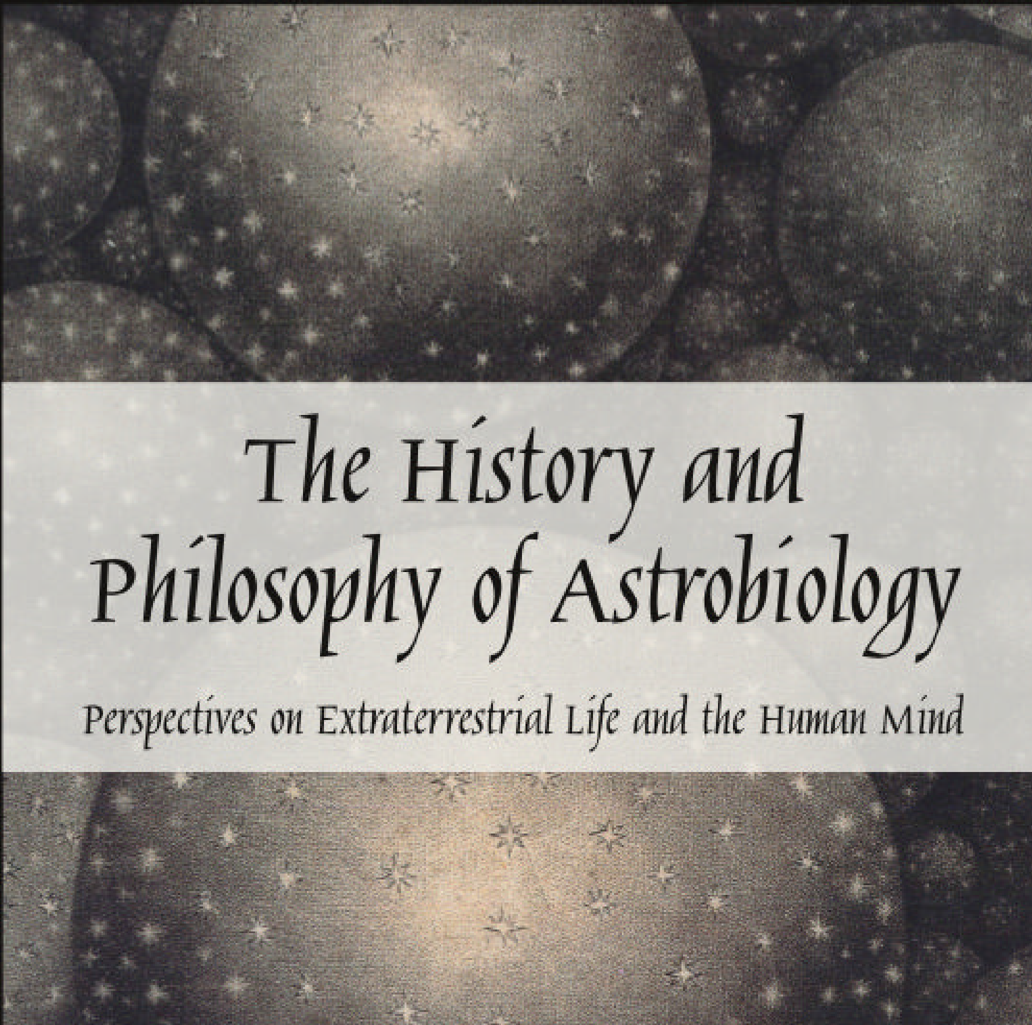


The History and Philosophy of Astrobiology

Perspectives on Extraterrestrial Life and the Human Mind

Edited by

David Dunér
with Joel Parthemore, Erik Persson
and Gustav Holmberg

The History and Philosophy of Astrobiology

The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind

Edited by

David Dunér
with Joel Parthemore, Erik Persson
and Gustav Holmberg

CAMBRIDGE
SCHOLARS
PUBLISHING

The History and Philosophy of Astrobiology:
Perspectives on Extraterrestrial Life and the Human Mind,
Edited by David Dunér, Joel Parthemore, Erik Persson, Gustav Holmberg

This book first published 2013

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-5035-7, ISBN (13): 978-1-4438-5035-3

TABLE OF CONTENTS

INTRODUCTION

EXTRATERRESTRIAL LIFE AND THE HUMAN MIND

DAVID DUNÉR

> … cette curiosité mutuelle avec laquelle les Planetes s'entre-considerent & demandent l'une de l'autre, *Quel Monde est-ce là? Quelles Gens l'habitent?*
> —Bernard de Fontenelle,
> *Entretiens sur la pluralité des mondes* (1686)

Conversations on the plurality of worlds

Perhaps, wonders the philosopher in Bernard de Fontenelle's *Entretiens sur la pluralité des mondes* (1686), there are astronomers on Jupiter; and perhaps we cause them to engage in scientific quarrels, so that some Jovian philosophers must defend themselves when they put forward the ludicrous opinion that we exist. Their telescopes are directed towards us, as ours are towards them, "that mutual curiosity, with which the inhabitants of these Planets consider each other, and demand the one of the other, *What world is that? What people inhabit it?*" (Fontenelle, 1701, p. 93; 1767, p. 207)

Human beings have wondered about the stars above since the dawn of the species. Does life exist out there? Are we alone? Questions of life in the Universe can be traced back to antiquity: to philosophers like Epicurus of Samos and authors like Lucian of Samosata. Since then, astrobiological questions have fascinated scientists and philosophers and have been discussed by religious thinkers and utopian authors. These questions have progressed from things of which we could only speculate upon into objects of practical study. When the cosmos was revealed during the scientific revolution of the sixteenth and seventeenth centuries, the super-lunar world – the Universe beyond the Moon – was no longer closed and unchanging but vast and evolving. When Copernicus put forward his heliocentric model in *De revolutionibus orbium coelestium* (1543), Earth was reduced to a planet like other planets. In 1609 Galileo trained his telescope on the Moon and, like the Earth, found it to be rugged and uneven, perhaps even having similar mountains and oceans (Galileo, 1610). Then scientists and philosophers wondered if these celestial bodies could also harbour life. Earth was no longer unique.

On 11 November 1572 when the Danish astronomer Tycho Brahe (1573) saw an extremely bright new star in the constellation Cassiopeia, the sidereal heavens no longer seemed unchanging and eternal, as Aristotelian cosmology had taught. His observation made it possible to think about creation and change not only with respect to the Earth, but also to the Universe. In 1575, as a gift in fief from the Danish king, Tycho received the island of Ven in the strait of Öresund. There he constructed the biggest, most advanced observatory in the world: Stjärneborg (*Stellæburgum*). With its help, he looked to the sky with his naked eye to search for distant stars and other worlds.

Tycho's measurements of the positions of the heavenly bodies (published by Kepler in 1627) were indispensable to his disciple Johannes Kepler's formulation (1609; 1619) of the planetary laws of motion. In turn, Isaac Newton (1687) brought Kepler's laws and Galileo's mechanics together into the gravitational theory of classical mechanics. This enabled the detection of the first exoplanet, 51 Pegasi b, which was announced on 6 October 1995 (Mayor and Queloz, 1995; Perryman, 2012).

In the late eighteenth century, scientists increasingly came to the conclusion that the Earth has had a long history composed of many geological eras. Contributing to the story of modern astrobiology are also the discovery of spectroscopy, Darwin's (1859) theory of evolution, the advent of genetics, the research into the molecules of life, the space programme, and, most recently, the discovery of extremophiles and exoplanets. All of these discoveries contributed to scientifically grounded arguments that the conditions for life need not be restricted to Earth.

Contemporary astrobiology is something more than the results of discoveries and theories. It is the product of societal factors. These include collaborations, institutions, and technological changes and of human inspiration and imagination. This volume captures this distinct history and diversity: the ideas and events that made it possible to think of other worlds and distant life. It traces the history of science and the development of new schools in philosophy. Its aim is to discuss the place of humanity in the Universe.

In 2010, eight scientists and scholars formed a research group at the Pufendorf Institute for Advanced Studies, Lund University, Sweden, under the direction of David Dunér. Two visiting professors joined the group in

2011. The project, entitled "Astrobiology: Past, Present, and Future", gathered astronomers, geologists, chemists, biologists, historians, philosophers, and other professionals interested in the multidisciplinary field of astrobiology. Research focus was, among other things, on the humanistic side of astrobiology, on the ethics of space exploration, on epistemological questions, and on the historical establishment and development of astrobiology as a scientific enterprise.

Over its first year, the group arranged four workshops and one international conference which addressed the emergence and evolution of terrestrial life, the possibilities for interstellar communication, the search for exoplanets, and the history and philosophy of astrobiology. A conference took place 27–28 September 2011 on Ven (now part of Sweden), a short walk from the remains of Stjärneborg (Dunér *et al.*, 2011). The "History and Philosophy of Astrobiology" conference gathered researchers from around the world to share their results and insights.

This volume is a product of that conference and the previous workshops (see also Dunér *et al.*, 2012). It seeks to establish *the history and philosophy of astrobiology* as a research field in its own right, such that research in the humanities is a necessary contribution to astrobiology and complements ongoing work from biology, chemistry, physics, and astronomy. Cognitive, linguistic, epistemological, ethical, cultural, societal, and historical perspectives on the development of astrobiology are necessary to understand what is meant by "life", "intelligence", "communication", and other phenomena in a universal perspective. It is also necessary to explore what ethical, epistemological, and societal problems are involved in space exploration.

The history and philosophy of astrobiology

What does it mean to speak of the history and philosophy of astrobiology? What lines of research are there and what has been accomplished? (See e.g. Dunér, 2012; Dick, 2012) And why history? Why philosophy? It is true that studies on the history and philosophy of astrobiology do not provide new empirical data on extraterrestrial worlds. Yet, what can be learned is the nature of what it is to be human. The search for life in the Universe touches on fundamental hopes and fears, on the essence of what it means to formulate a theory, grasp a concept, and have an imagination. This book aims to clarify why history and philosophy are important for the self-understanding of astrobiology. It will discuss how it has developed and what deeper fundamental problems it faces. History and philosophy of astrobiology is important for the self-understanding of the humankind itself. What does it mean to be a human being in the true universal sense?

In short, the history of astrobiology is concerned with the evolution of human conceptions of the plurality of worlds (for the history of the debate on extraterrestrial life, see Dick, 1982; 1996; Guthke, 1983; Crowe, 1986; 2008). In order to explain what the history of astrobiology is about, I will give a few examples of historical questions within this field.

Science: Recounting how and why astrobiology emerged and how it has become a respectable, scientific, empirical field of research.

Exploration: This involves the Earth, the Solar System, and extrasolar planets, along with the instruments and technologies that have made this possible. These include the optical telescope, radio telescope, microscope, and spectroscope, as well as manned and unmanned space probes. To a large extent, the history of astrobiology can be considered a history of technological change.

Theories: These include well-renown theories (and hypotheses and models) such as the heliocentric worldview, the theory of gravity, the theory of evolution, and the various theories of genetics. However, controversial and even refuted theories should also be studied. For example, Panspermia, known as the theory that life has spread through its seeds drifting in space, is still an optional explanation of life that has not been completely ruled out (Tirard, 2013, *this volume*; Brandstetter, 2012; Demets, 2012).

Scientific organisation: The organisation of institutions, laboratories, research groups, journals, space programmes, conferences, and international collaborations. The increasing levels of organisation within the last few decades reveal an ongoing institutionalisation of the field, in particular, the increasing number of scientific journals, workshops, university courses, and PhD programmes.

Science and society: Astrobiology does not exist in a vacuum but incorporates such factors which are sometimes regarded as extrascientific, such as politics, economics, religion, and public discourse (Billings, 2012; Race *et al.*, 2012). Since ancient times, debates over the place of life in the Universe have been strongly influenced by religious concerns and they have been equally influenced by the ways in which science is popularised.

Imagination: How human beings perceive the unknown and what they expect of extraterrestrials and distant worlds. Of course, human imagination says little about what actually is “out there”. However, it says a lot about people’s contemporary lives, culture, and world. It is about what was possible to think and the boundaries of their imaginations. Imagination seems to recombine projections of past experience in novel ways. What is entirely outside experience is likewise outside imagination. Studies into imaginary voyages might not be informative for the history and development of astrobiology as a science, but can be considered historically valuable documents in order to understand the dreams and imagination of a specific time period, as a key to the understanding the era they lived in, its conceptions, ideas, and endeavours (see e.g. Bjørnvig, 2012).

Astrobiology raises many other questions for historical research. In this volume we touch just a few questions that we consider most relevant, after all, it covers only one part of a much vaster territory. Much more remains to be said about the history and philosophy of astrobiology, such as how people in different times and places construct their conceptual frameworks, which is to say how they came to a structured understanding of the world around them and how they have used their cognitive capacities to locate the human place in the world and the Universe: a Universe in which the Earth is no longer taken to be the physical centre, but humanity remains very much – and for good reasons – the conceptual centre. This leads naturally into a discussion of the other main theme of this volume. Philosophy of astrobiology is, among other things, about:

Self-understanding: The philosophy of astrobiology is an ongoing, existential exercise in individual and collective self-understanding. That is, what it means to be human, where humankind's place in the Universe is, and how both of these conceptions have inevitably evolved. Since the time of the ancient Greeks, with their philosophers' motto *gnōthi seauton* ("know yourself"), philosophy, in particular the philosophy of the mind, has sought to capture what it means to be a mind, to have a mind, to live as an intentional agent in a seemingly non-intentional world. As Joel Parthemore writes (2013, *this volume*), space is "the ultimate mirror we hold up to ourselves." In confronting the depths of the cosmos, we confront the unseen depths in ourselves. In other words, astrobiology challenges our everyday conception of ourselves as human beings in the Universe.

Conceptual analysis: How to define key terms and frame discussion. Constructing concepts in order to be able to think and talk about the new phenomena encountered is a major task for astrobiological research. The most debated and discussed philosophical question in astrobiology is the concept of life. If we are searching for non-terrestrial life, we ought to know what it is we are looking for and what characteristics it might have. If life is a recipe, what are the essential ingredients and which are optional? Should these criteria pertain to metabolism, entropy, genes, reproduction, or something else? So far, the debate has intuitively employed an Aristotelian conception of definition (Aristotle, *Posterior Analytics*, 2.3.90b30–31), in which a "definition" is a limited list of characteristics that are both necessary and sufficient for something to be of the type of object it is, and from which all the characteristics of the object originate. In our daily lives, however, we make relatively little use of Aristotelian-type definitions and depend much more on prototypes (Rosch, 1975; 1978). Dogs, cats, and horses may seem to be more typical representatives for "life" than arsenic microbes. The debate on the definition of life could benefit from the insights of contemporary philosophy and cognitive science about human categorisation. However, astrobiology deals with categorisation not only with regard to life, but to such concepts as *habitable zones* (see e.g. Kane and Gelino, 2012), *Earth analogues*, *exoplanets*, *gas giants*, and *dwarf planets*. Future discoveries in astrobiology will most likely challenge our categorisations and definitions, that is to say, our preconception of what the world is and not is. So, we should be prepared to re-categorise and redefine our concepts. Future exobiological systematics and taxonomy will face problems concerning categorisation, identification, and description. The taxonomy of future extraterrestrial fauna and flora will be a product of the human mind.

Ethics: The philosophy of astrobiology is a coming-to-terms with basic ethical issues and human values. These are, among others, how we should behave if we find extraterrestrial life, whether we have moral obligations to such life and, if so, which ones (see e.g. Persson, 2012). There is also the question as to whether and under what conditions terraforming is permissible (see e.g. Haqq-Misra, 2012), to what extent inhabited and uninhabited planets should be preserved in their pristine state or whether they could be mined for their resources. If this is the case, then who has the rights to those resources? Do we have an obligation to spread life or to avoid contaminating other worlds with life, including the microbes that we have accidentally or intentionally transported there? Related to the ethical questions of astrobiology are those that include political ideologies and considerations, economical concerns, and distributional justice. Astrobiological exploration is expensive and involves political decisions. Who owns the Moon and future scientific discoveries of astrobiology? Why spend money on astrobiology and the search for unknown life on distant planets rather than using the money on the only life that we definitely know exists?

Epistemology: A rigorous consideration of what is known, what is knowable in practice or in principle, and what is knowably unknown. Epistemology of astrobiology is a less explored philosophical territory concerning the limits of astrobiological knowledge. How long should we search without positive results before we give up? What is possible and not possible? The epistemological problems of astrobiology are somewhat similar to those of other branches of science, but with the exception that the limits of our astrobiological knowledge seem to be much more uncertain.

Semiotics and language: The construction and decoding of interstellar messages raise a number of semiotic and linguistic questions. How can we recognise and decipher incoming messages? This is not just about constructing a vehicle for information transfer. Its concerns are on what is needed for effective communication and the symbolisation of concepts, and the relationship between syntax, semantics, and pragmatics. Interstellar communication (see e.g. Sagan, 1973; Vakoch, 2011) is in fact less of a scientific-technological problem than a communicative-semiotic one. How extraterrestrials might communicate depends on biological and cultural factors, i.e. how their bodies are constructed and how they interact with their environment and how they have evolved through the bio-cultural co-evolution (Dunér, 2011b). The semiotic and cognitive problems of interstellar communication will be further discussed below.

Cognition: What is intelligence and cognition and are there universal laws for the evolution of intelligence? How would the human mind, which is entirely a product of a terrestrial environment, function in extraterrestrial environments? Astrocognition deals with questions concerning the cognitive challenges of the human mind when confronted with the unknown (Dunér, 2011a), and the question of how cognitive abilities emerge through the evolutionary processes in different habitable environments. I will soon return to these questions and go deeper into cognition from an astrobiological point of view.

This volume

The historical and philosophical topics previously discussed are just some of those that to a minor extent have been studied or need to be scrutinised further in the future. Many more historical and philosophical issues are waiting to be explored. In this volume we have gathered a wide range of studies on the thought-provoking, imaginative, and critical questions of astrobiology, and the search for life and intelligence in Universe. The book is divided into three sections: Cognition, Communication, and Culture.

The first section (*Cognition*) focuses on the human mind and what it contributes to the search for extraterrestrial life. It explores the emergence and evolution of life and cognition and the challenges humans face as they reach to the stars. Erik Persson (Chapter 1) introduces some additional philosophical questions for consideration, the ethical dilemmas involved, and asks whether astrobiological research is justifiable at all. Mathias Osvath (Chapter 2) takes a cognitive zoology approach in formulating an astrocognitive theory of the universal principles of intelligence. Joel Parthemore (Chapter 3) considers the limits of human conceptual abilities and posits astrobiology as an entry into what he sees as truly the *final frontier*: “the unmapped territory of the human mind”. Astrobiology research threatens to challenge our cherished conceptual frameworks and provoke a radical re-conceptualisation of what it means to be human. In a similar vein, Per Lind’s (Chapter 4) interest lies with the human mind’s encounter with the unknown (see Dunér, 2011a) and consequent experience of what he describes as *cognitive derailment* and perplexity in front of the unspecified that challenges trusted frames of reference and interpretation. His framework is Pyrrhonean scepticism which establishes irresolvable cognitive conflicts as a means of overcoming dogmatism. Finally, Jean Schneider (Chapter 5) raises classical issues concerning the definition of extraterrestrial life and intelligence.

The second section (*Communication*) examines the linguistic and semiotic requirements for interstellar communication. It considers what is needed for successful communication and if there are universal rules of communication. Michael Arbib (Chapter 6) asks: if space aliens resembled octopuses, how would they communicate? Arthur Holmer (Chapter 7) uses comparative linguistics to discuss the possible features and restrictions of exolanguages. From semiotics, Göran Sonesson (Chapter 8) discusses the problems involved in recognising a message as a message. What does it mean to transfer meaning or decipher symbols? Finally, Maria G. Firneis and Johannes J. Leitner (Chapter 9) reappraise what is perhaps the most famous attempt to construct a logical interstellar language: Hans Freudenthal’s *lingua cosmica.*

The third section (*Culture*) considers the cultural and societal issues involved in astrobiological research. It inquires into astrobiology’s organisation as a scientific discipline, its responsibilities in relation to the public sphere, and its theological implications. Stéphane Tirard (Chapter 10) discusses the panspermia hypothesis, defended by French botanists and plant physiologists in the second part of the nineteenth and the beginning of the twentieth century. Gustav Holmberg (Chapter 11) studies the popularisation of exobiology, and the boundary work going on in science, with particular consideration to the Swedish astronomer Knut Lundmark. Urszula Czyżewska (Chapter 12) takes a sociological perspective on how an increasing number of scientific journals in the field reveal the ongoing institutionalisation of astrobiology. Christopher C. Knight (Chapter 13) and Ludwik Kostro (Chapter 14) discuss the interrelations, incompatible or harmonious, between theology and astrobiology. Finally, Jacques Arnould (Chapter 15) summarises the challenges facing astrobiology, in the present and future, and asks whether astrobiology is the next revolution.

In the remainder of this introductory chapter, I will further examine these three major themes – cognition, communication, and culture. Through addressing these questions on the history and philosophy of astrobiology, let us take the first steps in the exploration of the immense *terra incognita* of extraterrestrial life and the human mind.

Cognition

In the course of everyday events and encounters the human mind has been enabled, through an evolutionary process, to understand, interact, deal with and adapt to the environment of this particular planet and to the minds of other human beings. Thus, the human brain is well adapted to the biological, ecological, and physical characteristics of our planet, as well as the cultural, social, and cognitive characteristics of the tellurian species *Homo sapiens sapiens.* What if we extend this human perspective and turn our eyes towards the starry sky to ask ourselves: are there other thinking beings out there and, if so, what are they thinking? Then, we face the cognitive questions of astrobiology.

The multidisciplinary field of *astrocognition* – i.e. literally the "acquaintance of the stars", from the Greek *astron*, star, and Latin *cognitio*, knowing or acquaintance – was first proposed in 2009 (Dunér, 2011a) and could be generally defined as:

> the study of the origin, evolution and distribution of cognition in the Universe.

…Or, simply:

> the study of the thinking Universe.

This mirrors the 1996 NASA Strategic Plan's definition of astrobiology as "the study of the living Universe" (NASA, 1998; Chyba and Hand, 2005). While astrobiology searches for the necessary and sufficient conditions for life in the Universe, astrocognition goes further and seeks the conditions for awareness and self-awareness. The Universe is not only living, it contains not just self-reproductive entities, but it also has apparently led to self-reproductive organisms that are able to reflect on the Universe they live in, on their place in that world, and their own thoughts and existence. We humans are the only species here on Earth that is able, as we know it, to reflect upon the Universe. Through the self-conscious human being, the Universe can also be considered self-conscious and able to reflect upon itself. It seems, then, that the Universe is not only bio-friendly (Davies, 2007), but also *cogito-friendly*.

Like the astrobiology field of which it is a part, astrocognition requires a multidisciplinary approach which brings together cognitive science, philosophy of mind, animal cognition, semiotics, linguistics, anthropology, cultural studies, history of science, computer science, neuroscience, evolutionary theory, physics, astronomy, and space technology. If the general philosophical, ethical, and theological implications of extraterrestrial discovery have received only minimal discussion (see e.g. Davies, 1998; Jakosky, 2000; Bertka, 2009), then the cognitive implications have hardly been touched on.

The field of astrocognition concerns cognitive processes in space, the origin and evolution of cognitive abilities, and cognition in extraterrestrial environments (Dunér, 2011a; see also Osvath, 2013, *this volume*). In terms of astrocognition, *cognition* can be described as the ability of processing sensory inputs for action in the environment. If we are discussing the existence of extraterrestrial intelligence in space, then I maintain that we must take into account the research within cognitive science and affiliated research areas in order to find answers to these questions: What is needed for higher cognitive skills to evolve? What physical, biological, societal, cultural, and other environmental factors shape cognition? What cognitive abilities are needed for a living organism to be able to manipulate its environment or, in other words, to develop technology?

The astrocognitive paradigm states that exploration of extraterrestrial environments and contact with other forms of life and civilisations, will change our thinking, conceptual frameworks, and belief systems. This condition leads to a discussion of what we can know and cannot know about the extraterrestrial, or, put in another way, the limits of our human-based epistemology and the constraints of our evolutionary history. And further, what cognitive challenges we are likely going to face when we encounter the unknown, and the challenges our Earth-bound perceptual, cognitive, and psychological capacities face in a space context (Pálsson, 2009, pp. 79–80). In this way, as astrocognition reaches outward to the stars and to the minds we hope to find out there, it also reaches inward into the uncharted depths of the human mind. It is both about the human mind in front of the unknown and the evolution of an unknown potential mind in the unknown space. Thus, astrocognition tends to time and space, both temporal and spatial questions, i.e. tries to give answers to i) the origin and evolution of cognition, and ii) the distribution of cognition in the Universe.

Astrocognition concerns time and evolution. It starts from a fundamental, basic premise, *the evolutionary astrocognitive premise*:

> Cognition in the Universe develops through evolutionary processes of adaption to a specific, but changing environment, and the challenges it presents.

Thinking has an evolutionary origin (see e.g. Gärdenfors, 2006), and as such, cognition largely evolves as an adaptation to certain problems that the ancestors of a particular organism had faced during the evolution of the species. That is, the cognitive processor of the organism is adapted to, firstly, the physical and biological environment of their celestial body in order to understand and interpret, interact and deal with, and orientate itself in the particular physical and biological environment, in relation to its specific conditions, such as planetary orbit, gravitation, light conditions, atmosphere, radiation, temperature, chemistry, geology, ecology, and biota. Secondly, the cognition of an organism is also adapted to the mind and culture of its conspecifics in order to understand and interact with other individuals, to understand emotions, thoughts, and motives, etc., in a psychological and sociological interplay that forms that particular exoculture.

Under certain conditions, which we are only beginning to understand, the environmental pressures force the cognitive agent to evolve toward more complex and flexible cognition. On Earth we find that intelligence seems – in the same way as vision, aerial locomotion and other abilities – to have emerged several times, apparently independently, in the course of evolution and in separate evolutionary lines, i.e. convergent or parallel evolution (Seed, Emery and Clayton, 2009). The more intelligent or cognitively flexible species on Earth, such as primates, dolphins, and corvids, share some qualities. Firstly, they are social animals and have a high degree of social complexity. Secondly, they are adaptable to very different environments and diets. If we could better understand the processes behind the rapid brain evolution that began a few million years ago on Earth – the encephalization in the Phanerozoic (Bogonovich, 2011; Carter, 2012) – then we could use this knowledge to formulate astrocognitive theories on the evolution of intelligence in space (see further Osvath, 2013, *this volume*).

Astrocognition concerns space and spatial consciousness. Cognitive science can give clues to how we understand and think about the Universe and reveal new perspectives on human encounters with the unknown. In short, cognitive science studies how we and the external world are represented and how we use cognitive tools for our thinking, such as language, image schemas, mental maps, metaphors, and categories. Yet, it is also how we use and interpret signs, objects, drawings, images, etc., to enhance communication. It is about perception, attention, memory, learning, consciousness, reasoning and other things that we include in what is called "thinking".

Cutting-edge thought in cognitive science portrays the mind as embodied, extended, distributed, and situated. According to the theory of *the embodied mind* (see e.g. Varela, Thompson, and Rosch, 1991; Lakoff and Johnson, 1999; Krois, 2007; Thompson, 2007), the mind is not something independent and detached from the body. There is no "brain in a vat", to borrow a turn of phrase from Hilary Putnam's (1981) famous thought experiment. Instead we think with the body. Therefore, bodies of other kinds and evolutionary backgrounds, like those that might exist on other planets in the Universe will have other minds and ways of thinking. The brain does not only need the body, but also the surrounding world in order to function efficiently. The environment has an active role in driving cognitive processes. According to the *extended-mind hypothesis* of Andy Clark and David Chalmers (1998; see also Clark, 2008), the mind leaks in various substantive ways into the environment. The boundary between self and non-self, self and world, is one that is never fixed but constantly being re-negotiated. In a similar way, *distributed cognition* (Hutchins, 1995) stresses the non-localizability of much of cognition, that we are using our environment and tools for enhancing thinking, and that we place our ideas and memories in things, such as books, computers, etc.

So, where we are in time and space is essential for cognition. What the senses convey have to be interpreted through means of specific cognitive processes. The cognitive agent is never just a passive recipient of images and information from the surrounding world. Instead, the brain actively searches out patterns in what is conveyed to it through the senses and interprets them through a process that is determined by both biological and cultural factors. The world distorts our concepts, and the concepts distort our world. Striking examples of this *epistemic perception* are the maps of Venus and Mars from the seventeenth century and onwards, that delineated the surface of the planets (Dunér, 2013b). We see what we expect to find. In 1877 at the Brera observatory in Milan, Giovanni Schiaparelli recorded a detailed network of canals on Mars. This finding was confirmed by the American astronomer Percival Lowell, who detected hundreds of Martian canals that he interpreted as an artificial irrigation system.

Another cognitive ability is *categorisation* (Lakoff, 1990; Taylor, 2003). All living creatures seem to categorise the environment in terms of edible versus inedible, benign versus harmful, and so forth. Categorisation becomes more complex in humans. The human mind tends to categorise, seeing hierarchies and similarities between things, such as stellar spectra or species and genera in taxonomy (Berlin, 1992; Dunér, 2013c).

From these cognitive theories concerning the human mind we can conclude that encounters with the unknown outer space will: i) change our spatial consciousness; ii) change our thinking, conceptions, categories, belief systems, culture, and meanings of things. What we have come to believe so far through science and human cogitation will face anomalies. The old categories, systems, and beliefs will fall short when we try to understand these new unfamiliar things. Our thinking, science, and belief systems will then have to be revised, which will lead to adjustments, adaptations, and compromises. A task for astrocognition is to search for the limits of our bio-culturally evolved, earthly brains to try to find out what we can know and what we are likely to encounter in the future. As thinking beings we are earthbound and historically constrained. Our intellect does not transcend space. Instead, our cognition is situated in space.

Astrocognition puts discussions of cognition into a wider perspective. We will get further theoretical, scientific knowledge of i) how we encounter the unknown, how the human mind interacts with space and the environment around us; and ii) the evolution and prerequisites needed for cognition to emerge. These achievements will be valuable even though we might never go beyond our Solar System. From this we would learn more about human cognition and how it has been developed here on Earth. But we will also get iii) practical knowledge that will prepare us for future manned and unmanned space missions and terraforming, how to live in an extraterrestrial environment, and how to code and decode interstellar messages. Finally, vi) future comparative research on cognitive processes of extraterrestrial minds can reveal more about how humans think. Then will we know more on specifically human ways of thinking and sense-making, specific or typical characteristics of our species. Are we unique in the Universe? This approach can lead to the development of not just an anthropocentric, but also a cosmocentric perspective on cognition. An astrocognitive inquiry will give indications as to what a human being is from a truly universal point of view. Encounters with other minds here on Earth or on extrasolar foreign planets in deep space among billions of stars and galaxies can give a more universal answer to the question: "What is thinking?"

Communication

This is a message, a message addressed to you constructed in a code-system called English marked with Roman letters in ink on sheets of cellulose, or as liquid crystals on a computer screen. It is here and now, perceived by the senses, being interpreted by a being with a brain, body and history, living in the world. By using this code-system I hope to make myself understood, to awaken thoughts and ideas in the mind of the receiver which are similar to mine when I formulated this message. This hope of mine stems from the fact that we share the same human cognitive abilities that are a product of our common evolutionary history here on Earth, the planet Tellus. However, if we extend this communicational situation beyond Earth, the question naturally arises: how could communication be possible between intelligent beings of different environments that differ physically, biologically, and culturally, and have developed through separate evolutionary lines? This is the *cognitive-linguistic problem of interstellar communication* (Dunér, 2011b).

The circumstances in which we will have no kinship, that we will not share similar bodies or cultures, or even similar physical realities will have far-reaching consequences for how we will be able to construct and interpret messages from distant civilisations. The usual strategy to overcome the problem of interstellar communication has been to try to construct a message that is a universal symbolic information transfer, which is independent of context, time and human nature. This can be called *the universal-transcendental interstellar message objective*. However, this strategy and its requirements are not reconciled with what we presently know about cognition, communication, and evolution. Particularly, it presupposes the universality it is aimed at, and thereby ignores the facticity of evolution, and the situatedness and embodiment of symbolisation. It ignores the context that the living organisms, with their cognition and communication, are planet-bound and constrained by certain physical conditions. It leaves out time and history – the evolution, the phylogenetic, ontogenetic, and cultural-historical development – in which the organisms are evolving. Finally, it ignores the nature of the communicators – that they have bodies and brains evolved in interaction with their environment. Building on the basic observation in cognitive science that our cognitive and communicational skills are embodied, situated in and adapted to our terrestrial environment, we cannot exclude the context, the situation, space, time, and human nature if we would like to construct comprehensible interstellar messages.

In order to solve the interstellar communication problem we need the insights of cognitive science, evolutionary theory, semiotics, hermeneutics, and history. Inasmuch as interstellar communication is thought to be an exercise in coding and decoding signs, then the relevance of semiotics (Vakoch, 1998) should be obvious. Information transfers by means of symbols (or conventional signs), which will be explained below, are probably an ineffective and hopeless way of starting a communicative interaction. Thus, it could be argued that the problem of interstellar communication is not just a problem within natural science, but a true humanistic problem in its true sense, a human problem. It is we humans who will send and receive, and code and decode the messages. The communicative problem also has to do with history, understood on the most basic level, as the interaction of organisms with their environment over time. Our cognition and communication are results of time, of history, both evolutionary history and socio-cultural history. Communication is not something pre-given, but rather evolves in interplay with the environment, in dialogue of agent with agent. In our case, this process took millions of years (Donald, 1993; 2001; Christiansen and Kirby, 1997; Deacon, 1997; Tomasello, 2008). Human communication, whether it consists of lingual, symbolic, or bodily expressions, is dependent on the inner workings of our brains and how humans interact with their physical, biological, and cultural environments. Accordingly, communication is a bio-cultural hybrid, a changing product of the genetic-cultural co-evolution. Communication is therefore a situated practice. It is constrained by its surroundings and is adapted to specific circumstances. This means that we cannot exclude the situation where the message is performed, and the physical, biological, and socio-cultural context of the communicators. We are planet-bound creatures.

Intelligence could be seen as evolved mental gymnastics which is required for a particular organism to survive and reproduce within its specific environment. This includes the capability of representing activities and being able to make inner models of reality. If the extraterrestrials are intelligent then they probably have some kind of symbolisation abilities and abstract thinking detached from the environment, with which they can reason about things that do not exist and things that are not right in front of them, facing their senses, in a specific moment in time. A very effective tool for symbolising thought is our communicational devices. John Taylor (2002) describes language as a set of resources that are available to the language user for the symbolisation of thought and the communication of those symbolisations. It facilitates thinking about that which is not immediately in front of us, engaging with our senses; as well as that which, seemingly, could never exist outside of fiction. It frees us ever further from the here-and-now and lets us better contemplate the might-have-been. It allows us to share ideas and mental states. Yet it rests on the cognitive abilities that are a result of bio-cultural evolution here on Earth.

Like our spoken languages, interstellar communication is intersubjective, a system for sharing information and for socialising. Communication can be regarded as a sharing of mental states and the expression as information about a mental state (Østergaard, 2012). Semantics is based on a "meeting of minds", as Peter Gärdenfors (2013, *forthcoming*) puts it: "the meanings of expressions do not reside either in the world or (solely) in the mental schemes of individual users, but emerge from *communicative interactions* between the language users." The evolution of semantics could be seen as a co-evolution of intersubjectivity, cooperation, and communication (Gärdenfors, 2008a; 2008b). In linguistics and cognitive science, almost no one would deny that intersubjectivity (Zlatev *et al.*, 2008) plays a critical role in the acquisition of language, but in the context of interstellar communication these basic insights have often been overlooked (Dunér, 2013a).

The interstellar communication problem is very much a semiotic problem: how meaning can be transferred and interpreted. As Göran Sonesson (2013, *this volume*) points out, the first problem that arises in a situation of interstellar communication is realising that it really is a message at all. Some regularity and order and finding a repetition in the pattern is not enough. We have to understand that someone has an intention with it that we should understand as a message. Next is the problem of deciphering what the message means. Some kind of vehicle for transferring the mental content is needed. The problem with symbols is that they are conventional, or arbitrary, as Ferdinand de Saussure (1916) called them. They are detached representations and, as such, dependent on culture and human interaction. The sign (the expression) and the signified (the content) have no intrinsic connection. It is not impossible to imagine that the aliens would have certain knowledge about their environment that in its content is similar to our own knowledge of mathematics, physics, or chemistry. However, their expression of it, as Sonesson clearly points out, would most likely be very different from ours. This basic semiotic distinction between expression and content is neglected in most human attempts at interstellar messaging. For all its ingenuity, Freudenthal's (1960) *lingua cosmica* is probably an effort in vain. Sonesson's critical analysis shows the inevitable role semiotics must play in message constructions. How we, and the aliens, transfer meaning in different ways, I would say, is the result of our dissimilar evolutions, bodily and cognitive constructions, and socio-cultural histories. The construction and interpretation of the symbols are dependent upon how our brains work, what our bodies are like, how we interact with our environment, how our sensations are processed and the history of our culture.

So we can conclude that:

> Communication is based on cognitive abilities embodied in the organism that has developed through an evolutionary and socio-cultural process in interaction with its specific environment.

This is the case for human cognition and communication, according to recent research in cognitive science and cognitive linguistics. In other words, our communication is adapted to an earthly environment and for communication with our conspecifics. Our communicational and symbolic skills have evolved through an evolutionary and cultural-historical process here on Earth and are thereby constrained by our human bodies, terrestrial environment, and the socio-cultural characteristics of our species. So our human communication is, in fact, maladapted to interstellar communication. This understanding of human cognition is crucial for future interstellar communication and should be taken into account in order to be able to transfer messages to other minds in the Universe.

Culture

The human desire for exploration and man's encounters with the unknown are a fundamental part of the cultural history of mankind, from the first, stumbling steps on the African plains to the recent explorations of our globalised and urbanised world. From the very dawn of the hominids to the days of the modern man, this ever changing terrestrial being has expanded in ever increasing circles of spatial consciousness, in an endeavour to climb over mountains to the next valley, cross vast oceans, and fly through the air. The next small step for a man, or giant leap for mankind, that of going far beyond the atmosphere and gravity of the Earth to the unknown extrasolar space, is decisive, but that, too, is part of the long history of mankind.

A true universal history includes outer space and not only the human history on Earth, not only the short history of human civilisation but also the immense space, and, as we know it, the long history of matter from the big bang, through the formation of planetary systems and the evolution of life and cognition, to the future speciation and the final big crunch or the ever-expanding Universe and eternity of time. History is crucial in order to understand the perennial enigmas of who we are, where we come from, and where we are going. Historical narratives are essential in order to formulate possible answers to the big questions of science: the origins of the Universe and the development of galaxies, stars and planets, the evolution of life on Earth, the human species and the human mind.

