Sizing up prehistory: sample size and composition of artefact assemblages

Peter Hiscock

School of Archaeology and Anthropology, Australian National University

Abstract: A review of selected Holocene artefact assemblages in Australia demonstrates that assemblage composition is often dependent on the size of the sample. Rare types of objects, such as backed artefacts, are less likely to be present in small samples than in large ones. Although major questions in Australian archaeology focus on the presence or absence of such rare classes of object, the archaeological patterns have often been interpreted without regard to this sample-size effect. Consequently, interpretations of some temporal trends in Australian prehistory, such as the model that backed artefacts first appear 4000–4500 years ago, might be replaced with models that describe changing assemblage sizes through the Holocene. Similarly, some spatial differences in assemblage composition might also be adequately described in terms of varying sample size. The implications of this phenomenon are considered.

Introduction

Following the demonstration by Hiscock and Attenbrow (1998) that the manufacture of backed artefacts occurred in Australia during the early Holocene, it is important to re-evaluate our approach to the archaeology of this period. It had previously been thought that these distinctive stone artefacts were first made in Australia during the mid-Holocene, and had ceased to be made during the last one or two thousand years. This notion, that backed artefacts were markers of a restricted time period, 4500–1000 years ago, has been the basis for many propositions espoused by Australian archaeologists. For example, arguments that backed artefacts were introduced from outside Australia (e.g. Beaton 1977), that they arrived as part of a package (e.g. Bowdler 1981; Bowdler and O'Connor 1991), and that they spread rapidly across the southern two-thirds of the continent (e.g. Hiatt 1996:140) are all partly based on the mistaken view that backed artefacts appeared simultaneously in sites across much of the continent in the mid-Holocene (see Hiscock and Attenbrow 1998 for an extended critique). All of these propositions crumble under the weight of empirical evidence for backed artefacts at least 5000–9000 years old. The alternative model, of a late Pleistocene/early Holocene system of backed artefact production becoming much more frequently employed in the mid-Holocene as an adaptive response, must now be developed and evaluated (see Hiscock 1994).

There is ongoing dispute about other space–time patterns of backed artefacts in Australia. For example, it has been suggested that backed artefacts ceased being used in the last millennium or so (e.g. Morwood 1984) or, alternatively, that they continued to be made but with declining production rates during the late Holocene as a consequence of changed settlement organisation (Hiscock 1994). Another debate played out over the last two decades concerned the distribution of these artefacts across the continent, with a number of different northern boundaries being proposed (e.g. Flood 1983; Hiscock and Hughes 1980; Mulvaney 1975, 1985; Pearce 1974; Smith and Cundy
Since backed artefacts are often said to be one of the most distinctive implement types in Australia, and to be chronologically and spatially patterned, the perception of archaeologists as to their age and distribution has been a key factor in the interpretation of cultural change in Holocene Australia. For example, those archaeologists who perceived a relatively early and/or southerly distribution were inclined to advocate that backed artefacts were invented independently in Australia (e.g. Hiscock and Attenbrow 1998; White and O'Connell 1979), whereas those who perceived a relatively late and/or northerly distribution have sometimes argued that the notion of backed artefacts was introduced through external contacts (e.g. Beaton 1977; Hiscock and Hughes 1980; see also Dorch 1981). Clearly, in this context it is necessary to have an accurate picture of the distribution of backed artefact production through time and space.

But how are we to interpret the archaeological record? Much of this debate about chronology and distribution reflects the tendency for images of Australian prehistory to be constructed from the presence or absence of distinctive but rare kinds of stone artefacts. Although we now know that backed artefacts were made in the early and late Holocene, Australian archaeologists failed for so long to recognize this because (a) they interpreted the absence of such artefacts as an indication that they were not made or used, and (b) they sometimes discounted small numbers of these distinctive artefacts as an error of some kind, such as might be produced by stratigraphic disturbance (see Hiscock and Attenbrow 1998 for details). It turns out that both of these principles were misleading because the same patterns can be produced by other mechanisms. One example of this is the failure of Australian archaeologists to recognize the importance of the phenomenon of 'sample-size effect'. This article explores sample-size effect as a factor in the archaeological patterning of Holocene artefact assemblages in Australia.

Sample-size effects

A large body of literature now documents the effect of sample size on assemblage content and inter-site comparisons. This literature originally focused on quantifying the influence of sampling and assemblage size on measures of abundance within faunal collections (e.g. Casteel 1974; Grayson 1978, 1981, 1984). In more recent years these issues have been investigated for assemblages of material culture, including pottery, mobile art, and stone artefacts (e.g. Kintigh 1984, 1989; Kirch et al 1987; Thomas 1989). The consequence is a refutation of the idea that the variation between assemblages necessarily reflects site function in some simple way. In most sites it now appears that assemblage content and diversity may not directly relate to the diversity of prehistoric activities, but is dependent on the size of the sample (see Thomas 1989: 86). Unfortunately, Australian archaeologists have rarely incorporated this sample-size mechanism into their interpretations, the recent arguments by Hiscock (1993), James and Davidson (1994), Gorecki et al (1997), and Hiscock and Allen (in press) being notable exceptions. The purpose of this article is to illustrate the importance of these sample-size effects on our understanding of Australian prehistory. A number of case studies demonstrate that many of the chronological and spatial patterns of archaeological stone artefacts relate to, and may be a product of, the size of samples.

