, or types to form useful classes, or whatever we call this level of definition?

Moving on, we can see that effective communication requires knowledge of the viewpoint and feelings of those with whom one is trying to communicate. Other people are not necessarily wrong, they just see things differently and may have difficulties in grasping (for example) the generalising power of a formal argument.

Finally, and most contentiously, I argue that if we accept the need for quantification we must also accept the view expressed in 29.3.2 (5) — what one might call the sampling paradigm. A plausible (though in my belief fundamentally mistaken) case can be made that one should not quantify at all (29.3.2 (1)). But if we do quantify, I do not see how we can reasonably stop at a "half–way house" (e.g. 29.3.2 (3) or (4)). A quantified description of an archaeological assemblage that does not take into account the distortions possible in the archaeological process, and does not seek to isolate the factors that remain invariant, cannot be used for archaeological interpretation.

References

Bloice, B. J.
1971   Note. in G. J. Dawson Montague Close Part 2. *London Archaeologist* 1:250–1.

Bradley, R. J.
1973   Two notebooks of General Pitt Rivers *Antiquity* 47:47–50.

Burgh, R. F.
1959   Ceramic profiles in the western mound at Awatovi, Northeastern Arizona. *American Antiquity* 25:184–202.

Chase, P. G.
1985   Whole vessels and sherds: an experimental investigation of their quantitative relationship. *Journal of Field Archaeology* 12:213–8.

Cochran , W. G.
1963   *Sampling Techniques* (2nd edition) Wiley, New York.

Egloff, B. J.
1973   A method for counting ceramic rim sherds. *American Antiquity* 38:351–3.

Evans, J. D.
1973   Sherd weights and sherd counts: a contribution to the problem of quantifying pottery studies. in D. E. Strong (ed.) *Archaeological Theory and Practice*. Seminar Press, London. pp. 131–49.

Ford, J. A.
1962   A quantitative method for deriving cultural chronology. *Pan American Union Technical Manual 1*. Washington.

Glover, I. C.
1972   Excavations in Timor. PhD thesis, Australian National University, Canberra.

Hawkes, J.
1968   The proper study of mankind. *Antiquity* 42:255–62.

Hinton, D. A.
1977   'Rudely made earthen vessels' of the twelfth to fifteenth centuries AD. in D. P. S. Peacock (ed.) *Ceramics and early commerce*. Academic Press, London. pp. 221–38.

Hulthén, B.
1974   On choice of element for determination of quantity of pottery. *Norwegian Archaeological Review* 7:1–5.

Kroeber, A. L.
1916   Zuñi Potsherds. *Anthropological Papers of the American Museum of Natural History* 18:1–37.

Millett, M.
1979   How much pottery? *Institute of Archaeology Occasional Publication* 4:77–80.

Moorhouse, S.
1986   Non–dating uses of medieval pottery. *Medieval Ceramics* 10:85–124.

Orton, C. R.
1975   Quantitative pottery studies: some progress, problems and prospects. *Science and Archaeology* 16:30–5.

Orton, C. R.
1982   Computer simulation experiments to assess
        the performance of measures of quantity of
        pottery. *World Archaeology* 14:1–20.
Orton, C. R. & Tyers, P. A.
1990   Statistical analysis of ceramic assemblages. *Ar-
        cheologia e Calcolatori* 1:81–110.
Orton, C. R., Tyers, P. A. & Vince, A. G.
1993   *Pottery in Archaeology.* Cambridge University
        Press, Cambridge.
Solheim, W. G.
1960   The use of sherd weights and counts in the
        handling of archaeological data. *Current An-
        thropology* 1:325–9.

Author's address
Clive Orton
University College London
Institute of Archaeology
31–34 Gordon Square
GB–WC1H 0PY London
E–mail: c.orton@ts.bcc.ac.uk