[Special Papers in Palaeontology, 73, 2005, pp. 57–71]

# BIASES IN THE RECOVERY AND INTERPRETATION OF MICROPALAEONTOLOGICAL DATA

*by* LENNART JEPPSSON

Department of Geology, Sölvegatan 12, SE-22362 Lund, Sweden; e-mail: Lennart.Jeppsson@geol.lu.se, Lennart.Jeppsson@telia.com

**Abstract:** Bias caused by collecting and processing bulk samples is largely independent of what fossil clade or mineral is searched for. Instead, different methods bias the data to a different, frequently very large, degree. Furthermore, biases accumulate with each recovery step, and the sum may be extreme unless appropriate methods to minimize it are employed. The effects depend on what the data are used for, e.g. establishing range ends (zonal boundaries), taxonomy (frequencies as an aid to, for example, conodont apparatus reconstruction) and ecology (relative frequency in a fauna, frequency/kg, faunal diversity). However, the best published methods remove calcium carbonate and dolomite without bias. All rocks with such cement can be broken down without bias, and so can some claystones with little lime. The bias caused by concentration can be measured, kept low, and documented. Removal of clay is an exception: screening or decanting removes all small elements. Extraction methods should be stated in all publications so that the data can be assessed more fully and quoted properly. Information about the acid methods, screen hole diameter and collection size are especially important because these usually cause the greatest bias. Reliability of observed range ends increases with increasing number of specimens and with decreasing sample distance (recollecting near the boundary). Samples that are too small, yielding subadequate collections, can strongly bias placement of zonal boundaries and implied diversity. Not taking the uncertainty intervals of zonal boundaries into account may result in artificially extended observed ranges of other species. Methodological progress over the last 25 years has increased the potential average yield per hour of manpower over 100 times for samples yielding fewer than 100 elements/kg. This has made it possible to overcome most of the biases outlined herein. Similarly, taking the biases that are known or are likely to have affected data into account allows levels of precision in data to be evaluated and published.

**Key words:** bias, microfossils, conodont, collecting, laboratory work, analysis, methods.

FROM sampling to picking to final analysis, the methods by which conodont elements are recovered for study influence what is recovered. The observed record thus deviates more or less from the preserved one; that is, it is biased. Some kinds of bias are indiscriminate, affecting all elements to the same degree, but most discriminate strongly. In order to minimize bias and to document bias that is unavoidable, it is necessary to identify the cause and magnitude of bias. Most studies of conodont recovery have been directed towards maximizing yield, and our knowledge of bias is concerned largely with minimizing loss during recovery rather than as a topic that has been studied in its own right. Those parts of this account that deal with the destructive effects of unbuffered acids are based on controlled experiments and are thus empirically scientific. Anyone doubting the conclusions drawn from these experiments can and should repeat them (the critical test takes less than half an hour of work to repeat). In accordance with this principle, other workers (e.g. R. J. Aldridge, pers. comm. *c.* 1981/82) quickly confirmed my discovery of the destructiveness of unbuffered acetic acid (Jeppsson in Sweet 1982; Jeppsson *et al.* 1985). Similarly, Mawson (1987) tested and confirmed the increase in yield with buffered acetic acid. Dismissing the results of these experiments without repeating them is unscientific, as is the use of untested methods of acid processing.

Here I describe how the loss of conodont elements can be monitored and corrected at most stages of laboratory processing. A useful spike test (von Bitter and Millar Campbell 1984) measures element loss except that due to removal of fines. Other biases include fieldwork bias and small-collection bias. Apart from the acetic acid method of Jeppsson *et al.* (1999) and the buffered formic acid method of Jeppsson and Anehus (1995) the methods needed to cope with the various difficulties we face in extracting conodonts from rocks have not been subjected to rigorous testing. In the absence of such empirical studies, descriptions of bias in some methods are based on long personal experience, often including resampling and reprocessing of the same stratum. During 38 years of

research I have tried many different methods: most of them can be very efficient in destroying or losing conodont elements.

I also discuss how to determine the amount of bias in different methods. Some tests may be found useful as a daily routine. Others are important for quantifying the amount of bias in the routine application of a method and may, perhaps, not be used for every sample. Many published methods are not mentioned, either because they have been superseded by better techniques or because I lack experience of them. The methods recommended here have functioned equally well for samples from all geographical areas and geological intervals processed in the Lund conodont laboratory, not just Silurian residues from Gotland and Skåne. Hence, I expect that my conclusions will be applicable widely. Similarly, although this study is published in a conodont context, some of the types of bias identified are found in all kinds of collecting work, be it fossils or other particles, micro- or macrofossils of phosphate or other minerals or material.

Some of my conclusions are drawn from discussions with colleagues or resulted when we compared our collections and I tried to determine why they differed. Other conclusions are based on analysis of published data. I have treated any bias discovered in this way as due to honest mistakes, and I thus consider it inappropriate to cite the source of the data. Some examples of what can occur using the wrong method may therefore appear artificial and exaggerated. Unfortunately they are not.

Bias can be minimized but not avoided entirely: hence one should always discuss the methods used and evaluate the degree of bias caused by different techniques. In practice, the pressure for space in publications will usually limit any such discussion to references to standard published methods, with notes of any deviations from these methods and their effects on the data. However, using standard methods has a major advantage in that data in different papers can be compared. Different kinds of data will be biased to a different degree by different fieldwork and laboratory methods. Some of the more frequent causes of bias are listed in Table 1, together with what can and should be undertaken to minimize them and document their effects.

The methods advocated in this paper are those that I have found to be optimal in reducing the biases inherent in collecting and processing samples for conodonts (and other microfossils). I recognize that constraints of field logistics, laboratory time or other resources may conspire

**TABLE 1.** Some sources of bias, effects, possible improvements, and what documentation should be included in any publication; see text for details.

| Source of error (and unit) | Effect | Improvements | Documentation |
|---|---|---|---|
| Sampling distance (years; metres are only a local proxy) | | recollecting critical gaps | sampling specified, e.g. in a stratigraphic column |
| Collection size | abundance cut-off-limit (are rarer taxa recovered?) | adequate sample size; recollecting intervals where collections are too small | sample and collection size specified |
| Element destruction | distorted yield/kg; smaller collections | employ scientifically tested methods | references given for all methods; any deviation described |
| Selective destruction of elements | distorted relative frequencies | gentle handling: screening in water, etc. | references given for all methods; any deviation described |
| Selective recovery | distorted relative frequencies | routine testing of methods and result | references given for all methods; any deviation described |
| Screening (mm) | no specimens smaller than the cut-off-limit recovered | finer screen | sieve mesh diameter specified |
| Decanting | element size cut-off fuzzy and operator-dependent | use of fine screen | the smallest abundant specimen size discussed |
| Removal of non-phosphatic minerals | skewed proportions; smaller collections | use only the best methods available; testing all fractions; redo separations that are too biased | references given for all methods; any deviation described |
| Picking | skewed proportions; smaller collections | testing every tray picked | details of method discussed |
| Analytical error, e.g. no uncertainty intervals for frequencies, zonal boundaries | incorrect conclusions; unnecessary conflicts | analysis of confidence limits | confidence limits discussed |

to prevent optimal work in some circumstances. However, I hope that this paper can be used as a guide to how element recovery will be biased in such cases, and what steps can be taken to minimize bias or take it into account when interpreting the collections that result. Some descriptions of methodology may seem unnecessarily detailed, or too obvious to be repeated here. However, in order to compare the relative biasing effects of alternative methods it is essential that processing details are made clear.

## BIAS IN THE RECOVERY OF CONODONT ELEMENTS

### *The accumulation of bias during sampling, processing and interpretation*

If something goes wrong during processing a sample, it is usually impossible afterwards to identify at what stage it occurred (or even realize that something went wrong). In order to illustrate the progress in processing methods therefore two hypothetical samples are used in Table 2 to provide a summary of the biases introduced in processing and interpreting the resulting collections. The first of these hypothetical examples is based on methods of collection, processing and analysis that were the state of the art 30 years ago; the second is based on the best methods available today. The biostratigraphic implications of bias for each of the resulting collections ('Interpretation and zonal identification') compare well with two real sets of samples collected in 1973 and from 1989 onwards, respectively, from the Slitebrottet 1 and 2 sequence on Gotland, now *c.* 50 m of strata. The 1989 samples varied in size as a result of collecting constraints, and variation in what was found to be an adequate size for each interval (some samples exceeding 100 kg). Nevertheless, the two sets of samples differed in size by about two orders of magnitude. Similarly, muddiness, yield, etc., varied through the sequence.