'Sample', in this context, is used specifically to refer to the assemblage of artefacts recovered from a particular level in a stratified site, or from the site as a whole. These assemblages are samples because they are only a fraction of the discarded objects that exist in the unexcavated deposit. Furthermore, the artefacts that survive in the deposit are only a fraction of those that were used by the human occupants of the site, which is the target population (see Grayson 1984: 116; Orton 1992: 138). In the following analyses, 'sample size' represents raw counts of either all recovered artefacts or implements (depending on availability of data). Consequently, the sample-size effect mentioned above is the co-variation between the size of samples and any aspect of assemblages under examination. Two aspects of assemblage composition are of interest here and are compared with sample size. First is the number of specimens of a particular artefact type, such as the number of backed artefacts or the number of grindstones. The second measure of assemblage composition examined here is 'implement richness', defined as the number of implement types identified by the original excavator.

Fortunately, the identification of sample-size effects is often easy. Empirically derived correlations between the number of specimens and the number of classes in assemblages is the most common means of identifying sample-size effects. These empirical investigations not only confirm that very small samples often do not contain the full range of forms,
but also indicate that even comparatively large samples, containing hundreds of specimens, may not provide an adequate database for statistical manipulations (see Grayson 1984:117). This occurs because the probability of recovering at least one example of a rare (i.e. infrequent) type of object increases as the size of the sample increases, but a very large sample may sometimes be required before all varieties of artefact are represented.

A hypothetical example can be used to illustrate how this might happen. Even in sites where only one specific kind of knapping activity takes place, such as the manufacture of backed artefacts, the various objects employed and created will probably be discarded at different rates. For instance, many flakes will be rapidly discarded, cores are likely to be discarded less frequently, backed artefacts perhaps less frequently still, and the hammerstones may be very rarely thrown away. These differences in the likelihood of discard relate to a number of factors, including the length of ‘use-life’ of each kind of object. When only a few of these objects have been discarded it is likely that the assemblage will be dominated by only those classes of object that are discarded frequently, such as flakes and cores in this example. As occupation of the site continues, and the size of the assemblage grows with further discard of material, it is likely that objects such as backed artefacts or hammerstones may eventually be discarded. The end result of a simple process such as this will be that many small assemblages will not contain the rarer artefact forms which will commonly be found in larger assemblages.

The size of the sample typically needed to contain all categories that may have been discarded in a locality is proportional to the relative abundance of the rarest category. This means that there is no absolute sample size which in all sites or regions is sufficient to ensure the recovery of all categories (Orton 1992:138). Consequently, while in some sites or regions sample sizes of 100 may be adequate, in other regions sample sizes of 1000 or 10,000 may still be too small to yield the entire variety of forms. It is for this reason that the required level of sampling is often determined empirically.

A number of statistical procedures can be used to document sample-size effects. For example, Grayson (1984:18–20) employed analyses of rank, such as Spearman’s coefficient, to describe the relationship between sample size and the relative abundance of taxa. However, I have followed Jones et al (1989), Thomas (1989) and others in employing the standard Pearson’s product-moment correlation coefficient as a means of evaluating the strength of such relationships. The Pearson correlation (designated by the letter ‘r’) reflects the degree to which two variables are related in a linear manner, and the value ranges from +1 to -1, with a correlation of +1 when there is a perfect positive linear relationship between variables.

In most instances, the data presented below display strong linear patterns and can be simply treated. For those few instances in which there is a non-linear relationship, I have used the common approach of transforming one of the variables into another scale, such as changing the sample size into a log scale. Using this procedure, sample-size effects are recognised as a strong positive correlation between sample and implement numbers or richness. Such a correlation implies that, as the sample size increases, there will be increased numbers of any particular class and greater diversity of classes. Conversely, as sample size decreases, there will be a decrease in richness and the numbers of specimens in each implement class will diminish and in some instances become absent. Hence, the presence or absence of a particular class of implement in an assemblage becomes related to the size of the assemblage.

Establishing statistically significant correlations between the size of samples and the range and abundance of implement types does not, of itself, demonstrate that the assemblage composition is caused by assemblage size. It may be that a third factor is responsible, causing changes in both assemblage size and composition (see Grayson 1984:121). A typical example of this might be differences in prehistoric site function, in which occupation of a base camp could involve a wide range of artefact classes, whereas at an extractive site a smaller number of activity-specific tools become employed. In some circumstances, such differences in site function may also be partly responsible for differences in the size of the assemblage. While this alternative exists, and may be explored by researchers, I argue that in the sites discussed below sample size is a primary cause for assemblage variation. However, the purpose of the article is to demonstrate that sample-size effects are a common phenomenon in artefact assemblages within the Australian region and therefore need to be considered in interpretations of assemblage difference, rather than to argue that sample size is the sole cause of assemblage variation. Indeed, it is necessary to define
the nature of such sampling mechanisms before other, perhaps underlying, behavioural factors can be examined.

