WHAT ANIMAL STUDIES CAN TELL US ABOUT DETECTING INTELLIGENT MESSAGES FROM OUTSIDE EARTH.

A. Kershenbaum[1], T. Bergman[2], N. Carlson[3], F. da Cunha[4], L. Doyle[5], J. Elliott[6], R. Ferrer-i-Cancho[7], M. Gustison[2], G. Harp[5], B. McCowan[8], S. Semple[9],

[1]Department of Zoology, University of Cambridge, UK, arik.kershenbaum@gmail.com, [2]Department of Psychology, University of Michigan, USA, [3]School of Biology, University of St Andrews, UK, [4]Anthropological Institute & Museum, University of Zurich, Swizerland, [5]Carl Sagan Center, SETI Institute, USA, [6]School of Computing, Leeds Beckett University, UK, [7]Department of Computer Science, Universitat Politecnica de Catalunya (Catalonia), Spain, [8]Department of Veterinary Medicine, University of California Davis, USA, [9]Department of Life Sciences, Roehampton University, UK.

Understanding how intelligent life elsewhere in the universe communicates is constrained by the absence of any data from extraterrestrial communication systems. Indeed, the only language available for us to study is human language itself. However, humans are not the only intelligent and communicative species on the planet, and for many years, animal behaviour researchers have been studying the communication systems of different animal species in an effort to uncover the evolutionary origins of human language. We propose leveraging this experience in a multidisciplinary effort to understand the fundamental and universal principles and constraints underpinning the evolution of communication, and to search for "fingerprints" of intentional and intelligent communicative signals, whatever their origin. We identify six key questions that are important for SETI research, and that have already been partially addressed by the animal behaviour research community, providing significant new insights into the potential nature of alien communication. In addition, we propose the establishment of a uniquely comprehensive cross-species database of communicative signals, which will allow SETI researchers easy access to the only non-human messages available to us, in advance of genuine contact with extraterrestrial intelligence.

A key challenge in the search for extraterrestrial intelligence is identifying an extraterrestrial signal as being intentional communication, rather than random noise or some other natural phenomenon that appears communication-like [7]. This problem is particularly challenging, as human language is the only example available to us of truly understandable linguistic communication. Therefore, our interpretation of extraterrestrial signals is inevitably shaped by our knowledge of human language. The question posed by the SETI Institute call for white papers [3], "How does intelligent life communicate?" has hitherto been answered primarily with reference only to human intelligence, and the conclusions may or may not be relevant to communication by non-human intelligent life [11]. Furthermore, at first glance human language appears just as random and unstructured as non-intentional and naturally occurring sequences. A vital question therefore is, are there statistical or other "fingerprints" of intelligent communication that are truly universal, i.e. underpin interaction across species and planets, and can be used to distinguish communication formats from other patterns found in nature? Crucial to finding such universals is that we must be able to analyse and measure them without the benefit of the contextual information and semantic clues that we take for granted when studying other human languages [18].

The study of animal communication on Earth can provide important clues to answer these questions. Although no adequate test currently exists to determine if another species exhibits a true language, many cognitively sophisticated animal species use complex acoustic and other communication systems to convey intentional information to conspecifics [18]. By studying these communication systems, we can expand the scope of our enquiry beyond purely human language; providing us with the only way to test hypotheses about communicative universals among species other than our own. Furthermore, understanding the evolution of animal communication systems on Earth can give insight into evolutionary constraints on language complexity and structure, which may also apply in ecosystems on other planets in other solar systems [14]. Indeed, speculation to date about the nature of alien communication has, with a few notable exceptions [6], been wholly theoretical, and has not taken advantage of the wealth of research experience and knowledge among animal communication scientists, who have developed techniques and models for studying non-human communication for many decades.

In this white paper, we propose a multidisciplinary approach to answering the question, "how does intelligent life communicate?" via a consortium of biologists, zoologists, mathematicians, linguists, and SETI researchers. Using this multidisciplinary approach, we will synthesise new perspectives and provide shared resources for SETI researchers from around the world. In particular, we propose an open source database of animal communication corpora, to match the many human

language corpora currently available to computational linguistics researchers. We also propose providing libraries of analytical algorithms, drawn from the state of the art in both animal and human communication analysis, which can be applied to these corpora to further comparative studies and to encourage the development of testable hypotheses for detecting the presence of intelligent communication.

