ANIMAL BEHAVIOUR, 2005, **69**, F1–F7
doi:10.1016/j.anbehav.2004.09.002

Available online at www.sciencedirect.com
SCIENCE DIRECT®

# The appropriate use of Zipf's law in animal communication studies

**B. McCOWAN*, L. R. DOYLE†, J. M. JENKINS‡ & S. F. HANSER§**
*Population Health and Reproduction, School of Veterinary Medicine,
University of California, Davis
†SETI Institute
‡SETI Institute, NASA Ames Research Center
§Ecology Graduate Group, University of California, Davis

*(Received 3 September 2003; initial acceptance 18 September 2003;
final acceptance 23 January 2004; published online 26 November 2004; MS. number: AF-24)*

A Zipf plot (or statistic) is a log–log plot of the frequency of occurrence of signalling units (letters, words, phonemes, etc.) against their rank order (1st, 2nd, 3rd). Zipf's law emerges for almost all languages' letters and words as an approximate slope of −1 in this log–log plot, a result George Zipf (1949) stated was due to the 'principle of least effort' in communication systems, representing a 'balance' between the repetition desired by the listener, and the diversity desired by the transmitter. There have been many applications (some correct and some not) of this plot in animal communication studies (reviewed in McCowan et al. 1999) with a recent critique being that of McCowan et al.'s (1999) work by Suzuki et al. (2004), to which we herein reply.

The purpose of our reply is to clarify the inferences made in McCowan et al. (1999) as well as the correct use of Zipf's law in animal communication studies. There is much, on a one-by-one basis, that could be addressed from Suzuki et al. (2004); however, for the sake of brevity, we chose to paint this reply with a broader brush. We summarize the important points just below, which we then address in some additional detail in the main text.

(1) Suzuki et al. (2004) claim that McCowan et al. (1999) attributed linguistic properties to Zipf's law and used Zipf's law as a language detector. McCowan et al. (1999) 'never' attributed linguistic properties to Zipf's law; on the contrary, we outlined its correct application and the limitations of using this statistic in our paper (which we quote below). It appears to us that the conflict over the Zipf statistic having linguistic value (or semantic content, i.e. meaning) is really between Suzuki et al. (2004) and Cancho & Solé (2003) because Suzuki et al. (2004) state that '…Zipf's law is not an appropriate route to conclude anything about the linguistic nature or potential capacity for communication transfer' (page 11), while Cancho & Solé (2003) state 'Our finding strongly suggests that Zipf's law is a hallmark of symbolic reference and not a meaningless feature' (page 788). Again, Suzuki et al. (2004) state 'Zipf's law is not even a necessary condition for a data sequence to have semantic content…' (page 14), while Cancho & Solé state, 'Our results strongly suggest that Zipf's law is required by symbolic systems' (page 791). McCowan et al. (1999) only apply the Zipf statistic as an 'indicator of potential structure' in the distribution of signals, and then only in a differential sense with changes in the Zipf slopes being indicative of changes in the structural distribution of a signalling system repertoire (and then only at the repertoire level). Because Hailman et al. (1986) did attribute linguistic properties to bird calls using a Mandelbrot fit to Zipf's law, this paper also may be drawing conclusions contrary to those of Suzuki et al. (2004).

(2) Suzuki et al. (2004, page 14) imply that Zipf slopes cannot be used even in a differential sense when they state that it cannot be used 'as a comparison of two communication schemes'. While we are not certain what constitute 'two communication schemes' for Suzuki et al. (2004), we do apply it as an indicator of changes in the distribution of bottlenose dolphin, *Tursiops truncatus*, signals with age in McCowan et al. (1999) and the distribution of squirrel monkey, *Saimiri sciureus*, signals with age in McCowan et al. (2002). These results have also been confirmed by a bootstrapping of both infant (<1 month old) and adult signal data sets (see Fig. 1). These bootstrap analyses show that the differences in Zipf slopes

