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

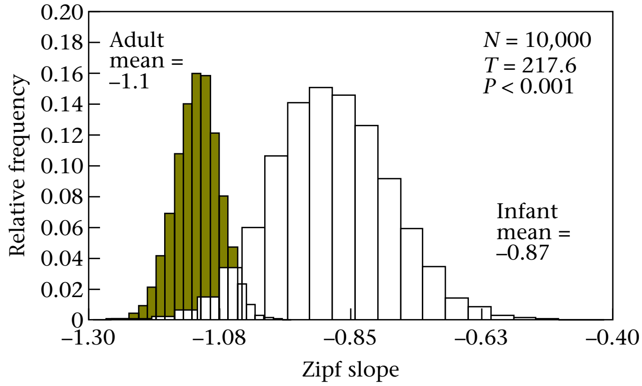

Figure 1. Distribution of values generated from a Monte Carlo simulation on adult and infant (<1 month) Zipf slopes in bottlenose dolphins. We conducted 10 000 iterations of Zipf slope calculations on the probability structure of the frequency of use of whistle types for adult and infant dolphins, respectively, 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 α 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