An Atlas of Communication Evolution Based on a Unified Database of Evidence. Anamaria Berea[1], [1]Center for Complexity in Business, University of Maryland, College Park, MD, 20742 (aberea@rhsmith.umd.edu)

Scientific Rationale:

Life laws follow physics laws, but they have a higher degree of granularity and variation (more "local" laws that fall into the "general" laws – i.e., cellular laws don't violate atomic laws), and civilization laws follow life laws, also with an even higher degree of granularity and variation (even more "local" laws – i.e., macroeconomic convergence does not violate evolution). Communication is life-dependent and intelligent communication is civilization-dependent, which means that our best way to find general patterns of communication in the universe is to look for general patterns of communication in life and civilization on Earth and to understand the limits of variations in the laws of communication within life and from those within civilizations. Essentially, this is a data science problem of cleaning, cataloguing, mapping and analyzing the evidence of all patterns (not observations, which most probably would be an impractical exercise) of communication that we know of today, but with an emphasis on "local" versus "global" transitions in communication patterns.

Communication is as ubiquitous as information, yet while there is an integrated scientific discipline of information science, with theories and applications, there is no integrated scientific discipline of communication science, that would study not only "local" patterns and observations of human communication, but the broader phenomenon of communication in the living world. In this paper I am outlining the rationale for curating and mapping our current but rich collection of communication evidence, scattered throughout multiple disciplines and sciences, into an *Atlas of Communication Evolution* from cells to societies, with a focus on local/global patterns, on means/meaning of communication and on natural/artifactual communication.

Concepts of Experimental Design – Methods and Protocols:

A comprehensive and broad "Atlas of Communication Evolution". Communication is being researched by the vast majority of disciplines, from biology (species specific communication) to linguistics (human languages), to communication theory (context specific human communication – i.e. parent-children, political, social communication) to business and economics (by organization specific communication). This domain specific research on communication has given us very rich details about *specific*, *local* communication patterns and the role of communication in other phenomena (i.e. predation, mating, immunotherapy, organizational functions, social movements, a.s.o.). This means that we now we have many pieces of a big puzzle to help us create a comprehensive map of communication on Earth, with taxonomy, ontology, phylogeny and evolution. Having a broad idea about how communication has emerged and evolved in the living systems on our planet, from cells to firms and societies, would enable us to highlight: 1.) the common patterns across species, ecosystems and civilization trends, and 2.) the differences/variations and the boundaries/limits where we know that communication *does not* and cannot occur. Given the large datasets and big data techniques available now from various research projects, harvesting and creating a database of such evidence can be achieved, once the protocols for storing and realigning the data and meta-data are set and an Atlas of Communication Evolution would help create the protocols for aligning and re-encoding the current evidence in a unified database.

a. Locality, globalism and transition phases in communication patterns. Such an Atlas would identify local patterns of communication of the living systems, from cells to cells interaction, to organism specific interaction, to human languages and human-computer interactions, paying attention to transition phases or boundaries between this phenomenon's "localities", within the same paradigm or framework of thinking. This paradigm is outlined by viewing communication through the following lenses: natural/artifactual; meaning/means; local/global.

b. Natural versus artifactual communication. While all living systems on Earth communicate through natural, genetically evolved means (chemical communication in non-vertebrate species, ultrasound in bats, olfactory, auditory and visual in most vertebrate species), only humans have developed artifactual means of communication – writing, drawing, and computational technologies. Moreover, humans have developed complex languages that combine natural (sounds) and artifactual (grammar) means of communication. The hallmark of an advanced civilization combines the natural and the artifactual means and meaning of communication in a way that would reflect *selection and adaptation* processes both in natural communication and in artifactual one.

c. The "wave-particle" theory of communication: communication as both *means* and *meaning* of information exchange. Shannon explored only the "wave" theory of communication – communication as a means of exchanging information, devoid of meaning. On another hand, semiotics and cryptography explore the "particle" theory of communication – the meaning

of communication devoid of the means of communicating. Therefore mapping communication from cells to societies in distinct terms of means and meaning of communication would help understand where the two aspects of the phenomenon of communication converge and where they diverge.