*Correspondence: B. McCowan, Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, 18830 Road 112, Tulare, CA 93274, U.S.A. (email: bmccowan@vmtrc.ucdavis.edu). L. R. Doyle is at the SETI Institute, 2035 Landings Drive, Mountain View, CA 94043, U.S.A. J. M. Jenkins is at the SETI Institute, MS 245-3, NASA Ames Research Center, Moffett Field, CA 94035, U.S.A. S. F. Hanser is at the Ecology Graduate Group, University of California, One Shields Avenue, Davis, CA 95616, U.S.A.*



0003–3472/04/$30.00/0

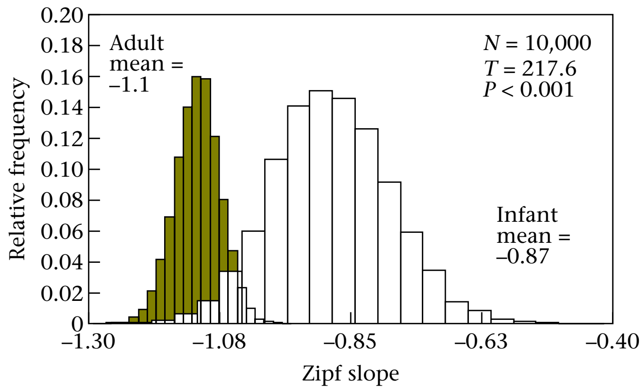


**Figure 1.** Distribution of values generated from a Monte Carlo simulation on adult and infant (<1 month) Zipf slopes in bottlenose dolphins. We conducted 10 000 iterations of Zipf slope calculations on the probability structure of the frequency of use of whistle types for adult and infant dolphins, respectively, and tested the results using a heteroscedastic *t* test (Law & Kelton 2000).

are significant and in the direction already indicated in the Zipf slope changes shown in McCowan et al. (1999, 2002). Suzuki et al. (2004) also object to our not fitting the Zipf plot with the additional constraint that the signal data set size be preserved because the constant *c* is constrained by the slope $\alpha$ in Zipf distributions. We discuss this important point in detail below.

(3) Suzuki et al. (2004) argue that because a nonlinguistic random process can produce a data set with a Zipf slope of $-1$, the Zipf slope cannot be used to investigate or characterize communicative repertoires. To demonstrate this, they reconduct a die-throwing experiment originally conducted by Li (1992). In this experiment, the components of 'words' are generated by a random process with each word length proportional to its probability of occurrence. This is essentially imposing on the uniform distribution of the die-throwing results a set of 'rules' that are purposely designed to sort the random data set into a form that mimics the most efficient structure for communication (i.e. the shortest sequences are the most frequent, and the longest sequences occur with the least frequency). The output of this experiment is similar to that of a Huffman coder, which is designed to optimize bandwidth use. Yet, one could choose another random process (e.g. randomly sampling from a uniform distribution) that would not have resulted in a Zipf slope of $-1$. More importantly, the fact that a random process can be designed to mimic Zipf's law does not alter the utility of the Zipf slope as a tool. This experiment simply demonstrates that the Zipf slope is not a 'sufficient' condition for the presence of higher-order structure. The experiment does not address whether it is a 'necessary' condition (nor does Suzuki et al.'s data compression example; see below).

(4) Suzuki et al. (2004) claim that whistles were defined in McCowan et al. (1999) by an intersignal interval of greater than 300 ms. We did not define bottlenose dolphin whistles as a predetermined time segment. That is, signal gaps had no threshold (300 ms or otherwise) for signal separation, and the signals themselves were categorized using an iterative *k*-means cluster analysis protocol based upon whistle contour (see McCowan 1995; McCowan & Reiss 1995), which was well referenced in McCowan et al. (1999).

Below we discuss these points and a few others in a bit more detail, while still trying to keep the discussion broad enough so as not to lose the main points about the correct versus incorrect applicability of the Zipf slope.