To avoid exaggerating the perils of the 5-kg hypothetical sample, the losses described for each step in Table 2 are at the lower end of the probable range. The differences in the real results were even greater than those in the hypothetical set: in the Slitebrottet samples from 1973 none of the zones named by Walliser (1964) at Cellon was identified (Jeppsson 1983*a*). The second set of samples, however, yielded both *K. patula* (at an average frequency of one platform element per 30–40 kg) and *Ozarkodina sagitta sagitta* (at a frequency in the per mille range). The latter taxon has $P_1$ elements (sp) that are thin-walled, and all its other elements are small and fragile, so recovery requires very gentle handling and very low discriminating bias in spite of the large volumes of mud etc. that must be removed during processing. With their ranges known reasonably well, two other zones could be characterized, the Sheinwoodian conodont zonation revised, and the new standard Sheinwoodian zonation applied to published sequences, including identifying one of the new zones and a major gap at Cellon (Jeppsson 1997*b*).

Another case of resampling illustrates the improvement in yield up to the 1990s: samples Kl67-32 and Kl67-33 from Skåne gave eight and 12 specimens from 0·468 and 0·485 kg, i.e. 17 and 25 elements $kg^{-1}$, respectively (Jeppsson 1975; the methods used are described). Samples Sk93-4LJ, 18·0 kg, and Sk93-3LJ, 37·2 kg, from the same beds (treated with most of the methods recommended herein) yielded *c.* 1200 and *c.* 4000 conodont elements, respectively. This equates to 67 and 107 elements $kg^{-1}$, respectively; compared with the earlier samples this is four times as many per kilogram for both beds and over 100 times as many specimens per sample.

### *Fieldwork*

*Sample size and yield.* There is a relationship between sample size and yield that through the effect of 'rounding off' can have an important bearing on sampling strategies. For example, a unique specimen in 100 g does not equate to ten such specimens $kg^{-1}$. Because of the statistics of small numbers, the true frequency is probably much less, and in my experience a new sample ten times as large would not be expected to yield more than approximately two or three specimens (Anders Martinsson, working with ostracodes, had a similar 'rule'). This has implications where resampling is undertaken in order to find more specimens of biostratigraphically important taxa, for example to increase the reliability of ranges (see below and Table 2). Recollecting a sample less than ten times as large will in most cases be a waste of time and other limited resources.

*Sample collection.* Collecting standard size samples from a lithologically diverse sequence (ranging, for example, from low-energy fine-grained marly lithologies to higher-energy, rapidly deposited coarse reef-detritus limestones) results in collections of different size. The yield can vary by a factor of much more than ten, and all or most of the rarer taxa will consequently be absent from the smaller collections. Strong apparent variations in diversity will arise as an artefact of the sampling, and stratigraphic ranges for most taxa will be artificially shortened. Only with correspondingly larger bulk samples can the expected lower yields be compensated for, although bias due to winnowing, crushing and other factors associated with samples from high-energy environments will remain and will need to be taken in account.

Collecting in quarries is usually easiest near the drill holes resulting from past blasting. However, blasting not only splits the rock but also creates microfractures within it and the fossils it contains (building-stone quarries use other methods). Hammering also causes microfracturing of rock and fossils, introducing a bias that is, for a number of reasons, unnecessary. Firstly, during collection and transport it is much more efficient to clean two small surfaces with a wire brush, and, when dry, label them with a marker pen, than to break a sample and pack a lot of

**TABLE 2.** Comparison between two hypothetical samples to illustrate the differences in bias introduced during collection, processing and analysis using methods that were the state of the art 30 years ago and those advocated herein. The characteristics of the two hypothetical samples are as follows: both from the same stratum and of average lithology [a limestone containing 10–20% mud (but not so much mud that less than 90% of the sample breaks down in acid), 2‰ quartz sand, 5‰ ore minerals (mostly pyrite and weathered former pyrite)]; containing 100 conodont elements per kg, representing 16 species, the most abundant at 30 per cent, and each of the subordinate taxa at *c.* 30 per cent of the remainder (i.e. 30, 21, 15, 11, 7, 5, 4, 3, 2, 0·8, 0·5, 0·3, 0·2, 0·1 and 0·05%). This is more favourable than most of my collections in the frequency of the tenth species being as high as 0·8 per cent, the drop-off being regular and the diversity higher. The two hypothetical samples belonged to two hypothetical sets of samples, the first set spaced *c.* 1 m apart, the second at *c.* 3 m apart except where recollected (to compensate for the doubled processing time of the larger samples and to leave resources for recollecting critical intervals).

| Methods of 1973 | Methods advocated here |
|---|---|
| *Sample size and theoretical max. element recovery* | |
| 5 kg | 50 kg |
| 500 elements | 5000 elements |
| *Rock processing* | |
| Crushed (breaking larger elements and many others beyond recognition). Sample suspended in colander in a bucket. | Not crushed. Sample suspended in a 'colander' in a 165-litre vessel (Jeppsson 1987, p. 48). |
| Ten changes of acid solution (100 min working time). | Ten changes of acid solution (150 min working time). |
| Spent acid decanted to point when fine sediment almost spilling out, leaving all insoluble residue and elements in bucket (possibly a few of the smallest elements lost; if clay removed by decanting, loss would have been higher). | Spent acid solution siphoned off through 53-$\mu$m screen (except for amount needed as buffer) leaving all insoluble residue and elements in vessel. |
| Vigorous addition of water from tap to dilute the solution to *c.* 10 per cent, resulting in a complete mixing during nine of the changes. Spent acid remaining in the bucket resulted in a ±buffered acid solution. | Vigorous addition of new acid and water through a hose reaching down into the bottom sediment. |
| pH not measured (but happened to be on safe side of 3·6 after nine of the ten changes). | pH measured, and any deviation from a *c.* 7 per cent solution and a pH slightly above 3·6 adjusted with more buffer or acid (Jeppsson *et al.* 1999). |
| | When repeated pH measurements indicate that remaining lumps are decalcified, most clay and silt automatically siphoned away under water through a 63-$\mu$m screen. |
| *Residue processing* | |
| Acid-insoluble residue hand-washed on a 63-$\mu$m screen. | Acid-insoluble residue hand-washed gently on a 63-$\mu$m screen, dried, treated with boiling water with some sodium carbonate, and washed. Treatment repeated until all clay removed. |
| Reduced using Franz Isodynamic Separator Model L1 (not available in Lund in 1973). | Reduced using Franz Magnetic Barrier Separator Model LB1. Traces of tap water carbonate and any dolomite removed with buffered formic acid (Jeppsson and Anehus 1995). |
| Density-separated using bromoform (a quartz crystal was used to judge density, density was high enough when it was floating on the surface). | Two-stage density-separation using sodium polytungstate at densities of 2·84 and 3·04, respectively (Jeppsson and Anehus 1999). |
| Result: large residue consisting mainly of pyrite that was not oxidized enough to be removed magnetically. | Result: residue consisting essentially of phosphatic fossils only; smaller amount than that from the 5-kg sample. |
| *Picking* | |
| Large residue spread thickly; each tray wet picked once (here in Lund, electrostatically). | Rich residue spread thinly; each tray picked electrostatically (Barnes *et al.* 1987) twice. |
| *Loss of elements* | |

