INTERPRETING◇THE◇PAST

RADIOCARBON DATING

Sheridan Bowman

University of California Press/British Museum

University of California Press
Berkeley and Los Angeles

Designed by Andrew Shoolbred
Cover design by Slatter-Anderson

Printed in Great Britain

Library of Congress Cataloging-in-Publication Data
Bowman, Sheridan.
Radiocarbon dating/Sheridan Bowman.
p. cm. - (Interpreting the past series ; 1)
ISBN 0-520-07037-2
1. Archaeological dating. 2. Radiocarbon dating.
I. Title. II. Series.
CC78.B68 1990
930.1'028'5—dc20 89-48681
CIP

Cover illustration: Section of the Belfast high-precision calibration curve for the recent past (courtesy of Gordon Pearson; see p. 46). The curve is superimposed on a section of a wooden timber which shows the annual growth rings of the tree. Tree rings are the basis of dendrochronological dating, which is used to provide the horizontal calendar-time axis of the calibration curve, here in years AD; the vertical axis is radiocarbon time BP (see ch. 4).

— 5 —

Radiocarbon and Archaeology

The archaeological record is an incomplete and fragmentary version of past human activity. What was deliberately, inadvertently or incidentally left behind is only a part of the material aspects of that activity, and this partial record has itself been subject to the vagaries of preservation and subsequent natural or human activities. The archaeologist is therefore faced with an incomplete and unrepresentative set of data from which a coherent whole must be inferred. A process of logic is used to link past events with contexts and features, such as stratigraphic levels and post holes, and to link these with artefacts found within them. If the artefact is organic it can be radiocarbon dated, but it is rare that a date for the artefact *per se* is required; instead it is assumed that the radiocarbon result will also date the event.

In many cases this may not be an unreasonable assumption. In the dating of a bone from an articulated skeleton in a grave, the assumption of association of sample and context (i.e. bone and grave) and of contemporaneity of sample and event (i.e. bone and burial) are good. All too often, however, if the samples submitted for dating are even to begin to answer the chronological questions being posed, the stages of inference linking event with context and context with artefact need more careful examination, together with the implications of what is represented by the ^{14}C activity of a sample. Liaison between archaeologists and radiocarbon scientists is therefore required from the planning stage of an excavation in discussing what radiocarbon can and cannot do, as well as practicalities such as sample size and packing. The better the liaison before and during excavation, the more likely it is that a useful series of samples will be processed.

The following sections elaborate on these points for the user, or potential user, of radiocarbon dating.

The axiomatic sample–context relationship

Deposition of any organic material in the ground obviously postdates the formation of that material and the cessation of its exchange with the biosphere. All radiocarbon age offsets make samples older than their usage or removal from the biosphere, and some, such as marine and 'old-wood' effects, make them substantially older. The exception is contamination, which can make samples appear older or younger, but pretreatment is designed to remove this. Furthermore, all depositional processes, other than downward movement as through animal burrowing or root action, are such that a date for a sample pre-dates the context in which it was found. Hence all radiocarbon samples provide a *terminus post quem* ('date after which') for their find context. How much they pre-date the deposit depends on both the nature of the sample and the taphonomic processes involved.

The 'old-wood' problem

Samples can appear to have a significant age at death due to reservoir effects such as hard water, or marine or volcanic origin of its carbon (see ch. 2). However, the more commonly encountered cause of an apparent age at death is when the organism ceased exchange with the biosphere before death, as in the case of wood (see ch. 1).