## Point 1. Linguistic Properties of the Zipf Slope

As stated above, McCowan et al. (1999) never attributed linguistic properties to the Zipf slope. The term 'linguistic' does not appear anywhere in McCowan et al. (1999), and in fact, we point out explicitly the limitations of using this relationship. Instead, we applied the Zipf statistic as an indicator of 'potential' structure in the distribution of bottlenose dolphin signals at the repertoire level, which is quite different. Examples from McCowan et al. (1999) are quoted below in which we use the terms 'potential' and 'comparative' specifically to avoid any misunderstanding about the scientific inferences we make.

*…we use the first-order entropic relation in a Zipf-type diagram… to illustrate the application of temporal statistics as comparative indicators of repertoire complexity…* **(McCowan et al. 1999, page 409)**

[Zipf's law] *measures the potential capacity for information transfer at the repertoire level by examining the 'optimal' amount of diversity and redundancy necessary for communication transfer across a 'noisy' channel (i.e. all complex audio signals will require some redundancy).* **(McCowan et al. 1999, page 410)**

(Note: we put the term 'optimal' in quotes here to indicate a balance between diversity and redundancy, which is clear from the context.)

*…we apply Zipf's statistic (Zipf 1949, 1968) to dolphin whistle vocalizations to illustrate its application as a comparative indicator of the structural complexity of vocal repertoires, as well as a potential indicator of acquisition/learning in animal vocal repertoires.* **(McCowan et al. 1999, page 410)**

*Such a function* [Zipf plot] *nevertheless remains a valid indication of both, the nonrandomness of a system as well as the potential capacity for communication transfer of such a system.* **(McCowan et al. 1999, page 411)**

(Note: an independent uniform random distribution will have a flat slope, and this term might have been better than the simple term 'random' in case there were any misunderstandings about what we meant.)

*The structure of the system is neither too repetitive (the extreme would be one signal for all messages) nor too diverse (the extreme would be a new signal for each message and, in practice, a randomly distributed repertoire would represent the highest degree of diversity). Thus, a system exhibiting such* *balance can be said to have a high potential capacity for transferring communication (which we term 'high potential communication capacity'). It only has the 'potential' to carry a high degree of communication, though, because Zipf's statistic only examines the structural composition of a repertoire, not how that composition is internally organized within the repertoire (i.e. higher-order entropies).* **(McCowan et al. 1999, page 411)**

*The Zipf statistic can measure and compare the structural complexity of vocal repertoires, but it has limitations in evaluating the actual relationship between signals in a dynamic repertoire. For instance, Zipf's statistic cannot examine how signals interact or are internally organized within a communicative repertoire.* **(McCowan et al. 1999, page 412)**

(Note: this essentially states that the Zipf statistic cannot really measure what might be called 'linguistic structure' or *n*-gram structure.)

[Zipf's] *statistic, applied to animal communication systems, can be interpreted only on a comparative and/or developmental basis…* **(McCowan et al. 1999, page 417)**

*A negative regression coefficient (negative slope) approaching −1.00 for a communicative repertoire suggests high potential for communication capacity in that repertoire (to be confirmed with higher order entropic analysis)…* **(McCowan et al. 1999, page 417)**

*As Zipf's statistic may provide an initial comparative indicator of repertoire structural complexity and its development, the slopes of higher-order entropies may serve as a comparative indicator of the deeper organizational complexity of animal vocal communication systems and its development.* **(McCowan et al. 1999, page 418)**