Some simple calculations on data available in the literature show that these sample-size effects may frequently be at work in Holocene Australian assemblages. A few examples of artefact abundance in archaeological sites in eastern Australia will illustrate that assemblage composition is often dependent on the size of the sample.

Examples of chronological change

One of the major questions in Australian archaeology has been the antiquity of the distinctive backed artefacts and points that are generally thought to appear in the Holocene. Much effort has been spent on identifying the date at which these implements first appear (e.g. Bowdler and O’Connor 1991; Johnson 1979; Jones and Johnson 1985; Pearce 1974). Broad theories involving social change and external contact depend on the contemporaneity of the appearance of these implement types and other changes in the Holocene archaeological record (e.g. Bowdler 1981). Consequently, reliable inferences about the age of these implements are crucial in constructions about Australian prehistory.

Recently, this debate has focused on the dating of stone points in northwestern Australia, at Nauwalabila 1 in Kakadu and at the Widgingarri Shelters in the Kimberley region (e.g. Bowdler and O’Connor 1991; Jones and Johnson 1985). However, it has been suggested that the vertical distribution of points within Nauwalabila 1 can be explained in terms of variable sample sizes (Hiscock 1993). It is my intention to now explore similar arguments for the vertical distribution of backed artefacts, and other implement types, in sites in eastern Australia, thereby revealing the general occurrence of the sample-size effect. Four rockshelters excavated over the past 15 years have been selected as examples: Cathedral Cave, Native Well 1, Shaws Creek KIL, and Capertee 3.

To introduce the sample-size mechanism, there is no better instance than Cathedral Cave in the central Queensland highlands.

Cathedral Cave

Beaton’s (1991) excavations at Cathedral Cave yielded a 2 m deep deposit, covering a time span of only 2000 years and apparently containing good chronological resolution. Since charcoal near the base of the sequence gave a radiocarbon date of 3560±80 (ANU 1762), the entire artefact assemblage has formed during the period when all commentators agree that backed artefacts were being manufactured in the region (Johnson 1979; Morwood 1981). And yet backed artefacts were recovered from some units but not others. Examination of data presented by Beaton (1991:60) demonstrates there is a strong positive correlation between the number of backed artefacts in each unit and the total number of implements in that unit (Figure 1). The strength of this relationship is measured by an $r^2$ value of 0.88 ($p<0.001$). When the number of implements in a unit falls below 50-100 there are few or no backed artefacts represented, even though the inhabitants of the region were thought to have been making and using that implement class. As discussed above this pattern is consistent with a direct sample-size effect.

Of course, at Cathedral Cave it is not only the backed artefact component of the assemblage that is subject to this effect. Other classes of artefact, such as burrens, tulas, hammerstones and discoidal scrapers, are present in only some strata. The absence of such types is not consistent with the simple loss of a particular technology. For example, tulas are found at the base of the sequence (unit 8) and at the top (unit 1), but not in some of the intervening strata. Since their absence is typically linked to low overall numbers of implements in the strata, it is likely to be a function of assemblage size. To illustrate this relationship, Figure 1 plots for each unit the total number of implements, plotted on a log scale, and richness as measured by the number of implement types. The positive correlation ($r^2 = 0.88$, $p<0.001$) between assemblage richness and size is typical of a sampling phenomenon.

At Cathedral Cave, the conclusion that sample size is affecting the assemblage composition is neither surprising nor radical in its implications. As Beaton has already pointed out, the intensity of occupation apparently varied between strata, resulting in assemblages of different sizes and thereby making sample-size effects likely. The absence of particular implement classes followed no strong pattern but, rather, alternated with levels containing such specimens. Consequently, the sampling-influenced assemblage composition creates no complications for Beaton’s interpretations of this site. At other sites, however, the effects of sample size may be dire.
Morwood (1981) has presented detailed interpretations of the vertical changes in rockshelter sequences from the central Queensland highlands. Much of his discussion focused on the timing of the apparent introduction and eventual disappearance of implement types, particularly backed artefacts. Variations in the abundance of a number of tool types are presented graphically in Figure 2, summarising the main elements of the illustration provided by Morwood (1981:29, fig. 21). In his consideration of these data, Morwood concluded that a number of major changes were apparent, and it is worth considering his discussion at length.

Changes in the range and abundance of each implement type were identified by Morwood (1981:28–9). In his view, the Native Well 1 sequence displays a growth and then a decline in implement diversity. The lower spits (8–16) contain only three to five different ‘tool types’, a number Morwood described as ‘restricted’. Above spit 7, he noted that ‘many tool types unrepresented in earlier levels were introduced’. Forms that become particularly abundant in these levels include backed artefacts, blades and grindstones. As a consequence of the presence of these forms, the richness (i.e. variety) of categories increases above spit 7 (see Figure 2). At the top of the deposit, in spits 1–2, he identified a reversal of this trend, with richness decreasing as some types disappear from the record.

Differences in the quantity of archaeological material in each spit was also discussed by Morwood. He identified the major trend as a marked increase in the number of implements per level from spit 7 until spit 4, when numbers decline.