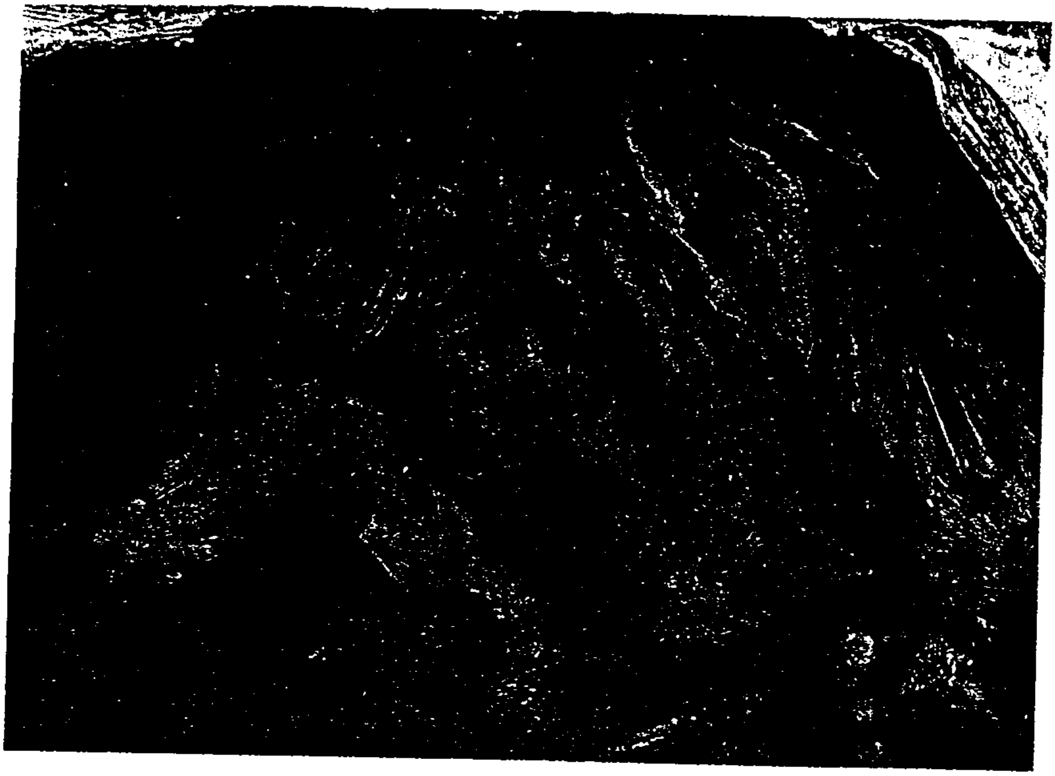
Great care must be exercised in the selection of wood or charcoal for radiocarbon dating. If the sapwood to heartwood boundary is identifiable, the age offset can be estimated using ring counts, or can be minimised by dating sapwood alone. Indeed, if sufficient rings of appropriate wood are present, dendrochronological dating may be better than radiocarbon (see ch. 4). Alternatively, twiggy material (identifiable if the complete cross-section is present by the presence of sapwood, the small number of rings and the curvature of the sample) is best since the age offset will then be small and seasoning or re-use of such material is unlikely. It is highly advisable that a specialist identify the tree species from which the wood or charcoal derived, since this will indicate whether the species was long-lived and hence whether a significant age offset is likely. If a mixture of species is represented, short-lived ones can be separated out and dated.

When there is no alternative to dating material derived from long-lived species, it is important to ask whether the result will be useful and therefore whether the sample is worth submitting. In some circumstances mature oak may be quite helpful in providing an approximate date for a monument. However, a sample of a long-lived wood species should not be considered if it overlies the context to be dated. Samples with an unknown age offset cannot provide a *terminus ante quem* ('date before which') for the deposition of the underlying context.

Quite often this 'old-wood' problem is inadequately considered by those who submit radiocarbon samples. Perhaps if bristlecone pines and yew trees, with potential longevities of about 4000 and 1000 years respectively, were to feature more in the archaeological record, the problems would be more readily appreciated!