The quotes above are essentially what McCowan et al. (1999) had to say about the applicability of the Zipf statistic. A Zipf slope of approximately −1 does appear to be a necessary condition for a full-fledged language, but it may not be sufficient. But this is old news. Cancho & Solé (2003) reference work by Mandelbrot, written in 1966, showing that random combinations of letters and blanks can reproduce Zipf slope, and Li (1992) showed that a similar pattern can be generated using dice. Therefore, Suzuki et al.'s assertion that McCowan et al. use Zipf slope as a 'language detector' is untrue and misrepresentative of the statements in McCowan et al. (1999). Suzuki et al. (2004) also state that 'one would expect a high rate of false positive decisions on the statistical test [from Zipf analysis]' (page 14), implying that, in McCowan et al. (1999), we are somehow thresholding our estimates of the Zipf parameter, which is not the case. In summary, it is clear that, contrary to the statements in Suzuki et al. (2004) such as 'when Zipf's law is used as a test for linguistic …processes…, as in McCowan et al. (1999)' (page 9), we never attributed linguistic properties to Zipf's law, but rather outlined (as quoted above) its limitations along with its correct application in animal behavioural studies.

Therefore, McCowan et al. (1999, 2002) agree that the Zipf slope is linguistically shallow. Zipf statistic is linguistically shallow because it tells us nothing about the relationship between signals in a repertoire (e.g. the higher-order entropies). But this fact has no bearing on its utility as a comparative tool at the repertoire (i.e. first-order approximation to the entropy) level or as an indicator of the 'potential' for higher-order structure in a system, again as we have already stated (McCowan et al. 1999, 2002).

## Point 2. Mandelbrot, Model Fitting and the Meaning of Zipf Slope

As Suzuki et al. (2004) note, Mandelbrot (1953) introduces two additional parameters into the fit of the frequency curve, but this is not justified nor is it what George Zipf did. Naturally, more parameters can fit a curve better, but this does not mean that the data automatically justify it, and can sometimes hide the trend one is looking for. Comments by Pierce (1980, pp. 246–247) regarding the work of Mandelbrot (1953) may be useful here: '… the lengths of "words" produced by the random process described don't correspond to the length of words as found in typical English text'. Pierce also points out that language certainly has nonrandom features because, for example, words like 'taxicab' get shortened with use to, 'cab'. To claim that a random production of numbers leading to the production of 'word-strings' that obey Zipf's law is responsible for language word frequencies, it would be necessary to show that the forces that shape language processes obey this random process, which has not been done in the 50 years following Mandelbrot's suggestion.

Suzuki et al. (2004) state that the results reported in McCowan et al. (1999) using Zipf-type slopes to demonstrate potential for increased complexity from infant-to-adult dolphin populations does not support the conclusion of meaningful differences because only formal errors for the fitted parameters were reported in McCowan et al. (1999). We therefore performed a bootstrap analysis of the original fit by generating a large number of Monte Carlo simulation realizations in @Risk software (Palisade Corporation, Newfield, New York, U.S.A.) based upon the probability structure of each population's (adult, infant) whistle repertoire and fitting a linear function (the Zipf plot) to thereby empirically determine confidence intervals on the fitted slopes. Monte Carlo simulations demonstrated that the difference in the Zipf slope between infants ($<1$ month) and adult bottlenose dolphins was significant using a two-sample heteroscedastic $t$ test (Law & Kelton 2000) at a $P$ value of less than 0.001 (see Fig. 1). The mean slope was −1.1 for adults and −0.87 for infants and the 99.9% confidence interval for the mean slope was 0.001 for adults and 0.003 for infants, respectively. Therefore, these bootstrap results on infant and adult dolphin whistle repertoire structure support the applicability of Zipf slope as both a 'quick-look' analysis of potential further *n*-gram structure and, more importantly, as a differential technique for easily indicating changes in communicative repertoires, as discussed in McCowan et al. (1999, 2002).

Suzuki et al. (2004) raise an important point regarding the appropriate way to fit a Zipf distribution to a data set. They claim that unless the fitted model is constrained to yield a valid probability density distribution, the resulting Zipf slopes are invalid. However, because Zipf (1949) himself did not constrain the fitted model and the purpose of our comparison was to preserve continuity with Zipf's original application, we think that this criticism is invalid. Having said that, however, Suzuki et al. do make the point that constraining the model might be a more robust approach. Yet, as the following discussion will show, this is not a particularly strong argument when applied to well-sampled data sets with appropriate model-fitting procedures.