It is clear from Morwood’s discussion that these vertical changes in implement number and implement richness were seen as parallel but separate trends:

It seems that two types of change are being demonstrated in Fig. 21: changes in the implement range, and changes in the rate of implement deposition. Both lines of evidence have implications for the culture history of Native Well 1.

However, a re-examination of the Native Well 1 data reveals that these are not multiple changes but a single phenomenon, and that these changes are broadly explicable in terms of sample-size effect.

The size of the total assemblage varies markedly. Below spit 6 and above spit 2 there are fewer than
500–600 specimens in each spit. In contrast, spits 2 and 6 have 500–1000 specimens, while spits 3–5 have more than 1000. These vertical changes in artefact frequency are closely matched by changes in implement richness and the abundance of grindstones, blades and backed artefacts. In each case, the mode occurs in spit 4, with spits 1 and 7–16 having comparatively low values. Strong visual coincidence between all variables in Figure 2 is indicative of a causal relationship between assemblage size and the other variables.

The relationship between these variables and the number of artefacts in each spit is depicted in Figure 3, using scattergrams, regression lines and linear correlations coefficients. The strength of each relationship is measured by the high r² value, and in all instances p=<0.0005. Clearly, the abundance and diversity in each spit is tightly linked to assemblage size, and it is likely that assemblage composition at this site is largely a reflection of the assemblage size.

This conclusion has a number of significant implications for Morwood’s (1981:28) interpretations of the Native Well 1 sequence:

- The restricted range of implement types in the lowermost and uppermost spits is merely a product of the small number of artefacts in those levels.
- Conversely, the increased number of implement types above spit 7 is merely a sampling phenomenon resulting from dramatically larger sample sizes in those levels.
- Fluctuations in implement numbers, and the abundance of individual classes of implements, through the sequence is a reflection of fluctuations in the total sample.
- In any spit, the absence of specimens belonging to rare implement classes, such as grindstones, pirri points or backed artefacts, may relate to a low sample size.

It is this final implication that is the most significant, because it has substantial ramifications for the dating of the introduction of backed artefacts at the site. Morwood assumed that the lowest backed artefact recovered from the deposit represented the first use of backed artefacts in the site. On the basis of the one specimen in spit 7, immediately below a charcoal sample estimated at 4230±90 years BP (before the present), Morwood concluded that backed artefacts were introduced into the region 4100–4300 years BP (1981:30, 43; 1984:354, 357).
Figure 3
Native Well 1: scattergrams showing the relationship between number of artefacts per spit and (A) number of backed artefacts, (B) number of blades, (C) number of grindstones, and (D) implement richness
That interpretation is rendered problematic because of the small sample sizes in the spits underlying the lowest recovered backed artefact. Spits 8–11 have samples of only 250–410 each (X = 340). All spits containing backed artefacts have sample sizes larger than this (X = 1087). Furthermore, the regression analysis predicts no backed artefacts will be present in samples of 350 artefacts or less. Hence, there may be a threshold sample size, perhaps 300–400, below which it is statistically unlikely that backed artefacts will be present. Consequently, the existence of this sample-size effect means that the backed artefact in spit 7 should not be interpreted as marking the introduction of that implement class. It is statistically feasible that there may be earlier instances of backed artefacts, perhaps as old as 6000 years BP, that have not been recovered from this site for sampling reasons. At Native Well 1, the lowest backed artefact should therefore be considered to represent the minimum age of backed artefact usage.

A similar argument can explain the small number of backed artefacts in spits 1 and 2. Both of these spits have samples of more than 500 but less than 1000. The presence of backed artefacts, but in small numbers, in these spits is entirely consistent with a sample-size effect demonstrated above. Consequently, the recovery of backed artefacts from the upper levels of the deposit can be seen as a reflection of their use at the time those levels may have been created. This proposition is diametrically opposed to the interpretation offered by Morwood (1981:28), who suggests that these specimens ‘were probably not produced at this time, but have been scuffed up from earlier levels’. Notwithstanding the evidence for extensive vertical movement in Australian sandstone shelters (e.g. Richardson 1992), it is not necessary to invoke large taphonomic processes to explain the archaeological patterns at Native Well 1: those patterns are easily accounted for as sampling phenomena.

Shaw Creek KII

Excavations at Shaw Creek KII rockshelter, in the Blue Mountains west of Sydney, identified six phases, with Bondi points (a kind of backed artefact) being recovered only from phases I–III (Kohen et al 1984:63). Stratigraphic, radiometric dating and assemblage data were used to define the phase which contained evidence for the initial manufacture of backed artefacts. Phase III contained a charcoal sample yielding a radiocarbon date of 2235±120 BP (Beta–1210), and is described as younger than 4000 BP by Kohen et al (1984:62). This level contained four backed artefacts. The underlying phase IV deposit contains charcoal dated to 7860±220 BP (SUA-1398) and 4140±180 BP (Beta-1211), and represents sediment accumulation during the early to mid-Holocene. No backed artefacts were recovered from this level. Because vertical movement seemed unlikely, Kohen et al (1984:69, 71) concluded that backed artefacts were introduced into the sequence less than 4000 years BP.