Association

Apart from the importance of dating adequately sealed and unmixed contexts, there are also various calibres of association between the sample and the event to be dated. These were elucidated in the early 1970s by H. T. Waterbolk, a Dutch archaeologist, but his sound ideas often seem to be overlooked in the



21 In 1984, following peat-cutting operations, the upper body of a man was found at Lindow Moss (near Wilmslow in Cheshire, England). Owing to the preserving properties of the peat, a range of forensic as well as archaeological techniques could be applied, and it was discovered that Lindow Man appeared to have been ritually murdered. He had been garrotted, his throat cut and he had also received two severe blows to the head. Radiocarbon dating was applied to small samples of various types from the body itself, and two techniques, accelerator mass spectrometry (AMS) and mini-gas counting, were used. Unfortunately, although the association between the samples and the event to be dated were good, the agreement between the two techniques was not. AMS suggests Lindow Man was killed sometime in the first century AD, whereas mini-gas counting suggests that the event occurred some three or four centuries later: this is surely a mystery equal to that of the motive for the murder itself!

pursuit of dates. The best association is obviously when a date for the sample itself is required and age offsets are small. For example, in the dating of a bog body such as Lindow Man (fig. 21), a date for the body is required rather than a date for the bog in which it was found. The most dubious of associations can arise because the processes by which the sample and deposit have been brought together are ill defined or poorly understood. This is exacerbated by situations where dispersed material is bulked together to provide a 'single' sample for dating.

Mobility of samples is also a factor that needs to be considered now that facilities exist for processing very small samples. A small fragment of bone is more susceptible to movement by natural and anthropogenic mechanisms than a large bone and should not be dated in preference simply on the basis of size. If there is some reason for not destroying the intact bone, then a small sample can be taken from it for AMS or mini-counting. On the other hand, there may be good reasons for dating single grains, despite the danger of mobility, if the grain is identified to species and its presence in the context is of major agricultural significance.

The archaeologist is of course best placed to judge the reliability of association of sample and context, using the guiding principles of definable archaeological processes, selection of coherent samples rather than bulked scatters, and assessment of the likelihood of intrusive material.

Delayed use, re-use and residuality

Age offsets inherent to the sample material have already been discussed (see p. 51). Here offsets are considered that are some function of past human behaviour. The effects of these depositional processes are by no means quantifiable, but each can result in a sample giving substantially too great an age for the context being dated, even when the apparent association is good.

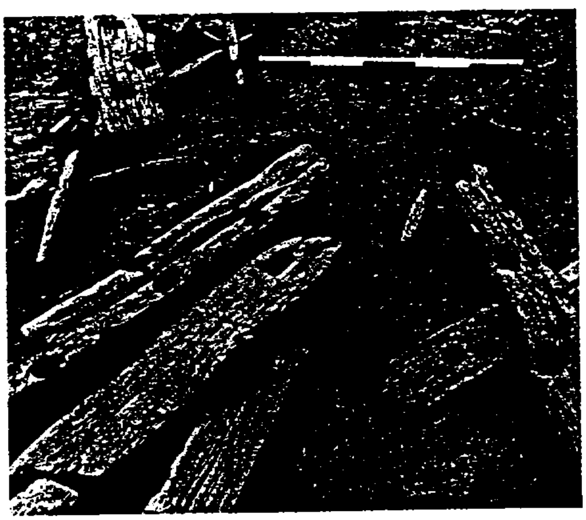
Delayed use

The idea of delayed use is familiar for wood, where seasoning might be involved prior to actual use of the timber. A less obvious example is the use of driftwood, particularly where indigenous building material is scarce. Here the identification by species might indicate the use of a foreign wood.

The custom of peat burning could also give large offsets, due to the use of aged material, if sediment samples from some sites were dated. The same would apply to coal, though here the radiocarbon age of the material is infinite, as it is for bitumen, a natural product of coal deposits. The use of bitumen is known at some Neolithic sites in the Near East, being used for decorative purposes as well as utilitarian ones such as the water-proofing of baskets.