Universal Markers, Signals and Systems:

a. Are there communication "pathologies" and what can we learn from them? In general, in science we aim to explain and understand the what, when, how, who, when and why of a phenomenon. We are less likely inclined to explain the what *does not*, when it *does not*, how it *does not*, who *does not*, when *not* and why *not* of the phenomenon – in other words to put the *non-occurrence* of the phenomenon in the same epistemological terms of study as the occurrence of it. The payoffs are higher and the search space is usually smaller for explaining a phenomenon than for explaining the *not*s of same phenomenon. For communication though, the search space might be similar in size to the search space for *non*-communication – explaining when there is no interspecies communication, why we cannot find meaning in certain messages, why we don't have a unified language, why computers cannot understand metaphors, a.s.o. The rationale for such an approach in an Atlas of Communication Evolution is that miscommunication is known to trigger cellular pathologies in molecular biology, but also useful mutations, to cause wars in the human civilization, but also to have languages converge and evolve. By the nature of information (information negation creates new information), communication pathologies lead not to the death of communication, but to the evolution and transformation of communication. For example, the Atlas of Communication Evolution would help identify the boundaries in interspecies communication, the life cycles of communication and the transition phases, and from these unrelated "local" communication failures we can potentially draw global patterns of communication pathologies.

b. A methodological reversal: social and collective behavior laws as proxies for intelligent communication laws. Can we extrapolate from the empirical laws and statistical regularities of the social and collective phenomena into communication phenomena? We know that Zipf law applies both to languages and to cities or companies sizes; we know that marginal analysis in economics (cost/benefit) also applies to metabolism and chemical communication. Can we test for other laws from collective and social behavior of animals, humans and computers to see how they would "behave" in the case of communication? For example, the "edge of chaos" model in complex systems behavior shows us where noise becomes signal or signal becomes noise, and that the "edge of chaos" is where both the means and meaning of communication happen. The "edge of chaos" model is both global and local for different complex systems phenomena, therefore an "edge of chaos" model of communication can potentially give us insights into the transition from local to global communication patterns.

c. A bridge between semiotics and data science: transition from locality to globalism. In data science and machine learning, natural language processing techniques are notoriously hard at identifying meaning in clusters of words and semantic distances. But they have the power of search for signals through large sets of noises, as long as we identify these signals a priori. On another hand, computational semiotics techniques are better at identifying signals in small data and reconfiguring algorithms based on semantic identifiers. A computational tool that would combine natural language processing with computational semiotics would be able to search through large datasets of noise and adapt the signal search based on dynamic semantic identifiers, thus enhancing the probability of identifying signals globally, not only locally.

Target Identification:

Patterns of "trivial"[1] *communication as a signal of successful communication evolution, that started local and became global.* Intelligent communication is a coevolutionary process where innovation in artifactual communication (writing, telegraph, Internet,…) and the type of collective behavior that forms a civilization intertwine and produce patterns such as transportation systems signs, brand logos (symbols of global organizations that transcend specific human languages), universal computer languages, traffic signals, universal data storage protocols, universal data encoding and analyses. The Atlas would help identify which are the truly large and "trivial" systems of communication.

The current "age of data" in our civilization is a result of two major forces: the global spread of artifactual means of communication (the "triviality" of human-computer interactions and the globalization of the English language) and the global spread of physical systems of recording and storing communication (the long evolution from Mesopotamian cuneiform tokens to current organizational servers and supercomputers). But data only as *means* is global, its' *meaning* is local. Examples of systems that use truly global *means* and *meanings* of communication are transportation infrastructures, such as the airspace, maritime, financial systems or Internet protocols. These transportation systems for communication are products of selection and adaption mechanisms emerged from the economic systems and requirements of our civilization. At the

[1] "trivial" means ubiquitous and of common and widespread use.

large scale of the civilization, these planetary systems of interconnectivity show artifactual, global patterns of signals/communication of the largest magnitude known so far. As writing began in Mesopotamia[2] in order to record bookkeeping for crops and animals, the evolution of our civilization's economy and the emergence of artifactual communication have been intertwined ever since.

