Acta Astronautica 68 (2011) 406–417

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Information theory, animal communication, and the search for extraterrestrial intelligence

Laurance R. Doyle[a,*], Brenda McCowan[b], Simon Johnston[c], Sean F. Hanser[d]

[a] *SETI Institute, Carl Sagan Center for the Study of Life in the Universe, 515 N. Whisman Road, Mountain View, CA 94043, USA*
[b] *California National Primate Research Center, One Shields Avenue, University of California, Davis, CA 95616, USA*
[c] *Australian Telescope National Facility, CSIRO, P.O. Box 76, NSW 1710, Australia*
[d] *Department of Evolution and Ecology, One Shields Avenue, University of California, Davis, CA 95616, USA*

ARTICLE INFO

Article history:
Received 7 April 2009
Accepted 23 November 2009
Available online 23 December 2009

Keywords:
SETI
Information theory
Animal communication
Zipf's Law

ABSTRACT

We present ongoing research in the application of information theory to animal communication systems with the goal of developing additional detectors and estimators for possible extraterrestrial intelligent signals. Regardless of the species, for intelligence (i.e., complex knowledge) to be transmitted certain rules of information theory must still be obeyed. We demonstrate some preliminary results of applying information theory to socially complex marine mammal species (bottlenose dolphins and humpback whales) as well as arboreal squirrel monkeys, because they almost exclusively rely on vocal signals for their communications, producing signals which can be readily characterized by signal analysis. Metrics such as Zipf's Law and higher-order information-entropic structure are emerging as indicators of the communicative complexity characteristic of an "intelligent message" content within these animals' signals, perhaps not surprising given these species' social complexity. In addition to human languages, for comparison we also apply these metrics to pulsar signals—perhaps (arguably) the most "organized" of stellar systems—as an example of astrophysical systems that would have to be distinguished from an extraterrestrial intelligence message by such information theoretic filters. We also look at a message transmitted from Earth (Arecibo Observatory) that contains a lot of meaning but little information in the mathematical sense we define it here. We conclude that the study of non-human communication systems on our own planet can make a valuable contribution to the detection of extraterrestrial intelligence by providing quantitative general measures of communicative complexity. Studying the complex communication systems of other intelligent species on our own planet may also be one of the best ways to deprovincialize our thinking about extraterrestrial communication systems in general.

1. Introduction

Little more than 400 years ago the human eye was first supplanted by the telescope for astronomical observation. Only over the past century did those telescopes grow to such enormous sizes as to be able to see billions of light years into space and back into pre-Earth formation times. It has been only over the past half a decade or so that long wavelength radio and spaceborne shorter-wavelength UV-to-gamma-ray telescopes have allowed the extension of detection to non-human ranges of the electromagnetic spectrum. We have also only recently ventured to the other planets in our Solar System using robotic reconnaissance spacecraft, while only over the past decade and a half have planets around other stars been detected. We are now on the verge of the detection of the first

---

[*] Corresponding author. Tel.: +1 650 960 4533; fax: +1 650 961 7099.
E-mail addresses: ldoyle@seti.org (L.R. Doyle), bjmccowan@ucdavis.edu (B. McCowan), Simon.Johnston@atnf.csiro.au (S. Johnston), sfhanser@ucdavis.edu (S.F. Hanser).

0094-5765/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.actaastro.2009.11.018

Earth-sized planets within the circumstellar habitable zones of other stars and, as these are found, the detection of exobiology elsewhere in our galaxy will be pursued over the following decades.

Photosynthetic plants play a huge role in the regulation of our own atmosphere, providing a basis for other macro-biological systems. Plants produce oxygen, which is so reactive that its presence in a planetary atmosphere is almost certainly indicative of active photosynthetic systems—the ozone absorption feature at 9.6 μs being the most obvious remote biomarker of such an ecosystem. But with all such detections it is a good idea to have backup estimators for these determinations. For example, ozone can saturate a planetary atmosphere abiotically if that planet is undergoing a runaway greenhouse effect, as well. A historic example of a biomarker false alarm was the detection of seasonal albedo variations on Mars. It was entirely reasonable to suspect that this was due to seasonal plant growth. However, it turned out to be due to seasonal dust storms on the red planet. So it may be with the search for extraterrestrial intelligence (SETI). There should be as many criteria as possible applied to any SETI signals detected to ascertain if they are sidereal, of course, and technological (if this is determinable) but also if the content of the signals themselves genuinely constitutes an "intelligent" communication system. Information theory can help us with the quantification of this latter determination.