Re-use

As the historic buildings of the relatively recent past demonstrate, hardwoods in particular are resilient to decay and the re-use of large timbers in rebuilding



22 Re-used timber found in association with a Bronze Age trackway at Withy Bed Copse in the Somerset Levels (England). Morticed timbers and worked wood have been found as 'make-up' beneath Bronze Age trackways of both the Somerset Levels and Ireland. Working would have had no utilitarian function in laying or stabilising the track; rather it indicates re-use of material from defunct structures. Often, however, re-used materials do not bear recognisable signs of a former use and might mistakenly be assumed to be contemporary with the context.

Waterlogged samples are an interesting problem: should they be dried before submission or sent wet? Charcoal should be dried. Bone if not too friable should be washed and dried, but remember to do this at normal ambient temperatures: drying in an oven will degrade the collagen component without which the bone cannot be reliably dated. Both waterlogged wood and peat (fig. 24) should be submitted wet, and if there is likely to be a long time between the collection of the sample and submission for dating, it should be frozen to avoid unpleasant infestations. If peat is dried it becomes impossible to distinguish the modern rootlets from the structure of the sample. The reason for keeping a large quantity of wood wet is that it is very hard to break down into small pieces if it has once been waterlogged and is then dried, thus making the pretreatment procedure more difficult. However, some laboratories may not mind this.

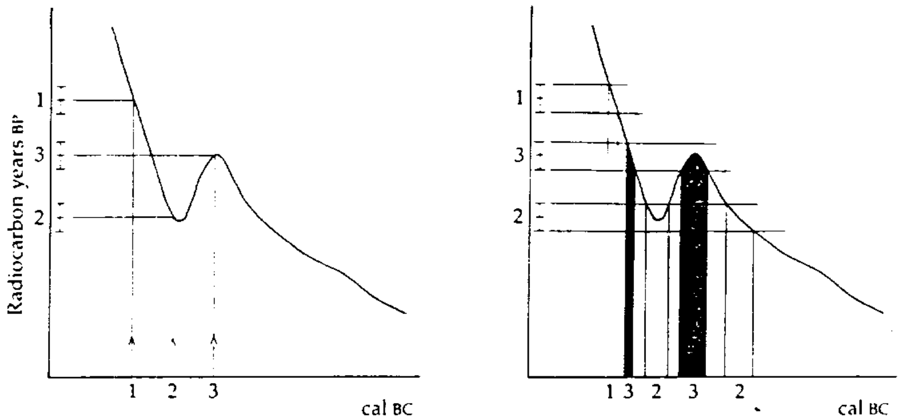
Remember that a laboratory will look more kindly on samples if they are not accompanied by a ton of soil. A certain amount of physical precleaning of samples can be done before they reach the radiocarbon laboratory, such as concentrating charcoal by extracting it from earth using metal tweezers, but if it is a widely disseminated sample, the first question is whether it is even worth dating. Perhaps most important to check with a laboratory that is likely to be dating the samples is what they would prefer for each type of sample.

Using radiocarbon results

Rarely is the interpretation of radiocarbon results completely straightforward. Occasionally a sample is dated simply to determine roughly whether an object is modern or of considerable antiquity; in essence, an authenticity test. Even then the answer may not be clear cut if, say, an old timber has been recently carved to produce an authentic-looking sculpture! In archaeology, the questions are often quite complex, involving non-contemporary samples. The difficulty arises from the necessity to calibrate radiocarbon results and the form of the calibration curve, precluding both the use of normal statistical tests to answer such questions and the use, other than in the broadest sense, of radiocarbon dating as a relative dating method.

Radiocarbon and relative dating

Prior to an agreement on which calibration curve to use, many archaeologists took the pragmatic approach of working in uncalibrated radiocarbon results rather than calibrate only to find that recalibration was necessary the next time a new curve was produced. This approach has, however, led some users of radiocarbon results to hold a rather spurious belief in a radiocarbon timescale that can be used as a relative dating technique. Unfortunately, this is only the case in a rather limited sense. There are several periods in the calibration curve where events that are separated in calendar time by several centuries appear contemporaneous from their radiocarbon results. The worst of these (see fig. 23) is for the period corresponding to the British Early Iron Age (*c.* 800–400 BC). There are also periods where the curve is steep, so that an apparently large difference in radiocarbon results arises from events separated by relatively small amounts of real time. There could even appear to be an inversion of events if the calibration curve is particularly wiggly and the error on the results is sufficiently low (fig. 25). Radiocarbon results can also appear