Suzuki et al. (2004) give an explicit formulation for the probability $p(r)$ of a symbol of rank $r$ drawn from a Zipf distribution:

$$p(r) = cr^{\alpha}, \tag{1}$$

where $\alpha$ is the Zipf slope and the constant, $c$, is constrained so that

$$\sum_{1}^{\infty} p(r) = 1. \tag{2}$$

One consequence is that for infinite, discrete probability spaces, the Zipf distribution is undefined for $\alpha = -1$, since no choice for $c$ exists that will ensure convergence of the series implied by equation (2), as Suzuki et al. (2004) note. However, there does not appear to be any reason to restrict Zipf distributions to infinite, discrete probability spaces. Indeed, any human language contains only a finite number of distinct words, phonemes, and so forth. So it appears reasonable to fix the upper limit of the summation in equation (2) to be $N$, the number of distinct elements in the data set. For such finite element probability spaces, Zipf distributions with $\alpha \geq -1$ are perfectly valid.

Now one of Suzuki et al's major objections to the results of McCowan et al. (1999) is that the fits to the data did not constrain the fitted distributions to be valid probability density distributions. Although this point is important philosophically, it is less cogent when considered in the context of empirical data fitting. That is, if the data are indeed well described by the model being fitted, then a numerical model with additional parameters should yield an equivalent fitted distribution, to within the experimental uncertainties. That is, the inclusion of a parameter in the model that is constrained by prior knowledge should not perturb the fit so long as there are an adequate number of data points to fit to. Empirically, fitting a Zipf distribution to a data set in the manner described by McCowan et al. (1999) is essentially a least-squares line fit to the log of the observed frequencies versus the log of the ranks. Suzuki et al. (2004) criticize this approach because it might distort the results. To investigate this matter, we performed constrained log least-squares fits to the repertoires of the adult dolphins and of the infant dolphins. For the adult dolphins, the constrained fit yielded $\alpha = -1.27$, in contrast to the unconstrained result $\alpha = -0.95$. For the infant dolphins, the constrained fit yielded $\alpha = -0.825$, which is essentially the same as the unconstrained value of $\alpha = -0.824$ (see Table 1). From this we would infer that the infant repertoire is better modelled by a finite Zipf distribution than is the adult repertoire. An examination of the adult repertoire reveals that the frequency of the first rank whistle is the culprit.

Can we modify the fitting procedure to reduce the sensitivity to the extra parameter represented by $c$? One characteristic of the fitting procedure used by McCowan et al. that is not criticized by Suzuki et al. (2004) is the implicit weighting of each data point in the log least-squares fit. The resulting fit is equivalent to a chi-square fit where the standard deviation of each bin is proportional to the frequency of that symbol. We note that the distribution of each bin in a histogram should follow a Poisson distribution, where the variance is equal to the average value of the number of counts in that bin. A better approach to fitting the data sets might be achieved by performing chi-square fits where the uncertainties of each data point are more accurately represented by the underlying Poisson distributions. We conducted such fits to the adult and infant dolphin repertoires, with the following results. For the adult dolphins, the one-parameter chi-square fit yielded a Zipf slope of −1.628, and a two-parameter chi-square fit yielded the same value to three decimal places. For the infant repertoire, the one-parameter chi-square fit yielded $\alpha = -0.807$, and the two-parameter chi-square fit yielded the same value to three decimal places (see Table 1). The slope of the line fitted to the adult distribution was even steeper because the first rank symbol was given more weight in the fit relative to the original fit.

