

COMMENTARY: ESTIMATING SEXES OF HONEYEATERS FROM HEAD-BILL MEASUREMENTS

D. I. ROGERS and K. G. ROGERS

340 Ninks Road, St Andrews, Victoria 3761

Received: 21 August, 1993

In many species of bird, the sexes look similar but one sex is larger than the other. When a series of measurements is available from unsexed samples of such species, a situation that often arises from banding studies, analysis of biometrics can be used to estimate (attribute) the sex of individual birds. Recently, Pyke and Armstrong (1993) applied a novel method of analysis in sexing New Holland and White-cheeked Honeyeaters on the basis of head-bill measurements. In this note we draw attention to some serious problems with their approach.

EXISTING UNIVARIATE METHODS

The histogram of a single type of measurement for a sexually size dimorphic species will often be bimodal (double-humped) like that in Figure 1. A variety of methods exist for analysing such data-sets (e.g. Harding 1949; Cassie 1954; Hasselblad 1966; Griffiths 1968; Day 1969; Macdonald and Pitcher 1979; Rogers 1995 and unpubl.; Rogers *et al.* 1986; Batty 1993), all of which could be applied to the problem of estimating sexes of honeyeaters from head-bill measurements. In this issue of Corella, K. G. Rogers presents a suite of user-friendly computer programs to deal with these situations. The above techniques assume that the measurement is normally distributed for each sex; the 'double-humped' histogram results from sampling from a pair of distributions which can be represented by normal curves like those superimposed on Figure 1. This is a reasonable and generally accepted assumption when applied to birds (e.g. O'Connor 1985; Fowler and Cohen n.d.; Borowski and Borwein 1989) and supported by our experience of analysing linear measurements (Rogers *et al.* 1986, 1990; Marchant and Higgins 1990, 1993).

When applied to sexually size dimorphic birds, the above methods separate the component Normal distributions and provide estimates for each sex, of number of birds, mean, and standard

deviation. The methods of Griffiths (1968) and Rogers *et al.* (1986) are the only ones to have been widely applied to Australian birds (Rogers 1984; Rogers *et al.* 1986, 1990; Barter 1985, 1986, 1989, 1990; Fry 1990); Rogers *et al.* (1986) did not use the methods of Day (1969), *contra* Pyke and Armstrong (1993). When comparisons have been possible, parameters estimated from these methods for each sex of a species have been very similar to parameters calculated directly from sexed birds in other studies (see above references).

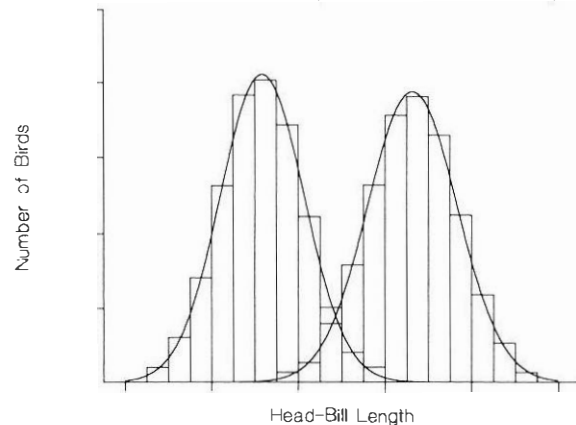


Figure 1. Example of a histogram of head-bill measurements for adults of a sexually size dimorphic species. The two component normal distributions (bell-shaped curves) are also shown.

Rogers *et al.* (1986, 1990) also calculated sexing criteria (rules by which individual birds can be sexed) for sexually size-dimorphic species. An estimate of the value at which 95 per cent of birds will be from the larger sex was given, as was the value at which 95 per cent of birds will be from the smaller sex (see Figure 2). Birds with measurements greater than the higher 95 per cent limit, or smaller than the lower 95 per cent limit, are even more likely to be correctly sexed (*contra* Pyke and Armstrong 1993); the proportions of birds correctly sexed, unsexed and mis-sexed by these criteria were also presented.