In addition to the appearance of backed artefacts in phase III, this sequence displays parallel changes in artefact abundance. Those phases containing backed artefacts also contained thousands of artefacts (X = 2115), whereas the lower phases without backed artefacts contained only hundreds of artefacts (X = 481). A correlation of 0.88 between assemblage size and backed artefact numbers indicates a pronounced sample-size effect. Because the relationship between these variables is not linear, the strength of the relationship is better judged by the correlation between assemblage size and the square of the number of backed artefacts (r = 0.95, p < 0.025). On the basis of these coefficients, the contrast between phase III, which contains backed artefacts, and phase IV, which contains no backed artefacts, is explicable in terms of sample size. While phase III yielded 1239 artefacts, the underlying phase IV contained only 460 artefacts (Kohen et al 1984:63). On this basis alone, few backed artefacts would be expected in phase IV. It is therefore possible that backed artefacts may have been in use towards the end of phase IV, and yet not be represented in the excavated assemblage because of the small sample which was acquired from that level.

Capertee 3

Johnson (1979) has described the dating of backed artefacts at Capertee 3, in the Blue Mountains west of Sydney. The sequence in the upper metre of the deposit in square Q13, representing the last 5000 years, consists of higher artefact densities (>10/kg) in the top 40 cm and below 60 cm. Between 40 cm and 60 cm there are low artefact densities. The lowest backed artefact occurred 40–53 cm below the surface, in a level described by Johnson as ‘transitional’ between levels without backed implements (Capertian) and levels with backed artefacts (Bondaian). This transitional level contains only a few backed artefacts (Johnson, pers. comm.). Both the transitional level, and the level immediately below it, coincide with the
zone of low artefact density. It is therefore possible that the low abundance of backed artefacts 40–50 cm below the surface reflects low artefact densities (7.5/kg), while the absence of backed implements 50–60 cm below the surface may reflect the very low densities (4.7/kg). Hence, backed artefacts may have been in use prior to the formation of the transitional zone but be absent from the recovered archaeological assemblage because of sample-size effects. Johnson (1979:93) acknowledged this general problem, noting that for backed artefacts in the ‘transitional’ zone:

the lack of a large sample from excavation units in the ‘transitional’ zone renders their attribution to one or other industry rather difficult and thus hinders precise dating of the transition in my excavation.

Implications for dating backed artefacts

Throughout the 1980s it was common for Australian archaeologists to argue that (a) a relatively precise date, of approximately 4000–4500 BP, could be assigned to the appearance of backed artefacts, and that (b) in many regions backed artefacts disappeared from the archaeological record within the last 2000 years (see Bowdler 1981; Johnson 1979; Kohen et al 1984; Morwood 1984:355). Although dissenting views were occasionally published (e.g. Hiscock 1986; Hughes and Djohadze 1980), many of the carefully excavated sites dug since the mid-1970s produced sequences conforming to those interpretations. The four sites discussed above, and especially Capertee 3 and Native Well 1, were considered key evidence in favour of these propositions. However, for each of these sites, vertical variations in the abundance and presence/absence of backed artefacts and other rare items are strongly related to the size of archaeological samples. Absence of backed artefacts in levels immediately below the lowest backed artefact, dating to more than 4000–4500 BP, cannot therefore be taken as evidence that such implements were not used in earlier times. Nor can the lowest backed artefact recovered from these deposits be unambiguously interpreted as the first backed artefact made by occupants of the sites. As Hughes and Djohadze (1980) correctly observed, the likelihood that pre-5000 BP backed artefacts would be recovered in many sites is statistically low (see also Hiscock and Attenbrow 1998).

In a similar way, the demonstrated sample-size effects deny the common proposition that backed artefact manufacture everywhere ceased 1000–2000 years ago (e.g. Morwood 1984:359–60). As noted elsewhere (Hiscock 1994), there are many sites in which small numbers of backed artefacts have been recovered in recent levels. In the four sites reviewed here, the small number of backed artefacts found in near-surface spits can be explained in terms of the small assemblage sizes in those recent levels. This sample-size effect may also explain why excavated assemblages from some sites contain no backed artefacts, while assemblages from other sites do. Consequently, the absence or minimal representation of backed artefacts and other rare items in recent assemblages may largely be due to relatively low artefact densities and hence small samples.

Examples of spatial variation

In addition to issues of chronological change in the numbers of particular implement types, Australian archaeologists have examined intra- and inter-site spatial variations in the abundance of implements such as backed artefacts. Because these implements are rare components in samples of varying sizes, the spatial patterns are also subject to sample-size effects. To illustrate this effect, two examples have been selected: Smith and Cundy’s (1985) attempts to define the boundary of backed artefact and point distributions, and the identification of activity variation within the Shaws Creek KII rockshelter by Kohen et al (1984:67–8).