Additionally, we carried out bootstrap analyses to determine the probability that the measured slope for the infants would be as high as that for the adults for each fitting procedure. This was carried out in two steps. First, we obtained bootstrap estimates of $\alpha$ by constructing artificial data sets from the original repertoires via sample and replacement. We constructed and fit 10 000 artificial adult and infant repertoires using all four fitting techniques described here (one-parameter/two-parameter, log

**Table 1.** Estimated Zipf slopes using different fitting techniques

| Fitting method | One-parameter log least-squares | Two-parameter log least-squares | One-parameter chi-square | Two-parameter chi-square |
|---|---|---|---|---|
| Adult repertoire | −1.2783 | −0.9534 | −1.6277 | −1.6277 |
| Infant repertoire | −0.8245 | −0.8240 | −0.8071 | −0.8070 |

**Table 2.** Significance of observed differences in $\alpha$ between the infant and adult data sets

| Fitting method | One-parameter log least-squares | Two-parameter log least-squares | One-parameter chi-square | Two-parameter chi-square |
|---|---|---|---|---|
| Significance of observed differences in $\alpha$ | $2\times10^{-6}$ | 0.00522 | $1\times10^{-7}$ | $2\times10^{-7}$ |

least-squares/chi-square). Next, we used these estimates for $\alpha$ to generate 10 000 000 pairs of estimates for each repertoire by bootstrap. The results show that the constrained log least-squares fit, and both unconstrained and constrained chi-square fits had less than $2 \times 10^{-6}$ chance of yielding an $\alpha$ for the infants that fell at or below the value for $\alpha$ fitted to the adult repertoire. In the case of the unconstrained log least-squares fit, there was a 0.005 chance that this would occur. In summary, the estimates for $\alpha$ from chi-square fits showed little sensitivity to whether or not the probabilistic constraint on the fitted model was enforced. Although there was some sensitivity to the constraint for the adult repertoire in the log least-squares fits, there was almost none for the infant repertoire. Regardless, bootstrap analyses showed that the differences in the fitted $\alpha$s for the adults and infants were highly significant, even for the two-parameter, log least-squares fit (see Table 2).

## Point 3. Die Throwing, Huffman Coding, Compression and Zipf Slope

Suzuki et al. (2004) have simulated in their die-rolling example a process similar to Huffman coding, which optimizes bandwidth use, a distribution that Suzuki et al. (2004) generated when they imposed the 'double-six' word-length generating rule that ordered the random distribution so that word length became inversely correlated with frequency of occurrence. So, while the Zipf slope is not a 'sufficient' measure of higher-order conditional probabilistic processes, it is an indicator of a process that is trying to maximize the use of available bandwidth, such as a communicative-like process.

Suzuki et al. (2004) have thus repeated an experiment done by Li (1992) that basically shows that if one organizes a uniformly distributed random process in a way that optimizes channel efficiency, one can obtain a $-1$ slope in a log–log plot of the probabilities arranged in rank order. They argue that the die-rolling game generates 'words' with frequencies that obey Zipf's law and that because the generated sequences are devoid of semantic content, this voids the use of Zipf's statistic in analysing communicative repertoires. It is interesting to ask why the sequences from the die-rolling game follow Zipf's statistic: the answer is that 'they are designed purposefully to do so'; that is, shorter words are favoured over longer words. A similar argument is used in justifying Zipf's law by his principle of least effort: if it costs energy to communicate, then it makes sense to have the most common words be short (Mandelbrot 1953 uses a similar cost analysis approach). Not much would be said if pronouns and participles like 'the' were replaced by words such as 'supercalifragilisticexpealidocious'.

Yet, Suzuki et al. (2004) could have just as easily chosen a different random process, the simplest being a random sampling from a uniform distribution. For example, if one instead threw a 27-faced die and simply tallied up the frequency of occurrence of each face, one would instead get a Zipf slope of 0. So, the outcome of the Zipf slope from such simulations is entirely dependent upon the random process (and underlying 'rules') one chooses. In Suzuki et al.'s example, while each die throw is completely uncorrelated with the next throw (i.e. independent), the 'rule' imposed on the data, of having the throw of a six determine the truncation of a 'word', produces an inverse correlation between the word size and that word's frequency of occurrence, which happens to mimic Zipf's law.