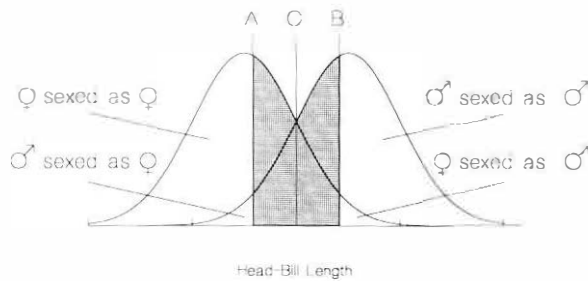


Figure 2. Example of Normal male and female head-bill distributions in a sexually size-dimorphic species in which males are larger. Birds in the grey zone are treated as unsexed. A. Value at which 95 per cent of birds are female (lower 95% limit). B. Value at which 95 per cent of the population is male (higher 95% limit). C. Value at which 50 per cent of the population is male and 50 per cent is female.

Pyke and Armstrong (1993) claimed that analytic methods for separating a bimodal distribution into two normal components require a computer and appropriate software, and that they do not provide estimates of sex for all individuals. Neither of these criticisms is valid. Several of the techniques for unscrambling these distributions are graphical (Harding 1949; Cassie 1954; Griffiths 1968; Rogers 1976) and can be done by hand. The methods calculate the sex ratio and the mean and standard deviation of the component normal distributions. These parameters can be used to develop sexing criteria, the analyst selecting the minimum probability of correct sexing; Rogers *et al.* (1986, 1990) followed statistical convention in using a minimum probability of 95 per cent. It is possible to estimate the sex of all individuals by using 50 per cent as the minimum probability of correct sexing; the implications of such a choice are discussed later.

THE APPROACH OF PYKE AND ARMSTRONG (1993)

One criticism of methods for separating bimodal distributions into two normal ones has been that they are difficult to use (Pyke and Armstrong 1993). To some extent this is true, simply because the problem is a difficult one. With practice, the graphical types of analysis take a couple of hours to perform and require only an understanding of probability paper and normal distributions. The numerical methods are generally to be preferred (they require fewer subjective judgements and allow the accuracy of estimates to be determined) but are considerably more

complex. In contrast, the method employed by Pyke and Armstrong involves two very simple steps. The first is to plot a histogram, manipulating the measurement intervals so that the double-humped nature of the distribution is clear. The second step is to locate the trough between the two distributions (i.e. the histogram interval between the peaks which has minimum frequency) by visual inspection (that is, by eyeballing). The midpoint of the trough is considered to be the sexing threshold, with the larger sex (usually male) lying to the right of this value, and the smaller sex (usually female) lying to the left.

Our concerns with this procedure are discussed below; they fall into three categories. First, no single threshold value will allow reliable sexing of all birds of a species in which males and females overlap in size. Secondly, the procedure does not produce results which can be used by other workers. Finally, for a variety of reasons we doubt that the procedure calculates the sexing threshold accurately.

Does a single value used as a sexing threshold have any value?

In most species of bird, including nearly all passerines, the sexes overlap to some extent in size. For these species, there is no single critical value above which a bird is certainly from one sex and below which it is certainly from the other. The nearest approach to a single 'sexing threshold' is the value (of, for example, a head-bill measurement) at which 50 per cent of birds will be from the larger sex and 50 per cent will be from the smaller sex. Birds with a head-bill measurement larger than this threshold value are more likely to be from the larger sex, and those with a shorter head-bill measurement are more likely to come from the smaller sex. However, those birds with head-bill measurements close to the threshold value cannot be sexed with great confidence since the probability that they are from one sex will be little different from the probability that they are from the other.

As an example, we have used data published on White-cheeked Honeyeaters from Beverley, Western Australia (Congreve, in Rogers *et al.* 1986). At this site (which is not in Victoria, *contra* Pyke and Armstrong 1993), the 50 per cent sexing threshold for head-bill can be calculated as 44.09 mm, males averaging larger. Using the parameters for head-bill measurements published

in Rogers *et al.* (1990) and a table of standard normal probabilities (and assuming, for the purposes of example, that sex ratios are equal) it is easy to work out the probabilities of mis-sexing birds of given head-bill lengths. For example, 14 per cent of birds with a head bill measurement of 44.6 mm will be mis-sexed females, as will 29.9 per cent of those with a head-bill of 44.4 mm and 40.5 per cent of those with a head-bill of 44.2 mm. In effect, there is a 'grey zone' around the sexing threshold in which sexing is unreliable. This concept is illustrated in Figure 2.