Implement distribution maps

On a continental scale, Australian archaeologists have often focused on inter-site comparisons involving the distribution of distinctive implement types that occur in the Holocene (e.g. Davidson 1983; Flood 1983; Hiscock and Hughes 1980; Mulvaney 1969; Pearce 1974). While there has been disagreement about the accuracy of these distributional maps, there is general acknowledgment that bifacially flaked stone points are restricted to northern Australia, and backed artefacts are generally found only in the south and east. Much of the debate concerns the location of the distributional boundary and the nature of that boundary. In a valuable attempt to place such inferences on a sound footing, Smith and Cundy sought to determine the location and nature of distributional boundaries by quantifying the abundance of points and backed artefacts within
assemblages in the Northern Territory. They concluded that backed artefacts were found only as far north as 20°S latitude and that points were found only in small numbers below that latitude. In addition, they raised issues of the boundary characteristics of these geographical distributions, suggesting that, while stone points have an abrupt boundary, backed artefacts gradually decline in density towards the northern margin of their distribution (Smith and Cundy 1985:36).

These conclusions were based largely on counts of implement types from either controlled surface collections or from excavated assemblages (Smith and Cundy 1985:32, table 1, and 33, table 4). A re-examination of the data provided by Smith and Cundy reveals that the size of assemblages varied greatly, and the abundance of either backed artefacts or points is closely related to assemblage size. Data from eleven sites south of the 20°S latitude line have a Pearson’s correlation coefficient between the number of backed artefacts and sample size of 0.90 \((p<0.001)\), while twelve sites north of that line have a Pearson’s correlation coefficient between the number of points and sample size of 0.96 \((p<0.001)\). This reveals a pattern in which the presence/absence or density of backed artefacts and points in any particular locality may not be accurately perceived from small samples. This is particularly the case with backed artefacts, which are rare or absent in Central Australian assemblages of less than 1000 specimens, even in the region below 20°S where they are known to occur. Hence, large samples are needed in order to accurately define the location and characteristics of the boundaries of backed artefacts and points.

Figure 4 shows the north–south variation in the size of assemblages used by Smith and Cundy (1985) to define the boundaries of these implement types. This figure is constructed by plotting all of their samples on a single north–south axis some 1700 km long. Each data point represents a 1:100,000 map sheet approximately 55 km long. The data points are connected by a smoothed line to illustrate the trends. Portrayed in this way it is clear that the assemblages immediately to the north and south of the 20°S line are extremely small. None of the map sheets within 300 km south or 450 km north of that 20°S line have samples above 1000 specimens. Furthermore, in the 250 km north of the 20°S line, Smith and Cundy (1985) had a total sample of only 92 artefacts. In view of small samples from these regions, the absence of rare implement types such as backed artefacts is not surprising; indeed it is to be expected, even if those kinds of implements had been made and used by local people. Consequently the location and form of distribution boundaries perceived by Smith and Cundy may be heavily affected by sample size. Until

![Figure 4](attachment:figure4.png)

*Figure 4*  
North–south variation in sample size employed in Smith and Cundy’s (1985) analysis of artefact distributions in the Northern Territory.
much larger assemblages, ideally containing tens of thousands of specimens, are studied the existing inferences concerning implements distribution boundaries must be regarded as tentative and potentially inaccurate.

Shaws Creek KII

Sample-size effects are also evident in spatial patterns within individual sites. For example, Kohen et al. (1984:67–8) have discussed the intra-site variation in assemblage form within the upper levels of Shaws Creek KII rockshelter. Their analysis broke the excavated squares into a transect of four categories running outward from the rear of the shelter, with the categories labelled: rear, central, dripline, and outside. Artefact densities, and consequently assemblage size, vary between these zones, being highest in the dripline and central zones, and lowest in the rear zone. In addition to differences in assemblage size, they observed differences in the number of specimens in each artefact type across the four zones. For example, they note that ‘back-blunted tools’ were more numerous at the dripline and outside the shelter. They also argue that ‘blades, pointed flakes and faceted buttts are slightly more in evidence in the dripline zones’.

To Kohen et al., these spatial differences in the abundance of artefact classes suggested activity differences within the site. They concluded:

The outer zones show an emphasis towards making small tools and, to a lesser degree, on more sophisticated flaking techniques. This may be taken as a sign of ‘liberal activities’ in stone working ... with the focus of attraction being the light and warmth of the sun at the front of the shelter. (1984:71)

This reference to ‘making small tools’ at the front of the shelter is based on the perception of a dual trend of greater artefact numbers and greater quantities of backed artefacts and backed artefact manufacturing debris near the dripline. A re-examination of these data indicates that this pattern is not a dual trend. Instead, there is a single trend of spatial variation in artefact densities, and the compositional differences are merely sample-size effects reflecting the differences in artefact density.

This conclusion is evidenced by a strong positive relationship between the abundance of each class of artefact and sample size. For example, the Pearson’s correlation coefficient for artefacts per zone and backed artefacts is 0.86 (p<0.025), for artefact numbers and blade numbers r=0.92 (p<0.05), for artefact numbers and pointed flakes r=0.89 (p<0.05), and for artefact numbers and flakes with faceted butts r=0.95 (p<0.025). These correlations suggest that, while the rate of artefact discard may have varied across the site, differences between zones in the quantity of specimens in each artefact class should not be seen as activity differences. Such variations that exist may simply result from dissimilarities in the size of artefact samples.