25 If radiocarbon results are used for relative dating, they can falsely suggest approximate contemporaneity of non-contemporary events and vice versa; even apparent inversion of events is possible. In the first diagram a perfectly feasible, if perhaps infrequent, situation is illustrated. Three events, equally spaced in calendar time, are radiocarbon dated with reasonable precision (say, ±40 years). Due to the wiggly nature of the relevant portion of the calibration curve, these events appear in a different order on the radiocarbon timescale. Only when the radiocarbon results are calibrated, as shown in the second diagram, is it apparent that confusion over the order of events 2 and 3 is possible (for the sake of clarity, the calibration of the result for event 1 is not shown). If these events have an archaeologically unequivocal stratigraphic relationship, then it would be possible to eliminate some of the calendar ranges. Such a stratigraphic relationship would also have demonstrated, even prior to calibration, that radiocarbon results do not necessarily offer relative dating, particularly for sequences covering a short time span.

to bunch around a temporal hiatus. The radiocarbon timescale continually compresses or stretches real time so that great care has to be exercised in using it for relative dating, particularly over a timescale of only a few centuries.

Combining results

Replicate measurements

If more than one radiocarbon measurement is made on a single sample, these replicate results can usually be combined. Of course, the sample itself should not represent an age span; if it does, then the various measurements will only be true replicates if the same age range is measured, for example, taking sections of the same tree rings from a large timber.

To quantify possible non-consistency of the results, a chi-square test can be done. This tests whether or not the variability of the results amongst themselves is consistent with the individual quoted error terms. If the variability is substantially larger, then, assuming the errors are correct, the results are not consistent with dating of a single sample and it is not valid to combine them. This could happen if the measurements were on different chemical fractions of a contaminated sample and then perhaps, strictly speaking, they should not be considered as replicates.

The test statistic is

$$\chi^2 = \sum \frac{(t_i - t)^2}{\sigma_i^2}$$

where t is the pooled mean of the individual radiocarbon results t_i and is given below, and the symbol Σ denotes summation of all the terms. The calculated value of χ^2 is looked up in a set of chi-square tables to determine if the variability is too great to be attributed to chance, given the errors involved. If the test is passed, in other words if the results conform to a normal distribution potentially representing a single radiocarbon 'age', then the results can be combined.

The formula for combining a number (n) of replicate results ($t_1, t_2, \ldots, t_i, \ldots, t_n$) is

$$t = \frac{\Sigma\, t_i/\sigma_i^2}{\Sigma\, 1/\sigma_i^2}$$

This formula 'weights' results according to their associated error term σ_i; more weight is placed on results with small error terms than on ones with large errors.

The error σ on the pooled mean is given by

$$\sigma = \sqrt{\left(\frac{1}{\Sigma 1/\sigma_i^2}\right)}$$

In the situation where all the σ_i values are the same, the formula reduces to the familar one for averaging

$$t = \Sigma\, t_i/n$$

and for σ

$$\sigma = \frac{\sigma_i}{\sqrt{n}}$$

Replication should not be undertaken simply to achieve higher precision (see p. 40). Often multiple dating of a sample is done where contamination is suspected and different chemical components are extracted.