**TABLE 2.** Continued

| Methods of 1973 | Methods advocated here |
|---|---|
| Estimated at *c.* 63 per cent. After fifth acid change (when 50% of the sample was dissolved) pH in insoluble residue dropped to 3·0 and all elements in that residue were dissolved, i.e. 50 per cent of the original elements. Without feedback from bias control and with a less efficient magnetic-separator an additional *c.* 5 per cent of elements were lost in each of magnetic separation, density separation and picking. Small and thin-walled elements preferentially lost (due to e.g. rafting with quartz sand during density separation). | Estimated at *c.* 4 per cent total. No elements were lost in the bias-free acid treatments. With feedback from bias control during all other steps, correction of any too biased step, and an efficient magnetic-separator, less than 1 per cent the elements was lost in each of magnetic separation, density separation, and picking). Small and thin-walled elements preferentially lost. |
| *Processing summary* | |
| 187 specimens recovered (37 per cent; but varying strongly according to which acid change buffering failed). | At least 4802 specimens recovered (96 per cent). |
| Working time: 5 units of time (1 unit of time per kg), 37 specimens per unit time (= 1 unit of yield per unit time). | Total working time twice that of 5-kg sample, i.e. 0·2 units of time per kg; 480 specimens per unit time (13 units of yield per unit time). |
| *Faunal recovery (based on hypothetical frequency distribution)* | |
| 11 species (three rarest absent due to crushing, biased recovery of small elements, and the randomness affecting small numbers, respectively). | 16 species (rarest represented by two elements). |
| *Interpretation and zonal identification* | |
| None of the defined zones could be identified in the collections from the 5-kg samples; all collecting and processing effort wasted if that was the only purpose. | All index species recovered; recollecting gave the zonal boundaries with a precision of better than 0·3 m. Further, the collections together yielded 14 specimens of the rarest species, indicating that all species had been found. |

sharp-edged pieces in several bags (rock travels better than bags). More importantly, however, breaking a sample into small pieces or crushing before dissolution results in only slight improvements in processing efficiency. These are due not, as one might expect, to increasing reactive surface area but to decreasing the maximum distance from the centre of the thickest piece to the nearest surface (Jeppsson *et al.* 1999, p. 969). Crushing results in pieces of different sizes, and the increase in bias due to breakage is disproportionately much larger than the decrease in thickness. The position and orientation of the sample in the vessel (slabs should be approximately vertical) influence processing time more than crushing (Jeppsson *et al.* 1999). Today my samples are cleaned, weighed and dissolved as they are, except when a piece does not fit into the 'colander' (0·6 × 0·4 m). Occasionally, evident bedding plane weakness indicates that the thickest pieces in a sample will split easily and without much hammering into slabs, and these provide the only exception.

Although many conodont workers may collect samples much smaller than those I suggest here, this is by no means standard among microvertebrate specialists. Ward (1984), for example, drew the boundary between small and large samples at 100 kg and described methods for efficient handling of samples above that limit. Similarly, the vessels for dissolution of limestone described by Cooper and Grant (1972) and those used by Barry Fordham (pers. comm. 1996) are larger than mine, although mine are approximately 75 times larger than the standard vessels I used before 1972. My point here is that many of the methods for handling large samples were in part developed and tested some time ago. The cost of scaling up processing can be small (most of the required equipment can be obtained second hand, from supermarkets, or home made), and doubling the size of a dissolution vessel effectively halves the amount of time spent changing acid, so after only a few samples the vessel has paid for itself. With such adjustments in methods and equipment the increase in total manpower per sample only doubles for a ten-fold increase in collection size (= yield). I have found this scaling relationship to hold true over two orders of magnitude (samples 100 times larger requiring only four times the work).

*Sampling strategies.* Most palaeontologists know that different lithologies can yield different faunas, and collect accordingly. Therefore, such facies-controlled faunal differences are not of major concern here (although where facies are preferentially sampled or avoided this should be

documented). Lithologies that are difficult to process are often under-sampled, because processing can be prohibitively expensive and frustrating (how this is exacerbated by using methods that yield inconsistent results is discussed below). The resulting bias will be most severe when working with condensed sequences or in quickly changing intervals where facies-related faunas and taxa may easily be missed. Similarly, external forcing may trigger a change to both a unique fauna and a lithology that is generally low-yielding or otherwise difficult. Using a model for what happened during a time interval, how that affected different taxa, and how later diagenesis and weathering have affected the fossils as a guide in collecting can yield new data by highlighting layers that would otherwise probably have been left unsampled. In my own work, I have found my model for oceanic cycles (e.g. Jeppsson 1990, 1998) extremely useful. For example, together with colleagues I have identified an interval representing probably only a few thousands of years at several localities on two palaeocontinents.

Owing to economic constraint, not every bed or even every 10,000-year interval can be collected adequately. A more or less unconscious eye for subtle lithological differences can yield better collections and better results, and sampling at standard distances may not be the best alternative in most work. We need to be aware of this and to find ways to document it. Apart from this, sampling for microfossils does not include any of the biases inherent in picking fossils by hand in the field.

### *Laboratory methods*

The effectiveness of a laboratory method can be rated according to how well it removes a given mineral and what percentage of the original conodont element content remains. An unbiased method removes 100 per cent of a specific mineral and leaves 100 per cent of the fossils that are searched for; in short, a 100/100 method. Some widely used concentration methods yield <100/≈100 and repeatable results; that is, the degree of bias is about the same for every sample if the worker and laboratory set up are unchanged. Other methods yield strongly varying results (<<100/0–90), even if worker and laboratory variables are constant. Some methods still in use may even destroy all fossils (Jeppsson *et al.* 1999).

*Rock dissolution and disaggregation.* Chemical methods tested and found to be safe for phosphate reach 100/100 effectiveness in routine work. The buffered pH-measured acetic acid method (Jeppsson *et al.* 1999) and the buffered formic acid method (Jeppsson and Anehus 1995) are of this kind. Hence, limestones, dolostones and all lithologies with such cement can be broken down with 100/100 efficiency, if: (1) the method is tested for all kinds of phosphate that one wants to recover (some taxa require raising the minimum pH; Jerre 1994; Siverson 1995; D. Fredholm pers. comm. *c.* 1986–1990); (2) the laboratory work is done as prescribed; for example, the pH meter is correctly calibrated and measurements are undertaken properly. Hydrochloric acid may also be used (Jeppsson *et al.* 1985), but the method is not yet well enough developed to be efficient. Thioglycollic acid dissolves iron compounds (Ward 1984, p. 256), and a method based on controlled experiments needs to be developed. Many claystones break down with sodium carbonate or petroleum ether/kerosene methods (Jeppsson and Anehus 1999, p. 58); both approaches are very efficient, but the former is preferable for reasons of health, environment, economy and versatility.

Only these chemical methods reach 100/100 efficiency and the sum of processing bias will consequently be lowest if these methods are used whenever possible. It is possible to remove any remaining grains of calcite and dolomite using physical methods, but this will increase bias, and it is better to use buffered formic acid as the final acid step. Similarly, it is often best to finish all chemical preparations before applying physical methods other than screening.

Some methods yield inconsistent results. This includes unbuffered or incompletely buffered acids (herein referred to as ±unbuffered, for short) and methods involving 'gentle kneading' or freeze–thaw action to break down difficult sediment. The bias due to ±unbuffered acids varies strongly with minor methodological details and minor differences such as the muddiness of the sample, and the amount of carbonate per litre of pure acid used. There is, however, little size-related variation in the bias. If unbuffered acid is used but not exchanged and if the ratio of 'pure lime to amount of acid' happens to be optimal, then the efficiency can approach >>90/≈90 for a nearly pure limestone. Such an optimal ratio requires weighing the sample, determining its lime content and calculating how much acid to use. A complete exchange of the spent acid results in a total loss of all previously freed conodont elements and of those nearest the surface of the remaining rock pieces (some limestone must be dissolved before the acid is sufficiently buffered). Even an incomplete exchange of the acid may be disastrous if, for example, the acid was not completely spent (pH should be 4·8–5 or higher). In practice, such methods result in a strong increase in the average element loss with decreasing lime content, and element loss is probably essentially complete even in 'medium difficultly' lithologies (typically high-yielding with pH-measured methods). Results can be as poor as <<100/≈90, i.e. lots of limestone and other grains remaining, but very few or no conodont elements found after days of picking through the residue (pers. obs.).

Unbuffered formic acid reacts much faster than acetic acid, and results are generally worse and more unpredictable.