Discussion

Recognition that some aspects of assemblage composition are highly correlated with, and may be determined by, assemblage size has several obvious implications for the practice of artefact analysis in Australian archaeology. Many of the case studies employed here derive from excavations done during the last 20 years, rather than during the pioneering days of the 1960s and early 1970s. Detailed excavations of the late 1970s and 1980s have, in return for the smaller spits and more precise excavations, paid the price of smaller excavated volumes. Consequently, the assemblage sizes produced in recent decades are typically small compared with the recovered assemblages of an earlier generation of researchers. As a result, recent excavations are more prone to pronounced sample-size effects than earlier ones.

Clearly, for some research objectives the existence of pronounced sample-size effects will impose a minimum excavation area if there is to be a high probability of recovering rare items. For reasons explained above, this minimum ‘effective’ excavation area may vary regionally or through the archaeological sequence as the rarity of classes of material varies. No standard excavation area can be reasonably proposed on theoretical grounds alone.

In addition, sample-size effects of this kind imply that identification of rare elements is not solely dependent on excavation technique but may be proportional to artefact density in a site. Hence, the precision with which we can assign dates to the use of rare implements at a site is not simply limited by excavation precision, but in some sense is intrinsic to the nature of the archaeological deposit. Consequently, by virtue of their small assemblages, some sites may be poor choices for addressing questions concerning rare items, irrespective of the
care taken during excavation. Sites such as Native Well 1 might fall into this category because, at levels representing the early and late Holocene, artefact assemblages are small. A similar conclusion must apply to the search for examples of rare archaeological classes in surface sites with small and/or low density assemblages. As demonstrated above, the samples within a few hundred kilometres of 20°S used by Smith and Cundy (1985) are too small to have a high probability of containing rare elements such as backed artefacts.

Equally as important is the implication for the way in which archaeologists compare sites and time periods. As described above, a common procedure in Australia has been to examine inter-site differences by comparing samples of two or more sites, and to consider questions of chronology by comparing different levels within a single deposit. Where such comparisons focus on the presence/absence of rare kinds of artefacts, including measures of the diversity of the assemblage, interpretations must consider the impact of samples of dissimilar sizes. For instance, interpretations of site function, so commonly based on the richness of artefact categories and expressed in terms such as ‘base-camp’ or ‘dinner-time camp’, must accommodate Thomas’ (1989) observation that it is not richness but the relationship between sample size and richness (the regression slope, to be precise) that will be a sensitive measure of the range of activities represented at a site. Without such considerations it will be difficult to avoid unproductive discussions of patterns which are merely a construct of our sampling strategies. Frankel (1988) has explored this phenomenon in Australian archaeology.

One obvious response to the existence of sample-size effects is to focus archaeological analyses on proportionately numerous rather than rare classes of material. In the case studies examined here, that might involve quantifying the debris created in manufacturing points and backed artefacts, rather than only the implements themselves. This possibility has been recognised by a number of Australian researchers. For example, Jones and Johnson (1985) attempted to identify point manufacture through the retouched flakes found in the Nauwalaliba 1 deposit, although the characteristics they examined were incapable of distinguishing bifacial thinning flakes from other debris (Hiscock 1993).

While technological studies of assemblages might be used to overcome such recognition problems, it is likely that a single implement class was manufactured in a variety of ways (see Hiscock 1986), making it difficult to derive a single debitage characteristic which can reliably indicate implement manufacture in the absence of the implement. Furthermore, there is no reason to think that manufacturing backed artefacts or points was more than a minor form of stoneworking at many sites, and hence debris from that manufacture may still be relatively infrequent. As a result, questions concerning the presence/absence of rare items, such as stone implements, might not easily overcome sample-size issues by simple studies of other elements in the assemblage.

Another possible response to the existence of a pronounced sample-size effect is to increase the volume excavated or the surface area collected, thereby increasing the size of the sample. The extent to which an expansion of the area of excavation/collection might be necessary would be determined by the rarity of the classes of objects of interest, and would need to be determined empirically in each instance. Although this may be a costly solution to the problem, it may often be effective. However, this response has implications for cultural resource management. For example, where site destruction is imminent, the scale of salvage investigation ought to be sufficient to identify the nature of sample-size effects at that site and to recover rare elements if that is the objective. Since future examination of the site may be impossible, these investigations must be pursued during the salvage phase if the results are to yield reliable interpretations.

The existence of sample-size effects has a second implication for cultural resource management. Site recording during archaeological consultancies in Australia has often relied on the measurement of small samples, perhaps fewer than 100–200 specimens. Indeed, Australian archaeologists often act as though a sample of 30–100 specimens will yield an accurate depiction of assemblage composition. Where this practice has been employed, it is possible that rare classes of archaeological material that are actually present within a site or region may not have been recognised, thereby creating a false impression that a number of archaeological assemblages are extremely similar. Conversely, it is possible that, with small sample sizes, variations in the relative abundance of rare classes might be portrayed as presence of those classes at some sites and absence at others. Since significance assessments are often based on the uniqueness of assemblages, focusing on the
degree to which a site can contribute knowledge which no other site can (Bickford and Sullivan 1984:23), sample-size effects of these kinds may affect the perception of scientific significance.