Combining results from different laboratories

If different laboratories, whether conventional or AMS, produce a radiocarbon result for parts of the same sample, their results can be combined as outlined above provided none of the laboratories has a systematic bias and each evaluates its error terms in approximately the same way. Significant systematic errors will become evident on applying the chi-square test, as might differences in error evaluation. If there is some doubt about how the laboratories have estimated their quoted errors, the variability of the results relative to each other can be used, without weighting, to provide a mean result, a standard deviation and hence a standard error on the mean. The usual error estimation process (see p. 38) treats the error as if known. However, in this situation, uncertainty in the error evaluation can be taken into account by use of Student's t-distribution, rather than the Gaussian. Where n results are involved, the mean is

$$t = \sum \frac{t_i}{n}$$

The standard deviation is

$$\sigma_{sd} = \sqrt{\left(\frac{(t_i - t)^2}{n - 1}\right)}$$

and the error on the mean is

$$\sigma_{se} = \frac{\sigma_{sd}}{\sqrt{n}}$$

This is the error that should be used in conjunction with Student's *t*-distribution to provide the confidence levels for the true result.

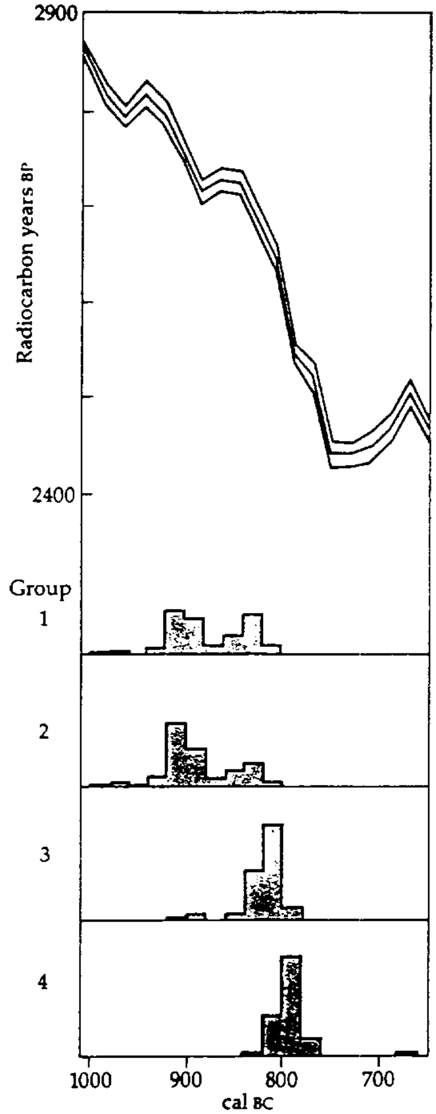
Combining results from different samples
If there are several samples from the same context and this is believed to represent a short-lived episode, then, with certain provisos, the results can be combined as outlined above. It must be remembered that the radiocarbon result for the sample is not the same as the radiocarbon age of the context. Re-use, residuality and age-offsets can all play a part. In combining results, those on large timbers, for example, should not be combined with those on bone from the same context. Usually this will be apparent from the radiocarbon results themselves, with the former being noticeably older than the latter (in fact, this situation should not normally arise, since the bone samples should be dated in preference to large timbers where the choice is available). However, even the results from several bone samples from the same context need to be considered with care; in particular, some samples may be residual. A chi-square test should be done and, if the results do not pass the test, they are not consistent with dating of a single episode and it is not valid to combine them.

If the test is passed, in other words if the results conform to a normal distribution potentially representing a single radiocarbon 'age', it must not be assumed that this proves the samples are from the same age population. Rather this evidence *together with* the archaeological evidence indicates comparability. It is also advisable to consult the calibration curve just to see if it is possible that non-contemporary events in calendar years might give effectively the same radiocarbon results, as for example in the period 800–400 BC (see above and fig. 23). Again, results from different laboratories, whether conventional, AMS or some from each, can be combined in this way if the conditions are satisfied and their errors are fully evaluated.

In any combining of results, reservoir corrections and corresponding adjustment of the error term should be done first, but the error term on the calibration curve does not enter these calculations because the assumption is that the samples would have the same ^{14}C content if in equilibrium with the atmosphere. The error on the curve must be taken into account, however, when the pooled mean result is subsequently calibrated.