What made astrobiology possible was not only those human cognitive and communicational skills that biological evolution has given us, but also human culture, the ability to learn from others. *Culture* can here be defined (Sinha, 2009, p. 292; see also Tomasello, 1999) as:

> the existence of intra-species group differences in behavioural patterns and repertoires, which are not directly determined by ecological circumstances and which are learned and transmitted across generations.

Characteristic of human social interaction is this ability to learn from others, i.e. culture, the transmission of learned behaviour and knowledge that is not biologically encoded, or in other words, the ability to transfer information from generation to generation that does not use the genetic code for the transfer but is learned, taught, and transferred by a multitude of communicative and cultural devices and artefacts, like language, signs, pictures, sounds, objects, etc. Accordingly, culture presupposes enduring joint beliefs or common knowledge.

Culture made technological change possible. It has given us the increasing capability to manipulate the environment in order to make it easier to live in it and to adapt the environment to fit us, instead of adapting ourselves to the environment. Technology could be described as ways of manipulating the environment, using objects in the environment outside the body in order to strengthen our genetically given capacities, such as body strength, perception, and cognition. Culture also enabled science. Astrobiology as a science is not something isolated from the living world, from culture, society, beliefs, imaginations, communication and interactions with other thinking, believing and feeling minds. The chapters of this book show the intense interactions going on in astrobiology, between mind and environment, between science and culture. Without social interaction, joint beliefs, intersubjectivity, information transfer – in one word, without culture – science and technology would not have arisen on this planet.

One day we might encounter another living planet. The travelogues and descriptions of these new worlds will inform us more about ourselves and our place in the Universe, how we interpret and understand the "reality" around us, than about the "real" or "objective" extraterrestrial world. An independent world outside us might exist, but we will never be able to reach it without filtering it through our minds. Anyhow, future discoveries and experiences of other worlds will change us and our culture forever.

The unknown

There are things we know. Even though life might not exist out there, it is we human beings with our brains, bodies, and cultures who are searching for it. The history and philosophy of astrobiology is centred on humans, or more specifically, the scientific endeavour's dependence on the human mind and the human culture. Astrobiologists have brains, for sure; they are using cognitive tools that are a result of the bio-cultural co-evolution of human cognitive abilities. Certain cognitive processes are at work when astrobiologists encounter unknown things, when they interpret their observational data, and when they gather and classify it. This does not go on in subjective isolation. Astrobiologists live in a culture, in a certain time in history, in a specific research environment, and collaborate with other thinking beings.

Through the history of astrobiology we find a certain common form of argumentation: the analogy, from what we know to what we do not know (Dunér, 2013b). An analogical argument could be explained as a search for similarities, i.e., a way of selecting features in the source domain that are to be mapped onto the target domain, and of transferring relevant properties from the source to the target. The challenge is then to select the correct and relevant salient features from an infinite number of possible ones in the source domain, which features will then be transferred to and mapped onto the target domain. In the *Sidereus nuncius* from 1610, Galileo showed, based on his telescopic observations and analogical reasoning, that the Moon had mountains and therefore had the same solid, opaque, and rugged nature as the Earth (Spranzi, 2004). In some sense, astrobiology as a whole is one single, great analogy. Starting from the one particular type of life we happen to know something about, namely life on

Earth, we proceed to search for life on other planets. We predominantly look for life as we know it: something needing oxygen, liquid water, being based mainly on carbon, inhabiting a planet of a certain magnitude that revolves at a certain distance from its sun and which in turn has to be of a certain size and capacity, and so on. Logically speaking, analogical arguments are invalid. However, in providing us with some point of departure, they still might hold some heuristic advantage in the search for life. What we are actually looking for is something that reminds us of ourselves, something similar to us. In fact, we are searching for ourselves. Though, life might be very different from what we imagine. The history of science is actually a history of surprises. The world we are living in turned out to be very different from what we first thought: richer, more complicated, more advanced, more peculiar and more astonishing than what we could dream of. This will also be true for astrobiology. Future discoveries in astrobiology will surprise us completely.

If we find extraterrestrial life, we can be sure that this will change our way of thinking and how we perceive the world and our place in the living Universe. It will change our culture and science. It could be said that this book is not on the *history*, but on the *prehistory* of astrobiology. Should the day arrive when we find extraterrestrial life on another planet in our Solar System, or on an exoplanet or exomoon orbiting another star, then that shall begin the new history of astrobiology. That occasion will be a historical turning point in our persistent search for life. The greatest discovery of all for a human in her life and for mankind itself in the history of its civilisation, would be the encounter with another thinking being.

Acknowledgements

Galileo dedicated *Sidereus nuncius* to his patron Cosimo II de' Medici and named the moons of Jupiter that he discovered the *Medicean stars*. Likewise, we dedicate this book to our own Medicean patron: the Pufendorf Institute for Advanced Studies, Lund University, which from 2010 through 2011 hosted the interdisciplinary research project "Astrobiology: Past, Present, and Future". My special thanks go to Per Lind, Erik Persson, and not least Joel Parthemore, who read and commented upon earlier drafts of this introduction.

References

Aristotle, *Posterior Analytics; Topica*, edited by H. Tredennick, and E. S. Forster. London: Heinemann, 1966.

Berlin, B. 1992. *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*. Princeton NJ: Princeton University Press.

Bertka, C. M., ed. 2009. *Exploring the Origin, Extant, and Future of Life, Philosophical, Ethical and Theological Perspectives*. Cambridge: Cambridge University Press.

Billings, L. 2012. "Astrobiology in culture: The search for extraterrestrial life as 'Science'." *Astrobiology* 12(10):966–975.

Bjørnvig, T. 2012. "The holy grail of outer space: Pluralism, druidry, and the religion of cinema in *The Sky Ship*." *Astrobiology* 12(10):998–1014.

Bogonovich, M. 2011. "Intelligence's likelihood and evolutionary time frame." *International Journal of Astrobiology* 10:113–122.

Brahe, T. 1573. *De nova et nvllivs aevi memoria privs visa stella, ...* Copenhagen: Lavrentivs Benedicti.

Brandstetter, T. 2012. "Life beyond the limits of knowledge: Crystalline life in the popular science of Desiderius Papp (1895–1993)." *Astrobiology* 12(10):951–957.

Carter, B. 2012. "Hominid evolution: Genetics versus memetics." *International Journal of Astrobiology* 11:3–13.

Christiansen, M. H., and S. Kirby, ed. 1997. *Language Evolution*. Oxford: Oxford University Press.

Chyba, C. F., and K. P. Hand. 2005. "Astrobiology: The study of the living Universe." *Annual Review of Astronomy and Astrophysics* 43:31–74.

Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* Oxford: Oxford University Press.

Clark, A., and D. Chalmers. 1998. "The extended mind." *Analysis* 58(1):7–19.

Copernicus, N. 1543. *De revolvtionibvs orbium cœlestium, libri VI. ...* Nuremberg: Ioh. Petreium.

Crowe, M. J. 1986. *The Extraterrestrial Life Debate 1750–1900: The Idea of a Plurality of Worlds from Kant to Lowell*. Cambridge: Cambridge University Press.

—. 2008. *The Extraterrestrial Life Debate, Antiquity to 1915: A Source Book*. Notre Dame IN: University of Notre Dame.

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection:*

or the Preservation of Favoured Races in the Struggle for Life. London: John Murray.

Davies, P. 1998. *Are We Alone? Philosophical Implications of the Discovery of Extraterrestrial Life*. New York: Basic Books.

—. 2007. *Cosmic Jackpot: Why Our Universe Is Just Right for Life*. Boston: Houghton Mifflin.

Deacon, T. 1997. *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

Demets, R. 2012. "Darwin's contribution to the development of the panspermia theory." *Astrobiology* 12(10):946–950.

Dick, S. J. 1982. *Plurality of Worlds: The Origins of the Extraterrestrial Life Debate from Democritus to Kant*. Cambridge: Cambridge University Press.

—. 1996. *The Biological Universe: The Twentieth-Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge: Cambridge University Press.

—. 2012. "Critical issues in the history, philosophy, and sociology of astrobiology." *Astrobiology* 12(10):906–927.

Donald, M. 1993. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition.* Cambridge MA: Harvard University Press.

—. 2001. *A Mind So Rare: The Evolution of Human Consciousness*. New York: Norton.

Dunér, D. 2011a. "Astrocognition: Prolegomena to a future cognitive history of exploration." In *Humans in Outer Space – Interdisciplinary Perspectives*, edited by U. Landfester, N.-L. Remuss, K.-U. Schrogl, and J.-C. Worms, 117–140. Wien: Springer.

—. 2011b. "Cognitive foundations of interstellar communication." In *Communication with Extraterrestrial Intelligence*, edited by D. A. Vakoch, 449–467. Albany NY: State University of New York Press.

—. 2011c. "Die Evolution kognitiver Prozesse im Universum." *Sagenhafte Zeiten* 13(6):14.

—. 2012. "Introduction: The history and philosophy of astrobiology." *Astrobiology* 12(10):901–905.

—. 2013a. "Interstellar intersubjectivity: The significance of shared cognition for communication, empathy, and altruism in space." In *Extraterrestrial Altruism: Evolution and Ethics in the Cosmos*, edited by D. A. Vakoch, 139–165. Dordrecht: Springer.

—. 2013b. "Venusians: The planet Venus in the 18th-century extraterrestrial life debate." In *Meeting Venus*, edited by C. Sterken, and P. P. Aspaas, 145–167. Brussel: The Journal of Astronomical Data 19.

—. 2013c. “The language of cosmos: The cosmopolitan endeavour of universal languages.” In *Sweden in the Eighteenth-Century World: Provincial Cosmopolitans*, edited by G. Rydén, 41–65. Farnham: Ashgate.

—. ed. 2011. *The History and Philosophy of Astrobiology: International Conference on Tycho Brahe’s island Ven, Sweden, September 27–28, 2011*. Lund: The Pufendorf Institute of Advanced Studies.

Dunér, D., E. Persson, and G. Holmberg, eds. 2012. “The history and philosophy of astrobiology.” Special issue of *Astrobiology* 12(10).

Fontenelle, B. Le Bovier de. 1686. *Entretiens sur la pluralité des mondes*. Amsterdam: Mortier; new ed. Amsterdam: Pierre Mortier, 1701; trans., *Conversations on the Plurality of Worlds*. London: Thomas Caslon, 1767.

Freudenthal, H. 1960. *Lincos: Design of a Language for Cosmic Intercourse*. Amsterdam: North-Holland.

Galileo Galilei. 1610. *Siderevs nvncivs magna, longeqve admirabilia spectacula pandens, …* Venice: Thomam Baglionum; trans. W. R. Shea, *Galileo’s Sidereus Nuncius, or, A Sidereal Message*. Sagamore Beach MA: Science History Publications, 2009.

Gärdenfors, P. 2006. *How Homo Became Sapiens: On the Evolution of Thinking*. Oxford: Oxford University Press.

—. 2008a. “Evolutionary and developmental aspects of intersubjectivity.” In *Consciousness Transitions – Phylogenetic, Ontogenetic and Physiological Aspects*, edited by H. Liljenström, and P. Århem, 281–385. Amsterdam: Elsevier.

—. 2008b. “The role of intersubjectivity in animal and human cooperation.” *Biological Theory* 3(1):1–12.

—. 2013. “The evolution of semantics: Sharing conceptual domains.” In *The Evolutionary Emergence of Language: Evidence and Inference*, edited by R. Botha. Oxford: Oxford University Press.

Guthke, K. S. 1983. *Der Mythos der Neuzeit: Das Thema der Mehrheit der Welten in der Literatur- und Geistesgeschichte von der kopernikanischen Wende bis zur Science Fiction*. Bern: Francke.

Haqq-Misra, J. 2012. “An ecological compass for planetary engineering.” *Astrobiology* 12(10):985–997.

Hutchins, E. 1995. “How a cockpit remembers its speeds.” *Cognitive Science* 19:265–288.

Jakosky, B. M. 2000. “Philosophical aspects of astrobiology.” In *A New Era in Bioastronomy*, edited by G. A. Lemarchand, and K. J. Meech, 661–666. San Francisco CA: ASP Conference Series, vol. 213.

Kepler, J. 1609. *Astronomia nova aitiologetos, seu Physica coelestis, tradita commentariis de motibus stellae Martis, ex observationibus G.V. Tychonis Brahe: ...* Heidelberg: G. Voegelinus.
—. 1619. *Harmonices mvndi libri V...* Linz: Johannes Plancus.
—. 1627. *Tabulae Rudolphinae: quibus astronomicae scientiae, temporum longinquitate collapsae restauratio continetur; ...* Ulm: Jonæ Saurii.
Kane, S. R., and D. M. Gelino. 2012. "The habitable zone and extreme planetary orbits." *Astrobiology* 12(10):940–945.
Krois, J., et al., eds. 2007. *Embodiment in Cognition and Culture*. Amsterdam: Benjamins.
Lakoff, G. 1990. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago IL: University of Chicago Press.
Lakoff, G., and M. Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
Mayor, M., and D. Queloz. 1995. "A Jupiter-mass companion to a solar-type star." *Nature* 378(6555):355–359.
NASA Strategic Plan 1998: With 1999 Interim Adjustments. http://www.hq.nasa.gov/office/codez/plans/NSP99.pdf
Newton, I. 1687. *Philosophiæ naturalis principia mathematica*. London: Josephi Streater.
Østergaard, S. 2012. "Imitation, mirror neurons and material culture." In *Excavating the Mind: Cross-sections through Culture, Cognition and Materiality*, edited by N. Johannsen, M. Jessen, and H. Juel Jensen, 25–38. Aarhus: Aarhus University Press.
Pálsson, G. 2009. "Celestial bodies: Lucy in the sky." In *Humans in Outer Space – Interdisciplinary Odysseys*, edited by L. Codignola, and K.-U. Schrogl, 69–81. Vienna: Springer.
Perryman, M. 2012. "The history of exoplanet detection." *Astrobiology* 12(10):928–939.
Persson, E. 2012. "The moral status of extraterrestrial life." *Astrobiology* 12(10):976–984.
Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
Race, M., et al. 2012. "Astrobiology and society: Building an interdisciplinary research community." *Astrobiology* 12(10):958–965.
Rosch, E. 1975. "Cognitive Representations of Semantic Categories." *Journal of Experimental Psychology: General* 104:192–233.
—. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by E. Rosch, and B. B. Lloyd, 27–48. Hillsdale NJ: Erlbaum.
Saussure, F. de. 1916. *Cours de linguistique générale*. Paris: Payot.
Sagan, C., ed. 1973. *Communication with Extraterrestrial Intelligence*.

Cambridge MA: MIT Press.
Seed, A. M., N. J. Emery, and N. S. Clayton. 2009. “Intelligence in corvids and apes: A case of convergent evolution?” *Ethology* 115:401–420.
Sinha, C. 2009. “Language as a biocultural niche and social institution.” In *New Directions in Cognitive Linguistics*, edited by V. Evans, and S. Pourcel, 289–309. Amsterdam: Benjamins.
Spranzi, M. 2004. “Galileo and the mountains of the Moon: Analogical reasoning, models and metaphors in scientific discovery.” *Journal of Cognition and Culture* 4(3):451–483.
Taylor, J. R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.
—. 2003. *Linguistic Categorization*. Oxford: Oxford University Press.
Thompson, E. 2007. *Mind in Life: Biology, Phenomenology and the Sciences of Mind.* Cambridge MA: Harvard University Press.
Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge MA: Harvard University Press.
—. 2008. *Origins of Human Communication*. Cambridge MA: MIT Press.
Vakoch, D. A. 1998. “Constructing messages to extraterrestrials: An exosemiotic perspective.” *Acta Astronautica* 42(10–12):697–704.
Vakoch, D. A., ed. 2011. *Communication with Extraterrestrial Intelligence*. Albany NY: State University of New York Press.
Varela, F. J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge MA: MIT Press.
Zlatev, J., T. P. Racine, C. Sinha, and E. Itkonen, eds. 2008. *The Shared Mind: Perspectives on Intersubjectivity*. Amsterdam: Benjamins.

PART I:

COGNITION

CHAPTER ONE

PHILOSOPHICAL ASPECTS OF ASTROBIOLOGY

ERIK PERSSON

Introduction: Is there a place for philosophy in modern astrobiology?

During antiquity, the astronomical questions of the day and the methods used to formulate and answer them were clearly within the realm of philosophy (Dick, 1996; Hoskin, 1999; 2000; Sikö, 1999; Högnäs, 2000; Bennett *et al.*, 2003; Dick, 2009; Bennett and Shostak, 2012). That changed most notably in the sixteenth century when Tycho Brahe turned astronomy into a modern empirical science by formulating (in principle) testable hypotheses, figuring out how to test them, building the proper instruments, and making – for that time – very accurate and systematic observations of the sky. These observations eventually led to the modern view of the solar system and ultimately the rejection of the classical Aristotelian worldview (Kaufmann, 1994; Wittendorff, 1994; Kaufmann and Comins, 1996; Hoskin, 1999; 2000; Sikö, 1999; Ferguson, 2002; Bennett *et al.*, 2003; Pedersen, 2003; Jönsson, 2004; Gingerich, 2006; Bennett and Shostak, 2012). Since then, the exploration of space has been located clearly in the realm of natural science – and, to an increasing extent, technology. In parallel to that, space has – at least since World War II – to a high degree been a place for military and political ambitions and activities. The scientists that have been involved in space exploration have mainly been physicists, together with some chemists and geologists. Lately, the exploration of space has engaged more and more biologists in the search for extraterrestrial life. The international space station, along with plans to send humans to Mars, has led to an increased involvement of medical science. If present speculations about space tourism and mining on planets, moons, and asteroids hold true, in the future we will probably see many new professions dealing not just with exploration, but also exploitation of space.

Philosophy has never totally left the realm of space exploration (Dick, 1996). As space exploration and exploitation become more and more interdisciplinary, and as space becomes a bigger part of our lives, it will be time once more for philosophy to turn its attention towards space. The mission for philosophers is not to indulge in speculations over questions that are better dealt with by empirical methods. There are, however, several questions connected to the exploration and exploitation of space that *are* of a philosophical nature and that deserve to be examined seriously from a philosophical perspective. I believe that philosophical scrutiny of these questions will be to the benefit of both space exploration and philosophy.

Interesting and, indeed, important philosophical questions arise in many areas of space exploration. Some of the most interesting and important ones relate to the search for extraterrestrial life, of which some will be presented here. I will not attempt to answer these questions here. Instead, this chapter should be seen as an introduction to some of the philosophical questions raised by astrobiology.

Is astrobiological research a justifiable use of resources?

The first question that must be considered is whether astrobiological research is justified at all. Like all other endeavours, it requires resources in the form of money, energy, and natural resources, along with the time and effort of the researchers involved. At the same time, we have many important problems in urgent need of solutions here on our own planet. If we decide that the search for extraterrestrial life *is* worth pursuing, this raises several further questions: e.g., how much will it cost and who should pay?

Critics argue that we should concentrate our efforts on solving problems on Earth before concerning ourselves with the question of life on other worlds. They point out that the probability of success is very small – even if there is life out there (see Tipler, 1981; 1982; 1993; Perlman, 1982; Dick, 1996; Garber, 1999; Cox, 2008; Lemonick, 2011, for examples and discussions).

Those in favour of the search for extraterrestrial life can point to the lack of any guarantee that resources saved by abstaining from astrobiological research will be allocated to solving problems like famine or environmental change; it is perhaps more probable that they will be used for e.g. military purposes. They can point out that astrobiology, in helping us understand the conditions for life on our own planet, can thereby play an important role in dealing with environmental problems here on Earth. They can point out that research motivated by pure curiosity sometimes turns out to be very useful. They can argue that both knowledge and the search for it have value as ends in their own right, while the question whether we are alone in the universe is an important existential one. They can also claim that curiosity and the drive to investigate new territory is an important part of our human identity and plays an important role in our achievements to date – good and bad (for further arguments in favour of astrobiological research, see Tough, 1998; Maurel and Zaccai, 2001; Shostak, 2009).

How to allocate resources and justify a project are partly matters of economy and management, partly of probability for success. They are also, ultimately, matters of value. This raises important philosophical questions. What has value? What does it mean to claim that something has value as an end in itself (*end value*)? How can we compare and prioritize between very different kinds of value: e.g., how can we compare the price in dollars of a new telescope with the end value of pursuing existential questions?

Justification is also a question of ethics: is it right to spend time, money, and effort on research motivated by curiosity while people are starving and nature is suffering? Who should pay is also partly a moral question: who should pay for basic research, how should responsibility be divided among nations, agencies, and private investors? What is a fair distribution of costs? Should all pay the same amount, should each pay according to her ability, should each pay according to what she expects to get out of the project (however that is measured), or should the expenses be distributed based on the participating agencies' and countries' formal and informal roles in the scientific community?

What is life?

Most of us have a general intuitive understanding of life. Even children can usually distinguish living from non-living objects in their surroundings. That said, in some situations things are not quite so simple. For example, it is not agreed whether viruses are alive. Viruses can multiply and undergo evolution; but they do not have their own metabolism, which many argue as a necessary property of being alive (for discussion see Choi, 2008; Hegde *et al.*, 2009; López-Garcia and Moreira, 2009; Ludmir and Enquist, 2009; Moreira and López-Garcia, 2009; Navas-Castillo, 2009; Raoult, 2009; Forterre, 2010; Herrero-Uribe, 2011). Nowadays, one finds computer programs that have some features in common with what we normally call life: they can e.g. reproduce and adapt to their surroundings. Some argue that they *are* alive (Levy, 1992) even though most would deny it. Many would further deny that living computer programs even are possible, arguing that life must have some form of chemical base or be the result of natural selection – not selection based on rules that humans have constructed.

Some things are hard to classify for practical or methodological reasons. The problem with life goes deeper. The big problem is a lack of consensus about how to define life. This is not for lack of suggestions: suggestions abound, but all have their difficulties. When we leave the Earth, the problem becomes much more difficult. We cannot expect life on other planets or moons to look exactly the same as life on our own planet. On the other hand, extraterrestrial life must have some things in common with life on Earth in order to be called life.

For astrobiology, the question of how to define life is particularly important – for three reasons. First, we need to determine if what we find is alive to know whether or not we have succeeded. Second, in searching for and attempting to understand the origin of life, in attempting to say when the transition took place from chemistry to biology, we need to know what life is. Third, when we look for life outside Earth, we need to decide how and where to look. We need to choose the right kind of instrument that responds to the right indicators (Cleland and Chyba, 2002). To do that, we need to know what characteristics to look for, and that presupposes knowing what life is. An informed discussion of how to define life is therefore very important.

The problems start already *before* we can start discussing the proper definition of life: we need to decide what kind of definition to use, a matter that is often overlooked. As I said in the introduction, my aim is not to provide answers; so I leave the production of the definitions themselves for future work (for examples and discussions of possible definitions, see Allen, 1899; Schrödinger, 1944; Johansson *et al*., 1978; Margulis and Sagan, 1995; Stenholm, 1997; Anbar, 2001; Bennett *et al*., 2003; NASA, 2003; Ruiz-Mirazo *et al*., 2004; Zhuravlev and Avetisov, 2006; Schulze-Makuch and Irwin, 2008; Carroll, 2009; Fry, 2009; Strick, 2009; Benner, 2010; Damiano and Luisi, 2010; Deamer, 2010; Gayon, 2010; Kolb, 2010; Popa, 2010; Ruiz-Mirazo *et al*., 2010; Tirard *et al*., 2010; Tsokolov, 2010; Weber, 2010; Herrero-Uribe, 2011; Bennett and Shostak, 2012).

Definitions can be of many kinds. One way to define life is to make a list of everything we consider to be alive: what is often referred to as a *list definition*. In this way, we can make use of our intuitive ability to recognize life when we see it (Føllesdal *et al*., 1988; Lübcke, 1988). The advantage of this strategy is that we need not specify the characteristics of life. We just see what we find, and once we have found it, we decide if it makes the list.

There are also obvious disadvantages to this method. One is that the list will never be complete; therefore, we will never have a complete definition of life. Another is that you cannot put anything on the list until you have found it: the list method can therefore not help us to choose how and where to look; neither can it help us decide whether something should be on the list or not. In particular, it cannot help us determine the really tricky cases, and these are precisely the cases where we really need a definition.

An alternative is the *ostensive definition* (Pap, 1950; Whiteley, 1956; Kotarbinska, 1960; Radford, 1964; Harris, 1986; Føllesdal *et al*., 1988; Lübcke, 1988; Ghiselin, 1995; Ruthrof, 2005). You point out examples of the category you want to define. To those who wonder what a book is, you can go to the nearest bookcase, point to a book, and say "this is a book". Such a method can work well for things that are relatively consistent. Books may differ a lot; but, compared to living things, one must nevertheless say that books are quite similar to each other. If you once have had a book pointed out to you, you should have no problem recognizing other books when you see them. Life is much more multifaceted. One of the most fascinating features of life is that it *is* so multifaceted. When we search for life on other planets, we can expect that what we find may prove very different from the life we know on Earth.

One can go a step further and select a particular book – or, in our case, a special organism – and designate it a model or *prototype* for all other books (or all other life). This is called a *prototype definition.* It differs from ostensive definition by using one particular instance of the category as the model rather than pointing out any instance of it. When we think about life, we can often see before us a mental image of a living thing that we are familiar with: maybe a dog or tree, or something else that we identify with life; such a strategy is intuitively appealing.

A problem with this strategy is that it will be difficult to agree which living being, we shall use as a prototype. Another problem is that life can look very different and have very different characteristics. Nor are all properties of living beings unique to living beings; we still need to decide which properties of the prototype are relevant. Let us assume that our prototype is a labrador. Labradors are typically black and have four legs; but that also applies to many dining room tables, which we do not consider alive. It does not, on the other hand, apply to the common egret, which we consider every bit as alive as a Labrador even though it is white and has two legs. On the other hand, common egrets and labradors have many things in common: e.g., metabolism and the ability to regulate body temperature. The question is: which properties are most relevant? If we do not want to include dining room tables among the living but do want to include common egrets, we must specify what properties in the Labrador are the relevant ones.

This brings us to a more common strategy for definitions: to try to determine which properties are necessary and which sufficient. This is sometimes called a *real* or *de re definition* (Føllesdal *et al*., 1988; Lübcke, 1988; Bernadete, 1993; Retana-Salazar and Retana-Salazar, 2004; Thompson, 2008). Applied to life, it assumes that all living things have one or more common characteristics that are unique to the living. These are sometimes called the core or essence of life. The ontological assumption behind this is called *essentialism*; it assumes that there is something or some set of things that constitute the essence of life. In some cultures and some periods of history, people believed that life itself is a special property or particular form of matter that could, in theory, be isolated and identified. Today, as we attempt to define life in a way consistent with science, we no longer believe in this kind of essence. Most attempts to define life do not even assume that there is one particular property that is unique for life. Instead it is usually a matter of finding the right combination of properties, where each property is not unique for life but the combination is.

Identifying these necessary and jointly sufficient properties can be done in different ways. The easiest is probably to just decide on a property or set of properties. One sometimes hears people say that something is "just a matter of definition." By that, one usually means that whether a statement is right or wrong depends on how one chooses to define it, which is more or less subjective. In the case of life, that means it is up to each of us, or maybe some authority, to decide how to define life for any particular purpose – just to get it over with, without worrying how others use the term or what connotations it might have.

This is called a *stipulative definition.* It can, in some cases, be useful (Lübcke, 1988). For the purposes of astrobiology though, it would hardly be recommendable. If we send an international expedition to Mars to look for life, we need to be fairly in agreement what to look for – both to choose the right strategy and instruments, and to be able to say whether or not we succeeded. If, in a paper in a scientific journal, someone claims to have discovered life, one reasonably expects that 'life' is used in a way that the rest of the scientific community accepts. Meanwhile, if we want to convince society at large that we have discovered life, we need to define life in a way that does not deviate too much from everyday usage. As discussed in the previous section, we also need to be able to justify our use of resources to society at large. To do that, we must aim for something that is widely agreed to be interesting. If we define life in a way that is convenient but differs substantially from people's intuitions about life, the matter of justification will be much more difficult – and rightly so. The question of the nature of life is an existentially important one that demands that we take the definition of life seriously.

Another alternative is the *operational definition* (Losee, 1980; Føllesdal *et al*., 1988; Fleischaker, 1990). Like the previous approach, it is not aiming to find *the* correct definition. An operational definition is based on the methods that happen to be at our disposal: i.e., on what we are able to find given technological and other limitations. To give 'life' an operational definition, we would start with the methods and instruments available to us. Based on those, we then decide on a definition of life. In some cases, this can be a very useful method; but it seems insufficient in this context. First, technology is constantly developing: we get continuously better methods and more sensitive instruments, so that a lot of things that cannot be counted as alive today – because we cannot find them, or because we cannot identify them as life – would count as alive in future. The number of living beings in the universe would increase in line with the development of our methods and instruments for identifying life. The question whether there is life outside our own biosphere is an important existential question for many of us; few of us would be content with a definition designed only to suit currently available tools. It might also seem strange that the question whether life exists (as opposed to the question whether we can find it, if it exists) elsewhere in the universe depends on technological development here on Earth.

However one seeks to define life – especially if the purpose is to answer questions like "are we alone in the universe" – it is probably important to have a definition that is consistent with our intuitive understanding of what life is: one that includes those things we agree are alive: e.g., sunflowers, labradors, and common egrets; excludes those things we agree are not alive: e.g., grains of sand, dining room tables, and space shuttles; gives us a practical method for determining the tricky cases; and – not least – explains why the issue is so important in the first place. If one defines life in a way that fits with existing tools, or in a manner that is otherwise convenient but more or less arbitrary, we run the risk of getting a definition that does not explain why it is so important to most of us to be able to tell life from non-life, to look for life outside our planet, to save lives, or to live. Taking these aspects into consideration should be a primary criterion for a successful definition.

Another alternative is so called *family resemblance* (Wittgenstein, 2009). According to this idea a concept is not defined by necessary and sufficient properties but by a cluster of properties that are associated with the concept in question. To say that something is alive then means that it has a number of these properties. No one living being has all of the properties and to particular living beings do not have to share the same properties as long as they have some of the properties in the cluster. This has the advantage that we need not decide on one particular property or set of properties as necessary and sufficient. It is enough to identify a number of characteristics that are associated with life. Not everyone agrees that this constitutes a proper definition, however. It might also be difficult to use in practice.

Some think that we should stop considering things as being *either* alive *or* not alive, and instead consider life a matter of degree (see e.g. Hazen, 2009; Bedau, 2010; Jager op Akkerhuis, 2010), in which case things may be more or less alive, based on e.g. how many characteristics of life they possess or to what extent they possess them. This seems to go against many people's intuitive understanding of life, however. It makes the hunt for alien life a matter of finding things that are more or less alive instead of finding Life – which might make the endeavour a bit less attractive.

One thing to keep in mind when looking for life criteria is that finding a criterion for being alive is not the same as finding a necessary chemical, environmental, or otherwise physical condition for life: the conditions follow from the definition; the two are *not* equivalent. If we conclude that metabolism is a necessary criteria for life, and then find out that metabolism requires a certain chemical composition, we have made an interesting discovery that will assist in our search for life, but we have not added a criterion for life. It is in this case still metabolism that defines life, not the chemical composition necessary for metabolism. This is important for two reasons. First, we want our definition of life to concentrate on what we find to be essential, not on what just happens to be the case. Second, we might be wrong in concluding that metabolism requires the chemical composition we think it does. When we take a closer look on another world, we might find something that has metabolism in a way we had never considered before. We want to avoid being stuck with a definition that is too dependent on how things are on our own planet.

The definition of life has been discussed since antiquity. Discussing it in relation to astrobiology does not just benefit astrobiology but is also useful to general biological and philosophical discussions of what life is, by forcing us to consider other options than the ones we are familiar with on our own planet, while raising the question: "what about life makes the question of whether it exists elsewhere in the universe so important and has given rise to so much speculation, fiction, and scientific curiosity?" Who would write a novel, make a film, or finance a scientific mission based on the question "do pebbles of 2–3 cm in size exist anywhere else in the universe?" Life is different; the trick is to capture this in a definition. Even if we cannot agree what life is, our attempts to define it sharpen our thinking about the properties of and conditions for life.

If we really are alone, how will we ever know?

If we find life on another planet, then we know extraterrestruial life exists; but if we do not find life, does that mean that extraterrestrial life does not exist? Clearly not. So far we have not found extraterrestrial life; but this cannot be taken as proof that it does not exist. Our search so far has not been extensive – certainly not thorough enough to show, convincingly, the absence of life out there. If there really is no extraterrestrial life, what would it take to show this? A different, more practically oriented way to ask the same question is: how long and hard must we look before it is time to give up? In philosophical terms: how does one verify a negative hypothesis? (see Achinstein, 2003; Chalmers, 1999; Noble, 1975; Popper, 1959).

Strictly speaking, it is impossible to *prove* a negative hypothesis; but one can be more or less certain. It is generally not fruitful to maintain that, having dropped a stone and seen it fall 1,000 times, we still have no idea whether, the next time, it will fly away instead. The more times we test and get the same result, the more certain we are – and rightfully so. Although we can never logically exclude the possibility that the stone will fly away on the 1,001st trial, it is more rational to bet it will not. To make progress, we must draw the line somewhere. We must, at some point, be able to say that we are certain *enough*.

If we continue to be unsuccessful in our search for life, where shall we draw the line and conclude that extraterrestrial life does not exist? The question needs to be approached differently depending on whether we are asking the general question "are we alone in the universe?" or want to know if life exists on a particular planet or moon.

The general question whether we are alone in the universe is special in three ways. (1) It is not possible to look through the entire universe for life. (2) The question whether we are alone in the universe is existentially important. (3) There is no need to reach a conclusion before any particular point in time.

In other words: the question is hard; it is important; and it lacks a deadline. Together, these features of the question tell us that we should not be hasty in concluding that we are alone. The harder a question, the more important it is; while the less in a hurry we are to find the answer, the more time we can and should allow.

The situation is different if we talk about a particular planet, moon, or other celestial body that is a potential candidate for exploitation, terraforming, or other invasive treatment that would affect any indigenous life. The area we need to look through is considerably smaller, even though it can still be considerable. Finding an answer is also arguably more urgent: we will face pressure from developers who want to proceed with the project. Finally the result of the search – though important indeed – does not have the existential implications of the question whether there is life anywhere beyond our planet. In this case, a different balance between deadline and certainty is needed. Finding that balance is of great philosophical interest.

The question of what it takes to verify (in any sufficiently fruitful sense of the word) a negative or *null* hypothesis is basic to the philosophy of science. The scientific community is largely in consensus about everyday research. The problem posed by the null hypothesis is often dealt with by deciding the degree of statistical significance needed to dismiss it, operationalized in questions about the sample size needed or how much the results must deviate from chance. In the search for extraterrestrial life, these questions are not of much obvious help: the population we sample from is, for all practical purposes, infinite; and we do not know what number to assign to chance. We must suffice by asking ourselves how thorough our investigation must be.

Perhaps different degrees of certainty are needed for different questions. Perhaps we should handle the question whether we are alone in the universe differently from the question whether a particular planet supports life. Different standards may also be needed depending on, for instance, whether we are just curious about some planet or are about to implant life – perhaps human life – there.

How should we behave towards extraterrestrial life?

In science fiction movies, the answer is often simple: if they look like us, we are meant to sympathize with them. If they look like insects or lizards, we are to fight them. In real life, things are a bit more complicated. No major contemporary ethical theory supports grounding moral concern in measures of similarity.

How useful will the ethical theories we use today be in our relations to extraterrestrial life? In principle, ethics is ethics; it should not make a difference whether the agent and the objects of concern come from the same planet. One should rather look at the properties of the particular life form than be preoccupied with which world it comes from.

Within contemporary terrestrial environmental ethics, one of the most important discussions is which kinds of life require moral consideration. Is it enough to be alive, or must it be sentient or even human? Must we talk only of individuals, or must we also show concern for colonies, species, and ecosystems?

The idea that *all* life has moral status is called *biocentrism* (see Schweitzer, 1976; Goodpaster, 1978; Taylor, 1986; Cockell, 2005; 2011, for some advocates of this theory). At first, it seems ideal for including extraterrestrial life; but, since it is modelled on life on Earth, it is not self-evident that extraterrestrial life will count.

The idea that all *sentient* life has moral status is called *sentientism*. It explicitly disregards everything but sentience life, which means if we find sentient extraterrestrial life it will count according to this theory. That it is extraterrestrial does not matter. All non-sentient life will be excluded however (Clark, 1977; Regan, 1983; Singer, 1993; 1995; DeGrazia, 1996; Regan, 2001; Bentham, 2005; Persson, 2008).

The idea that only human beings have moral status – historically taken for granted – is called *anthropocentrism* (see Descartes, 1960; Rawls, 1973; Carruthers, 1994; Kant, 1998; Smith, 2009, for some advocates of this theory). It might seem obvious that extraterrestrials will not count according to this theory but, in practice, it depends on what basis is chosen for anthropocentrism. If only humans count on the basis that, allegedly, we are the only organisms sufficiently intelligent to pass muster, then we must include extraterrestrials who are sufficiently intelligent, even though they are not human.

The idea that not only individuals but also species, ecosystems, and landscapes have moral status is called *ecocentrism*. It comes in different varieties; depending on the one we choose, extraterrestrial life might be fully included or excluded. On some accounts, we must also protect worlds without any life at all (Leopold, 1970; Callicott, 1980; 1985; 1987; 1992; 1999; Rolston, 1986; 1987; 1988; 1994; 1999; Johnson, 1991; 1992; Plumwood, 1991).

The discussion becomes both more complex and more interesting when considering extraterrestrial life that might have various combinations of properties, and where it might be unclear whether what we find is intelligent, sentient, or even alive; whether what we find consists of individual organisms or is some kind of collective. If we can categorize extraterrestrial life as having moral status according to one of the accounts outlined above, we have won something important: we will be able to apply the ethical theories already in use on Earth for deciding how to treat extraterrestrial moral entities (Persson, 2012).

This does not mean that all is said and done: important questions and competing theories remain. What it *does* mean is that ongoing discussions in moral philosophy are directly applicable to questions encountered in astrobiology. Meanwhile, we may be able to learn new things about interspecies relations on Earth by looking at interplanetary ethics.

What shall we do with uninhabited worlds?

The question of how to handle extraterrestrial life has many interesting philosophical aspects. But so does the question: “What are our obligations regarding worlds with no life?” The answer may not be as obvious as one might think. One aspect what we need to consider is that, in this scenario, we are dealing with a world that has, up until now, remained untouched by humans, and, perhaps, any other life. If nothing else, this makes such a world geologically interesting: it has value to scientists who might want to have, if not the entire planet, at least reasonably large or representative areas left untouched, for research.

Pristineness can also be important in itself. That an area is more or less untouched by humans is often seen as important when deciding what to do with areas here on Earth: many people assign special value to untouched areas. On Earth, it is difficult – if not impossible – to find areas that are totally untouched; but in the extraterrestrial case, we have whole planets that are totally untouched. Of course, it would not be a question of just one planet; the universe is full of untouched planets, each probably very different from the other. Each time, it may prove controversial whether to exploit this planet or that moon.