In addition, the fact that a random process can be designed to mimic Zipf's law does not alter the utility of the Zipf slope as a 'quick-look' tool. This experiment simply demonstrates that the Zipf slope may not be a 'sufficient' condition for higher-order structure, such as that found in human languages. This experiment, however, does not address whether it might be a 'necessary' condition.

To address whether Zipf's law is a 'necessary' condition (also see Cancho & Solé 2003), Suzuki et al. (2004) provide an example using data compression. They state that a compressed (gzipped) data set contains the same information and communicative content even though the values of the Zipf slopes for the uncompressed, compared with the compressed, data sets are not the same (uncompressed $-1.94$, compressed $-0.141$). They state, that because 'Both files contain the same information and communicative content…' (Suzuki et al. 2004, page 11), Zipf's law may not be even a 'necessary' condition for the presence of additional $n$-gram structure. This is not correct. The compressed data set 'plus' the compression algorithm/program itself, taken together, may have close to the same information content (which, incidentally is not the same as communicative content), but the gzipped data set by itself certainly does 'not' contain the same information content as the un-gzipped data set or there would be no point in compressing it in the first place. Thus, Suzuki et al.'s example here actually demonstrates the utility of the Zipf slope as an initial indicator of different amounts of information in the data sets. The compressed data set has a smaller negative slope, which means it contains much less redundancy and probably lower $n$-gram structure, so that this is actually quite a good example of the 'quick-look' utility of the Zipf plot.

## Point 4. Signal Definition and Categorization

Suzuki et al. (2004) incorrectly state that McCowan et al. (1999) defined dolphin whistle signals by an interwhistle interval of greater than 300 ms. We only stated that a 'typical' interwhistle interval was 300 ms. Whistles were defined by a start and stop in signal energy, regardless of the time interval between signals (common in studies of animal vocal communication). The contour classification method of McCowan (1995), based on *k*-means cluster analysis, was described and justified in both McCowan (1995) and McCowan & Reiss (1995), which are well referenced in McCowan et al. (1999). As stated in these papers, we agree that categories of dolphin whistles should be based upon (as much as possible) the perceptual boundaries of appropriately tested subjects. In fact, the contour classification method developed by McCowan (1995) was based upon concrete experimental evidence on how dolphins spontaneously imitate novel computer-generated whistles (Reiss & McCowan 1993), which to some extent reflect salient aspects of their perceptual system. In addition, each data set was analysed and categorized using the exact same methods and criteria, so the comparative usage of the Zipf statistic is still valid. Thus, the implication in Suzuki et al. (2004) that misclassified whistles distorted the outcome of the Zipf statistic in McCowan et al. (1999) is without basis.

## Point 5. Some Details

(1) Suzuki et al. (2004) state, 'entropy is a fundamental property of an information source, and as such is unordered' (page 10). Although we thought the terms 'zero-order,' 'first-order,' 'second-order', and so forth were clear in the context of their use, we could have been more explicit by stating that the 'entropic orders' were the 'approximations to the entropy'. For example, the term 'first-order entropy' is the 'approximation to the first-order entropy', the term 'second-order entropy' is the 'approximation to the second-order entropy' and so on, as Shannon & Weaver (1949) actually put it in their original volume on information theory (see also Chatfield & Lemon 1970; Yaglom & Yaglom 1983).

(2) In response to McCowan et al.'s (1999) statement that the Zipf slope '… remains a valid indication of both the nonrandomness of a system as well as the potential capacity for communication transfer…' (page 411), Suzuki et al. (2004) state 'In common terminology, they are stating that Zipf's statistic can test whether the sample points deviate from being equiprobable' (page 10). Yet, if one obtains a Zipf plot slope of $-0.6$ for infant babbling and a $-1.0$ slope for adult Russian phonemes, for example, one can say a lot more than just that these two distributions both deviate from being equiprobable. One knows, for example, that the babbling data set consists of phonemes that are more evenly distributed in frequency of occurrence. However, if the term 'random' data set in McCowan et al. (1999) was not clear from the context, it could be clarified as an 'independently uniformly distributed' data set, as suggested.