Destruction by acid is insidious in that damaged specimens are often rare in comparison with those dissolved or destroyed beyond recognition. Furthermore, damaged specimens are easily missed in the variation of preservation we find. I have repeatedly encountered colleagues who were unaware that their own material included such specimens and were hence confident that their acid method was safe. This is probably the main reason that some colleagues continue to use untested acid methods. (For further aspects of the variation in bias caused by such methods, see Jeppsson *et al.* 1985, 1999; Jeppsson and Anehus 1995.) Another frequent objection is that I have not taken into account the lateral variation in conodont faunas within a bed, the implication being that improved yield on resampling is the result of encountering part of the bed with more abundant conodonts, and not due to improved techniques and buffered acid. Although it is possible that conodont faunas may vary significantly in adjacent parts of a heterogenous bed, this cannot be the cause of my results. In all cases the better yield is obtained using the improved technique and enough localities have now been resampled to rule out statistically the alternative possibility. Furthermore, any conclusions regarding large lateral variation in conodont abundance within a bed that derives its data using processing methods that are demonstrably inconsistent in their results is, to say the least, open to question.

*Physical methods of residue reduction.* Physical methods, including sieving, decanting, density separation and magnetic separation, never achieve 100/100 performance in routine work. Some grains of the mineral that should have been removed will remain, and some of the phosphatic fossils will end up in the fraction removed. The degree of bias is strongly related to how these methods are used and is highest for small specimens (Jeppsson and Anehus 1999, p. 61). However, with proper care, the efficiency of some methods will be ≈100/≈100 and the bias will be more or less consistent. The degree of bias in the work can be evaluated by checking both fractions once processing has advanced so far that phosphatic particles can be seen in the residue. For example, if 100 phosphatic particles are seen per minute in one fraction but none in the other, over a similar period, then the loss is below or about 1 per cent if both fractions are of a similar size. If that is considered acceptable, the latter fraction can be stored without further processing. If loss is at an unacceptable level, processing is modified and evaluation repeated until an acceptable result is achieved. In this context, other phosphatic fossils are useful in two ways. Firstly, they increase the number of phosphatic particles seen in the conodont-containing fraction when testing a separation, and secondly, being on average more porous than conodont elements, such fossils tend to have more impurities and are more likely to be separated incorrectly. Their absence from the separated fraction thus provides an extra safety margin (see below).

A prerequisite when using physical methods of residue reduction is that the fossils are essentially free from adhering grains. Using sodium carbonate routinely during the washing of residues can help to achieve this.

*Sieving and decanting.* Removing the unwanted fine fraction from a residue by sieving or decanting causes the largest bias among all standard methods of residue reduction. It is the major reason why, to my knowledge, no collections contain the earliest growth stages of conodont elements. Furthermore, some taxa, e.g. *Decoriconus*, have elements that are so small that elements are found only rarely unless fractions below *c.* 150 μm are picked and many records are based on elements less than 125 μm in diameter.

Sedimentary, diagenetic and weathering processes may destroy the smallest specimens in a fauna, but generally the lower size limit of elements recovered is set by the choice of sieve. Decanting is often claimed to be better than sieving, but I have seen many resulting collections and the recovery limit is markedly operator dependent, rarely as low as 125 μm and usually considerably higher (the recovery limit is here defined so that essentially all specimens larger than that size are recovered; below that size the recovery percentage drops gradually to zero). The limit varies with the shape of the element and with their 'bulk density' (Jeppsson and Anehus 1999, p. 61). In contrast, it is easy to remove tens of litres of clay in routine work even if a sieve as fine as 63 μm is used (see below; a finer sieve has also been tested with good results). The recovery limit is much sharper than that for decanting, but occasional, smaller elements may also be recovered. Sieving has another major advantage: the bias can be expressed exactly, as in 'only elements over xx μm in diameter were recovered'.

Another kind of bias is also due mainly to the removal of the fines. The central parts of mature elements are relatively resilient, but other parts, especially denticles, distal processes and juvenile elements, are very brittle. Rough or extended washing intended to break down lumps of decalcified rock physically is much more destructive than generally thought. The effect is worsened if the residue contains heavy minerals such as pyrite, the effects of which resemble those of a ball mill or a tumbler. Methods for minimizing loss at this stage are detailed in the Appendix (section 1).

Washing under water is frequently used to recover brittle fossils. An 'explosive expansion' in the knowledge of Mesozoic metatherians (Mammalia) was 'in large part'

due 'to widespread use of underwater screen-washing and associated techniques' (Cifelli 2001, p. 1219). I have had similar experience with methodological improvements in conodont processing (Jeppsson *et al.* 1985, 1999; Barnes *et al.* 1987; Jeppsson 1987; Jeppsson and Anehus 1995, 1999; herein) resulting in a large increase in the number of conodonts extracted [average recovery of elements during 1967–1983, 1000 elements per year (1967–1973, see Jeppsson 1975); improvements during 1984–1995 gradually raised recovery to a few hundred thousand elements per year, including highly significant new data and previously unknown taxa]. This increase is similarly due in part to an underwater screen-washing machine (Jeppsson 1987, p. 49). It functions very well, routinely washing away most of the clay through a 63-$\mu$m screen. It was primarily constructed to save expensive working time and to decrease the amount of tedious work. This saving, invested in larger samples, together with the reduced element breakage, increased the yield. Very gentle removal of the remaining clay has now made it possible to benefit more fully from the latter effect (see Appendix, section 1, points 6 and 7) and has helped to maintain the yearly yield despite decreasing availability of manpower.

*Density separation.* Density separation in liquids with densities below and above that of conodont elements (Jeppsson and Anehus 1999) is the only concentration method needed to produce routinely a phosphate fraction that is *c.* 99 per cent clean. Sometimes such a residue includes fewer 'other phosphate' than conodont elements, making it faster to pick away that and pour the remaining conodonts into the slide. Densities very close to that of conodont elements give the cleanest residue but increase the risk of losing some elements, such as those with impurities in their basal cavity. If picking a residue directly after the first density separation would take more time than a second density separation, then picking it directly is a waste of limited time or money. Bias caused by density separation and how to minimize it were described by Jeppsson and Anehus (1999). Low-viscosity liquid causes an unacceptable loss at a density of 3040 kg $m^{-3}$ (G. K. Ahlberg, pers. comm. 2004), most probably because that density is so close to the density of the conodont elements that the rapid downward movement of heavier minerals, such as pyrite, is enough to drag them down, even when very small quantities are added at the time.

*Magnetic separation.* Magnetic separation removes paramagnetic material. Yellowish and grey pieces of hard clay are easiest to remove. (Different kinds of extra equipment permit removing ferromagnetic and diamagnetic minerals.) Both precision, in terms of degree of bias (i.e. the deviation from 100/100%), and the efficiency in terms of throughput vary with the magnetic-separator model. In the past I have obtained reasonable results using a Franz Isodynamic Separator L-1, but rebuilding it into a Franz Magnetic Barrier Separator Model LB1 increased both precision and efficiency enormously. If properly adjusted for each size fraction of a sample, bias can be kept low. However, with eight continuous variables to adjust on an LB1, the number of combinations is infinite and finding the best combination (with lowest bias) by random adjustments is impossible. It is necessary to know the effect of each possible adjustment in order to obtain good results (see Appendix, section 2).

Density separation is done in batches and the manpower requirement increases stepwise with size. Magnetic separation is therefore very useful for reducing large residues rich in paramagnetic particles prior to density separation. However, if residues are small or their unwanted parts consist mainly of quartz or other diamagnetic material, then magnetic separation may not save manpower. The cost per man-hour determines how many hours need to be saved to compensate for the investment in a separator because double density separation alone can yield a phosphate concentrate that is almost clean.

Sometimes even the clean phosphate fraction is too large to pick through and we have to choose between maximizing removal of unwanted material or minimizing the loss of conodonts. Impurities may make at least some of the other phosphatic fossils slightly denser and slightly more magnetic than the conodont elements, and fine-tuning the magnetic separator (see Appendix, section 2) or density separating at a density closer to that of the conodont elements can yield a fraction enriched in conodonts.