In most cases on Earth, pristineness is a matter of *living* nature more or less untouched by humans. That said, in some cases people want to save e.g. rock formations even though the organisms who live on those rocks could just as well live some place else. The rock formations end up protected because they have value in themselves, even though they have no moral status. Many national parks are dedicated, at least in part, to protecting geological formations: e.g., Grand Canyon and Monument Valley national parks in the US or Sånfjället and Söderåsen national parks in Sweden. Geological formations can be very beautiful.

Consider the highest mountain in the solar system: Olympus Mons on Mars. Imagine that some highly useful mineral is found there that would be technologically and economically feasible to extract. Would it be right to turn the solar system's highest mountain into a mine even though no one lives there?

Exploiting a previously untouched planet might tell us something uncomfortable about ourselves: it might e.g. confirm that we are a species of unappeasable colonisers who can never get enough. Perhaps we do not want to confirm this by sinking our teeth, so to speak, into yet another planet.

Others argue that we have a duty to spread life. If life is something good – which we generally believe it is – why not try to spread it as much as possible? This presents a dilemma. What should we do with totally lifeless planets or moons, such as our moon? Should we try to preserve them the way they are, or should we do what we can to implant life on them? If we have a duty to spread life as much as possible, what does that mean in practice? Should we turn as many planets as possible into human habitats? Should we implant primitive life and then let it evolve, to see what happens? Should we implant entire ecosystems on planets sufficiently similar to our own, and keep them as nature reserves in case we destroy the corresponding ecosystem on Earth? Should we move species threatened by human expansion to other planets, to save them from going extinct? Should we move humanity itself to a new planet and leave this one to nature?

Concluding words

Here we have seen some examples of astrobiology-related issues that have substantial philosophical dimensions. The list is far from complete; consider it an appetiser, an invitation for philosophers to get involved. I have argued for benefits to be won for both astrobiology and philosophy. For philosophy, astrobiological questions can give new perspectives to old questions and even pose new questions. For astrobiology, philosophical methods give new insights to questions that are outside the scope of empirical science but nonetheless central to the enterprise of understanding life: its origins, its distribution, and its future.

References

Achinstein, P. 2003. *The Book of Evidence*. Oxford: Oxford University Press.

Allen, F. J. 1899. "What is life?" *Proceedings of Birmingham Natural Historical and Philosophical Society* 11:44–67.

Anbar, M. 2001. "What is life?" *IEEE Engineering in Medicine and Biology* 20:15–19.

Bedau, M. A. 2010. "An *Aristotelian* account of minimal chemical life." *Astrobiology* 10:1011–1020.

Benner, S. A. 2010. "Defining life" *Astrobiology* 10:1021–1030.

Bennett, J., S. Shostak, and B. Jakosky. 2003. *Life in the Universe*. San Francisco CA: Addison Wesley.

Bennett, J., and S. Shostak. 2012. *Life in the Universe*. San Francisco CA: Pearson 2012.

Bentham, J. 2005. *An Introduction to the Principles of Morals and Legislation.* Boston MA: Adamant Media Corporation.

Bernadete, J. A. 1993. "Real definitions: Quine and Aristotle." *Philosophical Studies* 72:265–282.

Callicott, J. B. 1980. "Animal liberation: A triangular affair." *Environmental Ethics* 2:311–338.

—. 1985. "Bookreview of 'The Case for Animal Rights'" *Environmental Ethics* 7:365–372.

—. 1987. "The conceptual foundations of the land ethic." In *Companion to A Sand County Almanac – Interpretive Critical Essays*, edited by J. B Callicott, 186–217. Madison WI: The University of Wisconsin Press.

—. 1992. "Moral considerability and extraterrestrial life." In *The Animal Rights/Environmental Ethics Debate*, edited by E. C. Hargrove, 137–150. Albany NY: State University of New York Press.

—. 1999. *Beyond the Land Ethic: More Essays in Environmental Philosophy*. Albany NY: State University of New York Press.

Carroll, J. D. 2009. "A new definition of life." *Chirality* 21:354–358.

Carruthers, P. 1994. *The Animals Issue: Moral Theory in Practice*. Cambridge: Cambridge University Press.

Chalmers, A. F. 1999. *What is this Thing Called Science?* Buckingham: Open University Press.

Choi, C. Q. 2008. "It's alive?" *Scientific American* 299:18.

Clark, S. R. L. 1997. *Animals and their Moral Standing*. London: Routledge.

Cleland, C. E., and C. F. Chyba 2002: "Defining 'Life'." *Origins of Life and Evolution of the Biosphere* 32:387–393.

Cockell, C. S. 2005. “The value of microorganisms.” *Environmental Ethics* 27:375–390.
—. 2011. “Microbial rights?” *EMBO reports* 12:181.
Cox, B. 2008. “Low expectation from CNN.” *Herald-Tribune* Nov. 22 http://www.heraldtribune.com/article/20081122/blog32/811219994
Damiano, L., and P. L. Lisi. 2010. “Towards an autopoietic redefinition of life.” *Origins of Life and Evolution of the Biosphere* 40:145–149.
Deamer, D. 2010. “Special collection of essays: What is life?” *Astrobiology* 10:1001–1002.
DeGrazia, D. 1996. *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge: Cambridge University Press.
Descartes, R. 1960. *Discourse on Method and Meditations*. New York: The Liberal Arts Press.
Dick, S. J. 1996. *The Biological Universe: The Twentieth-Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge: Cambridge University Press.
—. 2009. “A historical perspective on the extent and search for life.” In *Exploring the Origin, Extent, and Future of Life: Philosophical, Ethical, and Theological Perspectives*, edited by C. M. Bertka, 167–185. Cambridge: Cambridge University Press.
Ferguson, K. 2004. *Tycho & Kepler, the Unlikely Parthership that Forever Changed Our Understanding of the Heavens*. New York: Walker & Company.
Fleischaker, G. R. 1990. “Origins of life: An operational definition.” *Origins of Life and Evolution of the Biosphere* 20:127–137.
Føllesdal, D., L. Walløe, and J. Elster. 1988. *Argumentasjonsteori, språk og vitenskapsfilosofi*. Oslo: Universitetsforlaget.
Forterre, P. 2010. “Defining life: The virus viewpoint.” *Origins of Life and Evolution of the Biosphere* 40:151–160.
Fry, I. 2009. “Philosophical aspects of the origin-of-life problem: The emergence of life and the nature of science.” In *Exploring the Origin, Extent, and Future of Life: Philosophical, Ethical, and Theological Perspectives*, edited by C. M. Bertka, 61–79. Cambridge: Cambridge University Press.
Garber, S. J. 1999. “Searching for good science: The cancellation of NASA’s SETI program.” *Journal of The British Interplanetary Society* 52:3–12.
Gayon, J. 2010. “Defining life: Synthesis and conclusions.” *Origins of Life and Evolution of the Biosphere* 40:231–244.
Ghiselin, M. T. 1995 “Ostensive definitions of the names of species and clades.” *Biology & Philosophy* 10:219–222.

Gingerich, O. 2006. “Astronomen.” In *Att låta själen flyga mellan himlens tinnar – Tycho Brahe och Renässansen*, edited by H. Håkansson. Stockholm: Atlantis.
Goodpaster, K. 1978. “On being morally considerable.” *Journal of Philosophy* 22:308–325.
Harris, J. F. 1986. “Language, language games and ostensive definition.” *Synthese* 69:41–49.
Hazen, R. M. 2009. “Emergence and the experimental pursuit of the origin of life.” In *Exploring the Origin, Extent, and Future of Life: Philosophical, Ethical, and Theological Perspectives*, edited by C. M. Bertka, 21–46. Cambridge: Cambridge University Press.
Hegde, N. R. 2009. “Reasons to include viruses in the tree of life.” *Nature Reviews Microbiology* 7:615.
Herrero-Uribe, L. 2011. “Viruses, definitions and reality.” *Revista de Biologia Tropical* 59:993–998.
Högnäs, S. 2000. *Idéernas historia – En översikt*, Lund: Historiska Media.
Hoskin, M. 1999. “From geometry to physics: Astronomy transformed.” In *The Cambridge Concise History of Astronomy*, edited by M. Hoskin, 94–124. Cambridge: Cambridge University Press.
—. 2000. “From geometry to physics: Astronomy transformed.” In *Cambridge Illustrated History of Astronomy*, edited by M. Hoskin, 98–141. Cambridge: Cambridge University Press.
Jager op Akkerhuis, G. A. J. M. 2010. “Towards a hierarchical definition of life, the organism, and death.” *Foundations of Science* 15:245–262.
Johansson, K., et al. 1978. *Ensamma i rymden?* Uppsala: Studieförlaget.
Johnson, L. E. 1991. *A Morally Deep World: An Essay on Moral Significance and Environmental Ethics*, Cambridge: Cambridge University Press.
—. 1992. “Toward the moral considerability of species and ecosystems.” *Environmental Ethics* 14:145–157.
Jönsson, Å. 2004. *Tycho Brahe: Världsmedborgaren från Ven*. Lund: Historiska Media.
Kant, I. 1998. *Groundwork of the Metaphysics of Morals*. Cambridge: Cambridge University Press.
Kaufmann, W. J. III. 1994. *Universe*. New York: W. H. Freeman and Company.
Kaufmann, W. J. III, and Neil F. Comins. 1996. *Discovering the universe*. New York: W. H. Freeman and Company.
Kolb, V. M. 2010. “On the applicability of dialethism and philosophy of identity to the definition of life.” *International Journal of Astrobiology* 9:131–136.

Kotarbinska, J. 1960. "On ostensive definitions." *Philosophy of Science* 27:1–22.

Lemonick, M. D. 2011. "ET, call us – just not collect." *Time* Apr. 28. http://www.time.com/time/health/article/0,8599,2067855,00.html

Leopold, A. 1970. *A Sand County Almanac*. New York: Ballantine Books.

Levy, S. 1992. *Artificial Life: The Quest for a New Creation*. London: Cape.

López-Garcia, P., and D. Moreira. 2009. "Yet viruses cannot be included in the tree of life." *Nature Reviews Microbiology* 7:615–617.

Losee, J. 1980. *A Historical Introduction to the Philosophy of Science*. Oxford: Oxford University Press.

Lübcke, P. 1988. *Filosofilexikonet*. Stockholm: Forum.

Ludmir, E. B., and L. W. Enquist. 2009. "Viral genomes are part of the phylogenetic tree of life." *Nature Reviews Microbiology* 7:615.

Margulis, L., and D. Sagan. 1995. *What Is Life?* New York: Simon & Schuster.

Maurel, M.-C., and G. Zaccai. 2001. "Why biologists should support the exploration of Mars." *BioEssays* 23:977–978.

Moreira, D. & P. López-Garcia. 2009. "Ten reasons to exclude viruses from the tree of life." *Nature Reviews Microbiology* 7:306–311.

NASA. 2003. *Astrobiology Roadmap*. http://astrobiology.arc.nasa.gov/roadmap

Navas-Castillo, J. 2009. "Six comments on the ten reasons for the demotion of viruses." *Nature Reviews Microbiology* 7:615.

Noble, C. E. 1975. "On Johnson's paradox: Hypothesis verification." *The American Journal of Psychology* 88:537–547.

Pap, A. 1950. "Ostensive definition and empirical certainty." *Mind* 59:530.

Pearson, H. 2008. "'Virophage' suggests viruses are alive." *Nature* 454:677.

Pedersen, O. 2003. *Tycho Brahe og astronomiens genfødsel*. Århus: Steno Museets Venner.

Perlman, B. A. 1982. "Letters." *Cosmic Search* 13:35.

Persson, E. 2008. *What is Wrong with Extinction?* Lund: Lund University.

—. 2012. "The moral status of extraterrestrial life." *Astrobiology* 12(10):976–984.

Plumwood, V. 1991. "Ethics and instrumentalism: A response to Janna Thompson." *Environmental Ethics* 13:139–149.

Popa, R. 2010. "Necessity, futility and the possibility of defining life are all imbedded in its origin as a punctuated-gradualism." *Origins of Life and Evolution of the Biosphere* 40:183–190.

Popper, K. 1959. *The Logic of Scientific Discovery*. New York: Basic Books.

Radford, C. 1964. "Ostensive definitions, coordinative definitions, and necessary empirical statements: A reply to Arthur Pap." *Mind* 73:270–272.

Raoult, D. 2009. "There is no such thing as a tree of life (and of course viruses are out!)" *Nature Reviews Microbiology* 7:615.

Rawls, J. 1973. *A Theory of Justice*. Oxford: Oxford University Press.

Regan, T. 1983. *The Case for Animal Rights*. Berkeley CA: University of California Press.

—. 2001. *Defending Animal Rights*. Urbana IL: University of Illinois Press.

Retana-Salazar, A. P., and S. Retana-Salazar. 2004. "Towards a simple logic in the determination of biological groups: The species and supraspecific groups." *Revista de biologia tropical* 52:19–26.

Rio Declaration on Environment and Development. The Earth Summit Rio de Janeiro 1992.

Rolston III, H. 1986. "The preservation of natural value in the solar system." In *Beyond Spaceship Earth: Environmental Ethics and the Solar System*, edited by E. C. Hargrove, 140–182. San Francisco CA: Sierra Club Books.

—. 1987. "Duties to ecosystems." In *Companion to A Sand County Almanac – Interpretive Critical*, edited by J. B. Callicott, 246–274. Madison WI: The University of Wisconsin Press.

—. 1988. *Environmental Ethics – Duties to and Values in the Natural World*. Philadelphia PA: Temple University Press.

—. 1994. *Conserving Natural Value*. New York: Columbia University Press.

—. 1999. "Ethics on the home planet." In *An Invitation to Environmental Philosophy*, edited by A. Weston, 107–139. Oxford: Oxford University Press.

Ruiz-Mirazo, K., et al. 2004. "A universal definition of life: Autonomy and open-ended evolution." *Origins of Life and Evolution of the Biosphere* 34:323–346.

—., et al. 2010. "Defining life or bringing biology to life." *Origins of Life and Evolution of the Biosphere* 40:203–213.

Ruthrof, H. 2005. "The semiotics of intercultural exchange: Ostensive definition and digital reason." *Semiotica* 157:387–410.

Schrödinger, E. 1944. *What is Life?* Cambridge: Cambridge University Press.

Schulze-Makuch, D., and L. N. Irwin. 2008. *Life in the Universe: Expectations and Constraints*. Berlin: Springer.
Schweitzer, A. 1989. "The ethic of reverence for life." In *Animal Rights and Human Obligations*, edited by T. Regan, and P. Singer, 133–138. Englewood Cliffs NJ: Prentice Hall.
Shostak, S. 2009. *Confessions of an Alien Hunter*. Washington DC: The National Geographic Society.
Sikö, A. 1999. *Efter ursmällen*. Lund: Studentlitteratur.
Singer, P. 1993. *Practical Ethics*. Cambridge: Cambridge University Press.
—. 1995. *Animal Liberation*. London: Pimlico.
Smith, K. C. 2009. "The trouble with intrinsic value: An ethical primer for astrobiology." In *Exploring the Origin, Extent, and Future of Life: Philosophical, Ethical, and Theological Perspectives*, edited by C. M. Bertka, 261–280. Cambridge: Cambridge University Press.
Stenholm, B. 1997. "Liv och civilisation – på jorden och bortom." In *Perspektiv på universum*, edited by N.-E. Cerne et al. Malmö: Gleerups.
Strick, J. E. 2009. "From Aristotle to Darwin, to Freeman Dyson: Changing definitions of life viewed in a historical context." In *Exploring the Origin, Extent, and Future of Life: Philosophical, Ethical, and Theological Perspectives*, edited by C. M. Bertka, 47–60. Cambridge: Cambridge University Press.
Taylor, P. W. 1986. *Respect for Nature: A Theory of Environmental Ethics*. Princeton NJ: Princeton University Press.
Thompson, M. 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge MA: Harvard University Press.
Tipler, F. J. 1981. "Extraterrestrial beings do not exist." *Physics Today* 34:9.
—. 1982. "Extraterrestrial intelligence: The debate continues." *Physics Today* 35:34.
—. 1993. "SETI – A waste of time." *SPIE* 28:28–35.
Tirard, S., et al. 2010. "The definition of life: A brief history of an elusive scientific endeavor." *Astrobiology* 10:1003–1008.
Tough, A. 1998. "Positive consequences of SETI before detection." *Acta Astronautica* 42:745–748.
Tsokolov, S. 2010. "A theory of circular organization and negative feedback: Defining life in a cybernetic context." *Astrobiology* 10:1031–1042.

Weber, B. H. 2010. “What is life? Defining life in the context of emergent complexity.” *Origins of Life and Evolution of the Biosphere* 40:221–229.

Whiteley, C. H. 1956. “Meaning and ostensive definition.” *Mind* 65:332–335.

Williams, D. C. 1936. “Analysis, analytic propositions and real definitions.” *Analysis* 3:75–80.

Wittendorff, A. 1994. *Tyge Brahe*. Copenhagen: Gads Forlag.

Wittgenstein, L. 2009. *Philosophical Investigations*. Hoboken: Wiley-Blackwell.

Zhuravlev, Y. N., and V. A. Avetisov. 2006: “The definition of life in the context of its origin.” *Biogeosciences* 3:281–291.

CHAPTER TWO

ASTROCOGNITION: A COGNITIVE ZOOLOGY APPROACH TO POTENTIAL UNIVERSAL PRINCIPLES OF INTELLIGENCE

MATHIAS OSVATH

Introduction: An astrobiological perspective on cognition

Astrobiology is a successful and maturing multidisciplinary field. In essence, it concerns itself with the universal principles of life – its origin, boundaries and potential expressions – to reason about extraterrestrial life and the possible signs of it. What sets it apart from more traditional biology is the broader questions it asks and the cosmological data it incorporates.

Finding life forms of any kind beyond Earth's biosphere would be one of the greatest scientific breakthroughs ever. At the same time, such a discovery would be dwarfed by any encounter with *intelligent* extraterrestrial life. Despite that the holiest of all astrobiological grails is non-earthly intelligence, cognitive science, the discipline studying intelligence, has been very poorly represented within astrobiology.

There could be several reasons for this dearth of cognitive perspectives. For example, it might appear misguided to focus on something that took roughly more than three billion years to evolve after the origin of life on Earth. Before we even know if there *is* life out there – in any form – it might seem pointless to consider cognition, which could well be a freak accident of life. Nevertheless, the unknown probability for extraterrestrial cognition is not a good reason for not engaging in an astrocognitive endeavour. Even if it somehow turns out that the rest of the universe is devoid of life, an astrobiological approach to cognition will not have been in vain. As will hopefully be seen in this chapter, it should teach us something about the principles of terrestrial cognition; it will incline us to think more deeply about what cognition is and how it arises.

There may be another reason for the lack of cognitive science in astrobiology: cognitive science has only relatively recently taken up a serious interest in biology. Cognitive evolution is now studied systematically, within the field of cognitive zoology – more commonly referred to as comparative or animal cognition, or cognitive biology. It is perhaps needless to say, but a biological framework is essential to astrobiology. In other words, the time is ripe for astrocognition.

There is little doubt that planetary conditions heavily influence the evolution of cognition. Sensorimotor processes, which are at the base of any cognitive system, function in accordance with gravity, lighting conditions, radiation, atmospheric density, and so on. The Earth has a 24-hour cycle of day and night. Most places on Earth have seasonal variations; they experience weather conditions of varying sorts. Such variation influences what types of biota can arise (and, indeed, the biota influence the variations). That, in turn, determines what types of cognition can arise within the various possible niches (and, indeed, the cognition influence the biota). Correlating terrestrial conditions with types of cognition throughout Earth's history raises many intriguing questions. A full-blown astrocognitive research program must study systematically the conditions necessary for cognition writ large and the expression of different cognitive adaptations, encompassing e.g. the origins of the nervous system, along with learning mechanisms, memory systems, attention, emotion, motivation, communication, and so on.

Astrobiology faces the daunting challenge that its subject of study – life – has only a single instance available for observation. This makes it extremely difficult to predict the likelihood of other instances elsewhere. Nevertheless, within this sole available instance, one finds abundant examples of biological expressions that have arisen, independently, time and time again: an evolutionary process called homoplasy, by which similar traits evolve in different species, independently of their common ancestor, in which no comparable trait was present. One of the most obvious examples is self-propelled flight, which has evolved independently in birds, reptiles (in now-extinct forms), mammals, and insects. Other examples include camera eyes, myelin, fingerprints, silk production, and electrical perception. Homoplasy usually comes about either through convergence or parallelism, the details of which need not concern us here. The point is that, given the necessary preconditions for life, selective environmental pressures can, and often do, produce highly similar traits. Homoplasy affords speculation on the ways the expression of life varies in relation to environmental circumstances. Its numerous examples have even been used to argue for the predictability of biological evolution (e.g. Conway-Morris, 2010).

Cognitive traits may also be subject to homoplasy. Arguably, they are more difficult to define than other traits; but they can be identified from a purely functional perspective: i.e., what the cognitively induced behaviour does. So cephalopods (e.g. octopuses and squids), which belong to the molluscs, seem to have learning and memory capacities similar to those of vertebrates (e.g. Mather, 2012). Object permanence – understanding that an object does not disappear when outside the sensory scope – is found in many mammals and birds but not in any reptiles, even though all three groups share a last common ancestor. I could provide further examples; but I wish to focus on what homoplasically evolved cognition can tell us about the ubiquitous term "intelligence": what we think intelligence is, what it could look like if found elsewhere, and how it arises.

Cognitive scientists might differ somewhat from astrobiologists in their view on intelligence: e.g., cognitive scientists are often reluctant to talk of intelligence in a broad sense. Probably even the way intelligence is used as a non-technical, everyday, term differs between representatives of the two disciplines. Finding an extraterrestrial equivalent of the Serengeti Plain – bursting with the analogues of ungulates, big cats, wild dogs, and birds – would, for the cognitive zoologist, present a boiling pot of complex cognition – and if pressed to give an everyday definition, the cognitive zoologist would most likely regard this as extraterrestrial intelligence. It is less clear what the equally awestruck astrobiologist would think; perhaps she prefers another of the common meanings of intelligence that includes well-developed mathematical or intra-species communication skills. So, the Serengeti Plain might be an example of complex life, but not of intelligence. Intelligence can be many things, even on Earth.

The terrestrial environment presents several independently evolved instances of complex cognition: what, from an anthropocentric perspective, one may call intelligence; remembering that, in the end, an anthropocentric perspective is the only one, one has when scanning the Earth or other planets for intelligence. That there are these instances might be taken to suggest that, given appropriate preconditions, complex cognition is not such an unlikely phenomenon in the universe. That provides strong motivation to look for those preconditions in extraterrestrial environments.

This chapter presents one possible framework for an astrocognitive program: one that raises more questions than provides answers. Along the way, a few theoretical corners are cut to make things more clear.

Thoughts on intelligence and how it relates to animals

The common view within cognitive zoology is that cognition is modular (e.g. Shettleworth, 2010). After all, cognition clearly evolves, and different species display different cognitive adaptations, which are easy to think of as modules. Intelligence, on the other hand, is traditionally regarded as a general cognitive capacity often referred to as the *g*-factor. It reflects a well-developed cognition in several, seemingly unrelated, domains: e.g. physical vs. social reasoning. The difficulty with the term "intelligence" is that it initially implied *human* capacities. It is typically measured in IQ-tests whose results are compared to the mean performance of a given *human* population; being intelligent means performing better than the average human. Measures of intelligence are measures between individuals within a species, not between species. Consequently, the term is usually avoided in the comparative perspective in cognitive zoology (although it is occasionally used in the search of a *g*-factor in a specific species). As will be evident, debate continues over whether certain forms of advanced cognition share underlying principles – principles that one might use to define what makes a species intelligent in a truly general and comparative sense.

Despite the reluctance among cognitive scientists to use the term when comparing species, most people – lay persons and researchers alike – agree that some species are (in some sense) more intelligent than others. Few would for example disagree on that a dog is vastly more intelligent than a slug. Of course, the search for extraterrestrial intelligence is in large measure the search for cognition, toward which a mildly cognitive space slug would, indeed, be a grand finding, but a celestial dog would be a far greater one. An obvious difference between slugs and dogs is that a dog has a wider array of cognitive capacities, or modules. The dog's brain has, by any measure, greater processing power, being larger in both absolute and relative size. All in all, the dog's behaviour is "more cognitive", making it more intelligent in at least one everyday sense of the word.

Another way to try to quantify cognition is to talk about degree of flexibility in an animal's behavioural repertoire. Those animals we like to view as intelligent are those exhibiting – among other things – high flexibility. However, again, flexibility implies some form a generality in cognition. And, admittedly, it is not easy to agree on an objective measure of flexibility. So bluntly put, those animals we think of as intelligent are those whose cognition is similar to our own in the domains we are inclined to fancy. Our cognition is tuned to recognize human-like cognition. It is quite conceivable that some complex behaviours pass under our radar; we fail to identify them because we do not share a similar set up of adaptations. We do, for example, not perceive magnetism nor electrical fields nor ultraviolet light. We have few clues on what it is like to produce our own perception, as e.g. cetaceans do through a cross-sense-modality integrated sonar, and what this would do for our cognition and communication. In any case, there is one seemingly objective measure, however crude, which correlate with those animals we think are smart: a large relative brain size. However, we do not precisely know what a big brain does that a smaller one does not, but we are acutely aware of that we ourselves have the biggest relative brain size that ever existed on Earth.

Such anthropocentrism often raises genuine concerns for cognitive zoology. Nevertheless, it need not pose too many difficulties for an astrobiological approach to intelligence – to the contrary. After all it is *we* who are looking for intelligence. If we find it, we must be able to recognize and appreciate it. Intelligence, to be intelligence, must be meaningful as intelligence, for *us*. Here, the astrocognitive search for intelligence diverges from the broader search for extraterrestrial cognition, since cognition need not be particularly overall human-like for us to recognize it as cognition. The latter search may well require a different set of tools from the former. However, this chapter concerns what *we* find intelligent – something that we now think has evolved several times here, on Earth.

Intelligent animals, and possible universals in the content and evolution of intelligence: Gouldian reruns

In the early 1980s, the famous palaeontologist Stephen J. Gould asked rhetorically whether intelligent life would evolve again if one could rerun the evolutionary tape, and whether intelligence would evolve on other planets sharing the same basic conditions (Gould, 1983). In the 80s, we were all but oblivious that more than one intelligent species existed – and exists – on Earth.

Over the past thirty years, our knowledge about the minds of other species has increased tremendously. For some time it has been understood that the great apes – chimpanzees, bonobos, gorillas, and orangutans – have a complex cognitive repertoire similar to our own. They seemingly plan for the future, solve problems through insight, grasp causality, innovate, possess a culture, understand the psychological states of others – and much, much more (see e.g. Seed and Tomasello, 2010). If their cognitive feats are impressive, we should not be surprised: they are our closest living relatives, with a long common history of selective pressures before our lineages diverged. Our similarities reflect our shared past.

Far more remarkable are the findings in recent years regarding the cognition of corvids (crow birds), parrots, and dolphins. When it comes to understanding their physical and social world, these groups have forms of cognition functionally highly similar to that the great apes: forms of cognition that were thought of as exclusively human not long ago. No complete, comparative catalogue of these groups' cognition exists as yet – not even for the great apes. And, we know more about some groups than others. Nevertheless, the discoveries to date are noteworthy. For all their evolutionary separation, these groups seem to share more similarities in certain of their complex cognitive capacities than they share with their closest living out-group relatives. They also have the largest brains in relation to body size of all animals, despite their sometimes radically different brains. Dolphin researcher Louis Herman thought that the cetaceans are so similar to us and the other great apes that he adopted them to our family as "cognitive cousins" (Herman, 2002). And, corvidologist Nathan Emery was so struck by crows' cognitive resemblance to hominoidea that he has called them "feathered apes" (Emery, 2006).

All these groups are but distantly related. The last common ancestor of the dolphins and great apes is around 90 million years ago, around the same time that the corvids and parrots share their last common ancestor. Meanwhile, birds and mammals have their last common ancestor as far back as 300 million years ago. These time scales are vast: birds and mammals have been subjected to 600 million years of separate evolution. This is roughly the same amount of time since the first animals appeared. Such time scales provide the ideal conditions for discovering the principles that underlie cognition in general and complex cognition in particular. Only a multiple-origin story can account for what these groups have in common with their complex cognition but do not share with their closest out-group relatives.

Behavioural similarities

The obvious question arises: what precisely *is* similar in the cognition of these groups? With the caveat that not all abilities have yet been tested in all groups, nevertheless, certain patterns emerge. However, before speculating about those patterns, it will be useful first to present a list – incomplete for reasons of space – of the similar behaviours one finds (for more comprehensive behavioural catalogues, see e.g. Seed and Tomasello, 2010; van Horik *et al*., 2012; Jaakkola, 2012).

Consider first the domain of social cognition: these groups are able to infer what others will do given the others' perception or state of knowledge. They can take the orientation of an animal's eyes and deduce something about what it perceives. They also show a high level of affective and affiliate behaviours, such as consolation and reconciliation. And, they cooperate in elaborate ways not found, or not commonly found, outside these groups. On the other side of the coin, they master deception and gain from fooling others. In essence, in an array of situations, they seem to show evidence of a *theory of mind*: i.e., the understanding of that the psychological state of the other causes its behaviour.

Consider next the domain of physical cognition. These groups show a flexible ability to use objects as tools in a way that they seem to understand the relevant properties and possible functions of the objects in relation to one another. They are also able to use what looks like insight when solving certain physical problems. These animals show the highest rates of innovation in nature. In general they seem to have a basic understanding of causality, instead of only using simple, associatively learned, correlations between physical events.

In the combination of the social and physical cognitive domains one sees an unusually level of culture. For example, innovations are spread over generations. Another example of abilities in the combination of the domains: corvids and apes have the ability to remember and plan for events, relative to their point of view, that lie beyond the current input of their senses and their current psychological state (abilities which in humans are based on the episodic system of perceptual simulations). These abilities have not yet been tested for in the other groups, although dolphins are known to have elaborate planning abilities within the immediate context.

I conclude this much abbreviated list with what appears, from an ecological standpoint, an odd ability: the ability to recognize oneself in a mirror. This has often been taken as evidence for some sort of self-awareness (which could be tied to social cognition).

Can one find any underlying cognitive principles for these groups' behaviours?

By now, any cognitive scientist reading this is likely to be objecting loudly. The theoretical issues surrounding the abilities described above have stirred heated debates. The disagreement boils down to whether the animals *really* have e.g. causal reasoning, theory of mind, self-conscious awareness, planning abilities, etc.; or whether the behaviours result, instead, from much developed learning abilities or other “low-level” mechanisms (for reviews of these debates, see e.g. Call and Tomasello, 2008; Penn and Povinelli, 2007; Lurz, 2011). And, a considerable possibility arises that these groups are relying on different aspects of, or knowledge, about the world, aided by different sense modalities, to reach similar functional outcomes. However, here I evade these questions by a simple astrocognitive slip through the Gordian knot. What matters is that these groups *appear* to have these cognitive abilities in common; even to the extent we are still arguing over it. Few animals even appear to have such abilities. What these groups have, and many others seem to lack, is the ability (on whatever cognitive foundation) to take in certain features of the world that are hidden from immediate perception. However, many animals outside these groups have behaviours that seem to require cognitive processing of what cannot be inferred directly from sensory input. The cat does not believe that the mouse magically vanishes when it hides under the couch. So, there is more to it.

What happens in these groups is that the ability becomes a bit more elaborate, undergirding both social and physical cognition. One may call this more sophisticated ability theory-like reasoning or second- (or higher-) order representation (see e.g. Bluff *et al*., 2007; Shettleworth, 2010). This is what allows humans to excel both in social and physical contexts. It may best be described through examples.

Consider the ability to respond to cause and effect: to reason about them, one must have some understanding – some “theory” – of causation. The eighteenth-century British philosopher David Hume famously noted the lack of any logical entailment between two pool balls colliding on a pool table and their subsequent trajectories: the causal relationship is not directly perceivable but only inferred (Hume, 1739). Nothing excludes the possibility that, on any particular occasion, one of the balls will rocket to the ceiling or not move at all (in which case, we would immediately theorize about the cause for that odd behaviour). All one directly observes is that certain positions and movements of the balls reliably recur. In similar fashion, one cannot observe the cognitive states in others that cause their behaviour; one can only observe the results. To grasp those cognitive states or predict future behaviours that may result from them, one requires a theory about that cognitive state: a theory of mind. This does not apply only to other’s cognitive or psychological states but also to one’s own: e.g., “I don’t remember how to do this; therefore as much as I try I am not going to be able to solve this task” or “I’m not hungry now, but I will be later, so I’d best bring that hammer along so I can crack coconuts.”

Having a system that in many circumstances appears like a theory of causation and of mind is quite economical: one need not have experienced all the events in detail that one is trying to predict; one is better equipped to differentiate between mere correlation and causation. It makes one more flexible.

Such apparent theory-like reasoning probably is not the whole story of full-scale intelligence (in the everyday sense). At the same time, it does seem to be a crucial prerequisite for a range of other cognitive mechanisms. More detailed accounts will probably be, as it were, but icing on the cake. In conclusion: an extraterrestrial lifeform whose behaviour suggests theory-like reasoning is one we would, almost certainly, regard as intelligent.

Possible conditions for intelligence

We know that these cognitive adaptations are not the result of common descent but rather selective pressures that has existed in the lineages of each group. This opens up possibilities for evolutionary "reverse engineering", enabling us to unearth some of the environmental factors needed to develop intelligence.

Detailing the selective pressures is as vast an undertaking as understanding the finer points of the cognitive mechanisms involved. Some initial attempts have been made to map socio-ecological and other factors that these animals have in common, apart from their intelligence. It remains difficult to discern which of these factors are the result or cause of intelligence, or if they are part of co-evolutionary processes ratcheting up intelligence to ever complex levels.

On the individual level, these animals are characterized by a long lifespan and prolonged maturation period. This is typically explained by the requirements placed on the animal by a proportionally large brain. Brains are very costly in calories consumed per unit of brain mass: brains need a constant flow of energy to which they are especially sensitive during periods of growth. What this means is that the animal must be able to obtain high-quality nutrients irrespective of seasonal or other environmental changes. If a large brain is evolutionarily selected for, it must be because it leads to higher fitness: i.e., offspring who themselves reproduce successfully. Growing this large brain into maturity takes time, and in order to increase fitness the lifespan must be prolonged. The slowly developing brain relies on extended care from parents and, sometimes, older siblings (for a comprehensive account of this so-called life-history filter see e.g. van Schaik and Deaner, 2003). A long maturation period, characterized by a growing brain with significant plasticity, has various knock-on effects. In particular, it assists in learning additional survival skills, which in turn increases fitness.

It is not entirely trivial to separate cause from effect when discussing longer lifespans and bigger brains. With its propensity for higher intelligence, a larger brain might be what is needed to buffer out changes in food availability (for a detailed account of the so-called cognitive buffer hypothesis see e.g. Sol, 2009). On the other hand, an environment that is demanding on a recurring seasonal basis could also be addressed by a faster, shorter lifespan providing only a short period for brain development, as is the case in several species.

Meanwhile, a stable, resource-rich environment would also not, in itself, favour expensive cognitive adaptations that are not immediately needed. However, such an environment does relax certain constraints on brain growth, and some species in such environments *do* have larger brains than close relatives living under harsher conditions: e.g., aye-ayes compared to other lemurs; but not on nearly the same scale as those intelligent animals I have been discussing. The bottom line is that, even if a more changeable environment provides important conditions for the evolution of complex cognition, it is by no means a foregone conclusion that a complex or changing environment will produce intelligence.

Another salient characteristic of these groups is a more-or-less advanced social life at some, or all, stages of life. Indeed, the *social brain hypothesis* – the idea of social life as a cause of intelligence – is, in one version or another, currently the best supported explanation of intelligence evolution across these groups (see e.g. Jolly, 1966; Humphrey, 1976; Byrne *et al*., 1988; Dunbar, 1998). All variations have in common the notion that, among all the environments in which an agent exists, the social environment is particularly cognitively demanding – consisting as it does of other cognitive beings, who constitute a complex web to which one must relate if one is to increase one's fitness. Such an environment makes fertile ground for many a cognitive arms race and various ratcheting effects. It generates selective pressures both for outsmarting others and for aligning with them in cooperative efforts. This is not to say that just any social environment will do: herd animals along with many fish and insects live in sometimes huge congregations without showing any signs of particularly well-developed individual intelligence: they seem to rely on comparatively simple rules for how to act, given the behaviour of their nearest neighbours.

In contrast, the groups I have been discussing – among others – live in groups and communities within which they reliably identify and re-identify each other as individuals. They are aware of kinship and other affiliative group relations; they show a keen understanding of dominance hierarchies. These are commonly taken to be among the basic conditions for developing complex cognition.

In any of its variations, the social brain hypothesis faces problems in accounting for these groups. Again, other species exhibit this type of sociality seemingly without reaching the same level of intelligence; even though e.g. wolves and baboons are viewed as rather intelligent animals, it is on a different order to the great apes, corvids, dolphins, and parrots, who appear to have domain-general cognitive abilities. They are skilled both in social and physical understanding.

The complex environment *per se* – social or otherwise – can only be part of the story. Van Schaik and colleagues advocate what they call the *cultural intelligence hypothesis*, which attempts to combine some of the important characteristics of the other approaches to the evolution of intelligence (for review see van Schaik *et al*., 2012). This approach is also based on sociality; its distinctive emphasis is on transmission of skills, through learning, over generations. Learning abilities in the broad sense are key. The intelligent animal must be able to learn from its physical environment and then innovate based on what it has learned. A solitary individual, acting on its own, cannot reach the level of flexibility seen in the large-brained animals I have been discussing. Learning *must* spread out into the social sphere, allowing innovations to pass between individuals: yet another example of a ratcheting effect, as social and individual learning interact in synergistic fashion, as the learning skills are probably neurologically closely related.

Van Schaik and colleagues do not discuss what underlying cognitive adaptations are needed to make learning effective. Part of the answer might lie in the difference between rote learning and "understanding" what you are learning: learning guided by some type of understanding makes it more efficient, especially when it comes to physical and behavioural innovations which can be developed over the course of generations. Being able to focus on what is more and what is less relevant in any learned task – what amounts to theory-like reasoning on the underlying causation of events and individuals – surely gives an advantage to solving similar tasks, affording various cognitive short cuts.

Van Schaik and colleagues' hypothesis – along with the connection I have suggested to theory-like reasoning – assumes a certain type of culture. On closer examination, their reasoning appears somewhat circular, given their claim that their hypothesis applies only to animals with domain-general cognitive abilities. The causal chain of evolution is, again, not entirely obvious.