We specifically propose a series of projects intended to address the following six questions:

1. Are there general principles of communication that are evolutionarily constrained or convergent?
Certain patterns recur both in languages and in the communication of other species: Zipf's law for word frequencies [19], Zipf's law of abbreviation [8], Menzerath's law [10], that have been interpreted as indicative of universal principles of communication. Recent studies have suggested that language is constrained between complexity (for the sake of information content), and simplicity (to reduce cognitive burden) [8,9]. Mathematical modelling can investigate the evolutionary advantages of balancing these two constraints, and may point to a universal indicator of language. We will use such models to look for hard constraints that may apply to any communication system evolved on Earth or elsewhere. By considering principles of communication that for their abstractness can be valid across modalities, we can remain agnostic to the physical modality itself, thereby remaining open to the possibilities of what might appear otherwise to be unexpected communication channels.

2. What is the biological basis of communicative "meaning"?
We take it for granted that we understand the concept of meaning in communication. However, animals have evolved to communicate as part of a manipulative game-theoretical process, where information is a resource just like food or shelter [2]. Other species may perceive information very differently from the way that we do [18]. What implications does this have for understanding how alien intelligence may need to encode information, given their different evolutionary histories? Our proposal includes examining known and novel examples of animal communication and reassessing our understanding of intentionality in the context of different ecological constraints that may be relevant to extraterrestrial species.

3. How and why did communicative sequences evolve from simple signals?
Only a few animal species communicate using complex sequences that superficially resemble human language [18], whereas many more animals use a small number of simple signals [1]. Mathematical modelling and comparative studies can help us to understand what drives the evolution of complex sequences, and what adaptive and informational advantage these sequences provide. In particular, is the advantage of communicating in complex sequences a universal, or a peculiarity of human language? We suggest addressing this question through models of evolutionary dynamics, and examining phylogenetic relationships between species with different tendencies to signal complexity.

4. Are there fingerprints of intentional or semantic content across the communication systems of different species?
Key to our investigation of extraterrestrial signals is the identification of indicators that a particular signal is either intentional, or that it contains semantic (i.e. contextual) information. Some sequences (such as DNA) contain information without being intentional [21], whereas other sequences (such as most birdsong) is intentional, but with little contextual specificity [4]. Several statistical indicators have been proposed to measure the language-like nature of communicative signals, such as complexity [19], entropy [16], and grammatical metrics [22]. However, are these indicators truly universal? Do they arise from physical or mathematical constraints that apply to all intelligent species in the universe? We propose a broad-ranging research program to look at multiple different indicators and draw conclusions from their results on corpora from multiple communicative and non-communicative sources.

5. Is there such a thing as proto-grammar in non-human animals? Can we measure the degree of "languageness" in a communication system?
In recent years, a number of studies have shown evidence of complex syntax in the acoustic communication of several different species [5,15,17,23]. Yet, most researchers still believe that among living organisms on this planet, only humans possess the complex grammatical elements necessary for true language [12]. But are the syntaxes of animals "proto" grammatical, in the sense that they represent a stepping-stone on the evolutionary path to language? If so, then perhaps we can measure where a particular communication system lies on the spectrum between language and non-language. A quantitative measure of linguistic tendency in a signal would allow more discriminating classification of extraterrestrial signals, and provide a graded assessment of the likelihood of intelligent origin, rather than a yes/no determination. Our research will use an extensive database of animal and human communications, together with phylogenetic and behavioural measures, to determine whether such a quantitative approach is applicable to natural communication systems.

6. Segmentation: how do we determine the correct units of analysis of a signal without any contextual information?

Any signal received from extraterrestrial intelligence will likely be presented without context and without indication of the natural units of composition (e.g. words and letters) used in the information stream. Inappropriate segmentation of the data could lead to erroneous interpretation, or even to dismissal of the signal as being non-communicative. Similar problems face human [20] and animal communication researchers, who rarely have access to information about how other animals perceive and interpret the communicative stimuli of conspecifics [18]. Further, zoologists often have to deal with animal communication that makes use of multiple modalities (e.g. acoustic, visual, electric, and chemical), either independently or synergistically [13]. We will develop new methods, illustrated with examples from established animal behaviour research, to overcome this problem and to present potential solutions when dealing with putative signals from extraterrestrial sources.