*Picking.* Picking is not free of bias. If particles are of different size, smaller specimens may be concealed beneath larger particles, but this bias can be reduced through size fractionation of a residue in a sieve-stack. Picking effectiveness and bias can be evaluated as follows: some material is spread thinly and evenly on a picking tray, which is then held over a sheet of paper and gently tapped horizontally a few times so that any perched particles are moved down by the horizontal vibrations; the tray is picked under a microscope at low magnification; the tray is again tapped over the sheet and picked at higher magnification; comparison of the numbers of elements obtained from first and second pickings provides a measure of picking bias. For example, if the two pickings gave 1000 and 100 elements, respectively, then a third would be expected to give *c.* ten specimens. If, on the other hand, the two pickings gave 1000 and 400 elements, 160 would be expected in the third, and so on. Even after the fifth picking, the remaining bias would be nearly twice that of the second picking in the first example, suggesting

that bias would have been reduced (and picking time decreased) if the material had been spread more thinly in the tray.

The electrostatic picking method (Barnes *et al.* 1987) allows 50 or more conodont elements $min^{-1}$ to be picked from a more or less clean phosphatic residue (R. Anehus and G. K. Ahlberg, pers. comm. 2002). The technique takes only a few minutes to learn and in addition to increased picking speed allows even the smallest specimens to be handled without breaking them. Sorting of the resulting stack of loose specimens is also much faster. Picking with a wet brush is inferior in all these respects and increases the bias, especially for small specimens (which are more brittle and easily destroyed during wet handling).

### *Other causes of bias*

Previous work has discussed upwards reworking and identification of such range end bias (Jeppsson 1997*a*, pp. 472, 488), and contamination (Jeppsson 1997*a*, p. 472). Mislabelling or mixing sample numbers during handling is part of an old problem the cause of which is well known: *errare humanum est.*

Observed breakage of elements is the sum of breakage caused by usage during life (Jeppsson 1976), sedimentary processes, compaction of sediments, tectonics, sampling, processing, concentration and handling of the residue (von Bitter and Purnell 2005 discuss biases caused by some of these factors in more detail). Breakage of different origin has different characteristics, allowing appropriate measures to be taken to minimize that caused by sampling and processing (see Appendix, section 3). [Wear (Purnell 1995) and abrasion due to sedimentary processes (Broadhead and Driese 1994) also affect conodont elements but, in my experience, breakage is the main destructive effect of our handling of conodont remains.]

## BIAS IN INTERPRETATION OF MICROPALAEONTOLOGICAL DATA

Interpreting how the observed micropalaeontological record deviates from that preserved in the studied strata is a two-stage process. Firstly, the various biases discussed above must be taken into account, their relative importance depending on the nature of the conclusions being drawn. Secondly, common sense or intuition is used to decide what conclusions can or cannot be drawn from the available data, or data are subjected to more rigorous analysis, including statistical techniques, to determine the degree of error or uncertainty in the results. This more rigorous approach is clearly superior in that results are not dependent on the subjective evaluation of the operator. Marshall (1990, 1994, 1997) has described methods for statistical evaluation of discontinuous records, and these aspects of analysis are not discussed here. His 1995 paper showed the power of such methods and clearly illustrated the danger in taking range ends at face value. Here I concentrate on problems that are more or less specific to bulk sample collecting and also comment on how the number of specimens in a collection influences the robustness of an identified range end.

### *Relative frequencies*

Conclusions regarding the relative frequency of different species in a community, the ratio between different elements in an apparatus, the ratio of juvenile to mature elements, etc., are strongly affected by sieve mesh diameter. Similarly, using any method in a way that discriminates against small specimens has strong effects. Two examples illustrate this effect. All species of *Panderodus* have a short, small, symmetrical element (with a true furrow on both sides). In collections of the more gracile species the frequency of this element may be as low as 1 per cent (Jeppsson 1983*b*) even when a 63-$\mu$m screen is used. In the comparatively robust *P. greenlandensis*, on the other hand, recovery of this element approaches the expected frequency of approximately 7 per cent (assuming an apparatus of 15 elements). This is because the smallest individuals will always be represented only by their largest elements, even with a fine sieve, and the bigger elements will be over-represented as a consequence.

A similar example is provided by some Ordovician 'platform conodont genera', which until recently were thought to lack ramiform elements, despite being studied by many different authors through the years. As a result, they have even been separated from their closest relatives as a distinct order. Adopting new methods has gradually resulted in better collections of these taxa (Löfgren and Zhang 2003), and these authors could show how a varying combination of decanting, different sieve sizes and different acetic acid solutions yielded collections of very different composition [A. Löfgren, pers. comm. 2003; Löfgren's acid dissolutions performed in Lund used the buffered acetic acid method (Jeppsson *et al.* 1985) until the pH-measured technique was developed (Jeppsson *et al.* 1999)]. By processing with an acetic acid solution with a pH known to be more than 3·60, and sieving with a 63-$\mu$m screen Löfgren and Zhang (2003, p. 722 and pers. comm. 2003) were able to obtain the collections 'essential for firmly establishing the apparatus reconstruction' and for separation of all ramiform elements of closely related species.

*Range ends, faunal diversity and variations therein*

The perceived range end position is based on the absence of the fossil on one side of the observed end, and its presence on the other. Its presence is relatively unproblematic; once the possibility of reworking, contamination, incorrect labelling, etc., has been ruled out, no bias need be taken into account. However, the adjacent sample that does not contain the fossil is more difficult; absence of evidence is not evidence of absence, and it follows that the probability limits of the absence must be analysed. Many taxa are more or less rare, and most publications, for various reasons, include some collections that are too small to give a complete species list (my own papers are no exception). Graphic correlation has its own special biases (for discussion see Jeppsson 1997*b*, p. 104).

The first step in documenting the degree of bias in the position of a range end is to record every find. If only first and last occurrences are marked, the reader can only use these in an analysis of range end bias (= the difference between true and observed range ends), and the result is a range end bias many times larger than a calculation based on all occurrences. The result of such an analysis can be given as statistical confidence limits (Marshall 1994, pp. 467–468). The reason for such an analysis can be illustrated with an example: taxon A occurs in every sample from its known range but taxon B in every tenth sample. Both statistical methods and common sense tell us that the true range ends of B are probably many samples outside the known range and may well be much more than ten samples away (see Marshall 1995 for a very illustrative example). In contrast, the range ends of A are probably at the most only one or two samples away.

Narrower confidence limits require the number of specimens to be known. For example, with over ten specimens of taxon A in every sample, the range end uncertainty is not wider than the interval between samples, with the same provisions.

The next step in analysis is to determine the robustness of the observed range end based on frequency changes near the true range end. For example, a species with an average of 100 specimens per sample has an observed range end that is an order of magnitude less sensitive to frequency fluctuations near the range end than a species averaging only ten specimens per sample. Large collections from the Ireviken Event interval have shown that some species maintain a roughly similar frequency through the youngest collections (spaced by <0·1 m, calculated to correspond to < *c.* 2000 years), until they go extinct. Others, such as *Distomodus staurognathoides* and *Pterospathodus procerus*, 'faded away' and the position of the observed range end differs markedly when based on smaller and larger collections (Jeppsson 1997*a*: compare data in the main text and the supplementary note written 2·4 years later when larger collections were available).

*Zonal boundaries*

First and last records of index fossils are widely used for defining and identifying zonal boundaries [Lazarus gaps (Flessa and Jablonski 1983) can be important but have been used only rarely]. Hence, analysing the degree of bias in the position of a zonal boundary is a special case of range end bias. No taxon has the same frequency across all facies and all latitudes, and although all index fossils should be very widespread and may be frequent in some areas, they will be rarer in many areas [most of the Wenlock and Ludlow index fossils proposed by Walliser (1964), for example, have frequencies below 1% in most areas, in some places below 0·1%]. Other rare taxa may be ignored without significant consequences, but because the burden of providing the best zonation frequently falls on the conodont specialist, we are under pressure to do our best, even if an index fossil is rare. The significance of this is clear: any kind of bias that influences zonal boundaries has far-reaching consequences. It is also worth noting that analysis of range end bias in only a single section is not enough; any correlated section should be subjected to similar analysis (unless previous authors have carried out such an analysis).