At the end of the day, despite these different hypothesis which all have empirical support, we still do not know precisely under what conditions evolutionary processes lead to complex cognition and intelligence. Nevertheless, it seems reasonable to assume that our intelligent extraterrestrial possesses both something similar to theory-like reasoning and some sort of culture, given the way the two appear to go hand in hand in the terrestrial environment. Note that culture – or, at least, a rudimentary social transmission system – need not necessarily imply theory-like reasoning. Culture may be possible without theory-like reasoning; theory-like reasoning, on the other hand, probably does co-evolve with culture.

Miocene: The cradle of intelligence?

Since the appearance of the first animals around 600 million years ago, the maximum ratio of brain to body size has increased exponentially. According to Russell (1983), this reflects an overall trend toward increase in relative brain size – a trend in which the appearance of the human brain fits quite well. On the assumption that larger brains produce more complex cognition, Russell implies that complex cognition was – given the terrestrial environment – bound to happen. The idea is intuitively appealing.

To play devil's advocate, the impressive increases in relative brain sizes could well be the incidental by-product of chance events. Using hindsight on a single system to derive principles can be treacherous. Without re-running the whole evolution of animalia – *per* Gould's suggestion – it is difficult to adjudicate just how universal or inevitable the pattern is. Perhaps, instead, one should take a narrower, more modest, geological perspective on what happened among the groups I have been discussing – all of whom have more or less the same relative (large) brain size. This would offer a miniature version of a Gouldian re-run.

Interestingly, the brains of the great apes, corvids, and dolphins seem to have reached their modern proportions at roughly the same time: around the middle of Miocene, about 15–18 million years ago. The relevant dates are necessarily imprecise and depend on the measures used; nevertheless, on a geological scale, they are effectively simultaneous. So e.g. Marino and colleagues (2004), working with the fossil remains of toothed whales, showed that Delphinoidea achieved its modern brain size around 15 million years ago.

The case of the corvids is based on genetic deductions: bird bones, being hollow and delicate, do not fossilize easily. Jønsson and colleagues

(2012) conclude from genetic analyses that corvids of the genus *Corvus* – all of which have essentially the same relative brain size – had a last common ancestor approximately 17 million years ago. From this one can reasonably conclude that they achieved their common brain size by that point. Corvids not of the genus *Corvus* – e.g., *Pica*, *Garrulus*, and *Nucifraga* – show similar behavioural flexibility and large (though relatively less large) brains. The last common ancestor of all corvids is estimated to have lived around 20 million years ago.

The case of the great apes could in principle be solved both through fossils and genetics. Using the same methodology applied to the *Corvus* corvids, one can conclude that the lesser apes, who have smaller relative brain size, diverged from the great apes around 20 million years ago. The orangutans split off around 18 million years ago (see e.g. Israfil *et al*., 2011) – meaning that the great apes achieved their brain size at roughly this point. Meanwhile, the fossil record on parrots is lacking for the same reason it is for the corvids. The genetic analysis needed to determine when they achieved their brain size remains to be done.

What does all of this tell us? Does it – as I have claimed – constitute a Gould-style re-run of evolution on a miniature scale? Perhaps, one cannot look at these groups in isolation. A lot happened during the Miocene: animals diversified, and several species we are familiar with today took on their modern form. The climate had cooled relative to the preceding Oligocene epoch. Just under 15 million years ago one finds the *Middle Miocene disruption*: a massive wave of extinctions, linked to a major cooling event.

The evidence could lead one to conclude that the bigger brains – or, more accurately, the accompanying advances in cognition – reflected adaptation to a more demanding environment. One might also be tempted, given the uncertainty with all of the dates I have mentioned, that the big brains of these groups were more precisely in response to the Middle Miocene disruption. If that is right, then the relevant causal factor was perhaps not inevitable cognitive evolution as *per* Russell but a chance event. Or, is this again an oversimplification? Could there be some principled elements that react in a predicable fashion – given some pre-conditions at the time – to major geo-climatological events?

At least the dolphins and great apes – and maybe also the corvids – appear to have experienced a previous increase in brain size that cannot be attributed to events in the Miocene. Their ancestors *already* had a larger brain size than body size would have implied. Consider again the toothed whales: their brain size increased radically around 35 million years ago; current speculation implicates the evolution of echolocation. Subsequently though, only Delphinoidae increased its brain size during Miocene; that of the other toothed whales remained the same. Meanwhile among the primates, monkeys already had more residual brain mass than similarly sized mammals – in this case, generally attributed to their relatively complex social life. Subsequently though, only the apes – in particular, the great apes – showed a significant increase in brain size. Whatever selective pressures there were for intelligence during the Miocene, they did not affect the brain size for apes' and dolphins' closest relatives.

We do not know what the Miocene increase in brain size means. Nevertheless, it offers an inviting starting point for investigating the preconditions for that increase. What were the brains and what were the lifestyles of these groups like when the Miocene began? What changed? What new niches did these groups come to occupy? Were the relevant causal factors *only* that of chance events, or did general principles of cognitive evolution also play a role? Perhaps catastrophic chance events only influence the speed of an inevitable cognitive evolution? Or, do they act as necessary ingredients?[1]

According to the best available evidence, it took a little less than 600 million years from the origin of the nervous system until the advent of something that appears as theory-like reasoning – more or less simultaneously in at least four separate lineages. A planet with a similar length of history of animal life, geological periods, and climate events might possibly be expected to harbour intelligent life. What would be helpful to know though is *why* it took Earth 600 million years to arrive at theory-like reasoning. Knowing that might well help in estimating the likelihood of intelligence on other planets based on something more than blind intuition.

The future of astrocognition

I have outlined a possible – potentially fruitful – area for astrocognition research. In doing so, I have perforce greatly simplified matters; had I been able to deliver a highly detailed account, then the research area would already exist. To go from studying the multiple origins of intelligence to studying astrocognition, one must, among other things:

(1) Examine closely the underlying proximal mechanisms behind seemingly similar behaviour across these animal groups. How many ways can intelligent behaviour be produced: i.e., how do differently structured brains arrive at similar solutions? Cognitive zoologists are necessary but not sufficient toward this endeavour; one requires neurobiologists, computationalists, geneticists, and followers of other disciplines as well.

(2) Examine closely the terrestrial ecological factors that have produced intelligent behaviour across these groups. How many different environments, how many different biota can give rise to intelligence? What are the preconditions for intelligence to arise? Views are needed from behavioural ecology, palaeontology, geology and other disciplines.

I am describing here two, overarching approaches. Naturally, they must work in tandem. So far, many of the relevant questions have only been approached in a piecemeal and disparate fashion, if they have received attention at all. One ought to find this surprising, given the importance of complex cognition in humans, and given what a more integrated approach could teach us about it. It becomes all the more surprising given the significance of intelligence in the wider search for extraterrestrial life. I have already suggested one explanation: many of the most important questions could not even be raised until recent empirical findings made them possible. Those findings have opened up a wealth of new methodologies affecting a wide number of disciplines.

A disclaimer: I have offered here only *one* possible astrocognitive program. Intelligence, as I have defined it, is not the only relevant topic for investigation; other questions – e.g., on the origins of cognition – likewise deserve attention. In addressing them, without preconditions, cognitive science takes its rightful place in the future of astrobiology.

Acknowledgements

I would like to thank David Dunér and Joel Parthemore for valuable comments. I also wish to express my gratitude towards Crafoord Foundation for its support.

Notes

[1] See Bogonovich (2011) for a comprehensive analysis of the evolution of encephalisation in the history of animalia. Following Russell (1983), Bogonovich uses an astrobiological framework, and focuses on general brain size increase. This differs from the approach in the current chapter where intelligence has been associated with particular behaviours and animal groups. Bogonovich's preliminary conclusions lean towards a true, predictable, trend for bigger brains in evolution.

References

Bluff, L. A., A. A. S. Weir, C. Rutz, J. H. Wimpenny, and A. Kacelnik. 2007. "Tool-related cognition in New Caledonian crows." *Comparative Cognition & Behavior Reviews* 2:1–25.

Bogonovich, M. 2011. "Intelligence's likelihood and evolutionary time frame." *International Journal of Astrobiology* 10:113–122.

Byrne, R. W., and A. Whiten. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans.* Oxford: Clarendon.

Call, J., and M. Tomasello. 2008. "Does the chimpanzee have a theory of mind? 30 years later." *Trends in Cognitive Sciences* 12:187–192.

Conway-Morris, S. 2010. "Evolution: Like any other science it is predictable." *Philosophical Transactions of the Royal Society B* 365:133–145.

Dunbar, R. I. M. 1998. "The social brain hypothesis." *Evolutionary Anthropology* 6:178–190.

Emery, N. 2006. "Cognitive ornithology: The evolution of avian intelligence." *Philosophical Transactions of the Royal Society B* 361:23–43.

Gould, S. J. 1983. "The wisdom of Casey Stengel." *Discover*, March: 62–65.

Horik, J. O. van, N. S. Clayton, N. J. Emery. 2012. "Convergent evolution of cognition in corvids, apes and other animals." In *The Oxford Handbook of Comparative Evolutionary Psychology*, edited by J. Vonk, and T. K. Shackelford, 80–101. Oxford: Oxford University Press.

Herman, L. M. 2002. "Exploring the cognitive world of the bottlenose dolphin." In *The Cognitive Animal*, edited by M. Bekoff, C. Allen, G.M. Burghardt, 275–283. Cambridge MA: MIT Press.

Hume, D. 1978 [1739]. *A Treatise on Human Nature*. Oxford: Clarendon.

Humphrey, N. K. 1976. "The social function of intellect." In *Growing Points in Ethology*, edited by P. P. G. Bateson, and R. A. Hinde, 303–317. Cambridge: Cambridge University Press.

Israfil, H., S. M. Zehr, A. R. Mootnick, M. Ruvolo, and M. E. Steiper. 2011. "Unresolved molecular phylogenies of gibbons and siamangs (Family: Hylobatidae) based on mitochondrial, Y-linked, and X-linked loci indicate indicate a rapid Miocene radiation or sudden vicariance event." *Molecular Phylogenetics and Evolution* 58:447–455.

Jaakkola, K. 2012. "Cetacean cognitive specializations." In *The Oxford Handbook of Comparative Evolutionary Psychology,* edited by J.

Vonk, and T. K. Shackelford, 144–165. Oxford: Oxford University Press.

Jolly, A. 1966. “Lemur social behavior and primate intelligence.” *Science* 153:501–506.

Jønsson K. A., P. H. Fabre, and M. Irestedt. 2012. “Brains, tools, innovation and biogeography in crows and ravens.” *BMC Evolutionary Biology* 12: doi:10.1186/1471-2148-12-72

Lurz, R. W. 2011. *Mindreading Animals – The Debate over What Animals Know about Other Minds*. Cambridge MA: MIT Press.

Marino, L., D. W. McShea, and M. D. Uhen. 2004. “Origin and evolution of large brains in toothed whales.” *The Anatomical Record Part A* 218A:1247–1255.

Mather, J. A. 2012. “Cephalopod intelligence.” In *The Oxford Handbook of Comparative Evolutionary Psychology*, edited by J. Vonk, and T. K. Shackelford, 118–128. Oxford: Oxford University Press.

Penn, D. C., and D. J. Povinelli. 2007. “Causal cognition in human and nonhuman animals: A comparative, critical review.” *Annual Review of Psychology* 58:97–98.

Russell, D. A. 1983. “Exponential evolution: Implications for intelligent extraterrestrial life.” *Advanced Space Research* 3:95–103.

Seed, A., and M. Tomasello. 2010. “Primate cognition.” *Topics in Cognitive Science* 2:407–419.

Shettleworth, S. J. 2010. *Cognition, Evolution and Behavior*. Oxford: Oxford University Press.

Sol, D. 2009. “The cognitive-buffer hypothesis for the evolution of large brains.” In *Cognitive Ecology II*, edited by R. Dukas, and J. M. Ratcliffe, 111–134. Chicago IL: The University of Chicago Press.

Van Schaik, C. P., and R. O. Deaner. 2003. “Life history and cognitive evolution in primates.” In *Animal Social Complexity*, edited by F. B. M. de Waal, and P. L. Tyack, 5–25. Cambridge MA: Harvard University Press.

Van Schaik, C. P., K. Isler, J. M. Burkart. 2012. “Explaining brain size variation: from social to cultural brain.” *Trends in Cognitive Sciences* 16:277–284.

CHAPTER THREE

THE "FINAL FRONTIER" AS METAPHOR FOR MIND: OPPORTUNITIES TO RADICALLY RE-CONCEPTUALIZE WHAT IT MEANS TO BE HUMAN

JOEL PARTHEMORE

Introduction

Contra someone like Roger Penrose and *per* e.g. Douglas Hofstadter, one finds real and, to a certain extent, knowable limitations to human conceptual abilities: there are things we do not know now; there are things we may never know; and there are things we can be fairly certain we can never know, because they fall outside our abilities: so e.g. one can describe the mathematical properties of a tesseract and build a model of a tesseract in three dimensions; but it seems a necessary limitation of being human that no one has ever pictured a tesseract in all its four-dimensional wonder, and no one who is human ever will.

At this stage in the development of the species, the budding fields of astrobiology and astrocognition offer an unmissable opportunity to explore the territory between what we do not know now, what we may never know, and what we *think* we can be certain we can never know. They represent a research area that is, as close perhaps as is possible to get, truly unknown – but which, at least in substantive part, need not stay so. In the absence of evidence, powerful intuitions rule; and otherwise respectable scientists proclaim the near certainty or unlikelihood of finding life elsewhere in the solar system or beyond. Without more substantial evidence, it is impossible to assign *any* probabilities. Largely trapped within our exceedingly narrow Earth–bound perspective, we begin – as we probably must – by looking elsewhere for conditions that foster life here.

Yet we already know from the explosion of research in so-called extremophiles just how many seemingly unlikely places life turns up on Earth. Almost simultaneously, the existence of other planets beyond the solar system has gone from theoretical possibility to known certainty, with new planets being discovered on an increasingly routine basis.

Regardless of whether we find intelligent life among the stars – "intelligent life" I will take as shorthand for "life possessing recognizably human–like intelligence" – the discovery of life in any form (underneath the surface of Mars or inside Europa, we expect to find at most single–celled organisms – but, again, we do not know!) will lead, almost inevitably, to a radical re–conceptualization of both the world – the Earth that has been nearly our entire universe, for nearly our entire existence – and what it means to be human. Our current conceptual frameworks in which we embed our understanding of ourselves are built on two million years of being – until the last fifty years or so – entirely Earth bound. Earth's sun is but one minor yellow star in an outer arm of a galaxy containing, by some estimates, nearly half a *trillion* stars. As we step up the pace of space exploration and the search for extraterrestrial life, our perspective on the galaxy – let alone the universe – of which we are a part cannot help but change fundamentally, yielding up fresh understandings of ourselves and our place in existence. By exploring the frontier that is outer space, we explore what is truly the final frontier: the unmapped territory of the human mind.

Exploring the limits

> …There are grounds for cautious optimism that we may now be near the end of the search for the ultimate laws of nature (Hawking, 1988, p. 172).

Philosophy of science

Philosophy of science presents two longstanding and competing traditions, going back – in one form or another – for centuries. The one, reflected in the quote from Stephen Hawking, sees science as revealing the time- and mind-independent truths of the universe. For any scientific question of substance, there is one, and only one, answer – one that *any* intelligent agent (terrestrial or extraterrestrial) must come to – and finding it is only a matter of time. The other is most strongly associated, perhaps, with Albert Einstein, as the following quote in reference to him suggests:

> In reply to my question what problem he was working on now, he said he was engaged in thinking. Giving thought to any scientific proposition almost invariably brought progress with it. For, without exception, every scientific proposition was wrong. That was due to human inadequacy of thought and inability to comprehend nature, so that every abstract formulation about it was always inconsistent somewhere. Therefore every time he checked a scientific proposition his previous acceptance of it broke down and led to a new, more precise formulation. This was again inconsistent in some respects and consequently resulted in fresh formulations, and so on indefinitely. – from the diaries of Count Kessler, quoted in (Stachel, 1982, p. 96), *emphasis added.*

On the one view, a move away from ignorance is a move closer toward time-independent truth. On the other, truth is a moving target, and the prime benefit of a move away from ignorance is greater awareness of one's own ignorance. On the one view, science is all about revelation, and there are no limits (or no important limits) to what it can reveal. On the other, science is all about exploration: exploring the limits of what is known as well as *what is knowable.*

Philosophy of mind

The often fierce debates in philosophy of science reflect a more fundamental debate in philosophy of mind over the nature of human conceptual understanding and cognition. Does human reason reveal the world, or does it explore a world that ultimately if not already quite rapidly outstrips its capacity to understand? *Are* there limits to what the human mind can know – and not just know but *know* that it knows; or are there places where the human mind can be fairly sure it cannot go: are there things it can know (or reasonably assume) that it *cannot* know?

Although one might be forgiven for thinking that this sounds like so many epistemological word games, it has practical consequences for how one views oneself, one's world, and one's universe. Does the universe amount, in practice, to one's conceptual universe: not to say that mind is world (as a Berkeleyan idealist would have it), but rather that conceptual perspective on the world and world itself do not cleanly pull apart? …Or is conceptual understanding the means by which one transcends the limits of one's conceptual perspective *to take in the world as it really is?*

In *Gödel, Escher Bach*, cognitive scientist Douglas Hofstadter takes the former view; in *The Emperor's New Mind* and *Shadows of the Mind*, physicist Roger Penrose takes the latter. Both men target Kurt Gödel and his Incompleteness Theorem (actually two, closely related, theorems);

Hofstadter would surely agree with this much of what Penrose has to say:

> It rapidly became accepted as being a fundamental contribution to the foundations of mathematics – probably the most fundamental ever to be found – but I shall be arguing that in establishing his theorem, he also initiated a major step forward in the philosophy of mind. Among the things that Gödel indisputably established was that no formal system of sound mathematical rules of proof can ever suffice, even in principle, to establish all the true propositions of ordinary arithmetic (Penrose, 1994, p. 64).

In something approach layperson's terms, what this means is that no *sufficiently expressively powerful* formal system can be both complete and consistent:

> There will be one or more statements, expressible within the system, that cannot be proven within the system;
>
> If the system is extended so as to be able to prove them, then it will, by that very extension, become inconsistent.

The problem is with self–referential statements: the translation into mathematics of Epimenides' Paradox:

> All Cretans are liars.

Epimenides was, of course, a Cretan. Epimenides' Paradox can be expressed more simply and bluntly as "this statement is false".

The question is whether any sufficiently expressively powerful messy, *informal* system like the human mind is limited in this way or not. In *Shadows of the Mind*, Penrose provides a succinct proof (1994, pp. 72–77) – or so he believes – that the human mind is *not* so bounded. In an appendix of my doctoral thesis (Parthemore, 2011a, pp. 192–195), I follow Hofstadter's reasoning and, in particular, the arguments of Daryl McCullough (1995), in analyzing where Penrose goes wrong.

In the end, the upshot is that, *pace* Penrose, what one *knows* that one knows may be somewhat less than what one *thinks* that one knows or actually knows. More to the point, what one knows (regardless of whether one knows that one knows it) depends critically on what one does *not* know: i.e., it depends on one's knowledge being incomplete. This is what Penrose's "proof" inadvertently reveals.

Philosophy of astrobiology

…Which brings me to the subject of this volume: astrobiology, a domain in which human understanding could, seemingly, almost not be *more* incomplete – our perspective almost not *more* myopic. Barring (much disputed) evidence of microfossils in a meteorite from Mars (McKay *et al*., 1996), the circumstantial evidence of geologically recent surface water activity on Mars (Clark, 2012), the discovery by the Cassini spacecraft of complex organic compounds being shot out from beneath the surface of Saturn's moon Enceladus (McKie, 2012), and the like, our understanding of and speculation about life in the universe is based, nearly in its entirely, on our often painfully limited understanding of life on Earth: consider that extremophiles have been known about for only three decades, even as the list of extremophiles (mostly, but by no means exclusively, limited to single-celled microbes) continues to grow. Likewise consider the field of *comparative cognition* (also known as *cognitive zoology*; see Osvath, 2013, *this volume*), which has gone from nothing to a major research area in roughly the same period of time.

No theory of anything starts from a blank slate: one starts with some idea of the thing one is examining. The familiar science fiction motif of encountering the "complete unknown" is incoherent at best; if something were *completely* unknown, one would not even recognize it as a thing in need of explanation. For astrobiology, the starting point is – as it must be – terrestrial biology, however limited a foundation that provides.

Certainly, our knowledge of our own galactic backyard is growing by leaps and bounds: from (next to) nothing, there is nowhere to go but up. When I was a child, scientists debated whether we would *ever* find planets outside our own solar system. The first exoplanets to be discovered were gas giants, nearly the size of stars themselves. Now, with improving techniques, the discovery of even Earth-sized exoplanets has become a routine (if not yet quite everyday!) thing: all in just over two decades.

The pace of discoveries thrills and inspires, as it should. At the same time, it can be easy to forget just *how* big the universe is: the Milky Way is made up of several hundred *billion* stars and is, itself, one of several hundred billion galaxies in the visible universe: pick a number between one and a thousand, add 24 zeroes behind it, and you will get a rough estimate on the number of stars that we *think* there might be. The scale of the visible universe truly challenges the capacity of the human mind even to begin to take it all in; and that is just the universe *we can see*.

Einstein, famously, speculated that the visible universe is a hypersphere. (A hypersphere is to a sphere as a sphere is to a circle; just as a hypercube or *tesseract* is to a cube as a cube is to a square.) One can describe some of the mathematical properties of a hypersphere, even anticipate some of the consequences if the universe *is* a hypersphere; but one cannot picture a hypersphere *as* a hypersphere, nor build even a small hypersphere. Either is, truly, beyond human capacities. If the universe *is* a hypersphere, that raises the possibility that it exists within a larger (hyperdimensional) space; but what the nature or extent of that space may be is, again, beyond human capacities.

Certain quantum physicists favour a so–called branching model of the universe. Consider a subatomic particle such as an electron: one can describe its location at any given time in terms of certain probabilities. Alternatively, one can describe it as occupying *all* those possible locations simultaneously in terms of percentages of a larger universe: one in which, at every possible choice point, all paths are followed. In popular imagination, this is typically cashed out as parallel universes; but a better description might be in terms of a single universe composed of an infinite number of infinitely thin three-dimensional layers stacked one next to the other, like so many sheets of infinitely thin paper: a universe in which pretty much *everything* conceivable happens somewhere. Such a universe makes the vastness of the visible universe look infinitesimally small in comparison!

Mapping the boundaries

So, in astrobiology and astrocognition as in nearly any field of human empirical endeavour: first, there is what we can be *reasonably certain* that we know (even though – taking a page from Einstein – we need not feel any obligation to think that we *know* that we know). Currently – as with Pandora's box after she opened it – this is not much. Still, it is a place to start. Next, there is what we know that we do not know but can reasonably expect to know in the future: does life exist elsewhere in the universe?[1] Is it common (or even ubiquitous) even in our own solar system? (Will we know it when we see it?) Finally, there is what we can be reasonably certain never *to* know (for all our speculations): answers to questions like whether the visible universe is part of a branching universe.

The boundary between what we now know and what we hope to know is where most of the exciting work in empirical science gets done: mapping it out (what exactly *do* we know?), pushing it back, seeing how far we can push it. Then there is the boundary between what we hope to know and what we can expect never to know (either because it would make our beliefs problematically inconsistent, or because it simply lies beyond our capacities to do more than gesture at). Provided one allows for this second boundary, then here is where one can expect to find some of the most exciting work in epistemological philosophy. Even more than the first boundary, this second boundary is *really hard* to pin down (on risk of becoming inconsistent!); it is the conceptual equivalent of looking too closely at a Klein bottle at the point where the handle joins the base (see Figure 1) – or contemplating just what kind of multidimensional space makes an "impossible" Penrose triangle possible.

Figure 1: Klein bottle. (Picture by Clifford Stoll, reproduced under the Creative Commons license.)

Radical re-conceptualization

> *Pace* the causal theorists of reference, 'water' did not always refer to H_20…. In a similar vein, I've elsewhere pointed out… that the content of the Copernican statement, 'planets travel around the sun', cannot be expressed in a statement that invokes the celestial taxonomy of the Ptolemaic statement, 'planets travel around the earth'. The difference between the two statements is not simply one of fact. The term 'planet' appears as a kind term in both, and the two kinds overlap in membership without either's containing all the celestial bodies contained in the other (Kuhn, 1990, pp. 4–5).

Whether one understands "water" as referring to H_2O – what one takes "water" to refer to at all – has a lot to say about one's theory of reference and, more broadly, one's metaphysics. As Kuhn implies and I strongly advocate, there is nothing privileging one conceptual framework over another, *provided that both are consistent with the pattern of our interactions with the environment.* I will have more to say about this in the section "Re-conceptualizing the world". First, I must provide some background on what I mean when I talk about concepts and theories of concepts.

A primer on concepts

Concepts are the means by which we structure our understanding of ourselves and our world. Looked at the one way round, they are the building blocks of systematically and productively structured thought: they can be used *systematically* across unboundedly many contexts, and a finite number of them can be used *productively* to create an unbounded number of complex concepts. Looked at the other, they are the abilities required for systematically and productively structured thought. These notions of systematicity and productivity, and the ambivalence they suggest between viewing concepts as abilities and viewing them as reifications, I borrow freely from Gareth Evans and, in particular, his so-called Generality Constraint on (structured) thought. The Generality Constraint is neatly summarized in (Evans, 1982, pp. 100–104), where he offers the following semi-formalized description:[2]

> There must be a capacity which, when combined with a knowledge of what it is in general for an object to be *F*, yields the ability to entertain the thought that *a* is *F*, or at least a knowledge of what it is, or would be, for *a* to be *F*.... If a subject can be credited with the thought that *a* is *F*, then he

> must have the conceptual resources for entertaining the thought that *a* is *G*, for every property of being *G* of which he has a conception (Evans 1982, pp. 103–104).

The “thought that *a* is *G*” illustrates both systematicity and productivity: the applying of *a* to a new context wherein it retains (most of) its meaning from the old one (systematicity); the creation of a new, perhaps never previously entertained, thought out of an existing set of conceptual elements (productivity).

Evans recognized the logical primacy of concepts as abilities over concepts as reified entities; even as he implicitly acknowledged, I think, the inevitability of thinking of concepts *as* reified entities. I prefer to make this tension explicit: if concepts as abilities are *logically* prior to concepts as reified entities, then concepts as reified entities are *conceptually* prior, precisely because of our inability to set our conceptual nature aside, even for a moment, “to see things just as they are”. I believe that – on the subject of concepts – we toggle between these two perspectives constantly, depending on whether our attention is on the concepts’ application (making the concepts themselves transparent) or on the concepts themselves (making the concepts opaque).

Conceptual frameworks

Regardless of whether concepts are thought of as reifications or abilities, they do not ever – unless you happen to be Jerry Fodor – come individually, but always in groups:[3]

> …It’s plausible prima facie that “a” might refer to *a* even if there are *no* other symbols. The whole truth about a language might be that its only well-formed expression is “John” and that “John” refers to *John*. I do think that uncorrupted intuition supports this sort of view; the fact that “John” refers to *John* doesn’t *seem* to depend on, as it might be, such facts as that “dog” refers to *dogs* (Fodor 2008, p. 54).

Pace Fodor, most philosophers of concepts would allow that beliefs are partly constitutive of concepts just as (and here Fodor would agree!) concepts are constitutive of beliefs; and that, just as beliefs come together to form a belief system, so concepts come together to form an overarching conceptual framework: that is, a conceptually mediated perspective on the world. Because of one’s concepts, one sees things *this* way rather than *that* way; certain things effectively may not exist for us at all, because one has no concepts by which to identify them.

Concepts relate to beliefs in another important way: just as belief systems tend to become increasingly self-reinforcing over time; so, too, concepts increasingly box conceptual agents into *one* view of the world, to the exclusion of (all) others. (When it comes to popular discussion of subject matters like astrobiology[4], the tendency to use "world" and "universe" interchangeably is telling. The world becomes the universe; while the universe is bounded, in the end, by one's conceptual universe.)

Consider concept acquisition as constraint propagation: choices made early on – no matter how innocuous – constrain further choices. With each perspectival choice one makes, one paints oneself further and further into a conceptual corner wherein concepts – like beliefs – are open to change, but in an increasingly bounded domain. One might expect to reach the point – eventually – where one has no perspectival choices left. At that point, to paraphrase the late theologian J. B. Phillips: "your world is too small".

Conceptual breakdown

Every once in a while though, something remarkable happens, and a conceptual framework gets partly or largely swept away – either for an individual or, as described by the quote from Thomas Kuhn that opened this section, for a community or society. Kuhn, whose interest lay in frameworks of scientific thought, spoke of *paradigm shifts*; I prefer to describe the broader phenomenon as *radical re-conceptualization.*[5]

Radical re-conceptualization for a society can mean the wholesale replacement of one system of religious beliefs for another, as social anthropologist Mary Douglas describes so eloquently in *Purity and Danger* (2002) – brought on e.g. by environmental catastrophe or encounter with a new culture. In similar fashion, a new discovery (e.g., microorganisms) or theory (e.g., relativity) can lead to radical re-conceptualization in the scientific community, eventually filtering through (in somewhat distorted form) to the wider society. For the individual, the catalyst can be the proverbial life-changing experience, mid-life crisis, crisis of faith, or the like. Meanwhile, some researchers, such as Etzel Careña (2011b; 2011a), believe that altered states of consciousness and mystical experiences constitute, in certain cases, the *temporary* replacement of one conceptual framework by another.

The key to radical re-conceptualization is the incommensurability between the old and the new – noting that the notion of *complete* incommensurability (often pinned on the earlier Kuhn (1970)) is – albeit implicitly – oxymoronic (as the later Kuhn, reflected in the opening quote, was keen to point out): conceptual incommensurability is critically dependent on conceptual continuity. *Some* part of the old system carries over into the new: *that* is precisely where the confusions arise. Old ideas get used in new ways even as people forget, for the moment, that the supporting structure has changed. Were there *no* overlap, communication of any kind between old and new would be impossible; indeed, it would be impossible to recognize the old framework as a framework at all! Communication is founded on the presupposition of *something basic in common*; as Göran Sonesson writes in a different context (2013, *this volume*): “…in a situation of communication, the first problem is not to find out *what* the messages means: it is to realise that there is a message.”

Re-conceptualizing the world

Here is where astrobiology and astrocognition earn their philosophical pay. When one’s world becomes too small, it is time to break down the walls of that world or push those walls outward.

When I was first learning to read, my teachers used the legendary (within the US educational establishment) SRA Reading Lab Kit. I still remember well one of the more advanced exercises in the series, which raised the question: if humanity discovers an extraterrestrial intelligence among the stars, what will it look like: bug-eyed monster, snake-like lizard, or something else again? On the argument that our two eyes, two ears, two arms, two legs, etc. adapt us perfectly to our environment, the writer concluded with the punch line that any aliens we meet mean would almost certainly have to look just like us, to the extent that we might not even be able to tell them apart from us!

The dubious assumption that the human body *is* perfectly adapted to its environment aside – in keeping with many biologists, one might argue that the human body is merely *sufficiently* adapted to its environment to ensure reproductive success, and that learning to walk upright created certain problems even as it created opportunities – the more glaring assumption is that life, or at least intelligent life, would have to develop in an environment very similar to our own.

The Earth is the only planet in our solar system that falls within the so-called *Goldilocks zone,* where distance from the star makes the presence of surface water possible. It is commonly believed that the Goldilocks zone around any star is a very narrow region indeed: so narrow as to make the likelihood of finding an Earth analogue seem remote. Yet we now know with something close to certainty that, at some point in the past, Mars had considerable surface water: so e.g. the NASA Curiosity rover is currently traveling through what appears to be a dry riverbed. It may be that, at some point in the past, the Goldilocks zone was located further out, so that surface water was possible on Mars even while it was not yet on Earth; or it may be that, for reasons not currently understood, the Goldilocks zone was significantly wider in the past. Only further exploration of Mars will be able to tell.

At the same time, it is not obvious that an Earth analogue is required. One finds increasing evidence for subsurface liquid water further out in the solar system: most untendentiously beneath the surface of Enceladus, even though its diameter is only 500 kilometers. Nor is it obvious that water is the only solvent capable of nurturing life: other suggested candidates include ammonia and methane. What is toxic to even the extremophiles on Earth may be a condition for life elsewhere.

…Which raises the question: how should one define life? Erik Persson (2013, *this volume*) raises just this question, without coming to any definite answers. One could stipulate that life must be carbon based or use water as its solvent; but this seems to beg the question. Stipulations are often useful, but by their very nature, they can hardly be informative. Persson suggests various desiderata as candidates for capturing life's necessary and sufficient conditions; I myself prefer the definition of life as autopoiesis, developed by biologists Humberto Maturana and Francisco Varela (see e.g. Maturana and Varela, 1980; 1992; Varela *et al*., 1991), which deliberately eschews any laundry list of requirements for a more functional definition: one that is intentionally open to non-standard biologies and, potentially at least, to artefacts.

An autopoietic system is a kind of *homeostatic machine*, where "machine" means only that the system is defined by its abstract organization, not by its physical realization: "…the organization of a machine is independent of the properties of its components which can be any, and a given machine can be realized in many different manners by many different kinds of components" (Maturana and Varela, 1980, p. 77). As Michel Bitbol and Pier Luigi Luisi write (Bitbol and Luisi, 2004), "…life is a cyclic process that produces the components that in turn self-organize in the process itself, and all within a boundary of its own making". This implies the three defining characteristics of autopoietic systems:

- They define their own, semi-permeable boundaries.
- They are *organizationally* and *operationally closed*: they may be perturbed by external forces, but the organization of the system is determined solely from within the system, even while its components are continuously being replaced.

> …The environment has its own structural dynamics and, although independent of the organism, it does not prescribe or determine the changes in it. It induces a reaction in the organism, but the accepted changes are determined by the internal structure of the organism itself. It is the structure of the living system and its previous history of perturbations that determines what reactions the new perturbation will induce (Luisi, 2003, p. 54).

- They are *autonomous* in a strong sense: not merely giving the appearance of autonomy that falls apart on closer examination. They are both "continually self-producing" (Maturana and Varela, 1992, p. 43) and adaptive (Di Paolo, 2005).

There is no *prima facie* reason why autopoietic systems must be based on deoxyribonucleic acid. Neither is there any requirement that they be able to reproduce. Life-as-autopoiesis deliberately avoids prejudicing matters toward "life as we know it".

For the budding field of astrobiology, autopoiesis strikes me as a step in the right direction. At the same time, it necessarily remains a largely terrestrially bound view on life, with e.g. its focus on the cell as the basic unit of life, "the main argument for this being simply that there are no other forms of life on Earth" (Luisi, 2003, p. 52).

Concepts, as I suggested earlier, do not exist in isolation but relate to each other as part of a complex web: pull on any one thread, and the whole system responds. Pull on the *right* thread – change certain key concepts – and the whole conceptual framework changes. *Life*, I propose, is one of those key concepts: basic to our understanding of ourselves is that we are *alive.* The discovery of extraterrestrial life, should that happen, will doubtless force us to reconsider all our definitions of life – traditional, autopoietic, and otherwise; at the same time it will, inevitably, broaden our perspective on many other things and expand our conceptual world. At the very least, we will no longer be quite so alone in the universe.

Re-conceptualizing the human

If what it means to be alive is pretty fundamental, then who we are individually (*what is it to be a self?*) and collectively (*what is it to be human*?) is even more so. Indeed, the distinction between self and non-self, self and other, self and world – three different faces to the same distinction – is, arguably, foundational to all the other conceptual distinctions we learn to make (see e.g. Zachar, 2000, pp. 144 *ff.*). The analogous distinction between human and non-human likewise shapes our collective identity: are we, as our religious heritage might suggest, a separate creation from the rest of the animal kingdom and the wider biosphere, fundamentally different from all else on the planet; or does one find important continuities to match the obvious discontinuities, including continuities of cognitive ability? How secure are we in our fabled position as the only intelligent species on the planet? If Osvath (2013, *this volume*) and others in the comparative cognition community are right – and I think they are – we lost that perch some time ago (and, indeed, held it only in our *own* conceptual world). Besides, a better understanding of what human beings have in common with other species is arguably critical to any proper understanding of the differences.

As human beings, we shift constantly and, for the most part, unreflectively, between two views of ourselves: physical organisms on the one hand, self-aware and intentional entities on the other: the so-called *mind-body problem* in philosophy, which arises, I believe, because of the widespread assumption that the two perspectives can and should be resolved into one, complete and consistent perspective. As I have suggested already, completeness and consistency may not sit together any better in human thought – outside sufficiently narrowly defined domains – than they do in formal systems.

This constant shift between perspectives reflects a broader duality in our view of the world. In one view, our role as observers is in the foreground; subjective and intersubjective perspective colour everything. We represent the world as being *one way* and not *another*; the symbolic structure of our conventionalized gestures and language facilitate certain thoughts and hinder others.

In the other view, the observer is pushed into the background or, seemingly, removed altogether. The assumption is often made, of course – especially in regards to "hard" science – that only the latter perspective is needed; as Inman Harvey writes, "the underlying assumption of many is that a real world exists independently of any observer; and that symbols are entities that can 'stand for' objects in this real world in some abstract and absolute sense. In practice, the role of the observer in the act of representing something is ignored" (1992, p. 5).

If this sounds like a slightly different version of the two perspectives on concepts I introduced earlier:

- Representations every time we stop and look at them.
- Something else – logic suggests non-representational abilities – when we don't (i.e., when we just get on with possessing and employing them non-reflectively).

…Then it should. The toggling between perspectives goes to the very heart of our conceptual nature.

If I am right, then any strictly biological definition of what it means to be human will be inadequate, even in biology; but even more so in areas outside biology. Why?… because while the boundaries of "human being" as a biological organism are relatively clear (although ask a biologist, and the biologist will admit that the definition of "species" is not so clear cut as it might appear at first blush); the boundaries of "human being" as a cognitive entity are much less so: *where does the self stop and the non-self/other/world begin?* Or rather, they are clear only if one begs the question and assumes that mind reduces to brain (the so-called *mind-brain identity hypothesis*) or emerges straightforwardly from brain, in a way that one can at least potentially reconstruct. They are clear only if one assumes a certain metaphysical perspective, as I have claimed (2011b) that Fred Adams and Ken Aizawa (2008; 2001) and Robert Rupert (2009b; 2009a; 2004) – fierce critics of Dave Chalmers and Andy Clark's *extended-mind hypothesis* (Clark and Chalmers, 1998; Clark, 2008) – all do. Regardless of whether metaphysical positions can be proven (I think it is intrinsic to their nature that they can neither be proven nor disproven), they *cannot* be taken for granted.