We think that in many studies (for example, when looking for sex-specific behaviours or sexual differences in the timing of moult) it is best to concentrate on observations taken from those individuals which can be sexed with some confidence. In the White-cheeked Honeyeater example above, emphasis could be given to the 80.7 per cent of birds which are sexed correctly to at least a 95 per cent level of probability. The choice of probability level is one for the analyst after taking into consideration the aims of the study, the 95 per cent level being a custom rather than a rule. Higher or lower levels of probability may be appropriate in some cases, but it is always necessary to find a suitable compromise between very high probability limits (which leave too many birds unsexed) and very low limits (which sex too many individuals incorrectly).

Non-usability of results

Two fairly obvious principles are well worth remembering when analysing or publishing measurements. The fundamental reason for measuring a bird is to find out how large it is, and 'the first task of biostatistics is to provide some form of summary description of the data' (O'Connor 1985). Surprisingly, Pyke and Armstrong (1993) have done neither; the output of their method does not describe the size of either sex. Further, their sexing thresholds cannot be used in other studies and their histograms are not given to a scale that allows readers to extract the information that they hold. These are serious problems. Measurements are genuinely useful and well worth publishing, but there is little point in doing so unless data-sets are described in sufficient detail for them to be used by others. In the discussion we suggest ways in which this can be done.

Does the intermodal trough give the required 50 per cent threshold?

The inter-modal trough located by Pyke and Armstrong (1993) and used as a sexing threshold is not always the same as the value at which a bird is equally likely to be male or female. The position of the inter-modal trough is determined by a combination of things, including size dimorphism, the sex ratio within the sample, the variation in size about the average in each sex, and the measurement intervals used to define the histogram.

The effect of unequal sex ratios. Sex ratios within a sample of birds will not always be equal, no matter how large the sample is. For example, virtually all Grey Plovers migrating as far south as Australia are thought to be females (Marchant and Higgins 1993; A. M. Dunn, pers. comm.); in European Sparrowhawks differential mortality of the sexes tips the sex ratio of adults strongly in favour of females (Newton 1986). If a sample contains a large proportion of birds from the larger sex, the plotted position of the inter-modal trough will be lower than it would be if the sex ratios were equal. This is demonstrated in Figure 3; each plot is from the same data-set (i.e. relative sizes of males and females are equal throughout) yet the location of the intermodal trough varies substantially with sex ratio.

The effect of unequal standard deviations. Even when the sex ratio is 50 per cent, the intermodal trough will differ from the 50 per cent threshold when the standard deviation (i.e. the scatter of measurements around the average) differs in each sex. This is a common situation because equal standard deviations will only occur when the larger sex is relatively less variable in size than the smaller sex. In general, standard deviations are larger for the larger sex (see, for example, measurements of independently sexed birds in Rogers *et al.* 1986, Marchant and Higgins 1990, 1993). In such cases the true 50 per cent threshold will be less than the value at the intermodal trough. If standard deviations are not estimated, the inter-modal trough will provide an estimate of the 50 per cent threshold value which has a bias of unknown size and direction.

The use of 'eyeballing' in drawing conclusions. Visual inspection of the data is a necessary, though not a final, step in any serious analysis, but it can be misleading. Pyke and Armstrong

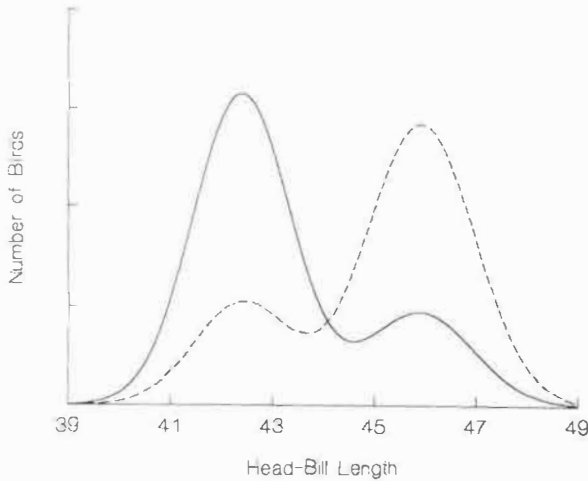


Figure 3. Example of head-bill distributions for two samples of mixed sex drawn from the same population of White-cheeked Honeyeaters, showing different inter-modal troughs. In the sample depicted by a solid line, 75 per cent of birds are female and 25 per cent are male; in that depicted by a broken line, 25 per cent are female and 75 per cent are male. Means and standard deviations are the same for both samples: 45.9 and 1.05 respectively in males, 42.4 and 0.95 respectively in females.