Comparison of results for different episodes/events

If results are obtained for different events or are shown to be inconsistent using the chi-square test, what statistical procedures can be done on them? Here the difficulty is that an age difference is indicated, but the true magnitude of that difference cannot be evaluated until the individual radiocarbon results have been converted to calendar years by calibration. Of course, as soon as this is done, the Gaussian probability distribution of the uncalibrated result is replaced by graphical date ranges such as those illustrated in figure 20. Statistical tests cannot then be done. It is not valid to perform the tests first and



26 Probability calibrations for different phases of the waterfront activity at the Runnymede Late Bronze Age settlement. These are based on twenty-six radiocarbon measurements relating to four discrete events stratified in one deep sequence; one event, the construction of the inner palisade, is represented in fig. 24. Each group of measurements is on material from *in situ* wooden structures or short-lived events, and virtually all measurements are on young wood (less than thirty years' growth), so that there is minimal risk of an age offset between calibrated dates and the true date of the events in question. Within each group the results are consistent both statistically and archaeologically with a single event. Probabilistic calibrations are therefore shown for the mean radiocarbon result for each event. The relevant portion of the calibration curve (illustrated at the top of the diagram) is steep, although a minor wiggle in the curve at 2700 BP leads to bimodal calibrated ranges for groups 1 and 2. It can be seen that the calibrated date ranges correspond well to the sequence observed in the stratigraphy and suggest a time lapse between events 1 (the earliest) and 4 (the latest) of approximately a century and possibly even as little as forty years.

then to calibrate. Nor is it valid to perform the tests and work only in radiocarbon results, because the radiocarbon year is not a true unit of time but is variable in length as previously discussed.

Graphical representations of the calibrated results will help, using either calendar date ranges if the intercept method has been used, or cumulative probability distributions if a probability method has been used (fig. 26). Unfortunately, these do not allow a succinct quantification of the data; for example, phase duration cannot be simply enumerated. However, even without the difficulties caused by calibration, questions such as phase duration involve problems that are inherent to some degree in all sampling of archaeological sites. The underlying assumptions are that the radiocarbon samples selected for dating are representative of the chronology of the archaeological record (for example, they are not biased to earlier or later periods) and that in turn the archaeological record is representative of past human activity; each inference may in reality be a considerable leap.

Rejecting radiocarbon results

In the datelists published in the journal *Radiocarbon*, submitters provide a brief comment on how the radiocarbon results compare with the archaeology and therefore with expectation. Comments such as 'archaeologically acceptable', while not very informative, are less frustrating than the bald 'archaeologically unacceptable' statements. Often there is no discussion of these 'unacceptable' results; they are simply rejected by the archaeologist when evaluating the chronology of the site. Such unexpected or anomalous results can, however, be of great value. For example, they might alert the user to a problem with the laboratory (or vice versa!). Alternatively, they might indicate one of a multitude of depositional problems, such as that the samples selected were residual or that there was unsuspected contamination. These 'unacceptable' results, perhaps more than any others, need careful consideration: they may provide the greatest information.

Sampling strategy

Radiocarbon dating anything and everything, just because it is there and because it is organic, is not a sampling strategy! The literature abounds with results that are of little or no use to archaeology as a result of this 'policy'. Some of the problems of radiocarbon dating and how archaeological depositional processes might affect selection of samples for dating have already been discussed. In summary, any strategy should:

- Involve the radiocarbon laboratory at an early stage.
- Ask how the context yielding the sample relates to the event that is to be dated, how the context was formed and what it means.
- Ask how the sample relates to a given context: is there good association, is the sample representative, is its deposition contemporary within reasonably narrow limits with the context?
- Ask if the ^{14}C activity of the sample is relatable to the time of death of the plant or animal from which it is derived, or whether there is an age offset and, if so, if it is acceptable.
- Ask if the contexts being sampled adequately represent the human activity that is being studied.
- Ask whether radiocarbon results after calibration can provide the resolution needed to answer the archaeological questions being posed.

Used well, radiocarbon is a very powerful and widely applicable technique, invaluable to our understanding of the unwritten past.