Consider the evolution of the concept of human in the context of e.g. recent US history: during the period of European settlement, the question whether or not Native Americans had souls was seriously debated (see e.g. Neusner, 2009, p. 93; Burns, 2005, p. 136). Clearly – to many people – the natives were less than fully human: an appellation best reserved to those of enlightened European extraction. Such thinking might be seen as having made its way into the US Constitution, where a non-free person (i.e., an African-American slave) counts as three-fifths of a person for purposes of tax distribution and Congressional representation (Article I, Section 2, Paragraph 3: a provision that remained in effect until ratification of the Thirteenth Amendment in 1865[6]); in any case, slaves were property (to be bought and sold) in a way that "full" human beings were not. Today, the idea of race having *any* biological basis is widely discredited; while the broader notion of race – the idea that there are human *races* (plural!) – having any usefulness beyond identifying historical inequities is in – one might hope terminal – decline.

If, in the past, the concept of human was limited to select sub-groups of the human species – something that few would be willing to advocate today, at least explicitly – then the animal rights movement of the past several decades may, on a certain view, be understood as pushing at least limited aspects of humanness *beyond* the boundaries of the human species:

other animals are to be granted at least minimal rights based, it might seem, on how much of *us* we see in *them*.[7] On this basis, a horse has more rights than a mouse; both have more rights than, say, a grasshopper; while the great apes, as our closest living relatives, are generally accorded the highest rights short of full membership in humanity. One finds this reflected in the animal welfare legislation and regulations that animal rights activists have pushed for and, in many cases, gotten adopted. What some would see as a false projection of ourselves onto other species – I have someone like Donald Davidson in mind (see e.g. Davidson, 1987); others – including those in the animal rights movement – would see as the acknowledgement of commonalities that had previously been overlooked or denied.[8] Remember the aforementioned dictum: *a better understanding of what human beings have in common with other species is critical to any proper understanding of the differences.*

If I am right, and we *are* pushing the (wider notion of) human beyond the boundaries of the biologically-defined species, then the search for extraterrestrial biology in general, and extraterrestrial intelligence in particular, promises to push things yet much further. Consider that – if the conceptual story I have told is anywhere even close to being right – then human and (non-human) world are defined, not in opposition to each other, but in intricate coordination. As with the individual agent and her environment, each arises out of the interaction: a key idea in the *enactive philosophy* that is also associated with Maturana, Varela, and their followers. Varela writes (Maturana and Varela, 1992, p. 255), "I have proposed the term enactive to… evoke the idea that what is known is brought forth…." Change the one, and one cannot help but change the other. Change the world enough, and what it is to be human changes.

First contact

In the section "Re-conceptualizing the world", I talked about how the discovery of new forms of life, elsewhere in the solar system or out among the stars, would, inevitably, re-shape the way we understand life on Earth. At the same time, recognition of extraterrestrial life *as* life in the first place requires making some substantive link back to life *as we currently know it.* Something substantive must be in common – even if we cannot say, in advance, what it will be.

On an even much deeper level, recognition of extraterrestrial intelligence as intelligence presupposes seeing the human in the alien: recognizing the *us* in *them* (and, by implication, the *them* in *us*). That recognition could be based on many things: on the artefacts that they build, on the recognizably message-like communication that they use (Sonesson's point comes again to mind), on the emotions that they *seem* to display under conditions in which we would experience those emotions as well; but it must be based on something. The idea of confronting an alien intelligence with whom we have *nothing* in common is conceptually self-contradictory.

Scientists and philosophers have speculated whether any alien intelligence we encounter will necessarily have a concept of e.g. prime numbers. It may well turn out to be that that *will* be the case – *not* because prime numbers have some platonic existence; not because prime numbers are necessarily anything besides a useful conceptual tool; but because any alien intelligence that did *not* build up its understanding of the world in part on a conception of prime numbers would be one we would simply fail to recognize as intelligent life. Its interaction with its environment would be too different for us to make the connection.

Blay Whitby has long argued (see e.g. 2003; 1996) against taking *human* intelligence to be representative of intelligence writ large; "science has to be interested in the whole space of intelligence" (2003). That may well be. Nevertheless, human intelligence is, unavoidably, the starting point by which we come to understand any intelligence – how could the starting point be anywhere else?; just as the Earth – however poorly it may be situated – is, unavoidably, our starting point for exploring our galactic neighbourhood (at least, if recent reports of a potential "warp drive" should bear any fruit (Moskowitz, 2012)); the Milky Way itself; and, perhaps, in time, the wider universe (or multiverse) beyond.

If and when first contact is made, the boundaries of the human – not the biological entity, but the broader notion – shall, subtly or not so subtly, shift. For all the uncertainties, what can be counted as near certainty is that, if and when first contact is made, both we and those we encounter shall come away substantively changed.

The final frontier

> The thing's hollow – it goes on forever – and – oh my God! – *it's full of stars!* (Clarke, 2000)

The thought I want to leave you with is that *we* are the final frontier, and space the ultimate mirror we hold up to ourselves: that every reaching out beyond our terrestrial confines is, simultaneously, a reaching inward; that, in contemplating the depths of the cosmos, we confront the unseen depths in ourselves.

It is, perhaps, characteristically human to downplay or outright deny the limitations of our individual and collective perspectives. Indeed, as I discussed in the section ‘Re-conceptualizing the human’, science has gone a long way by disregarding the role of subjective and intersubjective bias; on top of which, as the expression goes, scientists are “only human”. So it is not surprising when scientists make bold claims about the near certainty or near impossibility of life around other stars. It is not surprising when they speak, in the most certain of terms, about the nature of dark matter, the age of the universe (currently placed around 13.75 billion years), or the events of the first few nanoseconds after the Big Bang. It is only when we stop to consider the narrowness of our keyhole onto the rest of the universe that we remember that the Big Bang theory is a *theory*, albeit the best we have to date. Its position in scientific thinking is clearly dominant, but hardly universal; the astronomer (and science fiction author) Fred Hoyle, who died in 2001, famously spent a lifetime opposing it. To paraphrase Einstein, one might hope that, on further examination, inconsistencies will be revealed that will lead to more precise formulations (and so on, *ad infinitum*); or fresh evidence may push scientific thinking off in a new, previously unconsidered direction.

The universe – even just the *visible* universe – is almost unimaginably vast: the Milky Way’s neighbour Andromeda lies 2.5 million light years from Earth: the light we observe from its stars began its journey 2.5 million years ago, making one wonder what Andromeda looks like today (or to what extent it even makes sense to ask that question). The further we look out into the universe, the further back in time we go. Meanwhile, 13.75 *billion* years – if that *is* the correct age of the universe – makes the light from Andromeda look young. It is nearly 113 million times longer than the lifespan of the longest-lived human. The spatial and temporal scales are so vast that, if one were dropped off at some random location in the universe, even if one had the means to do so, one could never hope to find one’s way back home. On the scale of the universe, a human being – or the entire Earth for that matter – is far less than a speck of dust. A little reflection on how much we *know* that we do not know advises humility.

A human being may seem like an easily knowable thing in comparison – and, in many ways, it is. Yet how well do we know ourselves *really*: either individually or collectively… especially when it comes to our nature as self-aware and intentional beings, rather than our nature as biological organisms (keeping in mind my suggestion that these are not separate worlds to be brought mysteriously together but separate and irreconcilable perspectives on the same world)? The pioneering work of Sigmund Freud, Carl Jung, and others more than a century ago revealed just how much our motivations can remain hidden from ourselves. Although repressed memories are not always quite what they appear to be – as the recent discussions over so-called *false memory syndrome* have shown[9] – nevertheless, they can, and do, play a powerful role in shaping people’s lives. Contemporary research in the choice blindness paradigm (see e.g. Hall *et al*., 2012; Johansson *et al*., 2011b,a) suggests that many of our explanations for our own behaviour are *post hoc* reconstructions or rationalizations. Even such basic things as what it is to have or express an opinion or make a judgment turn out *not* to be so straightforward.[10]

Our assessment of our collective cognitive nature fares little better. Alien visitors to Earth might well wonder why, in the face of catastrophic human-mediated climate change and risk of truly global pandemics, the nations of the world continue to devote so much of their resources not to solving our collective problems but to reinforcing our military capabilities: to devising, in effect, better ways to kill each other. Could we offer them any satisfactory account? Or would we shrug our collective shoulders and say, that’s just human nature? In doing so, would we be offering them any explanation or only giving an acknowledgement of our own self ignorance? Time and again, we seem to act collectively in ways that leave us, individually and collectively, astonished or bemused.

Let me frame the issue a different way. Consider the nature of consciousness. On the one hand, consciousness – so deeply and even, seemingly, intrinsically subjective – seems not the sort of thing that objective science is set up to account for; on the other, as Penrose notes:

> A scientific world-view which does not profoundly come to terms with the problem of conscious minds can have no serious pretensions of completeness. Consciousness is part of our universe, so any physical theory which makes no proper place for it falls fundamentally short of providing a genuine description of the world (Penrose, 1994, p. 8).

Various attempts have been made, over the last two decades in particular, to reduce consciousness to *x*, where *x* might be any one of e.g. Daniel Dennett’s *multiple drafts theory*, from a book famously or infamously titled *Consciousness Explained* (Dennett, 1991); Guilio Tononi’s *information integration theory* (Tononi, 2008; Balduzzi and Tononi, 2008); Bernard Baar’s *global workspace theory* (Baars, 1996; 1988); or Penrose’s own preferred choice, *quantum consciousness* (Hameroff and Penrose, 1996), according to which consciousness results from the collapse of quantum-level wave functions in the microtubules of the brain.

At the same time, one might be forgiven for concluding, in company with David Chalmers (1996), that consciousness has *not* been explained; that, at best, one or another aspect of consciousness has been addressed.

Indeed, many a scientist and philosopher, and many a speaker at the annual conference of the Association for the Scientific Study of Consciousness, continue to echo Chalmers' concern that so little progress has been made in giving a scientific account of consciousness. Of course, it *could* be the case that so little progress has been made because the wrong approaches have been tried; but it could also be, to steal a line from Chalmers, that consciousness poses a truly *hard* problem: one with many, many layers to it; one that forces us to confront both our subjective and objective perspectives on ourselves and the world; one that raises serious questions about the ultimate capacity of any (self-)conscious entity to explain its own consciousness. After all, only a fully self-conscious entity would ever think to ask about the nature of consciousness in the first place; it might appear as if any account of consciousness is necessarily given from within an existing conceptual framework that *presupposes* consciousness. Putting this another way, one cannot give an account of consciousness in general without, simultaneously, giving an account of one's *own* consciousness. Questions of looming circularities aside, consciousness may simply be the sort of phenomenon that, the more we learn about it, the more we learn there *is* to learn.

Here, then, is the great hope of astrobiology: to discover not just life but *intelligent* life; by which I mean, intelligence that is recognizably like our own. Intelligence, in this strong sense, presupposes (self-) consciousness – *contra* those early AI researchers who wanted to set questions of consciousness aside; just as, I believe, consciousness presupposes life: a point I freely borrow from Jordan Zlatev (2009; 2002; 2001), though it is an article of faith among enactive philosophers in general.

The aliens – be they little green men from New Mars or bug-eyed monsters from Alpha Centauri – will, presumably, be no better able to explain their own consciousness than (to date at least) we have been able to explain ours. Here we can reasonably hope to help them, as they can help us, precisely because of those ways in which, inevitably, we will be *unlike* each other. For – by the very objective distance that our dissimilarities give us – we may give them insights into their nature, and they may give us insights into our nature, that neither we nor they could achieve on our own.

Even if the search for extraterrestrial life never uncovers intelligent life; even if it never uncovers life at all (though here, my own intuitions, as a non-biologist, are that, if we go looking, we will find life in one form or another in various corners of our solar system and conclude it a ubiquitous phenomenon in places further afield); it has, if I am right, a great deal to reveal to us about ourselves. Remember the claim from enactive philosophy that we are substantially defined in interaction with our environment: change the world – change the boundaries of the world – and we change ourselves. In the meantime, we will keep looking, with our space probes and robot landers and, perhaps, in time, with manned missions to Mars and beyond. We will keep listening with our radio telescopes. We will keep learning and re-learning what it is to be human in a universe that challenges us to explain it even as it continually outstrips our ability to do so.

Notes

[1] Note that this question is one of those questions that can only ever be answered definitely in the positive, never the negative.
[2] Not everyone agrees with the Generality Constraint: Charles Travis (e.g., 1994) has made a professional career out of opposing it, while Jacob Scott Beck (2007) objects to its *a priori* nature, contending that the matter should be decided by empirical investigation. However, most philosophers *do* accept the Generality Constraint; I think the burden is on its (few) opponents to explain how conceptual thought is not systematic and productive in the way Evans suggests.
[3] I set aside, for now, the relationship between concepts and language. Suffice that, for Fodor, most simple concepts are lexical concepts.
[4] Of course, within the field itself, the term "world" is given a more precise – or at least narrower – meaning, referring to planets, dwarf planets (as Pluto is now re-designated), moons, and possibly asteroids. This is not to say, however, that astrobiologists do not fall into the same conceptual trap as everyone else! It is, I think, a gross mistake to think that researchers do – or even can – place the limitations of lay discourse behind them when they do science.
[5] Compare Piaget's notion, from children's cognitive development, of a period of *accommodation* (rapid change and reorganization) following one of *assimilation* (slow, incremental change) – for a good discussion, see (Bitbol and Luisi, 2004, p. 103).
[6] The full text of the US constitution, including amendments, can be found e.g. at http://en.wikipedia.org/wiki/Us_constitution.
[7] To clarify: I am not claiming that the animal rights movement sees other species as *being* human; rather, that it is driven by looking for and finding that which was previously considered exclusive to humans in non-human species.
[8] This is precisely the position of the so-called *animal concepts* philosophers: I have in mind such people as Albert Newen, Andreas Bartels (Newen and Bartels, 2007), and Colin Allen (1999).
[9] Contemporary debate was kicked off by the founding of the False Memory Syndrome Foundation by P. J. Freyd in 1992; see also (Brainerd and Reyna, 2005).
[10] Note that the choice blindness findings are fully compatible with having free will even in the strong libertarian sense; it may be only that *where* we think we make real choices is not always, or maybe even often, where we *think* we do.

References

Adams, F., and K. Aizawa. 2001. "The bounds of cognition." *Philosophical Psychology* 14(1):43–64.
Adams, F., and K. Aizawa. 2008. *The Bounds of Cognition*. Oxford: John Wiley and Sons.
Allen, C. 1999. "Animal concepts revisited: The use of self-monitoring as an empirical approach." *Erkenntnis* 51(1):33–40.
Baars, B. J. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
—. 1996. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.
Balduzzi, D., and G. Tononi. 2008. "Integrated information in discrete dynamical systems: Motivation and theoretical framework." *PLoS Computational Biology* 4(6):1–18. http://dx.doi.org/10.1371
Beck, J. S. 2007. "The generality constraint and the structure of thought." Presentation at MindGrad2006 conference, University of Warwick, UK, and unpublished paper. Available online from http://www.webpages.ttu.edu/jabeck/GCST.pdf (accessed 21 December 2012).
Bitbol, M., and P. L. Luisi. 2004. "Autopoiesis with or without cognition: Defining life at its edge." *Journal of the Royal Society Interface* 1:99–107. DOI: 10.1098/rsif.2004.0012.
Brainerd, C., and V. Reyna. 2005. *The Science of False Memory*. Oxford: Oxford University Press.
Burns, W. E. 2005. *Science and Technology in Colonial America*. Santa Barbara CA: Greenwood Publishing Group.
Cardeña, E. 2011a. "Altered consciousness is a many splendored thing." *Journal of Cosmology* 14:207–218.
—. 2011b. "Neurophenomenology and the study of states of consciousness." Seminar presentation at the Centre for Cognitive Semiotics, 13 October 2011.
Chalmers, D. J. 1996. "Facing up to the hard problem of consciousness." In *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, S. R. Hameroff, A. W. Kaszniak, and A. Scott, 5–28. Cambridge MA: MIT Press.
Clark, A. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press. Also available as ebook through Oxford University Press (http://www.oxfordscholarship.com).
Clark, A., and D. Chalmers. 1998. "The extended mind." *Analysis* 58(1):7–19.

Clark, S. 2012. "Water on Mars has a long history." Available online from http://www.guardian.co.uk/science/across-the-universe/2012/oct/01/water-mars-history (accessed 4 November 2012).

Clarke, A. C. 2000 [1968]. *2001: A Space Odyssey*. New York: Roc Books.

Davidson, D. 1987. "Rational animals." *Dialectica* 36(4):317–327.

Dennett, D. C. 1991. *Consciousness Explained*. Boston MA: Little, Brown.

Di Paolo, E. A. 2005. "Autopoiesis, adaptivity, teleology, agency." *Phenomenology and the Cognitive Sciences* 4:429–452.

Douglas, M. 2002 [1966]. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London: Routledge.

Evans, G. 1982. *Varieties of Reference*, edited by John McDowell. Oxford: Clarendon Press.

Fodor, J. A. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Hall, L., P. Johansson, and T. Strandberg. 2012. "Lifting the veil of morality: Choice blindness and moral attitude reversals on a self-transforming survey." *Plos ONE* 7(9):e45457. Available online from http://dx.plos.org/10.1371/journal.pone.0045457 (accessed 4 November 2012).

Hameroff, S., and R. Penrose. 1996. "Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness?" In *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, edited by S. Hameroff, A. Kaszniak, and A. Scott, 507–540. MIT Press. Available online from http://www.quantumconsciousness.org/penrose-hameroff/orchOR.html (accessed 4 November 2012).

Harvey, I. 1992. "Untimed and misrepresented: Connectionism and the computer metaphor (CSRP 245)." *University of Sussex (UK) Cognitive Science Research Papers (CSRP) series*. Available online from http://www.sussex.ac.uk/informatics/cogslib/reports/csrp/csrp245.ps.Z (accessed 20 December 2012).

Hawking, S. W. 1988. *A Brief History of Time*. London: Bantam.

Johansson, P., L. Hall, and N. Chater. 2011a. "Preference change through choice." In *Neuroscience of Preference and Choice*, edited by R. Dolan, and T. Sharot, 121–141. Oxford: Elsevier Academic Press.

Johansson, P., L. Hall, and P. Gärdenfors. 2011b. "Choice blindness and the non-unitary nature of the human mind." *Behavioral and Brain Sciences* 34(1):38–29. Available online from http://www.lucs.lu.se/wp-content/uploads/2011/01/Johansson-Hall-Gardenfors-2011-Commentary-on-von-Hippel-and-Trivers-BBS.pdf

(accessed 4 November 2012).
Kuhn, T. 1970. *The Structure of Scientific Revolutions*. Chicago IL: University of Chicago Press.
—. 1990. "The road since Structure." In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2, 3–13.
Luisi, P. L. 2003. "Autopoiesis: A review and an appraisal." *Naturwissenschaften* 90(2):49–59.
Maturana, H. R., and F. J. Varela. 1980. *Autopoiesis and Cognition: The Realization of the Living (Boston Studies in the Philosophy of Science)*. Dordrecht: Kluwer Academic Publishers.
Maturana, H. R., and F. J. Varela. 1992. *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala.
McCullough, D. 1995. "Can humans escape Gödel? A review of Shadows of the Mind." *Psyche* 2:1–9. Available from http://www.theassc.org/files/assc/2327.pdf (accessed 21 December 2012).
McKay, D. S., Jr., E. K. G., K. L. Thomas-Keprta, H. Vali, C. S. Romanek, S. J. Clemett, X. D. Chillier, C. R. Maechling, and R. N. Zare. 1996. "Search for past life on Mars: Possible relic biogenic activity in Martian meteorite alh84001." *Science* 273:924–930.
McKie, R. 2012. "Enceladus: Home of alien lifeforms?" Available online from http://www.guardian.co.uk/science/2012/jul/29/alien-life-enceladus-saturn-moon (accessed 4 November 2012).
Moskowitz, C. 2012. "Warp drive may be more feasible than thought, scientists say." Available online from http://www.space.com/17628-warp-drive-possible-interstellar-spaceflight.html (accessed 4 November 2012).
Neusner, J. 2009. *World Religions in America: An Introduction*. Louisville KY: Westminster John Knox Press.
Newen, A. & Bartels, A. 2007. "Animal minds and the possession of concepts." *Philosophical Psychology* 20(3):283–308.
Osvath, M. 2013. "Astrocognition: A cognitive zoology approach to potential universal principles of intelligence." In *The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind*, edited by D. Dunér, J. Parthemore, G. Holmberg, and E. Persson. Newcastle: Cambridge Scholars.
Parthemore, J. 2011a. *Concepts Enacted: Confronting the Obstacles and Paradoxes Inherent in Pursuing a Scientific Understanding of the Building Blocks of Human Thought*. PhD thesis. Brighton: University of Sussex, Falmer.
—. 2011b. "Of boundaries and metaphysical starting points: Why the extended mind cannot be so lightly dismissed." *Teorema* 30(2):79–94.

Penrose, R. 1994. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford: Oxford University Press.
Persson, E. 2013. “Philosophical aspects of astrobiology.” In *The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind*, edited by D. Dunér, J. Parthemore, G. Holmberg, and E. Persson. Newcastle: Cambridge Scholars.
Rupert, R. 2004. “Challenges to the hypothesis of extended cognition.” *Journal of Philosophy* 101:389–428.
—. 2009a. *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.
—. 2009b. “Critical study of Andy Clark’s Supersizing the Mind.” *Journal of Mind and Behavior* 30:313–330.
Sonesson, G. 2013. “Preparations for discussing constructivism with a Martian.” In *The History and Philosophy of Astrobiology: Perspectives on Extraterrestrial Life and the Human Mind*, edited by D. Dunér, J. Parthemore, G. Holmberg, and E. Persson. Newcastle: Cambridge Scholars.
Stachel, J. 1982. “Comments on ‘some logical problems suggested by empirical theories’ by Professor Dalla Chiara.” In *Language, Logic and Method: Papers Derived from the Boston Colloquium in the Philosophy of Science 1973–1980*, edited by R. Cohen, and M. W. Wartofsky, 91–102. Dordrecht: Springer.
Tononi, G. 2008. “Consciousness as integrated information: A provisional manifesto.” *The Biological Bulletin* 215(3):216–242.
Travis, C. 1994. “On constraints of generality.” *Proceedings of the Aristotelian Society, New Series*, 94:165–188. London: Blackwell Publishing/The Aristotelian Society.
Varela, F.J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge MA: MIT Press.
Whitby, B. 1996. “The turing test: AI’s biggest blind alley?” In *Machines and Thought: The Legacy of Alan Turing*, vol. 1, edited by P. Millican, and A. Clark. Oxford: Clarendon.
—. 2003. “The myth of AI failure (CSRP 568).” *University of Sussex (UK) Cognitive Science Research Papers (CSRP) series*. Available online from http://www.sussex.ac.uk/informatics/cogslib/csr_abs.php?type=csrp&num=568&id=9340 (accessed 21 December 2012).
Zachar, P. 2000. *Psychological Concepts and Biological Psychiatry: A Philosophical Analysis*. Amsterdam: John Benjamins Publishing Company.

Zlatev, J. 2001. "The epigenesis of meaning in human beings, and possibly in robots." *Minds and Machines* 11:155–195.

—. 2002. "Meaning = life (+ culture)." *Evolution of Communication* 4(2):253–296.

—. 2009. "The semiotic hierarchy: Life, consciousness, signs and language." *Cognitive Semiotics* 4:169–200.

CHAPTER FOUR

PYRRHONEAN APORIAGOGICS AND THE ASTROCOGNITIVE CHALLENGE

PER LIND

I am made to sow the thistle for wheat, the nettle for a nourishing dainty
—William Blake, *The Wail of Enion*

Introduction

Can the human mind be prepared specifically for the unspecified; for encountering the unknown?[1] Can the extent of cognitive derailment and perplexity experienced in such encounters, when tried and trusted frames of reference and interpretation no longer seem to apply, be limited preemptively? This chapter sketches a possible line of research into this general problem area, in response to some issues recently raised regarding the possible challenge to human cognition posed by the explorative potentials of astrobiology.

Situated at the crossroads between classical philosophy and cognitive psychology, and proceeding from an expressly "aporiagogic" reading (see note 1) of the ancient Greek art-of-living philosophy of Pyrrhonism, the present chapter argues that the self-described Pyrrhonist method of establishing irresolvable cognitive conflicts as a means of overcoming dogmatism addresses the aforementioned issues. More precisely, it is argued that cognitive conflict is a natural impetus for metacognitive awareness, and that Pyrrhonism represents a systematic exploit of this cognitive modality, aimed at what is essentially a general metacognitive insight into the nature and limits of one's own cognition, an insight that in turn should yield a more open and flexible mode of cognition. Along these lines, a conceivable methodological avenue for preparing the human mind for the unspecified is ventured.

It is argued that this reading of Pyrrhonism holds an explanation to Pyrrhonism's apparent double commitment to truth *and* transformation, as well as its double construction of the suspension of judgment (*epochē*) as an automatic cognitive function *and* an intentional act of rational deliberation.

The astrocognitive question

The following "astrocognitive question" (AQ) is the focal point of a recent article on the potential challenges posed by space exploration to the human cognitive apparatus:[2]

> What will happen to human cognition when humans encounter a totally different environment, physically, biologically and culturally; an extraterrestrial environment that the human brain is not accustomed to and developed for? (Dunér, 2011, p. 118)

Drawing on recent theories of cognitive science, married to a pronounced evolutionary outlook, Lund University's David Dunér goes on to propose a new multidisciplinary field of research, *astrocognition*, dedicated to "the study of human cognitive processes in extraterrestrial environments" (*ibid.*). Impractical as such studies might seem, it is Dunér's sober suggestion that at present, empirical astrocognitive data should be looked for not in extraterrestrial settings, but in more accessible, albeit less exotic human dealings with the unknown. Besides experiments and simulations, Dunér lists some authentic areas of such interaction, the most elaborated of which is the history of Earth exploration (Dunér, 2011, pp. 125–126, 128–136).

The foundational idea of this research program is clearly that the basic principles of human knowledge formation and application – or at least most of them – apply regardless of object, so that that the explorative context specific to the AQ, in consequence, need not be significantly unlike any other explorative context. Dunér makes a number of observations on human knowledge formation and application along this main line of thought; in particular the observation that the unknown is interpreted in the terms of the known, or as Dunér puts it, "the unknown is a *reflection of the known*" (Dunér, 2011, p. 131. Italics in the original).

This general principle of human cognition forms the point of departure for the present chapter. In contrast to Dunér's predominantly descriptive and empirical approach, this investigation will proceed in a more philosophical direction. By "philosophical", I mean classically and prescriptively philosophical, as exemplified by the so-called art-of-living philosophies of ancient Greece. Specifically, the sceptical art-of-living philosophy of Pyrrhonism, a little-known sibling of the more illustrious art-of-living schools of Stoicism and Epicureanism, will be consulted regarding the human predicament of having to face and judge the unknown with cognitive recourse only to the known. First, however, some basic ramifications of this operational prejudiciality of human cognition must be sketched.

Cognition and metacognition

In everyday life, our ways of perceiving and understanding the world remain largely unchallenged. While this is particularly true of the most fundamental dispositions of reality interpretation, such as the robust separation of the self from the world, the imperative of basic ontological coherence or the stability of symbolic representation, the more mundane levels of our cognitive operational economy are more or less equally undisturbed. By and large, the world meets our expectations. Each day tends to resemble the one before. That is to say, in everyday life, the encountered unknowns are few in number, and more importantly, they resemble the known enough to pose no great challenge. For example, a shopping bag that suddenly and unexpectedly breaks open might ruin our entire day, but it does not stretch our cognitive frames of reference much. Yet, in its own limited way, it breaks our cognitive step in not conforming entirely to our expectations. Shopping bags do not normally break open for no apparent reason. It is, in short, an experience of cognitive contrast.

Theoretically speaking, most frames, models and projections of our cognitive system are similarly subject to the possibility of contrasting cognitions, in the sense that anything once mapped by human cognition retains the authority to redraw the map at a later point. Such involvements with epistemic novelties – rendering us surprised, amazed, confused, or baffled – are an integral part of human experience. On this subject, cognitive psychologist J. David Smith makes a crucial observation:

> Humans encounter situations daily in which their automatic and habitual behavior patterns are thwarted and they enter difficult and uncertain straits. It is worth reflecting on how cognition meets these challenging situations. Simply put, when the going gets tough, the tough get conscious. Times of difficulty and uncertainty seem to shift the mind into conscious overdrive as humans ask themselves that all-important question – so what do I do now? […]
>
> There is a structural, almost logical, reason why mind needs to be so composed that it responds to difficulty and uncertainty in this way. Difficulty and uncertainty imply that the well-learned behavior, the well-oiled habit, will not work and could be dangerous. Instead, a close

> cognitive call has to be made and a behavioral solution chosen online. (Smith, 2005, pp. 266–267)

According to Smith, a general cognitive pattern can be identified in these situations: Habituated patterns of cognition give way to conscious awareness. Conscious awareness of what, however? Following Smith, the object of this raised awareness is the situation itself, along with any behavioural factors, such as goals and strategies, that tie into it (Smith, 2005, p. 267). And indeed, everyday experience underscores that what we above all tend to become more conscious of when we find ourselves in cognitively rough terrain is precisely the terrain itself – that is to say, whatever patterns of interpretation, presuppositions, beliefs, and projections that combine to make up the current cognitive difficulty. In becoming aware of this terrain, we enter into *metacognition* in the classic, literal sense: "any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive activity" (Flavell, 2004, p. 275).

Should the difficulty prove to be easily solved, however, habituated cognition very soon reinstates itself, and our awareness quickly and unceremoniously returns to its everyday duties, much as if the difficulty had never existed. Yet, crucially, even the most simple and commonplace experience of being wrong about something involves passing through some amount of metacognitive awareness. Consider, for example, the cognitive shuffling act visible in the typical exclamation of surprise "What! *I thought that…*" Here, what used to do the cognitive work of reality up until the moment of surprise is already denounced as its own functional opposite, having been passed on to the relative unemployment of something merely thought. The cognitively real, to put things bluntly, has become the cognitively cognitive. A thought has been identified by thought *as* thought; or alternately, a cognition has been re-cognised as a cognition. In its stead, another cognition has taken on the job of reality representation ("Ah! *Here's* the book on Aristotelian metaphysics! I thought I left it in the car …").

From the standpoint of cognitive psychology, to speak of metacognition in this context may appear unorthodox, on the grounds that both the main aspects of metacognition – i.e. knowledge and regulation of cognition – are usually understood to concern cognitions in actual operation, and not cognitions that you have disowned as faulty. But my point is not that we have metacognition in any classic sense of the word with respect to cognitions we have decommissioned as faulty, but rather that the very process of decommission must acknowledge these cognitions as ours to regulate. As has been observed elsewhere, regulation of cognition and knowledge of cognition are related to each other (Schraw, 1998, p. 115). Through this connection, even the smallest, most insignificant cognitive mistake contains a seed of metacognitive growth.[3]

Regarding the kind of cognitive difficulties presently discussed, a final comment is in order. This limited kind of interaction with the unknown generates what cognitive psychologist Stellan Ohlsson (2011, p. 21) aptly calls “monotonic” learning. As the name suggests, this is a kind of learning where everything novel remains basically consistent with the already known. The “novelty” in monotonic learning is thus superficial, not genuine. In contrast, I would reserve the title *epistemic novelty* for what is involved in the more uprooting encounters with the unknown that fit Smith’s description. Epistemic novelty is set apart precisely by its way of challenging the cognitive system to which it is novel. Such instances of cognitive challenge demonstrate the problematic side of human cognition’s operational prejudiciality. Presently, these concerns will be looked at more closely, as the problem of epistemic novelty (PEN).

The Problem of Epistemic Novelty (PEN)

What, then, are the negative implications of facing epistemic novelty, given the inescapable human predicament of having to judge the unknown with sole recourse to the known? Clearly, it is a cognitive *modus operandi* of great economy for the known to cast its reflection on the unknown as long as the two do not differ much. As already indicated, this is the standard case. The apparent drawback of this arrangement comes alive when they do, however, and – needless to say – more so as the difference grows larger. In such contexts of facing epistemic novelty, two things of cognitive significance are likely to happen. We will be wrong, and we will be perplexed. The first point, that we most probably will

1) *make mistakes of interpretation and judgment regarding the unknown*

should be obvious enough, given that the unknown, as stated, here is sufficiently inconsistent with the known to pose a cognitive challenge. These mistakes of interpretation and judgment will then give rise to the kind of cognitive anomalies and contradictions that normally signify that mistakes have been made – in virtue of the cognitive ground rule that there can be no incoherence in that which is counted as reality.[4] Interestingly, the presence of anomalies and contradictions in the cognitive system does not in itself guarantee that we will stop and think things over, and even less that we do so in any progressive sense. It is well known that the psychological discomfort of non-fitting relations between cognitions is often regulated by means that serve to protect the current layout of one's cognitive map rather than revise and develop it (Festinger, 1957, pp. 18–24). The human cognitive system has, as Ohlsson puts it, a "disturbingly powerful ability to sweep anomalies and contradictions under the conceptual rug" (Ohlsson, 2011, p. 358).

That said, however, at *some* point any increasingly obsolete cognitive map or model will be brought to a local service halt precisely because of the emergence of such anomalies and contradictions – or more precisely, the cognitive conflicts that feed them. In my view, it is impossible to specify a single, decisive trigger mechanism for this process. Nevertheless, a number of partly overlapping key factors can be identified that clearly influence the liveness, or metacognitive pull, of a specific cognitive conflict to a given subject. These include, but may not be limited to, the following: a) its degree of irresolvability; b) the personal or general psychological importance of the reality claims involved in the conflicting cognitions; c) pressing extrapsychological circumstances pertaining to these reality claims; d) the amount of attention directed at the conflict; and e) the current level of general cognitive uncertainty.

Because they represent the presence of vested psychological interest, factors *b* and *c* may in fact, unlike the others, have an initial dampening effect, although they certainly contribute to the perceived liveness of any given cognitive conflict once the conflict is gaining foothold in the cognitive system ("Surely, I *cannot* have misplaced my wallet!").

Once a given cognitive conflict has gone live enough to command conscious attention, the habitual cognitive structures that are implicated by it – what Smith calls "the well-learned behavior, the well-oiled habit" – are in operational jeopardy.[5] This circumstance takes us to the second of the above points, namely that facing epistemic novelty, we will most probably also

2) *suffer cognitive derailment and perplexity*.

This derailment and perplexity is experienced in rough proportion to the liveness of the conflict, the nature and scope of the implied cognitive structures, and the extension of any lateral entanglements of the involved cognitions with other cognitions of importance. Perhaps most of all, though, the respective natures of the conflicting cognitions themselves determine the cognitive chaos; for, thanks to the conflict, they are simply unable to do any cognitive work. As they cannot both (or all) represent reality, none of them can, not as long as the conflict stands – for again, there can be no incoherence in that which is counted as reality. (This is not only why, but more importantly also how they cancel each other out.) In consequence, the area of cognitive contention becomes unreachable as far as the intellect is concerned. Thanks to the standing conflict, it cannot be intended in any specific way. There is only perplexity. Thus, the conflict can be said to hold intentionality in a state of suspended cognitive reality loss, in which it can literally go nowhere. (Hence the stunning "sting ray" effect of perplexity, famously featured in the early Platonic dialogues.) Any conflicting reality claims that resist a swift and sometimes even preconscious resolution simply go on cancelling each other out. The resulting state of inability and mental impasse is constitutive of the concomitant experience of derailment and perplexity.

The presence of live cognitive conflict clearly signals that the currently applied cognitive map is inadequate for the situation at hand. To my mind, there can be no greater impetus for the human cognitive system to give its metacognitive monitoring and regulatory faculties a more conscious focus than this perplexing loss of cognitive reality, for through it the decidedly cognitive rather than (operationally) real nature of cognitions is laid bare to the cognitive system that they normally serve *without* calling any attention to themselves as cognitions. Expressed in Smith's words, this involuntary upward shift of awareness is the "thwarting" of "automatic and habitual behavior" itself. Wherever cognitive reality projection fails in this way, there is a more or less forced shift of attention from the nature of the cognised – that is to say, reality – to the cognitions themselves. In terms of cognitive psychology, one could perhaps identify this process as an instance of "breaking set". Psychiatrist Richard Tuch writes

> Cognitive psychologists use the term *breaking set* to refer to the process by which an individual transcends the application of previously developed pattern-recognition schemas and so gains an alternative perspective on novel situations that are exceptions to the expected pattern (Tuch, 2011, p. 769)

Being an offshoot of research on cognitive habituation done at the University of Berlin in the 1920's, the cognitive phenomenon of breaking set remains relatively unexplored in contemporary cognitive psychology (McGaughey, 2008, p. 86). Nevertheless, it seems a promising theoretical resource for exploring the functional interplay between cognition and metacognition as informed by cognitive conflict, seeing that

> what distinguishes breaking set from the everyday realm of problem solving is the previous imbedding of a successful pattern which must first

> be broken before other options can be examined […] In terms of metacognition, this may be analogous to first pushing an override switch before other cognitive processes can be utilized.” (McGaughey, 2008, p. 86)

Needless to say, the perplexing loss of cognitive reality, in the sense discussed here, does not end all perception of reality in any general sense. It does, however, signify a loss of *intellectual* sense (see note 2). The experience is of a habituated process at the very margins of consciousness suddenly being halted and so brought to attention, in some ways resembling the waking up from sleep, in other ways not unlike what happens when a movie projector malfunctions in the middle of a picture show. Our conscious attention is commanded not by any clear and present effects of the halted process, but the abrupt and unexpected loss of these effects. The upshot is an evidently metacognitive motion of mind, an experience-based owning up to a previously less recognised instance of cognitive agency: Clearly, in so far as it can be brought to a halt, the intelligibility of reality must be *our* doing.

In the standard case, we tend to take a very instrumental and *ad hoc* kind of view on such momentary peaks of metacognitive awareness, and happily direct our attention elsewhere as soon as we can. Judging from this behaviour, it would appear that we generally do not like our intellectual command of reality to be negotiable, preferring any sense it makes to be its own doing. Yet, as I will demonstrate shortly, this very kind of metacognitive awareness was highly valued by at least one school of classical philosophy as a condition and motor of cognitive development, and ultimately, a life in existential balance.

Learning and the PEN

Taken together, the elements of error and perplexity discussed above constitute the Problem of Epistemic Novelty (PEN), which can be summed up thus: *facing epistemic novelty, we are likely to be wrong and perplexed*. Speaking broadly, the PEN represents something of a double-edged sword. One the one hand, that the human cognitive system possesses an innate adaptive capacity for dealing with irregularities and outright unknowns is, of course, a good thing. On the other, this cognitive functionality appears to come at the operational cost of error, perplexity and derailment. Ideally, one would like to be able to learn about the unknown without having to pass through error and perplexity on the way – especially in explorative contexts, where both may have critical repercussions. Clearly though, this can only be done in the case of monotonic learning, and explorative contexts, not least those pertaining to Dunér's astrocognitive question (AQ), are not likely to be monotonic in that sense.