In addition to erroneous correlations, placing zonal boundaries without consideration of the statistical uncertainty interval of first and/or last occurences can have other serious effects. For example, the ranges of taxa are related to what zone they are found in, so if a given taxon appeared everywhere at the same time, but in one area a zonal boundary is drawn too high (or too low), then that taxon will be erroneously recorded as appearing in another zone in that area. A correct analysis of the uncertainty in the zonal boundary position there would place the end record of the taxon in the uncertainty interval, resulting in the correct conclusion: that its range is not significantly different from the record elsewhere.

For a variety of reasons (biological, historical, bias against small specimens in many published collections) index taxa tend to have relatively large elements, and bias resulting from discrimination against small specimens is not a major problem for most index taxa. As shown above, however, the only insurance against a large range end bias is that the critical collection just outside the perceived zonal boundary is adequately large. Unless recovery bias is large (e.g. due to a destructive acid method), a smaller range end bias usually requires larger collections.

In principle, any zonation of a sequence includes an uncertain interval intercalated between any two intervals

confidently referred to zones, and this ought to be marked as such (for an example, see Kiipli *et al.* 2001). Once collection size is adequate therefore the next step to decrease zonal boundary bias is to resample the interval between the documented range end of the index fossil and the nearest adequate collection. The first fieldwork may often include only ten samples per $10^6$ years or even fewer. Thus, the range end uncertainty will be *c.* $10^5$ years or more. Strategic resampling can minimize the amount of effort required to reduce this significantly. For example, 15 samples are collected between the range end and the next adequate sample from the first sample set; these are processed and analysed one at the time, selecting the sample that lies in the middle of the remaining uncertainty interval for processing, so that each new collection reduces it by half. In this way, processing only four samples reduces the uncertainty by more than one order of magnitude (to $0{\cdot}5^4$, equivalent to 0·0625 of the initial uncertainty). In my experience, planning so that resampling of critical intervals is possible before publication decreases the bias in the data enormously, because it allows most kinds of bias encountered to be addressed. Thus, range end bias is decreased not by closely spaced small collections but by fewer, adequately large collections and resampling: more of the same *is* much better.

## SUMMARY

The hundreds of specialists on phosphatic fossils use many tens of more or less different methods of extraction. For every method in use there is at least one specialist convinced of the superiority (e.g. regarding efficiency) of that method for her or his own work. I am no exception. I am convinced about the superiority of bias-free and bias-controlled methods for all kinds of work, by all specialists. Hitherto nobody has proven me wrong, but, in contrast, many have reported considerable gains using one or more of these methods. I have also met objections and explanations from colleagues as to why they have not or did not plan to test my claims, most frequent among which are that my methods are idealistic and impractical. In short, they have assumed that such methods take too much time per sample to be useful in practice when hundreds of samples are processed. However, thousands of samples have been processed in the Lund conodont laboratory, using different methods. The bias-controlled methods now in use are by far the most efficient of those tried and my conclusion is that these are the most efficient ones existing (in addition, they yield reproducible results).

Each processing step adds to the total bias; hence, the sum can be considerable at the end. Bias is either systematic or random. Systematic bias can be statistical or complete. Not saving and picking a fraction finer than xx $\mu$m is close to a complete loss of every element below xx $\mu$m in diameter. Most physical methods discriminate more against small thin elements than against large ones. The degree of such statistical bias is strongly tied to the laboratory regime, including protocols for routine checking of the degree of bias. The person's degree of knowledge, awareness and carefulness, both regarding the proper use of each method and its peculiarities regarding bias, influence the degree of bias enormously. The statistical probability of finding a species, present in the sediment that was formed at a specific point in space and time, increases with the collection size to the level at which the collection can be described as 'adequate'. Some ways of using ±unbuffered acid methods result in random bias, and in such collections anything between 10 and 100 per cent of the original fossil content may have been lost in processing. There is also a statistical pattern in the degree of bias caused by ±unbuffered methods: samples that are muddier or otherwise more difficult to dissolve have higher probabilities of high bias.

The difference in working time between a fully bias-controlled approach and the most efficient uncontrolled approach is only a few per cent. In contrast, the many different less efficient alternative methods that have been suggested and probably are in use at many places require much more time per sample than a fully bias-controlled approach using the most efficient methods.

By choosing the best methods now available, including a protocol for routinely monitoring the amount of bias introduced, the degree of bias will be only a small fraction of that caused by methods not tested with controlled experiments. 'Adequate' samples are necessary to overcome the otherwise large bias of zonal boundaries and other range ends. Paying attention to every collecting step, from fieldwork to analysis of the resulting data, is, however, very profitable in another respect, too; the average yield per hour of work can easily be raised at least ten- to 100-fold when processing low or moderately yielding lithologies.

## RECOMMENDATIONS

1. Where alternative methods exist, bias should be kept as low as possible by using the method causing the lowest bias.
2. All methods used should be referred to, or described.
3. Data should be presented in a form permitting the reader to judge the amount of bias (the practice among conodont specialists of including a table with number of specimens of each element of each species in each sample is a very important step).
4. The total bias in the data must be taken into account when the results are analysed.

Bias control is, to a large extent, like any other new technique. The instructions may at first seem very impractical, but the control quickly becomes a routine. Furthermore, it soon becomes apparent which steps are the most risky for the kind of samples being processed (on average, the largest biases are caused by unbuffered acid, samples that are too small, incompletely buffered acid, and discrimination against small/fragile elements, in that order, but be prepared for some surprises). Gradually the results of the control at each stage of processing become more intuitively predictable, and at that point risk can be assessed and informed decisions made concerning where time and effort can be conserved without significant risk to element recovery (and when no corners can be cut). On the other hand, at that point the operator will probably, like me, have found that the bias control has not only given so much better results that the time taken to learn the new methods was well spent, but also that, once it is a routine, most of it takes so little time that it is like a low-cost insurance against a high risk, and that it continues to pay off. For me, developing bias control has also resulted in development of methods and a laboratory set-up that have saved far more working time. I suppose that paying attention to potential bias also results in paying attention to other possible improvements.

*Acknowledgements.* It is well known that one must learn from other people's mistakes because one lacks enough time to repeat all of them oneself. I have repeated too many, but the fact that some methods are very strongly person-related can only be found with help from colleagues. The content of this paper was influenced by what I have learnt during contacts with many colleagues, including Richard J. Aldridge, Rikard Anehus, Ondrej Babek, Chris Barnes, Stig Bergström, Mikael Calner, Phil Donoghue, Michael J. Engelbretsen, Barry Fordham, Git Klintvik Ahlberg, Peep Männik, Anders Martinsson, John Talent, Viive Viira and Otto Walliser. Ann-Sofi Jeppsson typed the first version of the manuscript and corrected many linguistic errors. Mark Purnell, Phil Donoghue, Peter Molloy and John Repetski (referee) provided useful comments on the manuscript and Mark and David J. Batten carefully re-edited it. Grants from NFR, and its successor, The Swedish Research Council, have financed my experiments with methodological improvements and other work and thereby most of the many years of experience upon which this paper is based. Last but not least, this paper would not have been written had I not been invited to do so by Phil and Mark. Only recently did I realize that I was not the first in Lund to use data from large samples. Hadding (1958, p. 23) reported a general analysis of major elements, in a 'boatload of limestone' from Wenlock strata in the Smojen quarry on Gotland.

## REFERENCES

BARNES, C. R., FREDHOLM, D. and JEPPSSON, L. 1987. Improved techniques for picking of microfossils. 55, 74–76. *In* AUSTIN, R. L. (ed.). *Conodonts: investigative techniques and applications.* Ellis Horwood, Chichester, 422 pp.

BROADHEAD, T. W. and DRIESE, S. G. 1994. Experimental and natural abrasion of conodonts in marine and eolian environments. *Palaios*, **9**, 546–560.

CIFELLI, R. L. 2001. Early mammalian radiation. *Journal of Paleontology*, **75**, 1214–1226.

COOPER, G. A. and GRANT, E. E. 1972. Permian brachiopods of west Texas, I. *Smithsonian Contributions to Paleobiology*, **14**, 1–231.

FLESSA, K. W. and JABLONSKI, D. 1983. Extinction is here to stay. *Palaeobiology*, **9**, 315–321.