Conclusions

Animal communication research is the closest we are likely to get to studying extraterrestrial signals, until such signals are actually received. Many of the challenges facing SETI research are similar to those already addressed in the investigation of animal behaviour, and the evolutionary origins of human language. Indeed, the evolution of language on Earth may in fact have been driven and constrained by similar principles to those operating on life on other planets. Therefore, many of the tools and data developed in different scientific fields can be used in a cross-disciplinary way.

Apart from ongoing research published in scholarly journals, we propose the establishment of a large cross-species database of communicative signals, made available to all SETI and animal behaviour researchers. In addition, tools, algorithms, and software that are used to analyse these signals will be made publicly available for application to these data sets, so that comparative studies can take full advantage of the expertise from the biological, mathematical, linguistic, and astronomical communities.

Additional Information:

(A) Our white paper is relevant to questions 2 & 3 of the *Alien Mindscapes* article: *How does intelligent life communicate?* and *How can we detect intelligent life.*

(B) Our project can make use of three types of Big Data sets: (a) human language corpora, including conversational as well as written texts, (b) animal vocal data sets, across modalities, and (c) data sets showing human activity in the electromagnetic spectrum, as transmitted away from the earth.

References

[1] K. Arnold and K. Zuberbühler, Nature **441**, 303 (2006).

[2] J. W. Bradbury and S. L. Vehrencamp, *Principles of Animal Communication.* (Sinauer, Sunderland, MA, 2011).

[3] N. A. Cabrol, Astrobiology **16**, 661 (2016).

[4] C. K. Catchpole and P. J. B. Slater, *Bird song: biological themes and variations* (Cambridge Univ Press, Cambridge, 2003).

[5] E. Clarke, U. H. Reichard, and K. Zuberbühler, PLoS ONE **1**, e73 (2006).

[6] L. R. Doyle, B. McCowan, S. Johnston, and S. F. Hanser, Acta Astronaut. **68**, 406 (2011).

[7] J. Elliott and E. Atwell, JBIS: Journal of the British Interplanetary Society **53**, 13 (2000).

[8] R. Ferrer-i-Cancho, A. Hernández-Fernández, D. Lusseau, G. Agoramoorthy, M. J. Hsu, and S. Semple, Cognitive Science **37**, 1565 (2013).

[9] R. Ferrer-i-Cancho and R. V. Solé, Proc Natl Acad Sci USA **100**, 788 (2003).

[10] M. L. Gustison, S. Semple, R. Ferrer-I-Cancho, and T. J. Bergman, Proc. Natl. Acad. Sci. U. S. A. **113**, E2750 (2016).

[11] S. F. Hanser, L. R. Doyle, B. McCowan, and J. M. Jenkins, Bioastronomy 2002: Life Among the Stars, Proceedings of IAU Symposium **213**, 514 (2004).

[12] M. D. Hauser, N. Chomsky, and W. Fitch, Science **298**, 1569 (2002).

[13] E. A. Hebets and D. R. Papaj, Behav. Ecol. Sociobiol. **57**, 197 (2005).

[14] R. Jackendoff, Language **87**, 586 (2011).

[15] D. Z. Jin and A. A. Kozhevnikov, PLoS Comput Biol **7**, e1001108. (2011).

[16] A. Kershenbaum, Bioacoustics **23**, 195 (2013).

[17] A. Kershenbaum, A. Ilany, L. Blaustein, and E. Geffen, Proc R Soc Lond B Biol Sci **279**, 2974 (2012).

[18] A. Kershenbaum, D. T. Blumstein, M. A. Roch, *et al*, Biol. Rev. **91**, 13 (2014).

[19] B. McCowan, S. F. Hanser, and L. R. Doyle, Anim. Behav. **57**, 409 (1999).

[20] J. R. Saffran, E. L. Newport, and R. N. Aslin, Journal of memory and language **35**, 606 (1996).

[21] A. O. Schmitt and H. Herzel, J. Theor. Biol. **188**, 369 (1997).

[22] C. ten Cate and K. Okanoya, Philos Trans R Soc Lond B Biol Sci **367**, 1984 (2012).

[23] M. Weiss, H. Hultsch, I. Adam, C. Scharff, and S. Kipper, Proc. Biol. Sci. **281**, 20140460 (2014).