As for the cases where there actually is something genuinely novel to face, one could argue that it is only through interacting with the unknown, in the tried and true form of learning, that the PEN can be addressed at all. This argument assumes that the PEN has no general character to address, and consequently, that it must be dealt with case by individual case, simply by making each instance of epistemic novelty less and less novel until, presumably, everything is known and no novelty remains. Such a line of argument represents a serious misunderstanding of the point I am making. Strictly speaking, the PEN is not at all alleviated by learning about the unknown. On the contrary, the PEN arises in the very interaction with the unknown that is the core of learning. Hence, the PEN is nothing if not a general problem. Unless addressed as such, it will reassert itself as long as epistemic novelty does.

The generality of the PEN is visible from the fact that it mirrors the generality of the AQ, as related in the introduction. It meets the "what will happen…" of the AQ head on. It seems a safe assumption, that "what will happen" should we "encounter a totally different environment" – such as, for instance, that of another planet – is that we will be wrong and perplexed (or at least more so than usual). The concern of the present investigation is what can possibly be done about this state of affairs at a general level – that is to say, to repeat the initial question, whether the human mind can be prepared specifically for the unspecified?

Why philosophy?

Assuming that staying ignorant is not an option, it might well be asked if there really is anything that can be done about the PEN. Indeed, if the greater cognitive challenges of everyday life are any indication, simply plodding on with whatever resources one happens to possess may seem the only way to go. Error and perplexity, on this view, simply come with the cognitive territory as the price of progress.

This, however, is where I suggest a turn to ancient philosophy. If knowledge does not address the PEN as much as it merely precludes it from arising again from the exact same conditions (through making a particular unknown known), perhaps something can be said for wisdom?[6] More to the point; if the AQ, as Dunér suggests, can be approached on the basis of what is already available to us as part of the human condition – our having to face the unknown with sole recourse to the known – then why should not any *schools of thought* addressing this predicament be equally relevant to the AQ?

So, what does philosophy have to say? Does it offer any instruction or advice on how we are to handle our epistemic provinciality? Reflections on this matter should be prominent, implicitly or explicitly, in any branch of philosophy that deals with the human condition as a practical problem in need of a practical solution. Such a system of thought, moreover, should obviously promote and ultimately form an integrated part of a certain way of life. Promisingly, these are known hallmarks of ancient Greek philosophy (see e.g. Hadot, 2002, pp. 1–6), where philosophical discourse and way of life are said to have been so closely and organically intertwined as effectively to forbid any modern expression in terms of opposites like *theory* versus *practice* (Hadot, 2002, p. 4). This was especially true of the Hellenistic art-of-living philosophies, the most famous of which are Stoicism and Epicureanism.

The aforementioned two are comprehensive systems of thought and instruction intended to bring about a good, happy life. Given their differing views on what constitutes the good life, their recommendations on its attainment understandably differ somewhat as well. For the present purposes, however, the important thing is that they share the centrality of such prerequisite theoretical inquiry. As basic assumptions go, this one certainly seems plausible – after all, without some general account of human reality, what is good in it, and what is bad, how are you supposed to improve your life? Unsurprisingly, this is the standard approach of ancient Greek philosophy to such matters.[7]

Pyrrhonean aporiagogics

Common sense notwithstanding, the lesser-known art-of-living philosophy of Pyrrhonism disagreed strongly with this line of reasoning.[8] It taught that the root of unhappiness in life lies not in any particular layout of our worldview – needful of wise and proper rearrangement and subsequent diligent implementation – but in our habitual confusing of this worldview with the world itself: our dogmatism. According to Pyrrhonism, it is precisely when we think we know how things really are that we invite unhappiness and inner turmoil into our lives. The reason is that this makes us partial and emotionally unbalanced toward matters of opinion, as well as things and occurrences we tend to consider good or bad by nature, such as wealth and sickness. The prescribed solution of Pyrrhonism is to rid ourselves of all dogmatism, shifting the weight of truth, as it were, from ourselves to the world, and contenting ourselves with the way things appear (Mates, 1996, pp. 89–93, especially sections 1.12, 1.25–30). That way, no judgements, past or present, colour experience; life is faced in its immediacy, with less expectation and reserve. By way of illustration, think of the way a young child, unlike most adults, can suffer a cold without any concomitant notion of having a bad, unwanted experience. Where Stoicism and Epicureanism sought their Archimedean points in reason and pleasure, Pyrrhonism's idea of existential stability and happiness lay in the neutral positionlessness of an utterly open mind.

To cultivate this ambitious end, Pyrrhonism presented an ingenious and highly unorthodox philosophical method, explicitly harnessing the condition of *isostheneia* ("same-strength") between conflicting cognitive elements: i.e., irresolvable cognitive conflict. This is the explicit rationale behind the allegedly all-destroying argumentation of Pyrrhonism, the misapprehension of which has fed the fiction of the hardcore, madcap sceptic so persistent in the western history of ideas. Such mis-apprehensions are uncalled for, seeing that the Second-Century physician Sextus Empiricus – our main textual source on Pyrrhonism – clearly writes:

> But the main origin of Skepticism is the practice of opposing to each statement an equal statement; it seems to us that doing this brings an end to dogmatizing (Mates, 1996, p. 90).[9]

Under the helpful headline "what is Scepticism?", this is elaborated somewhat:

> Scepticism is an ability to set out oppositions among things which appear and are thought of in any way at all, an ability by which, because of the equipollence [*isostheneian*] in the opposed objects and accounts, we come first to suspension of judgement and afterwards to tranquillity.
>
> […] Suspension of judgement is a standstill of the intellect, owing to which we neither reject nor posit anything. Tranquillity is freedom from disturbance or calmness of soul. (Sextus Empiricus, 2000, 4–5)[10]

Indisputably, Sextus Empiricus wants the notoriously contra-intuitive argumentative effort of Pyrrhonism to be understood in the context of a certain, specific series of events:

1) The setting up of cognitive opposition (marked by the *isostheneia* of its elements, forming an impasse, *aporia*, of the mind).
2) The suspension of judgement (*epochē*; from *epechein*: "to hold on" and so, "to hold back", "to check". As seen above, Sextus describes

the *epochē* as "a standstill of the intellect, owing to which we neither reject nor posit anything")

3) The end of dogmatism, yielding a calmness of soul (*ataraxia*).

This series of events describes *the leading* (agōgē) *to and through cognitive impassability* (aporia) that earns Pyrrhonism the designation *aporiagogic*.[11] Apparently, Sextus Empiricus felt the need to explain the origins of this unusual philosophical strategy, for he takes care to describe the Pyrrhonist approach as modelled on a chance discovery. This discovery was the unexpected, spontaneous attainment of *ataraxia* following a failed attempt to establish the nature of reality through philosophical inquiry, ending in undecidability and cognitive conflict (Sextus Empiricus, 2000, 10–11).[12] By this description, Sextus Empiricus communicates that the Pyrrhonean method should be understood as an attempt to re-create the end state of the earlier philosophical failure – i.e. "a standstill of the intellect" from cognitive undecidability and conflict (*epochē* from *aporia*) – with the ultimate ambition to reap the same benefits with respect to dogmatism (*ataraxia*).

It may seem a tad bewildering to try to create states of undecidability with the explicit purpose of progress. Certainly, the strategy does involve destructive elements. On reflection, however, it may not be so strange a notion after all. If our knowledge *is* limited, clearly we need to experience this limitation if we are to outgrow it. If we wish to befriend cognitive progress, holding on to perceived truth may therefore become our enemy. Tellingly, one of the central contentions of the research discipline variously known as pedagogy, learning theory, or philosophy of education – the ultimate object of which, incidentally, is also an *agōgē,* a leading or bringing of people somewhere – is that:

> The greatest obstacle to new learning often is not the student's lack of prior knowledge but, rather, the existence of prior knowledge. Most college teachers know from experience that it's much harder for students to unlearn incorrect or incomplete knowledge, than to master new knowledge in an unfamiliar field (Angelo and Cross, 1993, p. 132).

For good reason, the human cognitive system tends to err on the conservative rather than on the revolutionary side. First and foremost, it seems, the world has to make sense – and preferably, the sense it is already making, thank you very much. This inherent conservatism means that our cognitive structures are disabling as well as enabling. To some extent, they necessarily impede growth in staving off chaos. Yet, it is commonly observed that dogmatism tends to decrease with the accumulation of experience. Given the conservative slant of human cognition, what makes it so? More precisely, what *kind of experience* makes it so? It is certainly not the kind that conforms to the views already employed. It is rather the kind that provides a new angle; some kind of contrast; and, most forceful of all, the kind that thereby induces states of undecidability and perplexity.

Symptomatically, in the academic context where past knowledge is discussed as an obstacle to cognitive progress, a not insignificant portion of research has been directed at the phenomenon of cognitive conflict or contrast (e.g. Gorski, 2009; Tite, 2003; Atherton, 1999). In research on so called "conceptual change" (Posner *et al.*, 1982) – the restructuring of systems of cognition in order to accommodate new, contrasting material – this tendency is clearer still (Başer, 2006; Limón, 2001; Rea-Ramirez and Clement, 1998; Zohar and Aharon-Kravetsky, 2004). Indeed, Vosniadou (2007, p. 2) even speaks of a "'classical approach' to conceptual change", "a leading paradigm […] for many years" according to which cognitive conflict was seen as "a major instructional strategy for producing conceptual change".[13] A basically similar model of epistemic progress is found in Neighbour 2005 (p. 104). The present reading of Pyrrhonism should thus offer a fruitful philosophical framework for educational theories on unlearning and cognitive progress.

On the whole, the inherent conservatism of human cognition plays out as a strong tendency to see things the way we see them until an alternate view intrudes on us. Significantly, this general rule holds only so long as we do not choose to actively seek out and explore cognitive contrasts. The exceptional instrumentality of cognitive contrast, however, is not exhausted by this observation. Not only does it take a good deal of cognitive contrast – a clear and insistent conflicting claim on reality – to uproot a dogmatic view; it may take a good deal of cognitive contrast even to become aware of it in the first place as a *view* and not simply the nature of reality. Educational psychologist William Perry, known for his pioneering study on the cognitive development of Harvard students, recalls the following episode:

> A top student from a good school came to Harvard at a young age, possibly a year too young. Since he had won a regional prize in history, he enrolled in a section of Expository Writing that focused on writing about history. He consulted me in a state of some agitation, having failed three attempts to write a satisfactory response to the assignment: 'Consider the theory of monarchy implied in Queen Elisabeth's Address to Members of the House of Commons in 1601'. 'Look', he said, 'I can tell what she said […] all her main points. I've done it three times, longer each time. But he says he doesn't want that. What is this "theory of monarchy implied" stuff

> anyway? He says to read between the lines. So I try to read between the lines – and huh – there's nothing there.'
>
> The intellectual problem is not too obscure. The student cannot see a theory of monarchy because he has never been confronted with two. Until he sees at least two, a monarch is a monarch and who needs a theory? (Perry, 1988, pp. 148–149)

This is as good an example of everyday dogmatism as one gets. Who needs a theory when the nature of reality is clear as day? A monarch is a monarch! Even though the student may suspect that the Harvard course in expository writing is not entirely to blame for the cognitive anomalies and contradictions that brought him to Perry, his lack of metacognitive awareness holds him back from the developmental insight he needs to complete the course. Note that he cannot simply choose to rise above his predicament. Until he is presented with a contrasting view, he cannot see the view he possesses *as* a view at all. After all, a monarch is a monarch.

Pyrrhonism aims to raise our metacognitive awareness in much the same way an alternate theory of monarchy would raise that of the Harvard student. Through its aporiagogic approach of setting out "oppositions among things which appear and are thought of in any way at all", Pyrrhonism effectively undertakes a force-feeding of human cognition with just the kind of conflicting input that serves the natural development thereof toward mature cognitive flexibility. As Smith observed, conditions of cognitive difficulty and uncertainty wake us from habituated cognition, raising metacognitive awareness. We are strongly encouraged, sometimes even practically forced, to deal consciously with cognitions *as* cognitions in order to sort things out and end the cognitive conflict. Differently put, the presence of cognitive conflict often alerts us to an immediate need to break our current set – if the conflict does not break it *for* us. As will be clarified in the following section, it is capable of both.[14]

Cognition, metacognition and the double construction of the epochē

Recall Sextus Empiricus' claim that scepticism is the ability to counter any cognition ("things which appear and are thought of") with another cognition of equal strength (*isostheneia*). We have looked at the metacognitive effects of naturally occurring, relatively solvable cognitive conflicts, where coherence is quickly regained and cognition soon reverts to its normal, habituated mode of operation with regard to the concerned area of cognition. We have also looked at the more arresting effects of facing genuine epistemic novelty, where the metacognitive impetus is greater and the all-important return to coherence and normalcy is correspondingly more laborious. Now, what if coherence – through the deliberate cultivation of irresolvable cognitive conflict – were not allowed to return at all? What if a new opposition were set up immediately upon any resolution (however unlikely) of the first – and then another, and another? In other words: What if the Pyrrhonean aporiagogic method were to be rigorously applied?

In my judgement, we would eventually arrive at the "suspension of judgement about everything" (*peri pantōn epochē*) that Sextus Empiricus describes as being followed by tranquillity – the end of dogmatism – in discussion of "the general modes of suspension of judgement" (Sextus Empiricus, 2000, 11). This global "standstill of the intellect, owing to which we neither reject nor posit anything" seems the inevitable outcome of any strict application of the program of perpetual cognitive conflict: perplexity and undecidability can only grow until everything is encompassed by the cognitive stranglehold of perplexity. Most probably, though, the program will be deliberately abandoned at some point before this happens.

On this point, some observations are in order, which might seem at first merely tangential to the matter at hand. As it happens, these two different outcomes of the Pyrrhonist program may serve to illustrate a certain ambiguity inherent to the Pyrrhonean conception of the *epochē*, an ambiguity that any appreciation of the same program should offer some account of. Paradoxically, it turns out that this ambiguity allows for both outcomes to be regarded as instances of global *epochē* under certain circumstances. How can this be? Should not giving up the program be the opposite of staying on?

The heart of the matter is that Sextus Empiricus, as is well known, alternates between an active, normative construction of the *epochē* and a passive, causal one. On the one hand, he treats the *epochē* as an intellectual reservation, that is to say an act of intentional, rational deliberation; on the other, as a state of mental suspense quite inseparable from perplexity, resulting directly from the *isostheneia* of opposing cognitions (Thorsrud, 2009, p. 128). It appears, moreover, that he does not ascribe any great significance to the difference.

The mere acknowledgement of this practice does not make things much clearer, however. Rather, it underscores the interpretative difficulty: It is far from clear in what way an intellectual reservation could possibly be considered to be as functionally intertwined with the state of perplexity as Sextus Empiricus' seemingly indiscriminating treatment of them minimally suggests. For one thing, they are of fundamentally different orders of cognitive function – the one is an action, and the other a reaction. In addition, for all its apparent humility, an intellectual reservation still comes from an "outside" position of cognitive mastery. Holding reality in check by abstraction, it is congruent with a conscious exercise of intellectual authority and expertise. The state of perplexity, on the other hand, is in many ways the inverse. It is an "inside" experience, and one not of holding reality in check but being *oneself* held in check – by reality, as it were. There is, moreover, no abstract distance to the state of perplexity. It is reality *un*checked, reality thrown abruptly off its cognitive rails. Contrasting the noble, grey-templed air of intellectual reservation, perplexity merely signals plain inability. It is the timeless mark of the learner and the fool.

Thus, at first look, the two constructions of the *epochē* appear quite different: The one is in control, the other is out of it. Yet, as I will presently argue, they are indeed functionally intertwined. The model of their relation, I suggest, is quite simply the relation between metacognition and cognition: The active, normative construction represents the metacognition of cognitive conflict, and the passive, causal one represents the cognition thereof. This interpretation is supported by the clear textual presence of cognitive conflict on both constructions (see e.g. Sextus Empiricus, 2000, 10–11, especially the parable of Apelles the painter, paralleling the passive account given under the heading "what skepticism is", which has been quoted above). The double presence of cognitive conflict is reflected, moreover, in the statement that the *epochē* is said to come about "*to put it rather generally* – through the opposition of things" (Sextus Empiricus, 2000, 11, italics mine).

My suggestion is, in short, that the Pyrrhonean *epochē* comes either as a *direct effect* of cognitive *isostheneia* – what has been described above as the stunning "sting ray" state of perplexity – or something amounting to an *acknowledgement and regulation* thereof: "they came upon equipollent dispute, and being unable to decide this they suspended judgement" (Sextus Empiricus, 2000, 10). (The avid reader will remember that these two faculties – knowledge and regulation of cognition – are considered to be the two pillars of metacognition.) Importantly, though, while the active construction of the *epochē* involves metacognition, it is not metacognition *per se*. Metacognition, in alarming resemblance to the modern-day Husserlian *epochē*, can be initiated more or less entirely at will (see e.g. Husserl, 1973, §15).[15] The active *epochē* cannot, for it demands the supplementary presence of cognitive conflict. If matters were otherwise, not only would Pyrrhonism become incoherent with itself (the active and the passive construction of the *epochē* would no longer agree on the primacy of cognitive conflict), it would also make a claim that swears against the common experience of cognitive progress; namely, that one can simply will oneself out of dogmatism and become more cognitively flexible without any resort to contrasting views.

As the case of the Harvard student shows, things do not happen that way. On the contrary, the presence of cognitive conflict fills several indispensable functions. First, cognitive conflict is needed to signal that there is a problem at all. In the case of the Harvard student, this is the conflict between the *expected* outcome of an approach that had proven successful in the past, and its *actual* outcome in the new setting of Harvard. Next, cognitive conflict is needed to pinpoint the problem further. In this case, the critical element was the student's lack of awareness of the theory of monarchy he was already dogmatically employing in taking the nature of monarchy for granted. As Perry observed, "[t]he student cannot see a theory of monarchy because he has never been confronted with two." This illuminating confrontation is the second function of cognitive conflict.

The mere awareness of fact, however, does in no way equal a mature cognitive integration of the full reality that this fact may imply. The third function of cognitive conflict is to drive this process of cognitive integration forward. In the absence of a live cognitive conflict, there is simply no bite to the unlearning that such integration requires. It does not help just to pinch oneself mentally, telling oneself that "as an aware subject, I hereby reserve myself from this opinion/theory of monarchy". Rather, the situation is quite the obverse; the opinion needs a cognitive counterweight precisely because it *is* mine, a counterweight after the measure of its importance to me. The solipsistic heroism of the mental self-pinch is therefore, in the end, of little real consequence. Unless the self-pinch itself falls back on some state of cognitive conflict for purchase, it is completely immaterial to the issue of growing metacognitive awareness and fighting dogmatism. The crucial contribution of the exercise of cognitive agency in this context is therefore not the willful effort of performing intellectual reservations as such, but that of actively seeking out and setting up contrasting views with which to feed such reservations. My suggestion is that the active component of the active *epochē* in Sextus Empiricus is a reflection of this reality, and not any need for heroic wielding of willpower *per se* in the overarching struggle to overcome dogmatism.

On a more general note, regarding the functional intertwining as such of the two constructions of the *epochē*, it is beyond dispute that conscious thought is coupled to other, more habituated modes of cognition involving concrete, everyday practicalies.[16] It is almost trivially true, for example, that cognitive challenges raise our awareness in the way described by Smith at an earlier point in this chapter. But things can go in the other direction as well. As Perry's account suggests, what begins as a rational deliberation need not end as such. (Remember, the student consulted Perry "in a state of some agitation".) Recall that, as observed earlier, attending to a cognitive conflict may in fact *increase* its liveness. An agreeable session of armchair speculation on the nature of reality can suddenly turn into an ontological freefall of undecidability and perplexity. A lofty and clear-headed intellectual reservation inspired by the apparent relativity of morals may well lead to perplexity regarding *one's own* moral values. A Humean investigation into the nature of the self may be far less relaxing than a Cartesian one, and so on. These are scenarios well known to many a beginning philosopher and ones of great metacognitive growth potential – it seems that some ideas must be torn from one's hands before they can be recognised as having been held at all.

Evidently, as part of the same cognitive system, the faculty of rational deliberation is naturally "functionally intertwined" with the less volitional and more habituated subsystems of cognition. More importantly, metacognition is of course, in the end, nothing but a manner of cognition too. The cognitive act (or reaction) of stepping outside cognition into metacognition still remains "inside" cognition in a wider sense – for it is inescapably an act (or reaction) of cognition. Thus, the cognitions that are employed in metacognition may be thwarted just as well as those employed in cognition. It is only to be expected, then, to see this connection reflected in the double construction of the *epochē*. In the same way that metacognition can be said to distance itself from cognition in order to have it as its object of knowledge and regulation, the normative construction of the *epochē* can be said to distance itself from the causal one – which is the direct, "sting ray" effect of cognitive *isostheneia* – in order to deal with it. Likewise, just as metacognition is still a kind of cognition and as such always vulnerable to the emergence of cognitive conflict, so the perpetual possibility of the passive *epochē* is contained in the active one.[17] In fact, as will be demonstrated below, the Pyrrhonean program depends on it. If metacognitive conflict were not *cognitive* conflict, the program would not be able to start off.

More than anything, cognitive conflicts marked by the *isostheneia* of their elements are meant to be precisely the kind of cognitive conflicts that become more live to us the more we attend to them. This observation brings us right back to the riddle of the two different outcomes of the Pyrrhonist program that this section addresses: global *epochē* by staying on it, and global *epochē* by giving it up. The question, again, was: How can giving up the program possibly go under the same name as staying on it? The answer is simple: Giving up the program can also count as a "suspension of judgement about everything" (*peri pantōn epochē*) in so far as it involves the conscious, intentional acknowledgement of cognitive *isostheneia* (Sextus Empiricus, 2000, 10–11, especially the parable of Apelles the painter). Rather than offering refuge, the conscious act of giving up may then in fact effectuate a *return* to perplexity, for it is still a way of attending to the conflict in question. To escape this predicament, it is better to withdraw one's attention and simply walk away.

This ensnaring paradox of the Pyrrhonist method makes Heracles' meeting with goddesses Aporia and Eris (Impassability and Discord) in Aesop's *Fables* strangely instructive – especially considering that together, the pair make for a description of cognitive *isostheneia* as apt as any:

> Heracles was making his way through a narrow pass. He saw something that looked like an apple lying on the ground and he tried to smash it with his club. After having been struck by the club, the thing swelled up to twice its size. Heracles struck it again with his club, even harder than before, and the thing then expanded to such a size that it blocked Heracles' way. Heracles let go of his club and stood there, amazed. Athena saw him and said, 'O Heracles, don't be so surprised! This thing that has brought about your confusion is [Aporia] and [Eris]. If you just leave it alone, it stays small; but if you decide to fight it, then it swells from its small size and grows large.'(Gibbs, 2002, fable 534)

Speaking metaphorically, the evident morale is that sometimes the only way forward is backward; or in terms of cognitive progress, sometimes the only way to learn is to unlearn. This could well be made the Pyrrhonist slogan *par excellence*. It is the designated job of the aporiagogic method – Aporia and Eris, if you will – to make us experience the full extent of its truth.

In summary, the transformative potential of losing one's habitual cognitive ground to perplexity, as described in the earlier sections, is beyond dispute. Nothing sends such a forceful signal to the human cognitive system that something is awry with our cognitive map as the emergence and persistence of cognitive conflict. The Pyrrhonean *epochē* represents the methodological harnessing of this cognitive function; on the level of cognition *and* metacognition, on the inside as well as on the outside of the conflict. The picture is complicated, however, by the inescapable circumstance that the Pyrrhonean program by definition is metacognitively undertaken. The intentional setting up of cognitive conflict is after all a *method*, an intellectual exercise. The cognitions set up in oppositions are initially held in check by the awareness of this very context. Accordingly, the *epochē* necessarily starts off on the active construction. As observed above, this construction can go two ways. Simply put, it either catches hold of the cognitive system hosting it, seemingly reversing the initial order of agency ("the Heracles effect"), or it does not.

In the latter case, the active *epochē* comes in the overly cerebral form of a learned admission of ignorance or undecidability. The full implications of such admissions are typically kept separate from the actual cognitive reality of the admitter by the inherent "upward" momentum of abstraction – that is to say, by the subject staying stoutly *meta*cognitive – and this compartmentalization serves to dress these admissions up as expressions of expertise rather than inability. Here, the cognitive growth potential is very small, for there can be no real *thwarting* of cognitions that are already held in check by abstraction. Accordingly, there is no direct loss of, and subsequent owning up to, any cognitive reality. Strictly speaking, metacognitive awareness cannot grow where there is nothing to own up to except for that which is already mine by cognitive default; namely, my ideas, views and opinions. It is precisely as perceived realities, not perceived ideas (or views, or opinions), that they need to be owned up to if there is to be any growth pertaining to them. In other words, the cognitions have to be in a state of actual cognitive employment if there is to be any point in recognizing them as cognitions. Thus, just as conventional wisdom suggests, staying adamantly in the abstract is an equally formidable shield against cognitive growth as never venturing there at all.

In the former of the above cases, however, the one where the order of agency is actually reversed, the active *epochē* effectuates a tumbling out of metacognition and abstraction to cause a passive one. If the cognitive conflict migrates from the interplay of whatever cognitions are cognised in the active *epochē* to that of the cognitions doing this cognizing, or to that of whatever the cognitions cognised in the active *epochē* themselves cognise whenever they are not held back by abstraction in the active *epochē*, the intellectual reservation of metacognitive conflict is transformed into the perplexity of cognitive conflict. This, again, is the "Heracles-effect" depicted by parable 534. Whereas cognitive reality was initially held in check by abstraction, it is now returning the favour by holding abstraction in check by perplexity. For reasons already stated, this is the point where transformative work can be done – this is where human cognition comes to practical terms with its own boundaries. Were it not for this evident cognitive tie to the passive construction, the active construction of the *epochē* – and along with it, the Pyrrhonist method *as a method* – would be hard indeed to associate with any genuine unlearning, and thus, by implication, any genuine, non-monotonic cognitive progress.

Pyrrhonean ataraxia and the PEN

The main point so far is twofold. First, the cognitive system that forms and regulates our inner representation of reality handles difficulties, such as conflicting cognitions, by breaking out of habituated cognition and raising metacognitive awareness of the constituting factors of the problem. Second, the Pyrrhonean technique of cognitive conflict force-feeds this metacognitive faculty of the human cognitive system with fabricated cognitive conflicts to foster an increase of metacognitive awareness, with the explicit aim to put an end to our seemingly innate tendency to put cognitive map before cognitive territory: i.e., our dogmatism. (One could say, perhaps, that it provides the opportunity to learn from “fake” mistakes.) What remains to discuss is how a Pyrrhonean attack on dogmatism might address the PEN.

To recapitulate, the general idea is that any philosophic technique that arguably effectuated a non-dogmatic, open mindset for practitioners in the art-of-living context of Ancient Greece should be able to do the same for somebody analyzing, say, the mathematical configuration of a radio signal that just might be an interstellar message, or for somebody trying to make sense of a couple of bewildering Martian rock samples; or, for that matter, for somebody facing an artefact, environment or predicament more or less entirely unknown. Differently put: just as the AQ, as Dunér argues, can be scientifically approached on the basis of a general appreciation of human behaviour with regard to the unknown, predominantly informed by the recorded history of earth exploration, so should the PEN retain its general structure regardless of the nature and context of the epistemic novelty itself – and more to the point, so should any successful regulation of it.

Admittedly, the exploration of the double construction of the *epochē* took us on something of an exegetic detour, at least in the sense that the initial question of the past section is still alive: What if the Pyrrhonean aporiagogic method was to be rigorously applied? Notably, the Pyrrhonean recipe for conquering dogmatism involves, much like the PEN, cognitive conflict, perplexity, and cognitive reality loss on a defining level. Pyrrhonism, though, adds some extra twists, by virtue of the *intended* irresolvability of the conflicts induced, combined with the double construction of the *epochē* and the ensnaring paradox which results from it. Under such circumstances, the metacognitive pull of perplexity is bound to be more frequent, more powerful and more extended in time than in natural state encounters with epistemic novelty, making the aporiagogic method a virtual pressure cooker of metacognitive incitement.

Precisely how, however, does Pyrrhonism "arguably effectuate" a non-dogmatic mindset? Is there some kind of general anatomy to the transition from *epochē* to *ataraxia* that lends itself to description? Sextus Empiricus is frustratingly silent on this point.[18] This silence notwithstanding, it goes without saying that the transition, in so far as there is one, must pass through the state of global *epochē*. (How else could the calmness of soul that is *ataraxia* possibly obtain in the general manner consistent with the Pyrrhonean trademark of settling for what appears as opposed to going after the nature of reality?) But what, then, is the significance of the global *epochē*? More to the point, what is its *cognitive* significance – and by extension, what is its ultimate relevance to the PEN? In the present, concluding section of this chapter, an answer to this cluster of questions will be sketched.

First of all, one must recognise that the attainment of *ataraxia* is a subtraction from, not an addition to, one's cognitive makeup. If dogmatism may be summarized roughly as the cognitive habit of unwittingly treating one's worldview as the functional equivalent of the world itself, the attainment of *ataraxia* is simply the subtraction of dogmatism from cognition. Now, at the start of this chapter, I referred to the cognitive ground rule that "there can be no incoherence in that which is to be counted as reality". This ground rule may be restated thus: The only significant reason a cognition is at all open to opposition is because it is not itself reality. Accordingly, cognitions effectively demonstrate this their not being reality precisely in the activity of conflicting and cancelling each other out. The upshot is that everything implicated by the global *epochē* is thereby demonstrated to be *something other than reality*.

This does not mean that anything implicated by the global *epochē* is necessarily false. The phrasing "other than reality" here simply indicates the difference between cognition and reality that dogmatism serves to gloss over by its very nature. The global *epochē* thus pulls apart what dogmatism presses together. This, in turn, makes the cognitive habit of treating cognition as reality cognitively identifiable as what it is: a cognitive habit.

The cognitive significance of the global *epochē* may perhaps also be expressed in the following way. Elsewhere in this chapter, I have made the observation that if our knowledge is limited, we need to experience this limitation in order to outgrow it. In a way, the global *epochē* is precisely that: the needful experience of cognitive limitation as a means to pass beyond. The inherent limitation of any cognition is that it can be opposed by another cognition. On a larger scale, the inherent limitation of any worldview is that it can be opposed by another worldview. To varying degrees we are all aware of this fact *qua* fact; but we do not often experience the full reality of it on any serious or alarming level. When we do so, our relief in putting the experience behind us tends to be directly proportionate to the challenge it posed to our world. This is roughly the kind of experience the global *epochē* provides. It is an experience of cognitive undecidability that does not relent – the cognitive equivalent of a full frontal collision.

Needless to say, the global *epochē* is a psychologically challenging state to suffer. It can produce anxiety, denial, repression, perhaps even some sort of breakdown. While these risks should be acknowledged, in my view they offer no indication that the Pyrrhonean program misses its mark. On the contrary: they demonstrate that Pyrrhonism is doing what life itself is doing to us all from time to time – only a bit more so. Most of us know from personal experience that personal growth may be partly unpleasant or even hurtful, in that various attachments and pet conceptions about ourselves and the world may have to be left behind. As Aeschylus so masterfully observed, wisdom tends to come "in our own despite, against our will" (transl. Hamilton, 1942, p. 257). Please note that I am not saying that all pain is good pain. My point is rather that all pain is not all bad – that there is, if you will, such a thing as cognitive growing pains. As has been argued throughout this chapter, cognitive contrast, perplexity and cognitive reality loss is *how cognitive development happens normally*; this is the given order of things that the Pyrrhonist program hacks into and aims to manipulate to a developmental advantage.

In virtue of the separation of reality and cognition effectuated by the global *epochē*, the aporiagogic program of Pyrrhonism holds the potential of a truly global metacognitive insight into the nature and limit of cognition, infusing the "world-ness" of my worldview with the responsibility and ownership of "view-ness" and/or "mine-ness". This, then, would be the cognitive outlook of the Pyrrhonean end state called *ataraxia*. It is one perfectly balanced between the extremes of subjectivity and objectivity, not in the sense of having one foot in either kingdom, but in that of having no foot anywhere – for both these standpoints have by this time revealed themselves to be dogmatic in the global *epochē*. Differently put: The standpoint of *ataraxia* is a kind of subjectivity that does not explode subjectivity into a reified, objective position, because this would require the unfettered employment of a cognitive map immune to opposition. It is a subjectivity that is literally *not real*. However, it is not unreal either, for that, too, would require "the unfettered employment of a cognitive map immune to opposition." Thus, the ataraxian standpoint can assign no privileged epistemic status to anything, not even, for example, the nature and content of mental states.[19] Hard to envisage as it may be, it is a subjectivity unopposed to objectivity.

Now, it might well be asked how this "separation of reality and cognition" that is the global *epochē* can possibly make the "cognitive habit of treating cognition as reality" *cognitively* identifiable as a cognitive habit, seeing that cognition is supposed to be that which is locked in global perplexity by the equally global *epochē*. Are there thus two sets of cognition at large in the global *epochē* – one that is encompassed by it, and another that encompasses the *epochē* in turn, turning it into an insight? The short answer is that there is not, unless cognition and metacognition should count as separate entities. If we grant that a singular instance of *epochē* is likely to provoke metacognitive acknowledgement and attempted regulation, the same should go for the global one. The basic insight of the *epochē*, singular or global, is identical to that I have already claimed of cognitive conflict, namely that "in so far as it can be brought to a halt, the intelligibility of reality must be *our* doing." Similarly, the upshot is, in both cases, "an experience-based owning up to a previously less recognised instance of cognitive agency."

This understanding of the *epochē* and the state of *ataraxia* suggests an interpretation according to which Pyrrhonism is motivated less by truth than *the truth about truth*. Interestingly, this view retains Pyrrhonism's eudaimonist concerns (Vogt, 2011, p. 9), i.e., its transformative, developmental art-of-living focus, while simultaneously not abandoning its philosophical orientation toward truth in a general sense. Surely, life without dogmatism – the employment of a worldview that is fully and maturely owned up to – makes for easy, open-minded, cognitively flexible living. Although one could complain that such a goal improperly diverts Pyrrhonism's attention from the discovery of truth itself – the *tranquility charge* (as coined by Vogt, 2011, p. 1) – I am inclined to see such a charge as unsupported. After all, a promotion of metacognitive awareness can hardly be considered to come at any real expense of truth. The ultimate object remains the same, at least in a biographical sense (and what other sense truly matters?). Along the same lines, it also becomes less controversial how the attainment of the eudaemonist goal of *ataraxia* can be claimed by Sextus Empiricus (5–6, 10–11) to satisfy the appetite for truth. On this reading, *ataraxia* is related to truth as the truth of truth is related to truth; or, in the terminology of cognitive psychology, what was sought by cognition was found by metacognition. There is thus no real conflict between the orientation toward truth and the orientation toward *ataraxia*, just as there – for similar reasons – is none between the two constructions of the *epochē*.[20]

Once this hierarchal dynamics of cognitive development is recognized in the study of Pyrrhonism as it is in everyday life, an interpretation of Pyrrhonism as promoting cognitive flexibility is within reach – or at least within eyesight. While such an interpretation cannot be fully fleshed out here, it is certainly promising to note how cognitive psychologist Stella Vosniadou and Christos Ioannides explicitly associate increased cognitive flexibility with age and maturity on the one hand, and the development of metaconceptual awareness on the other:

> […] it is important not to overlook the fact that with increases in age […] we not only have a restructured system, […], but […] also a more flexible system, a system that makes it easier to take different perspectives and different points of view. One of the limitations of conceptual change research is that it has paid little attention to the development of cognitive flexibility and metaconceptual awareness. There is no doubt that adults […] have a different conceptual organization of science concepts than elementary school students, but, they also have a more flexible organization, one that allows them to take different points of view.
>
> An important determinant of cognitive flexibility is, in my opinion, the development of metaconceptual awareness. It is difficult to understand other points of view if you do not even recognize what your own point of view is. […] Increased awareness of one's own beliefs and presuppositions and of the fact that they represent interpretations […], is a necessary step in the process of conceptual change (Vosniadou and Ioannides, 1998, pp. 1226–1227).

As for the relevance to the problem of epistemic novelty (PEN) of this general understanding of the Pyrrhonist program, it is quite clear that when confronted by epistemic novelty, a person who has attained *ataraxia* should enjoy several advantages over someone who has not. Presently, we will turn to a brief review of these advantages. Remember, the general idea of coupling Pyrrhonism, via the PEN, to Dunér's astrocognitive question (AQ) was that

> any philosophic technique that arguably effectuated a non-dogmatic, open mindset for practitioners in the art-of-living context of Ancient Greece should be able to do the same for somebody analyzing, say, the mathematical configuration of a radio signal that just might be an interstellar message, or for somebody trying to make sense of a couple of

> bewildering Martian rock samples; or, for that matter, for somebody facing an artifact, environment or predicament more or less entirely unknown.

At the most basic level, the reason that this should be so is that a person having attained *ataraxia* will generally be more mindful not to treat her own views as the nature of reality. In unfamiliar circumstances, such a person will be less inclined to assume cognitive analogies between the present situation and past ones, and if she does make such an analogy, she will be more attuned to its inescapable epistemic provinciality. In the case of the radio signal, such a mindset may be of help in avoiding the interpretative pitfall of getting mentally stuck on any particular pattern of possible significance, as well as that of prematurely settling for the most obvious ones, viewed from the perspective of past Tellurian contexts of experience. Along the same lines, the awareness of epistemic provinciality should also bring about a moderation of the assertive scope of any "mistakes of interpretation and judgment regarding the unknown", i.e. it should cushion the negative effects that form the first problematic pillar of the PEN. In all likelihood, it will also, to some extent, make such mistakes less frequent. Regarding the second pillar of the PEN, there should, in consequence of the above, be less perplexity and cognitive derailment. If there is less dogmatism in the way of emotional investment and conceptual rigidity in the cognitive economy for the unknown to thwart, it follows that there is a corresponding decrease in the amount of bafflement experienced, should reality first seem to invite one interpretation and then suddenly start to behave in a "wrong", unexpected way.

The case of the bewildering Martian rock samples follows the same pattern. Here, it may be helpful, for example, to be able to cognitively stave off any obvious stone-ness of the samples – not as an expression of any factual concern, but as the practice of the interpretative virtue not to close the case prematurely. A predisposition to any one interpretation runs the risk of serving to gloss over elements that may be decisive from the standpoint of an alternative analysis. There is of course also in this case an advantage in knowing full well that one actually engages in cognitive analogies should one at some point choose to do so. As a result, the cognitive interaction with the unknown will be less dramatic: Anyone can lose a conception identified as a conception, or a theory identified as a theory. It is a different story to be robbed of a certain cognitive *reality*. If the level of cognitive ownership and responsibility can be raised beforehand, the latter should be less likely to happen, and consequently, so should states of perplexity and cognitive derailment.