(3) Finally, in response to Suzuki et al.'s compression example, there are special conditions where a very small Zipf slope (e.g. slope $< -0.2$) might contain 'hidden' complexity (i.e. *n*-gram structure with $n > 2$) but this occurs only for specifically coded signals such as Morse code, which is optimized for efficiency (as demonstrated by Suzuki et al. (2004), in their gzip data compression example). Highly redundant structures naturally contain *n*-gram structure, but are not optimized for information transmittal because repetition slows down the rate of information transmitted to allow error recovery.

## Conclusions

We conclude that Zipf's 'law' remains a useful tool to identify potential presence or changes in *n*-gram structure in a signalling data set if applied correctly. Noncoded natural communication systems may be expected to be structured in such a way that they approach such a distribution. That a Zipf slope of approximately $-1$ may be a necessary (although not necessarily sufficient) condition for human languages is not a large part of our work (see Cancho & Solé 2003), although such a condition highly suggests applicability to a first-look estimate of the complexity of animal communicative repertoires. We, therefore, still recommend its application to animal communication studies as a precursor to a more in-depth analysis based on Shannon entropic structure (i.e. subsequent analysis of the higher-order approximations to the entropy as a measurement of complexity in the Markov structure), as we originally stated in McCowan et al. (1999) and further discussed in McCowan et al. (2002).

We hope the dialogue between Suzuki et al.'s and our papers has helped to clarify the correct application of Zipf's statistic and we have appreciated the opportunity to respond and comment on this important topic in *Animal Behaviour*.

## References

**Cancho, R. F. I. & Solé, R. V.** 2003. Least effort and the origins of scaling in human language. *Publications of the American Academy of Sciences*, **100**, 788–791.

**Chatfield, C. & Lemon, R.** 1970. Analyzing sequences of behavioral events. *Journal of Theoretical Biology*, **29**, 427–445.

**Hailman, J. P. & Ficken, M. S.** 1986. Combinatorial animal communication with computable syntax: 'chick-a-dee' calling qualifies as 'language' by structural linguistics. *Animal Behaviour*, **34**, 1899–1901.

**Law, A. M. & Kelton, W. D.** 2000. *Simulation Modeling and Analysis*. Boston: McGraw-Hill.

**Li, W.** 1992. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, **38**, 1842–1845.

**McCowan, B.** 1995. A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphindae *Tursiops truncatus*). *Ethology*, **100**, 177–193.

**McCowan, B. & Reiss, D.** 1995. Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphindae

*Tursiops truncatus*): a re-evaluation of the signature whistle hypothesis. *Ethology*, **100**, 193–209.

**McCowan, B., Hanser, S. F. & Doyle, L. R.** 1999. Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, **57**, 409–419.

**McCowan, B., Doyle, L. R. & Hanser, S. F.** 2002. Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology*, **116**, 166–172.

**Mandelbrot, B.** 1953. An informational theory of the statistical structure of language. In: *Communication Theory* (Ed. by W. Jackson), pp. 89–101. London: Butterworths.

**Pierce, J. R.** 1980. *An Introduction to Information Theory: Symbols, Signals and Noise*. Toronto: Dover.

**Reiss, D. & McCowan, B.** 1993. Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*): evidence for vocal learning. *Journal of Comparative Psychology*, **107**, 301–312.

**Shannon, C. E. & Weaver, A.** 1949. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

**Suzuki, R., Tyack, P. L. & Buck, J. R.** 2005. The use of Zipf's law in animal communication analysis. *Animal Behaviour*, **69**, F9–F17. http://www.elsevier.com/locate/anbehav.

**Yaglom, A. M. & Yaglom, I. M.** 1983. *Probability and Information*. Dordrecht: D. Reidel.

**Zipf, G. K.** 1949. *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley.

**Zipf, G. K.** 1968. *The Psycho-Biology of Language: an Introduction to Dynamic Psychology*. Cambridge: Addison-Wesley.