combined measurements of immatures and adults on the grounds that the histograms of the separated age classes looked similar. This is hardly surprising given that they used histogram intervals which were larger than age-related size-differences previously reported by Rogers *et al.* (1986, 1990). Even when age-related differences in size are slight, they can affect sexing analyses of data-sets in which age-classes are combined; males and females will appear more variable in size than they really are and fewer birds will be sexed in consequence.

Eyeballing can be insufficient to judge whether or not a histogram demonstrates size dimorphism. For example, Pyke and Armstrong (1993) published a histogram of the head-bill measurements of Yellow-faced Honeyeaters. It does not show the clear peaks and trough associated with a double-humped distribution and Pyke and Armstrong concluded that 'the distribution was unimodal'. We suspect this histogram is consistent with a sample from a species in which the sexes differ slightly in size (see Rogers *et al.* 1986). In such circumstances the resultant histogram can be single-humped, even when there is a real size difference between males and females (see Figure 4). The characteristic bimodal histogram is only

seen when there is a fairly large and consistent size difference between the sexes. Pyke and Armstrong could have tested if their histogram of Yellow-faced Honeyeater head-bill lengths concealed slight size dimorphism by comparing it against a normal distribution, perhaps by applying a chi-squared test or by plotting the cumulative distribution on probability paper (Harding 1949).

Sampling error. A histogram of a sample measurement is unlikely ever to show the perfect regularity expected from distributional considerations. The numbers of birds in some intervals will be higher than expected, and in some intervals they will be lower; this random variation (sampling error) is most striking when samples are small. Accordingly histograms tend to look messy unless samples are very large; it can thus be difficult to locate the exact position of the intermodal trough in small data-sets. In practice the samples needed for an attractive histogram are larger than those needed to provide good estimates of mean and standard deviation. Over- or under-representation of intervals is especially likely to be striking in the region of the intermodal trough because smaller numbers of birds occur in that area. There is

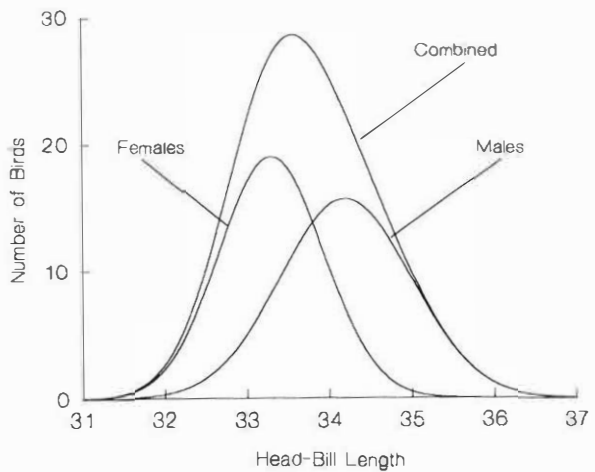


Figure 4. Distributions of head-bill measurements for adult Yellow-faced Honeyeaters; males and females separately and combined. The combined distribution is single-humped. Based on data published in Rogers *et al.* (1986):

	No. of birds	Mean	Standard deviation
Males	62	34.2	0.79
Females	60	33.3	0.63

therefore potential for the observed intermodal trough to differ from the 'true' one, especially when the data-sets analysed are smaller than those available to Pyke and Armstrong.

Choice of histogram intervals. Last but not least, the position of the inter-modal trough is affected by the histogram intervals selected. Large intervals may define the trough but the 'threshold' will lie in a relatively large range. Small intervals are subject to large sampling error, so the plotted position of the intermodal trough may be unclear or inaccurate. The definition of a histogram is at the choice of the analyst. Different choices of starting point and interval size can dramatically alter the appearance of a histogram, particularly with smaller samples (e.g. Silverman 1986). Different analysts working with the same data sets could make different choices and find different intermodal troughs.