While increasing sample sizes may be feasible in some circumstances, and may be especially desirable in the case of endangered sites, this will not always be possible. In situations where statistical analyses revealed strong positive relationships between assemblage size and the abundance of any class of artefact, it would be appropriate to presume that the variations being recorded were merely related to sample size. This proposition carries two implications. First, correlations between sample size and artefact assemblage composition should be examined before interpretations are developed (see James and Davidson 1994). And, second, where this effect cannot be removed by further sampling, or by more elaborate analyses, the safest course of action would be to exclude those sites from consideration when constructing interpretations of those aspects of archaeological change.

These considerations have other, equally fundamental, consequences for conventional approaches to the interpretation of assemblage variation and site function. Discussions of artefact use and site function in Australia have often been susceptible to the existence of sample-size effects for two reasons. First, many researchers have persisted in focusing on implement types and in equating these morphological forms with 'tools', despite evidence to the contrary (see Hiscock 1998), thereby constraining their discussion of artefact use to rare items. Second, discussions of site function have typically presumed that every assemblage is a comprehensive and representative record of activities carried out in or near the site from which it was recovered.

Together these propositions have made conventional interpretations of site function highly susceptible to sample-size effects. In particular, it could be expected that many smaller assemblages may not contain artefacts relating to all activities that were carried out in the vicinity. Variation in the size of recovered assemblages, or regional/chronological differences in the relative rarity of each class of object, may therefore be factors contributing to inter-assemblage differences in typological composition. The implications of this principle for existing interpretations of functional variations in Australian prehistory will need to be carefully evaluated.

Conclusion

Many of the patterns in the archaeological record on which archaeologists have focused are related to sample size. The issues emphasised here, such as dating of the time periods in which particular implement types were made and the geographic distribution of each type, cannot be adequately addressed without considering sample size, and in Australia that has not always been done. Consequently, many of the conclusions of researchers over the last few decades are in need of revision.

One revision suggested here relates to the typical interpretations of backed artefacts in the archaeological sequence of Australia. Throughout the last two decades it was commonly held that the comparatively small number of backed artefacts found in contexts dating to before 4500 BP or to recent millennia are actually in those stratigraphic positions as a result of vertical movement within deposits (e.g. Bowdler and O'Connor 1991; Johnson 1979; Morwood 1981, 1984). And yet, as demonstrated here, the levels yielding backed artefacts from 'early' or 'late' contexts often have comparatively small assemblages. This provides an alternative explanation for this aspect of the Australian archaeological record, in which backed artefact or point abundance in Holocene deposits is often tied closely to artefact density and hence sample size. In the light of such a correlation, backed artefacts or points associated with pre-4500 BP dates, or post-1000 BP dates, cannot be dismissed out of hand. While the stratigraphic position of such specimens may have resulted from vertical movement, this must be demonstrated rather than assumed. Furthermore, the common interpretation of the lowest backed artefact or point in a deposit, as being representative of the first manufacture/use events involving that implement type, is often without justification. Where underlying excavation units contain small sample sizes, the lowest backed artefact/point must be used to indicate the minimum antiquity of that type, unless evidence of vertical displacement is forthcoming.

Examples used here indicate that the influence of sample size on assemblage composition has generally been underestimated in Australian archaeology, especially by researchers investigating spatial and temporal change in Holocene stone artefacts and reconstructing site function. More sophisticated depictions of change in the archaeological record, incorporating an understanding of the effect of
sample size and a movement away from an exclusive focus on implement typology, are a prerequisite for sophisticated explanation of those changes.

ACKNOWLEDGMENTS

I thank Kim Akerman, Val Attenbrow, Peter Veth and Ian Walters for their discussions on the subject of sample size. I owe a special note of thanks to Barry Cundy for his incisive comments on the arguments presented here. Students in the ‘Understanding early technology’ unit at the Australian National University have patiently listened to my rantings on this subject and I appreciate their good humour.

NOTE

1. The backed artefact category used here incorporates four of Morwood’s (1981) classes: geometric microliths, Bondi points, backed points, and miscellaneous backed. The grindstone category used here incorporates both grindstones and mullers.

REFERENCES


Hiscock, P. 1986 Technological Change in the Hunter River Valley and Its Implications for the Interpretation of Late Holocene Change in Australia, *Archaeology in Oceania* 21, 40-50.


Hughes, P.J. and V. Djehardze 1980 Radiocarbon Dates from Archaeological Sites on the South Coast of New South Wales and the Use of Depth/Age Curves, Occasional Papers in Prehistory, Department of Prehistory, Australian National University, Canberra.


Richardson, N. 1992 Conjoin Sets and Stratigraphic Integrity in a Sandstone Shelter: Kurniff Cave (Queensland, Australia), *Antiquity* 66, 408–18.


62 *Australian Aboriginal Studies* 2001/1