The final case – facing "an artifact, environment or predicament more or less entirely unknown" – represents a more extreme version of the PEN than the previous two, given the lower level of general cognitive purchase on the object(s) of cognition. Here the benefit of a Pyrrhonean ability to stay with the appearances (or rather, the *appearance* of appearances, if you will) is perhaps less about how any emergent patterns of interpretative cognition are handled than seeing to that they are given a fair chance of emerging at all. To this end, minimising the duration, amplitude and number of instances of perplexity is paramount. The neutral, balanced positionlessness of post-ataraxian cognition fits this need hand in glove. Clearly, if no opinions or interpretations are added to experience as the nature of reality, there can be very little in the way of said reality for experience to uproot.[21]

Unlike the previously discussed non-strategy of simply trying to learn about the unknown in a direct, come-what-may fashion, the Pyrrhonean program addresses and reduces the problematic elements of facing epistemic novelty at their cognitive roots while in no way hampering the learning process that, given time, serve to make particular epistemic novelties less novel. The interaction with the unknown is unhampered, but the PEN is reduced. My conclusion, then, is that on the basis of the above observations and arguments, Pyrrhonism may well constitute something that in many respects amounts to a method for specifically preparing the human mind for the unspecified.

Acknowledgements

The author thanks David Dunér for initiating the discussion, and the Pufendorf Institute for Advanced Studies at Lund University for hosting the inspiring multidisciplinary conference "The History and Philosophy of Astrobiology" on Tycho Brahe's island Ven in September 2011. The author also sends his thanks to two anonymous referees for their comments and suggestions on an earlier version of this paper.

Notes

[1] In this chapter, I use the terms "Pyrrhonism" and "Pyrrhonian scepticism" to refer to the philosophy expounded by Sextus Empiricus (c. 200 C.E.) in his *Outlines of Pyrrhonism*. The term "aporiagogic" is constructed from the Greek *agōgē*, meaning a leading to or a bringing; and *aporia*, meaning impasse, difficulty, perplexity, or the state of being at a loss. "Aporiagogic" signifies a path into perplexity and obstruction as well as a path through the same to an ulterior end. The second clearly presupposes the first: these two steps of leading into and leading through *aporia* expresses the Pyrrhonian method *in nuce*.

[2] In this chapter, I use the word "cognition" and its derivatives to denote the same things as Festinger (1957, p. 3): "any knowledge, opinion or belief about the environment, about oneself or about one's behaviour". Cognition in this sense – and in keeping with the Pyrrhonean method – is operationally blind to any significant separation of thought from (interpreted) sensory experience. A similar view is employed in contemporary research on cognitive conflict and conceptual change (Heywood and Parker, 2010, p. 66).
[3] I want to clarify that the "cognitive shuffling act" spoken of here should not be understood as an expression of any philosophical assumption to the effect that reality has to be the one or the other; cognitive or real. Obviously, the very idea of metacognition constitutes a rejection of such a strict duality. Rather, this shuffling act is symptomatic of an apparent practical limitation of our system of cognition, namely that it has trouble focusing on both aspects simultaneously. When a given cognition is working, we typically tend to focus on its object rather than its nature or operational status, and when it is not, it is the other way around.
[4] This is obvious from the fact that the only possible way to detect cognitive errors "internally" – i.e., from within a cognitive system in actual operation – is through the cognitive conflicts these errors invite (Ohlsson, 2011, pp. 210, 338–342).
[5] I am well aware of the apparent contradiction in allowing that a given cognitive conflict can become "live enough to command conscious attention" by virtue of having been the object of such attention (Factor *d*). Nevertheless, this seems to me a very real possibility. It is common to our general understanding of problems that they tend to grow if brooded upon, and that this growth in turn feeds the brooding.
[6] The ancient Greek construction *filo-sofos* literally means "lover of wisdom".
[7] "The usual orientation of Greek philosophy is to try to achieve a desirable attitude to life by means of *successful theoretical enquiry*, in which one comes to understand certain of the ways in which things present themselves to us as true to their real natures, and certain others as false" (Bett, 2000, p. 178, italics mine). A similar observation is made by Martha Nussbaum (1994, p. 24): "all the theories we shall study (again, Skepticism excepted) insist that the ethical theories must cohere with our best theories in other areas of inquiry—inquiries about nature, for example, about psychology, about the relationship between substance and matter."
[8] The interdisciplinary exposition of Pyrrhonism that follows may come across as unorthodox, and perhaps even a bit unphilosophical. Nevertheless, I consider it a valid and necessary approach, if the transformative ("agogic") nature of Pyrrhonism is to be duly respected. That said, the present chapter is, for obvious reasons, neither exhaustive nor conclusive in its treatment of Pyrrhonism. The aim is merely to indicate possible lines of research. Whether I read something into Pyrrhonism that is not there is, in all probability, best judged by others. Pierre Hadot was undoubtedly correct in warning us that "it is impossible to remain faithful to a tradition without taking up again the formulas of the creator of this tradition; but it is also impossible to use these formulas without giving them a meaning that the previous philosopher could not even have suspected. One then sincerely believes that this new meaning corresponds to the deep intention of this philosopher" (quoted in Hadot, 1995, pp. 6–7). On a general note, a problem with the study of Pyrrhonism is that Sextus Empiricus, probably in order to evade

charges of inconsistency, endeavours to make his presentation of the method of cognitive *isostheneia* congruent with the claimed results thereof. This means that he necessarily speaks "without belief"; that is to say, from a *post-ataraxia* position where dogmatism is already eliminated. Thus, we arrive at the puzzling state of affairs, that *because of the method*, Sextus Empiricus cannot, in the eyes of a casual reader, wholly affirm the validity of the method. This is nowhere as clear as in the case of the connection between the *epochē* and the attainment of *ataraxia*, the connection that the Pyrrhonean practice arguably builds on. But why should the theoretical exposition of a method demand a textual consistency with the ultimate goal of that same method? This does not necessarily obtain even if the one doing the exposition has herself mastered the method. Everything else unconsidered, one might simply choose to speak in a certain way as opposed to another. As the case of Pyrrhonism demonstrates, the most goal-consistent exposition of a method need not be heuristically optimal from the standpoint of a listener. Thus, "staying true" to the end in the name of a private consistency that others find hard to fully fathom may well become an effective betrayal of that same end in a protreptic sense. In conclusion, any clear discussion of why Pyrrhonism actually might work as described in Sextus Empiricus does well, in my view, to do so from a position external to the method itself. Otherwise, it may find little or no method to discuss.

[9] The original Greek reads *systaseōs de tēs skeptikēs estin archē malista to panti logō logon ison antikeisthai; apo gar toutou katalēgein dokoumen eis to mē dogmatizein* (Mutschmann, 1912, Section 12). It is unclear why *archē* should be translated "origin" rather than e.g. "formative principle", given the presence of *systaseōs*. Likewise, it is at least questionable why, in the final clause, the Greek conjunction *gar* should not be translated "for", by virtue of its argumentative function. Nevertheless, this translation serves the present purposes by retaining the verb "dogmatize" from the Greek original.

[10] *esti de ē skeptikē dynamis antithetikē fainomenōn te kai nooumenōn kath'oiondēpote tropon af'ēs erchometha dia tēn en tois antikeimenois pragmasi kai logois isostheneian to men prōton eis epochēn, to de meta touto eis ataraxian…'epochē' de esti stasis dianoias di'ēn oute airomen ti oute tithemen. 'ataraxia' de esti psychēs aochlēsia kai galēnotēs* (Mutschmann, 1912, sections 8–10). I use two different translations of Sextus Empiricus' *Outlines of Pyrrhonism* in this chapter: Sextus Empiricus (2000) and Mates (1996). However, I use the latter only once. The reason for this exception is stated in note 9.

[11] See note 1 on the use of "aporiagogic". The instrumental nature of the Pyrrhonean method is underscored by its explicit declaration that the precise nature of the arguments is unimportant so long as they serve their purpose (Sextus Empiricus, 2000, 8, 216).

[12] As the quoted sources make clear, the unintended and unexpected occurrence – the attainment of tranquillity – is impossible to understand other than as an attack on dogmatism.

[13] Importantly, this speaking about the "classical approach to conceptual change" in the past tense should not be understood as indicative of a subsequent abandonment of cognitive conflict as an instructional strategy. Rather, some difficulties of implementation have, quite naturally, been identified. One such difficulty grows

out of the circumstance, that the meaning a given cognitive conflict is taken to have by students depends very much on their individual outlooks; another concerns more practical circumstances pertaining to the classroom format (Limón, 2001). Also, other major factors of conceptual change have been recognised as well, not least the role played by intention (see e.g. Sinatra and Pintrich, 2003). This development makes the Pyrrhonean double construction of the *epochē* a parallel of some heuristic promise, seeing that the latter assigns intention a certain sphere of influence as well.

[14] I am speaking here of the double construction of the *epochē*, which on my reading is closely tied to cognitive conflict, and more importantly, where one construction is active and deliberate while the other is not.

[15] While this point deserves a separate investigation, I am compelled to hold in advance that the Husserlian *epochē*, in disregarding the technique of *isostheneia*, stands degraded next to the original Pyrrhonean one.

[16] Otherwise, why would cognitive dissonance – as Festinger (1957, p. 18) observed more than five decades ago – be a motivator of human action comparable to hunger? Of course, this is not to say that the sphere of action and the sphere of rational deliberation are identical. Hume's (1777, SBN 39) famous separation of himself as agent from himself as philosopher is a classic move; yet, having made this separation, why would he still, as a philosopher, "want to learn the foundation of this [agential] inference" (*ibid.*), if the two were entirely separate? This cognitive hierarchy does not mean, by the way, that there can be no perplexity in a context of abstraction and general "outsideness", such as mathematics. Of course there can. The cognitive functionality that is thwarted by conflict does not have to be deeply habituated in a sense that rules out reflective contexts. Metacognition is "meta" to *any* employed cognitions thwarted by conflict, whatever the context. This includes higher-order cognitions cognizing other cognitions. Even though applied mathematics can be said to be "outside" its subject matter, it has an "inside" which may or may not give rise to cognitive conflict and perplexity: the employment and manipulation of a certain set of symbols in accordance to a certain set of rules.

[17] This is not to say, of course, that the active *epochē* always begins in the perplexity of the passive one. It always begins in *isostheneia*, though, and may well give rise to perplexity in turn. On a related note, by mentioning digestion Thorsrud (2009, p. 129) brings in automatic body functions as possible objects of comparison for the passive construction of the *epochē*. As should be obvious, I sympathise with this metaphorical line of thought. However, I myself would prefer to compare the *epochē*'s passive construction with the automatic body function of sneezing. The Pyrrhonian technique of setting up cognitive conflicts to produce the *epochē* then compares to the use of pepper to cause sneezing. Unlike digestion, sneezing has some definite voluntary dimensions. Use of pepper aside, you can mentally condition yourself for it by conscious intent. The Pyrrhonean equivalent would be something like reminding yourself that you have been wrong before. Moreover, you can also choose to blow your nose in such a manner as if you were actually sneezing. Such an action compares to the active construction of the *epochē* as a kind of intellectual reservation. Finally, and importantly, the two are

functionally intertwined: A consciously intended sneeze can initiate a sequence of events that ends up with real, automatic sneezing. In similar manner, the actively constructed *epochē* (intellectual reservation) sometimes initiates a sequence of intellectual "events" that ends in the state of a passively constructed *epochē* (perplexity).

[18] On this matter, see note 8.

[19] A similar reading of Pyrrhonism has been attributed to Hegel (Forster, 1998, esp. pp. 126–192). The exact nature and scope of the key concepts of Pyrrhonism are of course subject to some amount of academic controversy. This is not the place to embark on any full blown review of the relevant literature. Suffice it to say that my sympathies regarding the demarcation line between the Pyrrhonean concepts of beliefs and appearances (the so-called Frede-Burnyeat debate) which makes up the invisible context of my exposition of *ataraxia* in particular, lie with the view expounded by Michael Forster (2005). According to this view, the Pyrrhonist's "allegiance to the ordinary man and hostility to the philosopher or scientist" (p. 62) is motivated by a difference in the content of their views, and not any (non-existent) difference in dogmatic form. The demarcation between beliefs and appearances, therefore, is such that *no* beliefs are exempted from the Pyrrhonean machinery of equipollence. As a general rule, however, the views of the ordinary man can be accepted *as appearances*, for they are not dogmatic in content. However, I would like to suggest, to some extent *pace* Forster, that the apparent Pyrrhonean duplicity regarding the views of the common man should be understood as proceeding from two opposite points along the transformative ("agogic") axis of dogmatism – *epochē* – *ataraxia* before it is coupled to a reactionary response to a historically new form of belief (*dogma*, p. 64). Just to be clear: The Pyrrhonean criticism of the views of the common man concern a pre-ataraxian mindset, and the affirmation of the same is, correspondingly, of a post-ataraxian kind. Moreover, I find the compelling idea of the Pyrrhonists as philosophical reactionaries equally significant in another, more immediate historical context: that of *personal* history – or more to the point, that of developmental cognitive psychology. In writing the following, Forster could equally well be talking about the individual Pyrrhonists' *memory of childhood cognition* as about certain historical circumstances at the societal level: "achieving the end of *ataraxia* is really a displaced attempt to return to the type of belief which preceded the distinctive type of belief which they are attacking. It is really a displaced attempt to return to the archaic cognitive attitude of accepting fundamental principles without any consciousness of competing alternatives and in passive deference to communal tradition." (p. 76) "Unbeknown (or at least not clearly beknown) to the Pyrrhonists themselves, their yearning for a life by appearances and *ataraxia* was at bottom a yearning for something from the past. Beneath the conscious Pyrrhonian text of revolution lies a semiconscious subtext of reactionary vision and desire." (p. 70) I believe it to be a significant point, that making distinctions between ourselves and our non-selves is an acquired skill, steeped in psychological signification, and that, in consequence, so is the dualist framework of mind that it fathers. Hence, it seems to me not a great stretch of mind to assume that "the cognitive attitude of the Homeric age" (p. 68), is something we

all share as the lower layers of our personal ontologies. Unlike Forster's historicist angle, this one does not turn the Pyrrhonean effort in a nostalgic direction. Rather, it reflects the fact that cognitive progress has circular as well as linear characteristics. That is to say, it spirals.

[20] If there were, how could the unity of the two orientations be first presupposed, and then rediscovered in the developmental schema that comprises the self-description of the Pyrrhonist method (Sextus Empiricus, 2000, 10–11)?

[21] The mindset of post-ataraxian cognition therefore resembles that of a young child completely awake to the flow of experience, but in many ways utterly oblivious to the cognitive patternings of the largely unknown adult world. While the Pyrrhonist, unlike the child, possesses an adult awareness of these patternings, she does so without ontological allegiance to any one in particular. This is because through the global *epochē*, she has personally experienced that the world, contrary to the deep psychological dictates of dogmatism, is not upheld by any individual employment of human reason. What Scottish Enlightenment philosopher David Hume (1711–1776), after much fruitless rational deliberation, had to say about the connection between cause and effect, the Pyrrhonist could say about the general manner of human cognitive reality access: "it is not probable, that it could be trusted to the fallacious deductions of our reason, which is slow in its operations; appears not, in any degree, during the first years of infancy; and at best is, in every age and period of human life, extremely liable to error and mistake. It is more conformable to the ordinary wisdom of nature to secure so necessary an act of the mind, by some instinct or mechanical tendency, which may be infallible in its operations, may discover itself at the first appearance of life and thought, and may be independent of all the laboured deductions of the understanding." (Hume, 1777, SBN 56) The Pyrrhonean *agōgē* may be phrased thus: Fundamentally, the world does not need our ontological care. Such, however, is the grip of cognitive dogmatism, that in order to fully see this we must first step out of our ontological care for the world. Paradoxically, this stepping out is accomplished precisely through *exercising* our ontological care for the world with the help of the Pyrrhonean technique of *isostheneia*. Had we no desire whatsoever to know how things are, cognitive conflicts would not bother us, for the simple reason that nothing of value would be at stake in them. As it is, it is ultimately our own ontological care, as the energy or thrust of all cognitive conflict, that allows for the outside, metacognitive experience of itself and the concomitant "calmness of soul" that is *ataraxia*.

References

Angelo, T. A., and K. P. Cross. 1993. *Classroom Assessment Techniques: A Handbook for College Teachers*. San Francisco CA: Jossey-Bass.

Atherton, J. 1999. "Resistance to learning: A discussion based on participants in in-service professional training programmes." *Journal of Vocational Education & Training* 51(1):77–90.

Başer, M. 2006. “Fostering conceptual change by cognitive conflict based instruction on students’ understanding of heat and temperature concepts.” *Eurasia Journal of Mathematics, Science and Technology Education* 2(2):96–114.

Bett, R. 2000. *Pyrrho, His Antecedents, and His Legacy.* Oxford: Oxford University Press.

Dunér, D. 2011. “Astrocognition: Prolegomena to a future cognitive history of exploration.” In *Humans in Outer Space – Interdisciplinary Perspectives.* Studies in Space Policy vol. 5, edited by U. Landfester *et al*., 117–140. Wien: Springer.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston IL: Row, Peterson and Company.

Flavell, J. 2004. “Theory-of-mind development: Retrospect and prospect.” *Merrill-Palmer Quarterly* 50(3): 274–290.

Forster, M. 2005. “Hegelian vs. Kantian interpretations of Pyrrhonism: Revolution or reaction?” In *Kritisches Jahrbuch der Philosophie* 10, edited by B. Bowman, 53–70. Würzburg: Königshausen und Neumann.

—. 1998. *Hegel’s Idea of a Phenomenology of Spirit*. Chicago IL: University of Chicago Press.

Gibbs, L. 2002. *Aesop’s Fables*. Oxford World’s Classics. Oxford: Oxford University Press

Gorski, P. 2009. “Cognitive dissonance as a strategy in social justice teaching.” *Multicultural Education* 17(1):54–57.

Hadot, P. 2002. *What Is Ancient Philosophy?* Cambridge MA: Harvard University Press.

—. 1995. *Philosophy as a Way of life*. Oxford: Blackwell.

Hamilton, E. 1942. *The Great Age of Greek Literature*. New York: Norton.

Heywood, D., and J. Parker. 2010. *The Pedagogy of Physical Science*, Contemporary Trends and Issues in Science Education vol 38. Dordrecht: Springer.

Hume, D. 1777. *An Enquiry concerning Human Understanding*. URL= <http://www.davidhume.org/texts/ehu> Accessed 2011-11-24.

Husserl, E. 1973 (1931). *Cartesian Meditations: An Introduction to Phenomenology*. The Hague: Nijhoff; Dordrecht: Kluwer.

Limón, M. 2001. “On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal.” *Learning and Instruction* 11(4–5):357–380.

Mates, B. 1996. *The Skeptic Way: Sextus Empiricus’s Outlines of Pyrrhonism*. Oxford: Oxford University Press.

McGaughey, C. 2008. “To be or not 2 bee: An examination of breaking set.” In *Meta-Cognition: A Recent Review of Research, Theory, and Perspectives*, edited by M. F. Shaughnessy, M. V. E. Vennemann, and C. Kleyn-Kennedy, 85–108. New York: Nova Science.

Mutschmann, H. 1912. *Sexti Empirici opera*, vol. 1. Leipzig: Teubner.

Neighbour, R. 2005. *The Inner Apprentice*. Oxford: Radcliffe Publishing.

Nussbaum, M. 1994. *The Therapy of Desire: Theory and Practice in Hellenistic Ethics*. Princeton NJ: Princeton University Press.

Ohlsson, S. 2011. *Deep Learning: How the Mind Overrides Experience*. Cambridge: Cambridge University Press.

Perry, W. 1988. “Different worlds in the same classroom.” In *Improving Learning: New Perspectives*, edited by P. Ramsden, 145–161. New York: Nichols Publishing Company.

Posner, G., K. Strike, P. Hewson, and W. Gertzog. 1982. “Accommodation of a scientific conception: Towards a theory of conceptual change.” *Science Education* 66(2):211–227.

Rea-Ramirez, M. A., and J. Clement. 1998. “In search of dissonance: The evolution of dissonance in conceptual change theory.” *Proceedings of the National Association for Research in Science Teaching*. San Diego. (ERIC full text PDF link posted but currently disabled. Available upon manual request.)

Schraw, G. 1998. “Promoting general metacognitive awareness.” *Instructional Science* 26, 113–125.

Sextus Empiricus. 2000. *Outlines of Skepticism*, edited by J. Annas, and J. Barnes, 2. ed. Cambridge: Cambridge University Press.

Sinatra, G., and P. Pintrich, eds. 2003. *Intentional Conceptual Change*. Mahwah NJ: Erlbaum.

Smith, D. 2005. “States of uncertainty monitoring and metacognition in animals and humans.” In *The Missing Link in Cognition*, edited by H. Terrace, and J. Metcalfe, 242–271. Oxford: Oxford University Press.

Thorsrud, H. 2009. *Ancient Skepticism*. Stocksfield: Acumen.

Tite, P. 2003. “On the necessity of crisis: A reflection on pedagogical conflict and the academic study of religion.” *Teaching Theology and Religion* 6(2):76–84.

Tuch, R. 2011. “Thinking outside the box: A metacognitive theory of mind perspective on concrete thinking.” *Journal of the American Psychoanalytic Association* 59:765–789.

Vogt, K. 2011. “The aims of skeptical investigation.” In *Pyrrhonism in Ancient, Modern and Contemporary Philosophy*, edited by D. E. Machuca, 33–49. Dordrecht: Springer.

Vosniadou, S., and C. Ioannides. 1998. “From conceptual development to science education: a psychological point of view.” *International Journal of Science Education* 20(10):1213–1230.

Vosniadou, S. 2007. ”The conceptual change approach and its re-framing.” In *Reframing the Conceptual Change Approach in Learning and Instruction*, edited by S. Vosniadou, A. Baltas, and X. Vamvakoussi, 1–15. Amsterdam: Elsevier.

Zohar, A., and S. Aharon-Kravetsky. 2005. “Exploring the effects of cognitive conflict and direct teaching for students of different academic levels.” *Journal of Research in Science Teaching* 42(7):829–855.

CHAPTER FIVE

PHILOSOPHICAL PROBLEMS OF THE DEFINITION OF EXTRATERRESTRIAL LIFE AND INTELLIGENCE*

JEAN SCHNEIDER

Introduction

On Earth, Life is perceived via two aspects: organic life and psychical life. Organic life – the subject of biology – is shared by all living things, from bacteria to humans. Psychical life is the domain only of humans and some other animals. According to the common view, psychical life culminates in human intelligence, and there is no fundamental gap between human intelligence and animal psychology. Intelligence is just a skill: the ability to react to situations in the environment. In the context of astrobiology, the question naturally arises: are these approaches well adapted to understanding exo-life? Here I treat these questions with a philosophical approach. Note that there are two types of philosophy of knowledge: natural philosophy and critical philosophy; therefore, I must first clarify the differences between them, so that I can then explain why critical philosophy is to be preferred to natural philosophy. Finally, I attempt, using critical philosophy and an extension of it called *epistemo-analysis*, to define organic life and intelligent exo-life. I point out a basic difficulty in defining the latter.

As yet, we do not know of any life forms elsewhere in the solar system. That leaves plenty of room for life on extra-solar planets. I will deal here only with the possibility of life on these exoplanets. Throughout this chapter, the discussion is inspired by pragmatic considerations: what actions should we take to search for organic and psychical exo-life?

Natural versus critical philosophy

"Philosophy" covers a wide conceptual area, with unclear borders and poorly mapped regions. Among its prominent divisions are political philosophy, ethics, and philosophy of knowledge. The latter can be divided into natural philosophy and critical philosophy.

Natural philosophy is based on belief in an objective reality; the goal of knowledge is to catch its essence in statements of Truth. Knowledge acts as a bridge between the subject's mind and the essence of nature, based – like religious faith – on opinions and convictions rather than critical analysis. The resulting tendency is to project human feelings onto the external world, giving natural philosophy the flavour of animism – as pointed out by Bachelard (2002), who termed the obstacle "substantialism" in his book *The Formation of Scientific Mind*. Believers in natural philosophy are generally insensitive to the distinction between the two approaches, their convictions being based on a kind of faith.

Critical philosophy starts with analysis of the procedures by which we explain, essentially through language, our experiences in any domain. The explanation is called a theory or, more generally, a discourse. Kant developed the approach in his *Critique of Pure Reason*; Ernst Cassirer (1965) reassessed it in the light of modern science in his *The Philosophy of Symbolic Forms*. The approach is remarkably well summarized in the *Critique*'s introduction: "if our knowledge starts WITH experience, it does not prove that it only derives FROM experience, since it could well be that even our experience-based knowledge is a composite of what we receive from our perceptions and of what our power to know [i.e. concepts] produces itself".

More recently, various authors within the school of analytical philosophy, have pointed out the essential role of language in explaining experience. On their view, knowledge is always a construction, with the help of language, of a reality that does not pre-exist – not the discovery of pre-existing things. Reality as a source of perception is thus purely metaphysical; realism is the ideal, not the practical. In contrast to natural philosophy, critical philosophy has played a fruitful role in science. It has helped, at least indirectly, to rid physics of the notion of essences like phlogiston and aether. To summarize: critical philosophy deals with the processes of constructing a reality; natural philosophy deals with the essences of things: essences that are illusory from the viewpoint of critical philosophy.

Epistemo-analysis as extension of critical philosophy

To critical philosophy, concepts are operations upon the world of experience. Epistemo-analysis, as an extension of critical philosophy, analyses concepts' emotional roots. Borrowing from psychoanalysis, the neologism is a recent introduction (Schneider 2002; 2006); but the notion has been around for some time, discernible in e.g. Bion's (1962b) *Theory of Thinking*. Epistemo-analysis makes use of two key notions: *family romance* and *object relation* (Laplanche and Pontalis, 1974). Family romance is the way by which we construct abstract notions like "the past" from our own, fantasised past.[1] Its unconscious strength is manifest in the symptom-like persistence of presenting scientific results in the form of narratives in astronomy outreach papers.

Subtle and complex, the object relation derives from the Freudian notion of drive. The subject is embedded in a relation with the objects (more precisely, proto-objects) of his desire at the same time he is detaching himself from this embeddedness, thanks to independent, exterior forces, so that the (proto-)objects become external objects of desire. The paradigmatic object relation is detachment from the mother. In this conception, the object of the object relation is a construction.

The object relation is logically different from a relation *with* an object. The object must not be seen – as it is too often presented – as that to which the subject relates. On the contrary: in the object relation, the relation is *in* the object.[2] The relationship between the subject and the object is quite unlike the relationship between subject and predicate in grammatically correct linguistic utterances, which was analysed as early as 1662 in Arnauld and Nicolle's famous *Logic or Art of Thinking*. This makes it difficult to explain the object relation in language, whose structure is not suited to the task. As Sigmund Freud wrote (1989) in his final book, *An Outline of Psycho-Analysis*: "we are [unfortunately] obliged to translate our deductions in the language of our perceptions, disadvantage from which we can, for ever, not escape".

The primary root of embeddedness is affection: objects are "good" things; they are objects of love. This provides the unconscious root of all living objects and of life itself. The observable behaviour of living objects (as constructed via the object relation) is correlated empirically with another type of experience: i.e., physico-chemical experiments in modern biology or standard organic chemistry. As viewed from biology, life is an intellectual construction based on physical concepts. Exobiology tries to make similar observations beyond the Earth. At the same time, it offers no guarantee that we can have the same emotional relations to these observations.

Definition of exo-life

I wish to apply the constructivist approach to exo-life. First, however, I must discuss what should be expected from a definition of exo-life. What is a definition anyway? The definition of "definition" is the subject of a vast literature in philosophy: one that is impossible to summarise in a few paragraphs. It began with Aristotle, was developed in the Middle Ages by the nominalists, and more recently has been taken further by the various schools of logic. In his *Posterior Analytics*, Aristotle understands definitions as designating, for each thing defined, the collection of its attributes – as clearly characterized by the method of division (II, Section 3 [Theory of Definition], Chapter XIII). For the followers of nominalism – beginning with Roscelin of Compiègne; followed by Thomas Aquinus, Pierre Abélard, and others – a definition is a kind of name that creates a category without assuming any essence.

In contemporary philosophy of natural science, a definition may be of two kinds. First, it can be an arbitrary convention: e.g., the neologism "pulsar". Second, it can attempt to clarify the content of a pre-existing word for which people have certain preconceptions, however those preconceptions are grounded – to catch the (ultimately illusory) essence of the thing being defined. Definitions of the latter type use pre-existing plain language with *a priori*, pre-scientific content: the sort of content that can be revealed through epistemo-analysis. Unfortunately, these definitions are likely to sow much confusion.

Rosch (1973) uses the notion of prototype to offer a better account, even (or especially) for those cases where definitions are vague. Nevertheless, her approach ignores an important empirical fact: when they are *not* pure conventions, words – along with their unconscious (and, therefore, somehow obscure) content as revealed through epistemo-analysis, pre-exist any definition.

Modern language theory draws attention to the performative nature of words. They do not designate pre-existing things – not really; they create first what they designate as exterior and pre-existing second. It is an instance of *afterwardness*: one of the great conceptual discoveries of the 20th Century. If a definition constructs what it defines, there can be no such thing as an absolute definition – only definitions that depend on the procedure by which they construct the *definiendum*. This means that definitions are essentially relative. In the remainder of this chapter, I wish to deal with two definitions of life: one based on the object relation as I have described it, one on standard laboratory-based biochemistry (and, more generally, physics).

As I have argued above, life is not an objective property but always a construction, based – in the common-sense meaning of the word -- on the object relation, or else on physico-chemical concepts from biophysics. From an astrobiological perspective, life is not to be understood in the object-relation (i.e., everyday) sense; even while life, in that sense – an attribute of unconscious emotional relationships – cannot be constructed out of purely physico-chemical concepts.

As physico-chemists making celestial observations, astrobiologists repeatedly use the word "life" improperly, in a way that inevitably carries the emotional content of the object relation implicated in the most primitive sense of the word. They are being less than fully truthful to their audience – not unlike the astrophysicists who refer to the parameter *t* as time (Schneider, 2006). The questions "when does the human embryo become a human being?" and "when did pre-hominids become human?" offer a pertinent analogy. In either case, setting the time at which the transition takes place is – unavoidably – an arbitrary choice. I would like to suggest a similarity as well to quantum physics. In quantum theory, observables – represented by linear operations in a vector space – cannot be built from the state vector representing the structure of the measurement apparatus. They are *sui generis* (Ulfbeck and Bohr, 2001).

Life has no essence – not even when considered as organic life. The claim that life originates only from life is based on an arbitrary construction. As revealed by the object relation, experience consists only of relations with objects, constructed from observations. Some of those objects we declare as (and want to believe are) living. Astrobiologists want to limit the living to objects of sufficient complexity, whose complexity is stable and self-regenerated. That said, such properties are by no means only the domain of the living (in the common-sense sense): they apply as well to such matter-recycling objects as stars, which are not considered living. Stars exhibit a disposition in local entropy fluctuations toward lower entropy. Objects declared to be alive in the astrobiological sense of self-organizing structures are not necessarily alive in the object-relation (i.e., emotional) sense.

Let me offer another analogy: with light. Consider a community of speakers examining a strawberry and declaring it red. Physicists then make a spectral analysis and find a wavelength around 675 nm. Clearly, the everyday word "red" is correlated with the 675 nm. wavelength. Meanwhile, no colour is associated with wavelengths larger than ~750 nm. or smaller than ~400 nm. In similar fashion, life – in the everyday object-relation sense – may not apply to complex extraterrestrial structures too divergent from terrestrial organisms.

The question of intelligent life raises a further, paradox-inducing problem: we seek to define alien (non-human) intelligence, but we must do so using distinctly human concepts. It is a conundrum reminiscent of the Zeno paradox, which attempts to analyse motion in terms of a series of static steps. Motion must be something more. So it is with extraterrestrial intelligence: human intelligence is a kind of prison, from which we must escape. The difficulty is well known to those astrobiologists attached to the SETI project, who must interpret potential SETI signals using inadequate human concepts. Our only hope is to discover resources in ourselves that go beyond standard intelligence, as (psycho-analytic) unconsciousness extends beyond consciousness.

Operational conclusion

Astrobiologists inevitably begin with prejudices about exo-life, which they must use as guidelines for their observations; especially because, being dependent on space travel, these observations are very expensive. At the same time, they – and we – should keep open minds, making as many and as diverse of observations as possible. From those observations, we can then select those that offer interesting relations and correlations. As in bioethics, where the ultimate division between human and non-human embryo is arbitrary; so, too, will be the division between extraterrestrial life and non-life. Perhaps some day we will need exo-bioethical committees, as today we have bioethical committees.

Notes

[*] Drawn upon content from Jean Schneider, "Philosophical problems in the definition of extraterrestrial life and intelligence". *International Journal of Astrobiology*, 12, pp.259-262 (2013). Copyright © Cambridge University Press 2013.

[1] For its application to cosmology, see Schneider 2006.

[2] For an account of the object-relation, see Bion 1962, St. Clair and Wigren 2003, or the more difficult Lacan 1998.

References

Arnauld, A., and P. Nicolle. 1996. *Logic or Art of Thinking.* Cambridge: Cambridge University Press.

Bachelard, G. 2002. *The Formation of the Scientific Mind: A Contribution to a Psychoanalysis of Objective Knowledge.* Manchester: Clinamen

Press.

Bion, W. 1962a. "A theory of thinking". In *Second Thoughts: Selected Papers on Psycho-Analysis*, 111–119. London: Karnac 1967.

—. 1962b. *Learning from Experience*. London: William Heinemann.

Cassirer, E. 1965. *The Philosophy of Symbolic Forms.* New Haven CT: Yale University Press.

Freud, S. 1989. *An Outline of Psychoanalysis.* New York: Norton.

Kant, E. 2003. *Critique of Pure Reason.* New York: Palgrave Macmillan.

Lacan, J. 1998. *La relation d'objet.* Paris: Seuil.

Laplanche, J., and J.-B. Pontalis. 1974. *The Language of Psycho-Analysis.* New York: Norton.

Rosch, E. 1973. "Natural categories." *Cognitive Psychology* 4:328.

Schneider, J. 2002. "L'exobiologie, l'imaginaire et le symbolique." In *Epistémologie de l'exobiologie*, Cahiers François Viète 4:91. http://luth7.obspm.fr/~schneider/15mai.html

—. 2006. "On the 'history' of the Universe and the beginning of time." In *SF2A-2006: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, edited by D. Barret, F. Casoli, G. Lagache, A. Lecavelier, L. Pagani, 303. http://www.obspm.fr/~schneider/SF2A06-Schneider1.pdf

St. Clair, M., and J. Wigren. 2003. *Object-Relations and Self-Psychology: An Introduction*. Pacific Grove CA: Brooks Cole.

Ulfbeck, O., and A. Bohr. 2001. "Genuine fortuitousness: Where did that click come from?" *Foundations of Physics* 31:757.

PART II:

COMMUNICATION

CHAPTER SIX

EVOLVING AN EXTRATERRESTRIAL INTELLIGENCE AND ITS LANGUAGE-READINESS

MICHAEL A. ARBIB

Introduction: CETI, Communication with an extraterrestrial intelligence

In 1979, given the then current estimate (based on very tenuous evidence) that the nearest civilization capable of radio communication would be at least a hundred light years away, I raised the question of what we should say in a message for which it takes two hundred years to get a reply (Arbib, 1979) and argued that we should seek to distill the collective wisdom of our societies into an *Encyclopedia Galactica,* to be assessed in terms of scholarly interpretation by the aliens rather than an ongoing conversation. The present paper addresses the further challenge that we don't know who "they" are and thus statements that might seem clear to us (given that the code we use has been deciphered) may not be clear to them.

Consider the question that Christina Behme asked when she sent me the photo in Figure 1: "What direction is up?" At first you recognize the trees apparently going up into the sky but then you notice what appear to be pieces of floating wood and so recognize the photo as showing an air-water interface. Since we normally stand above pools we would conclude that this is a reflection and that the "up" is in the opposite direction to what is apparent. But if you were an underwater creature then you would interpret up as in fact what we thought when we first looked at this image. This example gives us a flavor of the knowledge that goes into an interpretation and how our particular world view may strongly influence the interpretation we give to a message, as in the case of the photo of Figure 1.

Figure 1. Where is Up? (A photo by Christina Behme: Erkner, near Berlin, April 2011. Reproduced with permission.)

In this paper, I will focus on *Communication with an Extraterrestrial Intelligence* – CETI, pronounced Cheti. CETI is not to be confused with SETI which is Search for Extraterrestrial Intelligence. One issue for CETI is "what is in it for *us* and what do we think in it for *them*." Are *they* trying to open up an enlightened conversation or are they planning to take over the planet to be anti-terraformed to their needs? Or, as Stephen Hawking fears, might such communication impact our civilization so negatively that we should dampen our terrestrial communications to minimize the chance that they will attract alien attention? Another problem is the technology gap. I have said that 200 years is one estimate of the round time for the message. Suppose we send detailed instructions on how to build a decoder. 200 years on Earth takes us back from building a fabrication facility to make VLSI chips almost to Benjamin Franklin flying kites to learn about basic properties of electrostatics. I am not suggesting that we will be able to communicate with anybody with science 200 years less advanced than ours. But even if they are around 200 years – let alone 2,000 or 200,000 years – "ahead" of us then what might seem to them like a set of elementary instructions to decode their messages might seem to us like unintelligible magic. But these are problems for other papers. Let us return to our own attempts at CETI.

I am going to assume that we are attempting to interact with *a society of embodied individual intelligences,* not a "swarm intelligence" like Hoyle's "Black Cloud". I will further assume that these are creatures on a planet but whether they live on the land or in the air or above or under water is open – and I will assume that they form a society of individually intelligent creatures forged by a Darwinian process of natural selection shaped by niche construction. In addition, it is a reasonable assumption that if we are communicating with these beings, they will have achieved a technological level which makes something like radio communication possible, and that this will involve a mathematics which rests in part on an understanding of the natural numbers. I thus endorse the key element of Hans Freudenthal's (1960; 1980) design of *Lincos* – namely rooting communication in a shared understanding of counting and operations like addition – even though I think later strategies he proposes are unworkable. But the key point I want to make is this: Lincos uses assumed "universalities" of mathematics and physics to *establish* communication, but new strategies seem required to *communicate novel ideas* about, e.g., embodied *social* experience and emotion.

This is not communication in the sense of Shannon's theory of reliable communication in the presence of noise (Shannon and Weaver, 1949) where sender and receiver are plucking messages from a pre-assigned ensemble – it is creating new Meanings that the receiver has not thought about before. Communicating the idea that seventeen is a prime number just lets the receiver know that "he" has found the right way to decode the message. But that's a long way from conveying the meaning of a discourse on the ups and downs of the history of the Middle East in the Twentieth Century, or how the torture death of one man 2000 years ago has religious significance for so many people today but yet does not have this significance for many more.