VALIDATION OF SEXING CRITERIA

Pyke and Armstrong (1993) tested sexing threshold values by looking at measurements of 77 New Holland and 47 White-cheeked Honeyeaters which had been sexed on the basis of brood patch or behaviour. They concluded that 8 per cent of New Holland and 11 per cent of White-cheeked Honeyeaters from these samples were sexed incorrectly by using head-bill thresholds. However it is not possible to say that these figures apply to the study population in general, for sampling error may well have affected their results. Very large samples of sexed birds are needed to ensure that the representation of very small males and very large females caught (i.e. the birds most likely to be mis-sexed) in the sample is similar to that occurring in the entire study population. In general it is more accurate to use means and standard deviations (which are calculated from the entire data set) to predict the nature of size overlap between males and females.

There are also practical disadvantages to the validation method used by Pyke and Armstrong. In many circumstances bird-banders will have few or no opportunities to sex individual birds directly, particularly if they are working from old data-sets. Provided that they know which sex is larger, analysts using procedures which unscramble mixed normal distributions do not need samples of independently sexed birds. The suite of programs presented by Rogers (1995) also allow the analyst to examine the effects of sample size

and measurement precision on the performance of sexing criteria.

DISCUSSION

In most species of bird, the sexes overlap in size and measurements will not allow reliable sexing of every individual. This does not mean that attempts to use biometrics in sexing birds will be fruitless; a soundly based approach that will permit a large proportion of birds to be sexed is certainly better than leaving all birds in a study population unsexed. In addition, the analyses involved are helpful in describing samples.

A normally distributed sample is described by its size (i.e. number of birds), mean and standard deviation. For species in which the sexes differ in size, these parameters ought to be presented separately for males and females. In such cases, combining measurements of both sexes is unsatisfactory because the value of the mean will depend on the sex ratio within the sample. For similar reasons, measurements for different age classes should only be combined when it is clear that the ages do not differ in size.

Mean, standard deviation and sex ratio **cannot** be calculated accurately for each sex if a single threshold value is used as the basis of sexing. This method would make the average measurements for each sex appear more divergent than they really are, except in those few species in which the sexes do not overlap in size. The parameters can be estimated with considerable accuracy by the analytic methods cited above. When authors are unable to perform such an analysis we think they would be justified in publishing their raw data in a form which allows other readers to use it. Tabular histograms with small intervals are a convenient way of doing this. For further information and an example, see Rogers (1995).

Pyke and Armstrong (1993) state in their discussion that their method is suitable for region-specific studies. This commentary has shown that their method is subjective, is subject to biases of uncertain size and direction, and produces no objective measures of performance. With these limitations, their method can only give a general indication of size dimorphism in a population, and then only if the size differences between the sexes are substantial enough for a histogram of their measurements to show an intermodal trough.

They also state that in cases where accurate sexing is not possible on the basis of a single type of measurement, analysis of a combination of measurements 'using multivariate techniques such as discriminant analysis' may be required. We agree that multivariate techniques are potentially better than univariate ones, but they should not be regarded as a cure-all. First, no statistical method will allow accurate sexing if the samples are too small or if the size differences between the sexes are too slight. Secondly, multivariate analyses are much harder to understand and perform than univariate ones and require larger samples because more parameters need to be estimated. Thirdly, published guidelines are rather inaccessible, for as yet ornithologists appear not to have used multivariate techniques appropriate for estimating sex from multivariate data when no independently sexed birds are available. Discriminant analysis requires a sample of sexed birds. There is an extensive literature on the use of discriminant analysis in sexing birds from measurements. Criticism of these papers is beyond the scope of this commentary but we would suggest that ornithologists using discriminant analysis describe the 'grey zone' in which sexing is unreliable, and consult Fatti *et al.* (1982) for formulae that deal with situations when standard deviations are not equal in each sex.

It has not been our intention in this commentary to imply that sexing birds by measurements is a problem that can only be addressed by sophisticated statistical techniques, nor do we want to suggest that there is no room for new analysis techniques. We doubt, however, that short-cuts will prove helpful, for any new method of analysis should fulfil the following two conditions: it should provide unbiased estimates of the size of each sex and it should find the limits of the 'grey zone' in which measurements do not provide a reliable sexing guide.