HADDING, A. 1958. The Pre-Quaternary sedimentary rocks of Sweden VII. Cambrian and Ordovician limestones. *Lunds Universitets Årsskrift. N. F., Avd. 2*, **54** (5), 262 pp.

JEPPSSON, L. 1975. [1974]. Aspects of Late Silurian conodonts. *Fossils and Strata*, **6**, 79 pp.

—— 1976. Autecology of Late Silurian conodonts. 105–118. *In* BARNES, C. R. (ed.). *Conodont paleoecology.* Geological Association of Canada, Special Paper, **15**, 324 pp.

—— 1983*a*. Silurian conodont faunas from Gotland. *Fossils and Strata*, **15**, 121–144.

—— 1983*b*. Simple-cone studies, some provocative thoughts. *Fossils and Strata*, **15**, 86.

—— 1987. Some thoughts about future improvements in conodont extraction methods. 35, 45–51, 52, 53. *In* AUSTIN, R. L. (ed.). *Conodonts: investigative techniques and applications.* Ellis Horwood, Chichester, 422 pp.

—— 1990. An oceanic model for lithological and faunal changes tested on the Silurian record. *Journal of the Geological Society, London*, **147**, 663–674.

—— 1997*a*. The anatomy of the mid-Early Silurian Ireviken Event. 451–492. *In* BRETT, C. and BAIRD, G. C. (eds). *Paleontological events: stratigraphic, ecological, and evolutionary implications.* Columbia University Press, New York, 604 pp.

—— 1997*b*. A new latest Telychian, Sheinwoodian and Early Homerian (Early Silurian) Standard Conodont Zonation. *Transactions of the Royal Society of Edinburgh, Earth Sciences*, **88**, 91–114.

—— 1998. Silurian oceanic events: summary of general characteristics. 239–257. *In* LANDING, E. and JOHNSON, M. E. (eds). *Silurian cycles: linkages of dynamic stratigraphy with atmospheric, oceanic, and tectonic changes.* New York State Museum, Bulletin, **491**, 327 pp.

—— and ANEHUS, R. 1995. A buffered formic acid technique for conodont extraction. *Journal of Palaeontology*, **69**, 790–794.

—— —— 1999. A new technique to separate conodont elements from heavier minerals. *Alcheringia*, **23**, 57–62.

—— —— and FREDHOLM, D. 1999. The optimal acetate buffered acetic acid technique for extracting phosphatic fossils. *Journal of Paleontology*, **73**, 957–965.

—— FREDHOLM, D. and MATTIASSON, B. 1985. Acetic acid and phosphatic fossils – a warning. *Journal of Paleontology*, **59**, 952–956.

JERRE, F. 1994. Anatomy and phylogenetic significance of *Eoconularia loculata*, a conularid from the Silurian of Gotland. *Lethaia*, **27**, 97–109.

KIIPLI, T., MÄNNIK, P., BATCHELOR, R. A., KIIPLI, E., KALLASTE, T. and PERENS, H. 2001. Correlation of Telychian (Silurian) altered volcanic ash beds in Estonia, Sweden, and Norway. *Norsk Geologisk Tidskrift*, **81**, 179–194.

LÖFGREN, A. and ZHANG JIANHUA 2003. Element association and morphology in some Middle Ordovician platform-equipped conodonts. *Journal of Paleontology*, **77**, 723–739.

MARSHALL, C. R. 1990. Confidence intervals on stratigraphic ranges. *Paleobiology*, **16**, 1–10.

—— 1994. Confidence intervals on stratigraphic ranges: partial relaxation of the assumption of randomly distributed fossil horizons. *Paleobiology*, **20**, 459–469.

—— 1995. Distinguishing between between sudden and gradual extinctions in the fossil record: predicting the position of the Cretaceous-Tertiary iridium anomaly using the ammonite fossil record on Seymour Island, Antartica. *Geology*, **23**, 731–734.

—— 1997. Confidence intervals on stratigraphic ranges with non-random distributions of fossil horizons. *Paleobiology*, **23**, 165–173.

MAWSON, R. 1987. Documentation of conodont assemblages across the Early Devonian-Middle Devonian boundary, Broken River Formation, North Queensland, Australia. *Courier Forschungsinstitut Senckenberg*, **92**, 251–273.

PURNELL, M. A. 1995. Microwear on conodont elements and macrophagy in the first vertebrates. *Nature*, **374**, 798–800.

SIVERSON, M. 1995. Revision of the Danian cow sharks, sand tiger sharks, and goblin sharks (Hexanchidae, Odontaspididae, and Mitsukurinidae) from southern Sweden. *Journal of Vertebrate Paleontology*, **15**, 1–12.

SWEET, W. C. (ed.) 1982. *Pander Society Newsletter 14*. Department of Geology and Mineralogy, Ohio State University, Columbus, Ohio, 40 pp.

VON BITTER, P. H. and MILLAR CAMPBELL, C. 1984. The use of 'spikes' in monitoring the effectiveness of phosphate microfossil recovery. *Journal of Paleontology*, **58**, 1193–1195.

—— and PURNELL, M. A. 2005. An experimental investigation of post-depositional taphonomic bias in conodonts. 39–56. *In* PURNELL, M. A. and DONOGHUE, P. C. J. (eds). *Conodont biology and phylogeny: interpreting the fossil record*. Special Papers in Palaeontology, **73**, 218 pp.

WALLISER, O. H. 1964. Conodonten des Silurs. *Abhandlungen des Hessischen Landesamtes für Bodenforschung zu Wiesbaden*, **41**, 1–106.

WARD, D. J. 1984. Collecting isolated microvertebrate fossils. *Zoological Journal of the Linnean Society*, **82**, 245–259.

## APPENDIX

### *1. Manual for routine samples*

In sequences with stable yield, adequate samples may be of equal size and a standardized treatment possible. Efficient handling requires tight control on the man-hours spent on each sample (independent of who processes the sample). For details of methods, see Jeppsson and Anehus (1995, 1999), Jeppsson *et al.* (1999) and herein.

1. Collect a sample of adequate size from the lithology expected to produce the best yield.
2. Clean, store a reference piece, and weigh the amount to be dissolved (needed for calculation of yield), do not crush (saves time and reduces bias).
3. Place sample in a colander in a vessel large enough to hold enough acid solution for a complete dissolution (20 litres per 1 kg sample). (Saves time otherwise spent changing acid; regarding smaller vessels see 5 below.)
4. Add acetic acid, acetate and water (vigorously to achieve a complete mixing); measure pH (at least until you have succeeded to mix the right pH, slightly above 3·60 for at least ten consecutive samples; and then now and then for succeeding samples to check that, for example, the concentration of the acetate solution has not changed). Provided that the amount of solution does not exceed what is needed for dissolving the whole sample the risk from the pH being too low at the start of processing a sample is moderate even if the solution is only partly buffered because every conodont is still encased by limestone. [If the solution is completely unbuffered, the pH will have increased to 3·60, the critical point, when *c.* 15 per cent of the sample has been dissolved (Jeppsson *et al.* 1999, fig. 2).] Further, the initial reaction is fast with a clean limestone sample, hence I estimate that less than 10 per cent of the conodonts freed before the reaction ends would be destroyed beyond recognition. However, measuring pH routinely takes so little time that it remains a routine here, even at the start of a sample.
5. If smaller vessels are used, the solution needs to be changed. This adds working time (5–15? minutes for each change). Even if clean limestones are processed and the pH has been slightly above 3·60 every time after ten such changes, I do not recommend abandoning measuring pH after a change of acid because the time saved per change (*c.* 1 min, when many samples are handled at a time) is small compared with the risk that, say one in 20 or 100 changes would go wrong, destroying all previously freed conodonts. ('Recover the fines after each change' is not an alternative since that requires 10 – >>100 min for cleaning the sieve, washing out the fines, cleaning them, transferring them to a filter paper and handling two or more residues with lots of calcite grains.)
6. Once dissolution is essentially complete, remove the larger particles first with a 1-mm sieve, perhaps also a 0·5-mm sieve. Then sieve the residue on a fine screen (I use 63 μm) and dry. To minimize loss use 'gently flowing water' and minimize washing time by only washing away the acid/acetate and the clay that easily pass the sieve. Store the remaining residue overnight with some sodium carbonate (or detergent) and wash again; repeat if useful.
7. If lumps of clay or dirty material remain: pour boiling hot water with some sodium carbonate over the residue,

wash and dry. Next, dry the residue thoroughly (e.g. in an oven at 50°C), pour boiling water (with some sodium carbonate) over the dry warm residue (if inefficient, gently boil it directly) wash, repeat if useful. Marl samples are frequently moderately difficult, requiring several cycles, interspersed with acid treatment to decalcify remaining lumps of rock. More difficult lithologies require many cycles. End sieving with rinsing in distilled water if you expect that density separation may be needed. Dry. Use a microscope to decide what is needed to get an easily picked residue.