2. Information theory measures

As distance between stars is so vast—even at the enormous speed of light—two-way electromagnetic communication between the vast majority of star systems is not practicable for individuals of most terrestrial species (and this is certainly true for the only species presently using radio telescopes on our planet). Thus we shall not be formulating information theory measures in terms of a two-way communication system, but rather quantifying the internal complexity *within* a given communication system itself. Rather than comparing, for example, transmitted with received messages (more like a cross-correlation), we shall be examining the internal complexity of a given signal-system within itself (more like an auto-correlation approach). For this we shall use the concept of different entropic orders, which measure different levels of conditional information within a message or communication system. In human language systems these measures quantify the degree of syntax, grammar, and any other structural rules that govern the use of that language. Such rules are essential for the transmission of what might be called "knowledge". For now we shall designate this rule structure the "communication complexity" of the signaling system (e.g., [1–3,32,33]). We note that such communicative structure or constraining "rules" on a given signaling system are essential for error recovery by the receiver, as well, and so—in species dependent on signals for functioning and propagation—such complex communicative constraints (rule structure) as the species can accommodate could have important survival value.

The equations for the formulation of information entropies are as follows (see, e.g., [1,3–6,33,34]). The zero-order entropy is:

$$H_0 = \log_2 N \tag{1}$$

where N is the number of different signal types. H_0 is the maximum number of bits (or highest degree of freedom) of the communication system. The first-order entropy (sometimes called the marginal entropy) is

$$H_1 = -\sum_{i}^{N} p(i)\log_2 p(i) \tag{2}$$

where, $p(i)$, is the probability of occurrence of a given signal type, i, from a message assumed sufficiently large to constitute a well-sampled data set. We can see that putting in a uniform distribution for Eq. (2), $p(i)=1/N$, will give Eq. (1). The second-order entropy introduces conditional probabilities into consideration and is given by

$$H_2 = -\sum_{i,j}^{N} p(i,j)\log_2 p_i(j) \tag{3}$$

where $p(i,j)$ is the joint probability of events i and j, and $p_i(j)$ is the conditional probability that event j will occur given that event i has already occurred. The different orders of entropy can thus be generally formulated for higher-order conditional probabilities as

$$H_n = -\sum_{i,j,k,\ldots n}^{N} p(i,j,k,\ldots,n)\log_2 p_{i,j,k,\ldots n-1}p(j,k,l,\ldots,n) \tag{4}$$

where n is the entropic order and events $i,j,k,\ldots n$ are again assumed to be sufficiently well approximated (well sampled) from frequencies of occurrence counts in the communication system data set. In this way probabilities may be derived from sampling n-gram (i.e., n-length) chains of signals to ascertain the connectedness (conditional probabilistic relationships) of signals to each other, which may be indicative, as discussed below, of the communicative structural capabilities of various species.

As an example, let us take the English alphabet where there are 27 characters possible (including a space character to indicate word length) giving a maximum information entropy of about H_0=4.75 bits according to Eq. (1). Taking into account the typical frequency of occurrence of English letters one would obtain (Eq. (2)) H_1=4.03 bits. Here the information entropy is less because more constraints are now being placed on the frequency distribution (probability of occurrence) of the letters. (One may consider that unmeasured entropy is "uncertainty" while measured entropy is "information" and that rule structure thus decreases the uncertainty within a given message. English writing, for example, is at least 75% redundant because of rule structure.) One can continue this process by essentially taking into account the di-gram (two-letter) frequencies of occurrence to see how much a given letter's occurrence depends on the previous letter. Using the frequency of all English letter di-grams as probabilities in Eq. (3) gives H_2=3.32 bits. One can find higher rule structure for English letters by inserting