Observing Scenarios and Exploration Strategies:

The rationale for collecting and cataloguing the current evidence of communication from cells to societies into an "Atlas" would be referential for understanding what can be classified as "common communication" in the living universe, what possible scenarios of communication can be developed with small variations in the living forms, and most importantly, to prune out the scenarios or possibilities of communication that can never lead to intelligent, decipherable communication.

a. Scenarios of relaxing/variation in assumptions from the bottom up. Given that we have a good map about how our civilization actually communicates globally, we can explore potential scenarios where we relax one or more assumptions from this larger picture, such as – what if cells on a different world do not communicate chemically, would that imply different forms of organization between higher organized intelligent species? What if organisms do not develop sight or visual communication cannot be recorded, how would a civilization without visual records communicate? Is intelligent communication always dependent on social or collective forms of behavior or can there be entities that do not connect with peers, that do not have local, bounded communication, but have only a global framework of communication?

b. Scenarios of nullifying assumptions from the top down. Thinking in null hypothesis terms, does a civilization need an economic system in order to achieve global systems of trivial, artifactual communication? Does a civilization need an economic system in order to emerge general, not culturally localized communication? Humans have developed a diversity of forms of organizing their social or economic or political behavior that cannot be reproduced or recognized in any other species. But the civilization as we know it comprises culture and social behavior that also gave us not only global communication but also ultra-specialized (large variation) communication in technology, art, or economics. We cannot find both global and high variance local communication in other living systems. Are these statistical patterns crucial for the development of a civilization and its communication?

c. Scenarios of variation in the natural-artifactual coevolution. The interplay between the natural and the artifactual worlds through our economic systems have determined, for better or for worse, the current state of our civilization. Therefore, we can design scenarios where we can explore: What types of economic systems with co-evolving artifactual versus natural forms of organized behavior are the ones that can lead to many types of sustainable civilizations? What types of civilizations that are based on the natural-artifactual duality can develop long term, sustainable communication? What is the interplay between communication of a species as a collective (whole), and the many types of organizing behavior (natural-artifactual communication topologies)?

These are only a few of the multitude of scenarios and possibilities that can be explored by formulating and exploring hypotheses, once a baseline or referential guide and a unified dataset of evidence are in place. But in order to reduce the search space of exploration with many "if"s into "probable" and "possible", a unified referential map would help guide these efforts.

Additional Information:

(A) This paper is answering Question no.2 from Alien Mindscapes - A Perspective on the Search for Extraterrestrial Intelligence, *How does intelligent life communicate?*

(B) Throughout this paper I have tried to show how given the current existing data repositories and new data science techniques can help with the creation of the Atlas for Communication Evolution. Specifically, we can now much faster align data from different repositories and different fields; we can use the universal data science protocols as proxies for organizing and creating the taxonomy of communication and we can use the latest machine learning techniques (particularly those from natural language processing and image processing) to explore the boundaries in communication and the evolution of communication between humans and machines at local and global scales.

References

1. Simon, Herbert A. The sciences of the artificial. MIT press, 1996.
2. Shannon, Claude E. "Communication theory of secrecy systems." *Bell Labs Technical Journal* 28.4 (1949): 656-715.
3. Kauffman, Stuart A., and Sonke Johnsen. "Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches." *Journal of theoretical biology*149.4 (1991): 467-505.

[2]The origins of wirting are currently being disputed by the recent trnaslations of the Tartaria tablets in Estern Europe.

4. Cabrol, Nathalie A. "Alien Mindscapes—A Perspective on the Search for Extraterrestrial Intelligence." (2016): 661-676