8. If calcite or dolomite grains remain: buffered formic acid dissolves both minerals. If the microscope control indicated that density separation would also be needed, rinse with distilled water directly after washing, otherwise tap-water washing is enough.
9. Use a microscope to decide if one or more further concentration methods are required, density separation at 2·84 (2·90 if white mica) or 3·04, or magnetic separation (see below), saves more picking time than the concentration method requires.
10. Pick, sort, identify and evaluate the result: are there enough conodonts for a statistically reliable conclusion or should the conclusion include 'perhaps' or 'probably' and a larger sample be collected?

### 2. *Separation with a Franz Magnetic Barrier Separator Model LB1, adjustments and evaluation of results*

Sliding a paper below and above the chute before starting the machine best checks its position in the pole gap (two nuts below the chute). If the chute is in contact with the magnet, the vibrations of its lower end are dampened, and transport out from it is obstructed (and there is more noise). The lateral position of the chute in that gap (screws on the underside of the chute holder) is best checked when separation results in a distinct stream of magnetic particles: the splitter should guide that stream into the upper (inner on an LB1) fraction. Forward slope and vibrations regulate the transport forward. Side slope regulates how large a part of the gravitation is used to get the phosphate out of the stream of magnetic material. The lowest degree of bias results from an adequate part of the gravitation being used with a magnetic field of adequate strength.

To minimize bias of different kinds:

1. Screen the residue in fractions: if small, 63–250 and 250–1000 μm, further if large, e.g. 63–180–250–500–1000 μm.
2. Use a high side slope to get enough gravitational effect on the phosphate; I usually use 30 degrees, i.e. 50 per cent of the gravitation.
3. Conodont elements are brittle, so the less vibration needed for moving the particles forward, the less breakage is expected. Thus, adjust forward slope so that a minimum of vibration gives a smooth flow through the chute: at least 25–30 degrees for the finest fraction and slightly less for the coarsest is often best; less only if the particles are round enough to roll. Adjust the lip of the feeder accordingly.
4. Start with a low magnetic field. If 0·1 A removes some highly magnetic particles, use that. If such a residue is started at a higher magnetic field, such particles will block through-flow, causing conodont elements to end up in the wrong fraction.
5. Once the machine is adjusted, put all trial fractions back into the feeder.
6. Run some material and analyse the result under a microscope (see below). If satisfactory, run the rest of the residue.
7. Particles magnetic enough to be removed with only a small increase will repeat the problem described in 4, if the field is increased more than *c.* 75–100 per cent. Increase the strength of the magnetic field so much (less if needed to keep the magnetic fraction below 80%) and run the conodont fraction again. There is a considerable risk that some phosphate is trapped in a steady stream of magnetic particles if, e.g., 99 per cent is removed in one run. Repeat up to maximum magnetic field.
8. If needed, continue by lowering side slope stepwise using a very low rate of feeding (to compensate for the lower gravitational effect) and maximum magnetic field.

In order to keep bias low, both fractions must be checked under the microscope after every run (see laboratory methods in the text). Three causes for a higher bias may be identified and remedied. If the magnetic fraction contains:

1. Preferentially small, light phosphatic particles. The cause is either too high a feeding rate or too little of the gravitation being used to pull them out from the stream of magnetic particles, or both. Remedy: increase side slope (= increase the gravitational pull) and/or decrease feeding rate. (Removal of too much material in one run may cause a similar bias.)
2. Phosphatic particles with magnetic grains adhering to, or lodged in them. Remedy: increase side slope. (Inefficient washing or washing methods may cause some of this bias.)
3. All sizes of phosphatic particles. Stop only the feeder and wait until the flow out from the chute has ceased. Decrease magnetic field (the ampere) to zero, and, if needed, increase vibration. Check if a group of further particles leave the chute. If not, the remedy is probably to decrease feeding rate. If the test revealed such magnetic particles, see 4 and 7 above. If some particles block transport even at zero amperes, the sample must be spread out on a paper and such ferromagnetic particles removed by a hand magnet. Keep the magnet within another sheet of paper (to permit easy cleaning) and pass it some millimetres above the particles. If such pretreatment does not function, increase side slope. (If the chute is not properly positioned in the pole gap, similar transport problems will occur; see above.)

### 3. *Separation of different causes of breakage*

Breakage during life was repaired, or the lost piece partly regenerated, except when breakage occurred immediately before the animal died.

Sedimentary processes caused the same kind of rounding as on other sediment grains. Breakage due to compaction was sometimes followed by deposition of phosphate between the fragments fusing them in oblique positions. Similarly, tectonic breakage and pull apart was often 'healed' by inorganic deposition of phosphate. The original pieces show standard CAI-darkening, but the inorganic phosphate lacks organic matter and thus remains hyaline (white if pitted/'frosted'). Breakage surfaces, resulting from any of these processes, exhibit the same set of microdamage (e.g. phosphate loss, crystal growth or imprints) as the outside of the last lamella; hence such breakage can be identified.

Breakage due to collecting and processing differs from that described above. Broken surfaces are clean and sharp-edged. (However, using improperly buffered acid etches the surface on all affected elements, including those due to such breakage, blurring this distinction. Only later breakage is 'unfrosted'.) Microfracturing (due to blasting or hammering) of fossils still encased in the rock can easily be identified; easiest on large blades or platforms. When the rock splits through a fossil, the latter can equally well happen to be split along any plane and most breakage is more or less oblique. Each specimen tends to be split in a unique way, and the pieces of large or relatively rare forms can be identified (and glued together). Sometimes the fracture plane has cropped the free part of two or more denticles along an oblique plane. In contrast, breakage of specimens no longer encased in the rock will be orientated along a preferred crystal plane. At least in those ramiform elements I am familiar with, breakage of freed specimens is usually perpendicular to the length axis of the cusp, denticle or process. Denticles are usually broken at different heights or, if very badly damaged, at the base.

When the pieces of the sample are dissolved, conodont elements will begin to protrude until, for example, only a denticle tip remains in the limestone. Thus, any movement of the lumps towards each other or the support will easily break the specimen. Breakage will be partly that of encased elements, partly that of freed ones. In order to avoid such damage on silicified brachiopods, Cooper and Grant (1972, p. 17) coated the lower surface of their blocks with cellulose acetate. Conodont elements are smaller; hence, the extent of such breakage will be less. Placing the slabs nearly upright, leaning against the side of the colander, decreases the surface in contact with the support and thereby the amount of damage (in addition, dissolution is much faster, because such an orientation results in a much more vigorous circulation).

Rough or extended washing is extremely destructive. The (sub)final result is characteristic; instead of say the 2000–5000 well-preserved elements expected from an argillaceous sample, the yield may be a few tens or hundreds of fragments (the rest were broken into fragments small enough to be washed away). Such destruction can be minimized; see 6 and 7 in section 1.

Magnetic separation is frequently seen as a major source of breakage. However, my experience has been that it has very small effects compared with other causes, even on high-quality collections with long, thin denticles, if properly adjusted (see section 2).

Breakage due to crushing is probably not affected much by the purity of the rock. However, breakage due to washing and other concentration work increases easily with decreasing purity. The laboratory-induced breakage is partly method-related but also to a very large extent person-related (laboratory assistant, professor, research student). Two persons might get very different results, especially when processing difficult samples, even if both used the same pH measured, methods and screen. The relative destructiveness of different methods can be evaluated by determining what percentage of the denticle tips remains in a collection from a fine-grained argillaceous limestone.