ACKNOWLEDGMENTS

We are grateful to Annie Rogers and an anonymous referee for commenting on the manuscript, and to Margaret Cameron and John McLaughlin for their help in locating some references.

REFERENCES

- Barter, M. A. (1985). Sex determination of bill length of live adult Curlew Sandpipers. *Silt* 7: 8-17.
- Barter, M. A. (1986). Great Knots partly undone. *Silt* 9: 5-20.
- Barter, M. (1989). Bar-tailed Godwit in Australia. Part 1: Races, breeding areas and migration routes. *Silt* 14: 43-47.
- Barter, M. (1990). Morphometrics of the Eastern Curlew *Numenius madagascariensis*. *Silt* 16: 36-42.
- Batty, L. (1993). Computer analysis of wader morphometric data. *Wader Study Group Bull.* 70: 23-27.
- Borowski, F. J. and Borwein, J. M. (1989). 'Dictionary of Mathematics.' (Collins: London.)
- Cassie, R. M. (1954). Some uses of probability paper in the analysis of size frequency distribution. *Aust. J. Marine and Freshwater Res.* 5: 513-522.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* 56: 463-474.
- Fatti, L. P., Hawkins, D. M. and Raath, E. L. (1982). Discriminant Analysis. In 'Topics in Applied Multivariate Analysis.' (Ed. D. M. Hawkins.) pp 1-71. (Cambridge University Press: Cambridge.)
- Fowler, J. and Cohen, L. (n.d.). 'Statistics for Ornithologists.' British Trust for Ornithology, Guide 22.
- Fry, G. (1990). Biometrics. Moulting and Migration of Terek Sandpiper *Tringa terek* spending the non-breeding season in north-west Australia. *Silt* 17: 24-28.
- Griffiths, J. (1968). Multi-modal frequency distributions in bird populations. *Bird Study* 15: 29-32.
- Harding, J. P. (1949). The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Mar. Biol. Ass.* 28: 141-153.
- Hasselblad, V. (1966). Estimation of parameters for a mixture of normal distributions. *Technometrics* 8: 431-444.
- Macdonald, P. D. M. and Pitcher, T. J. (1979). Age groups from size frequency data: a versatile and efficient method of analysing distribution methods. *J. Fish. Res. Board Can.* 36: 987-1001.
- Marchant, S. and Higgins, P. J. (Eds.) (1990). 'Handbook of Australian, New Zealand and Antarctic Birds. Volume 1, Ratites to Ducks.' (Oxford University Press: Melbourne.)
- Marchant, S. and Higgins, P. J. (Eds.) (1993). 'Handbook of Australian, New Zealand and Antarctic Birds. Volume 2, Raptors to Lapwings.' (Oxford University Press: Melbourne.)
- Newton, I. (1986). 'The Sparrowhawk.' (Poyser: Calton.)
- O'Connor, R. J. (1985). Biostatistics. In 'A Dictionary of Birds.' (Eds B. Campbell and E. Lack.) (T & AD Poyser: Calton.)
- Pyke, G. H. and D. P. Armstrong (1993). Estimating sexes of New Holland Honeyeaters from head-bill measurements. *Corella* 17: 43-46.
- Rogers, K. G. (1976). Sexing Criterion for the Greater Flamingo Based on Tarsal Length Measurements (United Nations Development Program, Office of Technical Co-operation, Iran). (Available from author.)
- Rogers, K. (1984). Bar-tailed Godwit morphometrics. *Victorian Wader Study Group Bulletin* 8: 23-25.
- Rogers, K. G. (1995). SHEBA: Computer programs for sexing birds on measurements using univariate data. *Corella* 19: 25-34.
- Rogers, K., Rogers, A., Rogers, D., Lane, B. and Male, B. (1986). 'Banders Aid. A guide to ageing and sexing bush birds.' (A. Rogers: St. Andrews, Victoria.)
- Rogers, K., Rogers, A. and Rogers, D. I. (1990). Banders Aid: Supplement Number One. RAOU Report No. 67.
- Silverman, B. W. (1986). 'Density estimation for statistics and data analysis.' (Chapman and Hall: New York and London.)