Materials and methods

My aim will not be to *solve* the problem of communicating novel ideas about embodied social experience so much as to argue for the relevance of imagining novel alien species with which we might try to communicate. Linguistics studies the commonalities and differences between human languages and in doing so establishes insights that aid the decoding of newly encountered human languages. Clearly, studying linguistics without have any languages to study would be a bad strategy, yet in CETI we are crippled by the fact that we have no LETIs (languages of extraterrestrial intelligence). Even if we insist with Freudenthal that each LETI will include means to discuss prime numbers, we still face the terrible challenge of anthropomorphism, shown in basic form in discussing Figure 1: how do we make our meaning clear when so much of our language has its roots in our embodied experience as human beings? What might be other LETI universals beyond the ability to discuss the natural numbers? Perhaps the ability of the members of the species to act upon objects and to have a communication system that expresses such acts would be universal, though it is controversial whether or not there would be something akin to proper names to characterize the *who* in *who did what and to which*.

The challenge for CETI, then, is that we need some languages, some LETIs, to work on but since we have no examples of real LETIs, we need to start creating them. Here we are trespassing on the field of science fiction as we try to think about how a different embodiment might underwrite the evolution of a very different language system. A linguistic anthropologist encountering a new human group can exploit basic human embodiment and needs and experience, aided by pantomime and pointing, to establish the basic vocabulary of their language and build from there a map of lexicon, grammar and social custom – though even with members of the same species, immense misunderstandings may arise. If we are already in physical proximity with an alien species, then we can learn much about their embodiment as a basis for beginning to decode their language – though students of animal communication will tell you how great that challenge is even with Earthly species whose communication systems are much simpler than languages. If you already know who you are communicating with and can gather some information about their situation then you may be better prepared to design messages specific to communication with them. But further challenges remain when no physical encounter has occurred.

Our method will be two-fold. First, we take a quick look at the evolution of the human language-ready brain as backdrop for imagining the evolution of an alien species. Then, for our example of how such evolution might yield a very different form of language, we turn to the octopus as the most alien (to humans) intelligent life form on this Earth. Octopuses are invertebrates, have eight arms, live underwater, and do not survive completion of the reproductive cycle. Our challenge will be to imagine creatures on other planets whose physical and social structure evolved from something like the squids or octopuses of present-day Earth to yield an intelligent species with a language fully expressive of their life-world. I call this creature the *octoplus*, and will describe it to create aspects of an extraterrestrial language that may lay the basis for subsequent hard work to determine what sorts of decoding might be appropriate when we are confronted with the disembodied language of an embodied species.

Evolving the language-ready brain: Humans as a "handy" comparison for aliens

Unlike many other theories, the Mirror System Hypothesis (Arbib, 2005; Arbib, 2012; Rizzolatti and Arbib, 1998) does not suggest an evolutionary path from the vocal apparatus of our last common ancestor (LCA) with monkeys that yields precursors to language which all take the form of primitive precursors to speech. Rather, it posits a different route which looks at our incredible dexterity and the fact that we – like monkeys and, presumably, our LCA – have neurons called mirror neurons that are active not only when we control our own hand movements but also when we recognize the hand movements of others, with specific neurons firing for a similar range of actions in both cases. The hypothesis offers the following evolutionary sequence:

1) A mirror system for manual actions, matching action observation and execution, was established in our LCA with monkeys (more than 25 million years ago).
2) A simple imitation system that could support transfer of a few simple manual skills, grounding a limited set of manual gestures was established in our LCA with chimpanzees (some 5 to 7 million years ago).

This was followed by evolutionary changes in the last 5 million years that led from that LCA to modern humans:

3) Our distant ancestors developed increasingly complex abilities for imitation supporting social transfer of an increasing repertoire of manual skills and communicative gestures.
4) A brain and social system that could support communication by pantomime. Whereas in imitation, you see me acting on an object and use that to figure out how to gain my skill, pantomime can indicate an object or an action or desire related to that object even when the object is not present. Pantomime has the great advantage that it yields an open semantics – communication is not limited to a previously established repertoire. The pantomimic can create novel messages with a reasonable chance that conspecifics will understand them. The drawback is that a pantomime may be costly (e.g., it takes a long time) to produce and may be very ambiguous.
5) Protosign then emerged through the conventionalization of pantomime to reduce the ambiguity of communication and increase

its speed *within a group that shared the resulting communicative conventions.* Whereas any member of the species can understand a pantomime, more or less, only two creatures who have both learned the protosigns can use them to communicate.

6) Protosign provided scaffolding for protospeech, and the two then co-evolved in an expanding spiral.
7) However, the conventionalization that makes protolanguage (protosign and protospeech) possible means that it takes considerable effort to invent new (proto)words and get them adopted by the social group. The breakthrough is the development of syntax to support ways to put words together freely to readily convey composite and novel meanings. The Mirror System Hypothesis holds that it was cultural evolution in *Homo Sapiens* (rather than further genetic changes in brain and body driven by natural selection) then supported the emergence of syntax and thus fully rich languages, with an interwoven increase in both cognitive and linguistic complexity.

Various aspects of the Mirror System Hypothesis are controversial, though no more so than other contenders, and it addresses data that the alternatives do not take into account (see the critiques and response in *Language and Cognition*, 2013, vol. 5, issue 2/3).

But arguing pro or con the hypothesis is not the point here. Rather, I call this hypothesis a “handy” comparison because it rests on a very specific kind of embodiment that supports manual dexterity, and reflects a very specific evolutionary history. It helps us understand a biosocial evolutionary process which led to a rich capacity for language in a way which reflects a particularity of our primate line – the importance of dexterity and the ability of protohumans to use the hand to imitate the shape of the object or pantomime actions.

Octopuses and squids

As background for our fictive evolution of the language of the octoplus, we here review some facts about *Octopus vulgaris* and some other cephalopods. *Octopus vulgaris* has very well defined eyes, but what might look to us like a chin beneath his eyes is actually his body, the mantle. Eight arms radiate out. An octopus can open a jar by using the pads on its arms, each of which is a hydrostat with no bones. The octopus is good at learning to do a variety of things. Indeed, my own introduction to the octopus as a creature worthy of scientific study was from the work of J. Z.

(Jay Zed) Young who was a professor in University College when I was a postdoc in London back in 1964. He then published a book (Young, 1964) called *A Model of The Brain*, which approached brain modeling through studying the brain of the octopus. This gave me a sense of the intelligence of the octopus and linked it to my concerns with computational modeling of brain function. But despite its intelligence, the octopus has one big disadvantage: the male dies after mating while the female lives only long enough to protect her eggs until they hatch and the paralarvae drift away. Thus the intergenerational transmission of information in the octopus is purely genetic.

Figure 2. An octopus camouflaged at top emerges from its background at bottom. (Photos courtesy of Richard Hanlon; extracted from the movie viewable at http://video.google.com/videoplay?docid=-9203354934457918497.)

Figure 2, taken from one of the many video observations of octopus behavior made by Roger Hanlon, shows the same octopus in the same position, camouflaged at the left but not at the right. The octopus makes himself "invisible" not by making himself transparent, so you can see through him but, rather, by the ability to assess the statistics of the texture and patterning of his immediate surroundings as a basis for controlling two amazing systems: One involves chromatophores: little patches of color with muscles which can pull back the skin to reveal the color or relax to obscure the color. He can also control papillae which can create intricate patterns of skin texture. In this way, the octopus we see on the right of Figure 2 was able to obscure himself as seen on the left.

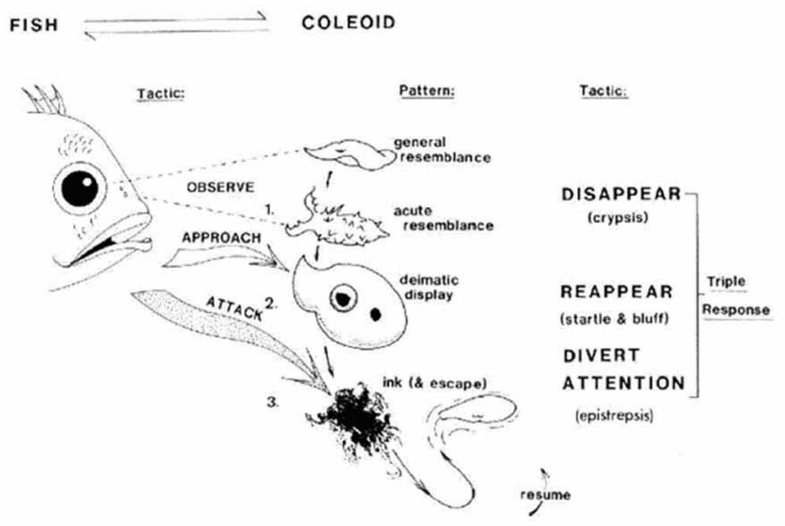


Figure 3. How the octopus evolved to fool the visual system of predators (Packard, 1988).

The person who has done most work on the chromatophore system is Andrew Packard, as exemplified in "Through the looking glass of cephalopod color patterns. A skin-diver's guide to the Octopus brain." (Packard, 1995) and "Visual tactics and evolutionary strategies" (Packard, 1988). Figure 3 gives his take on the fact that fish prey on cephalopods to ground his view that the octopus has evolved a whole range of strategies to avoid these predators, each of which can be deployed in a fraction of a second. The first strategy is that of crypsis, as exemplified in Figure 2: changing the color pattern and texture of his skin to fool the fish's visual system into not seeing him. The second strategy, deimatic display, is to suddenly abandon crypsis and reappear but displaying an unusual pattern like a false eye on its skin which may serve to scare off the fish. But if these fail, the third strategy is to release a cloud of ink which may attract the attention of the fish while the octopus swims away behind the cover of the ink.

In addition to presenting these three strategies, Packard is making the very important point that any species evolves within an environmental context and is therefore greatly shaped by predators. Thus, another item for our earlier list of universals is that if a species evolves to the point of being intelligent enough to want to communicate across the void, their intelligence will be shaped in part by the need to avoid predators by one means or another.

But let us place octopuses (and humans) in a broader context. In a classic paper of 1934, Jakob von Uexküll (1957 for the English translation) introduces the notion of the *Umwelt*, showing how different the world may be even for creatures embedded in the same region of physical space, based on the different capabilities for sensing and acting that different species may possess. The paper is an important precursor to theoretical ethology, with its careful study of the different worlds of different animals, even though much of its theory now appears dated in light of subsequent developments in neuroethology: the neuroscience of animal behavior. (Interestingly, where von Uexküll talked of a stroll through the worlds of animals and men, Ziemke and Sharkey (2001) talked of strolling through the worlds of robots and animals, applying von Uexküll's theory of meaning to adaptive robots and artificial life.) With this, let's look in some more detail at a comparison between the receptors and effectors of the octopus, as stand-in for the distant ancestor of our octoplus (Figures 4 and 6).

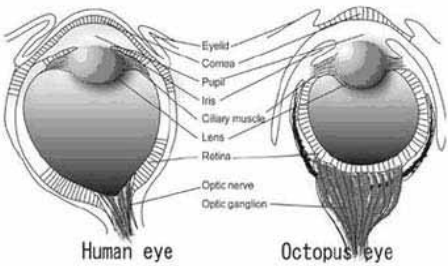


Figure 4. Convergent evolution of the human and octopus eye.

There is *convergent evolution* of human and octopus eye even though their distant ancestors had very different visual systems. Unlike many other invertebrates – think of a fly or a crab – which have compound eyes, octopuses have simple eyes like us: each eye has a single lens focusing patterns of light on the cellular array of the retina. But there are also intriguing differences. The human lens has muscles which adjust its focal length, whereas muscles serve to shift the lens in and out in an octopus; moreover, the receptors in the octopus retina are arrayed closest to the lens, whereas those in the human are furthest from the lens so that light must pass through the other layers of the retina to reach them. But, presumably, the visual system of the octopus brain must take the patterns of retinal activity and treat them quite differently to accommodate the life of the octopus in a 3D submarine world, whereas human locomotion occurs primarily on a 2D terrestrial surface. Action is the core.

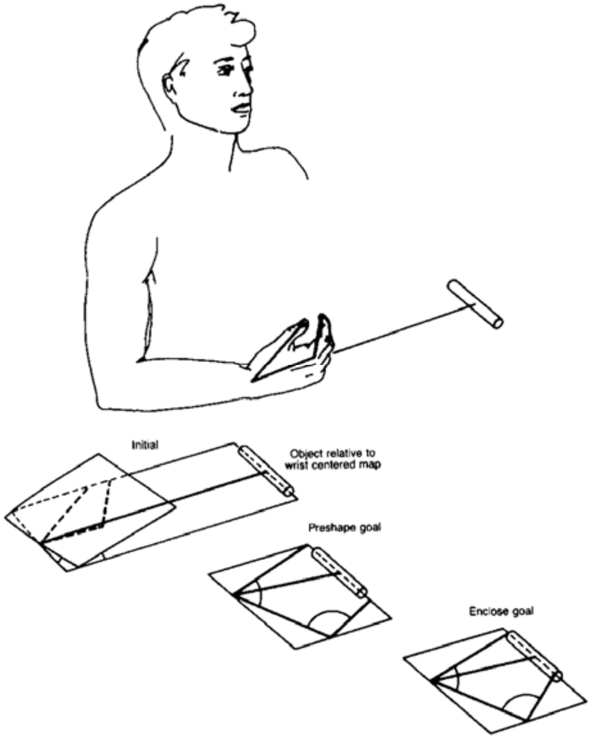


Figure 5. The object centered planning of the reach to grasp (Iberall and others, 1986).

Crucial aspects of cognitive systems, be they *robots*, *humans or octoplus*, are (a) spatial knowledge processing, (b) mechanisms for control of information processing, (c) behavior closing the loop that makes perception meaningful, and (d) the notion that behavior grounds an evolutionary view of *language built upon an embodied system*. The visual system of the octopus and the visual system of man are very similar, and yet their evolutionary history on this planet is far apart. The ability to locate objects in space, and maintain a map of the object's position in the brain (in the general sense of the physical substrate of the mind) argues for the evolutionary advantage of developing the ability to sense the reflected energy of the local star on an array of receptors.

But when it comes to arms, the evolution of human and octopus – and thus of our hypothetical octoplus – differ greatly. Neural mechanisms linking visual input to the control of arm and hand (Figure 5) were central to MSH. But what of octopus arms? Our arms are inherently angular, so for our embodiment the notion of angle is absolutely fundamental – you cannot get away from it. By contrast the octopus arm is a muscular hydrostat, akin to a human tongue up to a meter long! The octopus can locate objects in space and aim its arms to catch them, but no angles are involved (Figure 6).

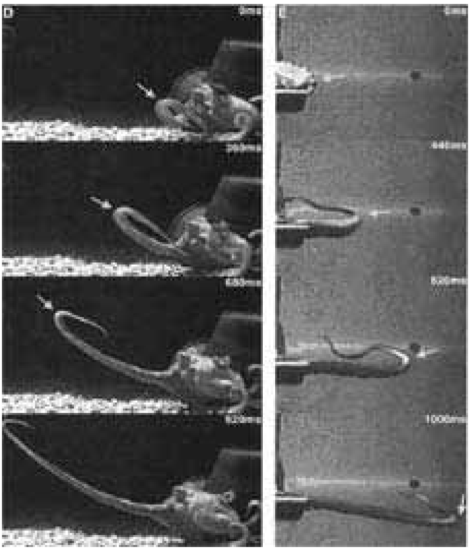


Figure 6. Control of an octopus arm (Sumbre and others, 2001).

Results: The evolution of octoplus

The previous section conveys some idea of a range of abilities very different from control of voice and hands (which is not to downgrade the remarkable use octopuses can make of their arms), which grounds the Mirror System Hypothesis that evolution of the human language-ready brain may have critically involved, at its core, a mirror system for manual dexterity. We now assess a comparable (but speculative) scenario for the evolution of the *octoplus*: a putative alien extraterrestrial intelligence. What might be some of the properties of a language evolved from the basis of chromatophores and body texture rather than visual control of the hand?

Figure 2 demonstrated the ability of the octopus to match chromatophores and skin texture with the statistics of what it sees around it: a complex sensorimotor transformation, parallel to the visual control of hand movements (Jeannerod and others, 1995), which formed the starting point for MSH. This observation grounds my invention of the LETI of the octoplus: what if communication were based on an expanded ability to control patterning of the body surface rather than on use of the hands or the voice?

The first postulate of our octoplus evolutionary scenario is that their distant ancestors surmounted the ancient hurdle of death through procreation or came to live much longer lives so that, in either case, mature creatures could provide a social environment for the development of the young. I make this postulate under the assumption that the development of rich languages and advanced technology is necessary for any species that can engage in CETI, and that cultural evolution is a necessary complement to biological evolution in making these developments possible. I add that the second scenario – in which death remains an immediate sequel to procreation – has vast implications for social structure which may greatly affect the structure of octoplus languages and provide extra risks of miscommunication with creatures such as ourselves for which the nuclear family is, more or less, the historical norm.

For MSH, pantomime establishes the shape of objects as the crucial property for pantomime, which is then postulated to guide the conventionalization that guides protosign and grounds the transition to language.

For octoplus evolution, the analog of pantomime would be the move from patterning to blend with a background object to displaying something of the texture and pattern of the object as the key for communicating about the object with an other. On this scenario, evolution would support the transition from exhibiting such a display in terms of current surroundings as camouflage to exhibiting such a display from memory to communicate to another octoplus about objects which have that texture.

The next step in the evolutionary scenario is based on the observation of cases where a male squid is courting a female as another male is approaching and exhibits a courting display on the side of the body near the female and an aggression display on the side nearer the male. A sub-hypothesis of MSH is that human evolution involved the fractionation of pantomime-like "holophrases" to yield separate words for object-like and action-like components of the pantomime, and the constructions (the elements of grammar) for putting the words back together in novel combinations (Arbib, 2008). This sets the stage for a "video screen" model in which the display is an assemblage of subdisplays. This scenario holds that the octoplus body evolved (biological evolution by natural selection) far beyond a single or dual display to yield a display of multiple patterns in diverse relationships that could support the cultural evolution of an inherently two dimensional form of communication – what we might call a *chromatophore language* – where the color or texture pattern is of prime importance in defining the components of any particular display.

I postulate that the basic notions of object and action would be universal, even though the categorization of what is a "noun" and what is a "verb" varies greatly across human languages; but I want to consider more fully the way in which the "world" of an organism is not "objective" but is a consequence of its specialized receptor and effector apparatus. The differences in species-dependent *Umwelt* charted by von Uexküll (1957) are only magnified by further differences in embodiment and in social structure – further reminding us of the challenges of communication across the *Umwelten* of humans and octopluses.

We seem to have established that any intelligence with which we are likely to establish communication will have vision, language, and a sophisticated knowledge of applied mathematics. But if an "octoplus" did have language, the words it uses will be different from ours. What may seem an obvious concept to them may be a very complicated concept to us. Our concern in speculating on communication with intelligences as alien as that of the octoplus is the question of whether, even if we agree that different embodiments in different (possibly constructed) environmental niches provide different grounding for language in human and octoplus, can we posit convergent cultural evolution in expressiveness? As we saw (Figure 6), there is no angle built into octopus arms, and this might yield a very different geometry as the basis for octoplus mathematics. Indeed, the etymology of geometry as geo + metry emphasizes its origins in the measurement of land, a far cry from a 3D submarine world. Marinometry?!

By "a space", we can mean a confined space, such as an arena, room or table top; or an extended space, such as the surface of the Earth. These support mathematical abstractions of space, of which the most basic *for humans* is the Euclidean plane, in which the behavior of points and lines can be fixed by axioms. One notion is that geometry emerged from the calculations of farmers pacing out their land (Room, 1967) – it starts with a certain form of action, so space is determined by the world of action, the way you walk or, for other creatures, the way you swim or the way you fly. Given our concern with alien intelligences, it is relevant that Nicod (1970) sought to understand how different forms of action and perception might yield quite different forms of geometry. However, when we look at the human brain, we find no single representation of space but, rather, a variety of neural representations of space adapted to varied realms of action – neural maps of the space for the control of the arm, the control of the eye, and the control of locomotion; the peripersonal space in close where we can grab objects and the extra-personal space further beyond (Arbib, 1997; Colby, 1998). Although we have the illusion of a single unity of conscious awareness, in fact our brains and bodies have evolved so that these different spaces are seamlessly integrated in our everyday experience.

I use the notion of *locometric space* as representing the highly species-specific aspects of space relevant to the modes of locomotion of the species. Here, space is determined by the world of actions measuring the world in terms of the actions (walking, swimming, flying) whereby we traverse it. The animal measures the world in terms of actions (e.g., how many steps taken) or *perceived measures of such actions* (e.g., the visual effect of an action such as the achievement of a goal). The notion of locometric space is particularly relevant to the notion of *path integration*: the ability of a wandering animal to keep track of the location of its home base relative to its current position

But if we think of e.g. the map of the London Underground or the metro map of any large city we find a representation that is very different from locometric space. It is not based on how long it takes you to walk from one place to another. Rather, it is based on the abstract notion of a public transport system that has a number of stops with multiple subway lines to get from one place to another. Perhaps here, rather than at the level of locometric maps, we find a level of abstraction for which any intelligent species would achieve a homomorphic representation. The history of humanity certainly reveals a great diversity of conventions in map making.

And what would a map be for an octoplus with an inherently 3D world? Nonetheless we may see here a movement towards the sort of representation that would support interspecies communication in a way that would not be possible for a grounding in such species-specific notions as the angle of an arm.

To change the focus, let us now consider that much of our communication is motivated and framed by emotion. Will alien species have emotions? Jean-Marc Fellous and I (Fellous and Arbib, 2005) edited a book called *Who Needs Emotion?,* which confronted knowledge of the neurobiology of emotions in animals and humans with the then current state of getting robots to display something that looks like a human emotion as a means to improving human-machine interaction. Today's robots, running unencumbered by problems of energy supply or reproduction have no need for emotions like hunger or fear. But what of the octoplus? An alien biological species shaped by competition with predators would, I suggest, certainly experience something akin to fear. But given the reproductive patterns of the octopus, and even if the octopus evolves to an octoplus that survives to take a role when the young begin to mature, would they know love? The general point, again, is that what may be unexpressed touchstones for our interaction with other humans may be inexplicable to an alien species.

With this background, we now paraphrase MSH as it might apply to the evolution of octoplus language:

1) A mirror system for arm actions, matching action observation and execution, was first established.
2) A simple imitation system that could support transfer of a few simple skills with the arms, grounding a limited set of brachial gestures, evolved millions of years later.
3) Distant ancestors of the octoplus developed increasingly complex abilities for imitation supporting social transfer of an increasing repertoire of brachial skills and communicative gestures.
4) Whereas MSH is based on the notion of the voice having little ability for pantomime, the evolution of octoplus is posited to involve convergence of two systems for "pantomime", one involving the arms – brachial pantomime – and one involving the display of chromatophore and papillary patterns that provided moving pictures of basic objects and behaviors: *chromo-pantomime*. A brain and social system evolved that could support communication by brachial and chromo-pantomime.

5) *Protosign* then emerged through the conventionalization of brachial pantomime to reduce the ambiguity of communication and increase its speed *within a group that shared the resulting communicative conventions*.
6) *Protodisplay* similarly evolved through conventionalization of the use of body pantomime; protosign and protodisplay then co-evolved in an expanding spiral.
7) The breakthrough to octoplus language came with the development of syntax to support ways to put words together freely to readily convey composite and novel meanings.

Discussion and implications

We have seen how MSH grounds language in the embodiment of pantomime, and our scenario grounds octoplus language in the embodiment of arm movements and their integration with voluntary, fractionated display evolved from chromatophore and papillary control for crypsis.

By charting the evolution of octoplus and its language-ready brain, I have laid down the challenge of how humans might decode it – as well as the challenge to CETI scholars (exolinguists) to invent further embodiments and scenarios for the evolution of language among creatures so embodied. When (i) we have designed a truly alien set of embodiments and LETIs evolved atop those embodiments and (ii) have been able to figure out what methods would have worked to translate such languages, then we will be more prepared to get going if we finally encounter a genuine LETI from across the void, and we will also be better prepared to design the messages we broadcast in the hope that alien intelligences can make sense of them.

It is encouraging that MSH charts a path that reduces the constraints of embodiment via the ability to make conventions shared by a social group to reduce ambiguity and speed communication. We cannot build a technology on a scientific basis without a language made of symbols which can be rearranged in arbitrary fashion, transcending imitation or transmission of pictures of actual events. Nonetheless, no matter how well we close the gap between an alien *Umwelt* and our own, the challenge of interpretation will remain. Just consider the challenges of understanding messages written in the Bible of a few thousand years ago or set forth in the American constitution from less than 250 years ago. In each case, even though these messages were recorded by beings who share our embodiment and many aspects of culture, we debate endlessly about what was really meant. Despite the great spans of time that separate us, these messages are nonetheless of value to us. Despite their ambiguity, they serve as invaluable reference points for us to debate about current issues. Let me recall, then, the conclusion of my 1979 article on reading the *Encyclopedia Galactica:*

> It is perhaps to be expected that, in the nature of humankind, many 'false prophets' will arise once we receive messages [from alien civilizations], who will forcefully argue for the adoption of various social structures long before they have been transmuted into a form adapted to human needs and history.
>
> But the give and take between fashionable cult and accumulated wisdom has always been part of the human condition, and there is no reason to expect that to change, no matter how much of the galactic wisdom should one day come to be ours.

Acknowledgments

This paper is based on a talk presented at the *Symposium on Interstellar Communication: Semiotic, Linguistic, and Cognitive Approaches*, Pufendorf Institute, Lund University, Sweden, May 5, 2011. My thanks to David Dunér for inviting me to the Symposium and for his hospitality in Lund. I also thank Roger Hanlon and, especially, Andrew Packard for invaluable discussion of the behavior of octopuses.

References

Arbib, M. A. 1979. "Minds and millennia: The psychology of interstellar communication." *Cosmic Search* 1:21–28.

—. 1997. "Modeling visuomotor transformations." In *Handbook of Neuropsychology, Volume 11, Section 16: Action and Cognition*, edited by M Jeannerods, 65–90. Amsterdam: Elsevier.

—. 2005. "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics." *Behavioral and Brain Sciences* 28:105–124; discussion 125–167.

—. 2008. "Holophrasis and the protolanguage spectrum." *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems* 9:151–165.

—. 2012. *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.

Colby, C. L. 1998. “Action–oriented spatial reference frames in cortex.” *Neuron* 20:15–24.

Fellous J.-M., and M. A. Arbib. 2005. *Who Needs Emotions: The Brain Meets the Robot*. Oxford: Oxford University Press.

Freudenthal, H. 1960. *Lincos: Design of a Language for Cosmic Intercourse*. Amsterdam: North-Holland.

—. 1980. “Towards a cosmic language.” *Cosmic Search* 2:35.

Iberall, T., G. Bingham, and M. A. Arbib. 1986. “Opposition space as a structuring concept for the analysis of skilled hand movements.” *Experimental Brain Research Series* 15:158–173.

Jeannerod, M., M. A. Arbib, G. Rizzolatti, and H. Sakata. 1995. “Grasping objects: The cortical mechanisms of visuomotor transformation.” *Trends in Neurosciences* 18:314–320.

Nicod, J. 1970. *Geometry and Induction*, trans. J. Bell, and M. Woods. Berkeley CA: University of California Press.

Rizzolatti, G., and M. A. Arbib. 1998. “Language within our grasp.” *Trends in Neurosciences* 21:188–194.

Room, T. G. 1967. *A Background (Natural, Synthetic and Algebraic) to Geometry*. Cambridge: Cambridge University Press.

Shannon, C. E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana IL: University of Illinois Press.

Sumbre, G., Y. Gutfreund, G. Fiorito, and T. Flash. 2001. “Control of octopus arm extension by a peripheral motor program.” *Science* 293:1845–1848.

Uexküll, J. von. 1957. “A stroll through the worlds of animals and men: A picture book of invisible worlds.” In *Instinctive Behavior: The Development of a Modern Concept*, edited by C. H. Schillers, 5–80. New York: International Universities Press. [Also in *Semiotica* 89(4):319–391. Originally appeared as von Uexküll. 1934. *Streifzüge durch die Umwelten von Tieren und Menschen*. Berlin: Springer.]

Ziemke, T., and N. E. Sharkey. 2001. “A stroll through the worlds of robots and animals: Applying Jakob von Uexküll’s theory of meaning to adaptive robots and artificial life.” *Semiotica* 2001:701–746.

CHAPTER SEVEN

GREETINGS EARTHLINGS! ON POSSIBLE FEATURES OF EXOLANGUAGE

ARTHUR HOLMER

Introduction

In analogy with "exobiology" (the biology of extraterrestrial species) and "exoplanets" (planets in other solar systems), "exolanguage" (possibly coined by Warner 1984) is a term applied to hypothetical languages of hypothetical extraterrestrial species. This paper addresses an issue that might seem more relevant to the domain of science fiction than to that of scientific inquiry. Nevertheless, the issues it raises are extremely relevant to linguistic theory today. Its purpose is *not* to provide a grammar of Martian or Sirian, nor even a prediction about what grammars of extraterrestrial languages could be like. Neither is it intended to suggest how we could go about attempting to establish contact with other civilizations.[1] Rather, its purpose is to sketch possible dimensions of linguistic diversity and discuss to what extent one can speculate, given information about a hypothetical species, on the kinds of restrictions to which their language might be subject.

Background

Since the Renaissance, humanity has been fascinated by the possibility of extraterrestrial life and has never ceased to speculate about extraterrestrial civilizations. Early science fiction pictured such civilizations as recognizably human and, indeed, improbably 18th Century-like.[2] Meanwhile, some science fiction has been concerned with the structure of exolanguages: e.g., the short story *Omnilingual* by H. Beam Piper from 1957, or Suzette Elgin's (1984) *Native Tongue* trilogy.

We do not expect extraterrestrials to speak in correct – if somewhat archaic or overly formal – English, any more than we expect them to be humanoid. Knowledge of biological diversity on our own planet suggests that, no matter what expectations we may have about the appearance of extraterrestrial life, we are bound to be surprised by what we find if and when we establish contact. Nevertheless, it seems possible to make educated guesses about possible features of exolanguages by extrapolating from known properties of human language, at the same time speculating on the cognitive properties of the alien intelligences who possess them and the nature of the lifeworld they inhabit.

How much can we surmise about life on other worlds? Presumably, any intelligent species is the result of evolutionary processes, which presupposes the (contemporary or historical) existence of other species. The types of co-existence patterns could, of course, differ greatly from what we know from our own planet where, from our point of view, other species function primarily as sources of protein, carbohydrates, and vitamins.

For any civilization to develop, one must first have intersubjectivity, in the sense of (Thompson, 2007) or (Zahavi, 2001): i.e., the awareness of others as agents. It is difficult to envisage a civilization *without* intersubjectivity, since the cooperation needed to build a society would seem to require it – a possible exception being species possessing collective consciousness: e.g., as technologically advanced versions of terrestrial ant colonies. It can be argued (e.g. Thompson, 2007, pp. 383 ff) that intersubjectivity is a necessary prerequisite to intelligence as we know it.

Whether or not intersubjectivity applies only within the species or embraces other species is another issue altogether. Presumably that question depends on another: whether the civilization has previous experience of interaction with other species – something which, as I have argued above, is fairly likely. One point is clear: for attempted communication with other worlds, one *must* assume intersubjectivity across species boundaries. Thus, the only scenario that could ever lead to our receiving a message from another civilization necessarily involves intersubjectivity on the part of the sending civilization.

My thought experiment is based on sheer guesswork about the alien lifeworld. However, given some reasonable guesswork, one can attempt some predictions as to which linguistic consequences one can expect to draw from alternate scenarios.

Exolanguage

It is natural to assume that any extraterrestrial intelligence must have a language: i.e., individuals – or individual cells within a collective organism – must have means of communicating with one another. The accumulation of knowledge, over generations, stretching beyond the lifespan of single organisms, is a necessary prerequisite for the development of civilization as we know it. Such accumulation of knowledge is unconceivable without some kind of communication system to transmit it.

This by no means implies that exolanguage will have any of the surface properties we as humans associate with language. I wish to consider various properties of human language and speculate on the extent to which they might carry over to exolanguage. I will look closely at the medium of communication and the function, structure, and components of language.

This kind of thought experiment forces one to address the issue of which properties of human languages result from innate genetic restrictions on cognition – or, indeed, *Universal Grammar* in the generative sense; which result from use of language in the human lifeworld; and which are realizations of universal principles of economy vs. communicative need: i.e., independent requirements for communication as such? The latter category is of particular interest to issues of exolanguage.

Medium of transmission

Communication on our planet takes place in several modalities. Insects – and, indeed, many other species, including humans – use chemical communication: e.g., pheromones. Such communication appears to be universally restricted (on Earth) to involuntary communication. A chemical communication system presents various functional problems, the most obvious being that signals tend to linger in the environment, making fast, complexly structured communication difficult. I will therefore ignore this option.

Visual communication seems to be an obvious option, given our lifeworld experience; but here we should be aware that we are really dealing with two entirely different issues. One concerns the perception of motion, gesture, or textural properties of surfaces such as the skin[3] by means of the reflection of some kind of waves. The appropriate waves could be electromagnetic radiation (for humans, light) or vibration (e.g., sound or ultra-sound). This type of communication requires a stable source of waves, be the waves electromagnetic, vibration-based, or otherwise. The most obvious candidate for a wave source, at least in the case of electromagnetic radiation, would be the primary star; although planetary rotation, leading to regular periods of darkness – almost a certainty on any homeworld of a hypothetical species – might make this less than optimal, at least if alternatives are available. Meanwhile, the exact properties of such communication would be highly dependent on e.g. the available gestural apparatus.

The other possibility is the direct emission of some kind of wave. Arthur C. Clarke's (1972) science fiction novel *Rendezvous with Rama* describes such a system: the inhabitants of Rama communicate by emitting electromagnetic radiation in various colours. For terrestrial species, however, there is a much more obvious candidate: sound waves.

With the exception of sign languages, most human languages – indeed, many non-human means of communcation as well – are based on the transmission of vibrations across short distances through the gases in our atmosphere. This might be considered to be a very terrestrial phenomenon, the result of the lifeworld in which we have evolved, but it should be noted that communication using sound waves is found in whales and other marine mammals as well, i.e., it is not necessarily dependent on the physical medium.[4] Any planet harbouring life as we know it can safely be assumed to possess an atmosphere, making the use of sound as a medium of communication conceivable.

Sound has advantages: it can carry around obstacles; it does not require a constant, independent source of waves for reflection; and it is easily generated. Any species capable of motion is capable of generating vibrations that can be used for communication – whether by rubbing forelegs together like crickets or vibrating vocal cords like humans.

Sound has disadvantages, too: it has a fairly low transmission speed, which could pose a problem over long distances; it carries poorly in thin atmospheres; and it is easily disturbed both by simultaneous signalling and non-communicative noise: e.g., in a stormy environment, it would not be the optimal strategy. All in all, however, the advantages outweigh the disadvantages; it would not be surprising to find extraterrestrial communication systems based on sound.

This does not imply that such communication systems would be reminiscent of human language in any other way.[5] Assuming that the medium is sound, we would have no way of knowing which sound distinctions were relevant to communication. The nature of the sound would depend on the anatomy of the beings involved; the relevant distinctions would depend on the their perceptual abilities. Human language makes distinctions primarily on the base of timbre: i.e., the spectral quality of the sound; as well as relative duration and, to a certain extent, relative pitch. Vowels are distinguished by the acoustic properties of the vocal cavity as a resonating chamber, with different shapes strengthening certain bands of overtones and suppressing others. Consonants are distinguished by other factors as well, such as the effect on adjacent vowels (especially in the case of stops: e.g., *p* or *t*). These properties are specific to the human articulatory apparatus.

Nevertheless, some dimensions of alternation are less likely than others: we can probably exclude absolute measures of sound amplitude as useful, since these would be affected by distance between the beings communicating. Of course, levels could be calibrated at the beginning of an utterance; but this would make communication difficult unless the communicators were subsequently stationary during the interaction. Otherwise, levels would need to be reset constantly.

Absolute pitch (or, more generally, vibration frequency) could be equally problematic for communication among beings whose locomotion is much faster than humans', since frequency patterns could easily be distorted by the Doppler effect at high velocities. More importantly, one might expect the produced pitch to lower as the speech apparatus of a being grows to adulthood, at least given a species where individuals increase in size in their development towards maturity, as is the case with all terrestrial species. In contrast, relative pitch – as in the artificial language Solresol – is a possibility.[6]

So far we have only discussed single features. However, a combination of several distinctive features will generally be more efficient than a wide variety of values along one single dimension of variation. One might therefore expect combinations of timbre, pitch, and duration to be distinctive features, as they are in human languages.

This discussion concerning medium is interesting when speculating about what actual exolanguages may be like. That said, the only communication medium likely to be relevant for our purposes, except in such science fiction scenarios as *Close Encounters of the Third Kind*, is electromagnetic radiation: e.g., radio waves. Whatever form the communication originated in, be that visual gestures, acoustic patterns, or purely abstract units of meaning, it must – if we are to receive it – be encoded in a radio message. We could then choose to visualize it as e.g. acoustic or visual patterns; but we would have no way to know in what way this visualization corresponded to reality.

In its originating form, an exolangage could also make use of two modalities at the same time. It is well known that, in human communication, the visual medium aids full comprehension through e.g. gesture and facial expression. The spoken medium itself contains several levels and types of information. Beyond the segmental level of consonants and vowels, which – for speakers of languages like English – fix the lexical meaning, one has also e.g. pitch, which is primarily used in English to indicate intonational variation but which also provides lexical information in languages like Chinese and Thai. Voice quality and timbre can provide information about gender, age, and emotional and physical state. How much of this information is linguistically relevant is a matter of debate. Certainly, it could vary between human and exolanguage, just as it varies between human languages: voice quality such as 'breathy' or 'creaky' is phonemically distinctive in certain Southeast Asian languages but not in English.

Structure

Exolanguage must involve some kind of compositionality: i.e., it must have an open-ended grammar, since a closed system can only transmit a finite set of predefined meanings. The development of civilization – in particular, the development of the technology necessary to attempt communication with other civilizations – must be capable of dealing with new situations and expressing new meanings. This implies recurring patterns of some kind, even if the complexity could be such that humans might find it difficult or impossible to recognize.

Such recurring patterns may have another source: double articulation (Hjelmslev, 1961; Hockett, 1966). In all human languages, a large number of meaning-bearing units – morphemes – can be derived compositionally from a quite restricted set of distinctive units: phonemes. The phonemes do not convey meaning in themselves; they only serve to distinguish between morphemes.

Double articulation serves an essential purpose in human language: our primary medium of communication – acoustic signals produced by the human articulatory apparatus – does not allow a great number of discrete elements. The maximal number of phonemes in a language rarely goes beyond 100; Khoisan languages like !Xóõ in Botswana stretch the limits of human articulation (Traill, 1986). The requirements of human communication demand a much greater number of signs. Human language is impossible without double articulation. That need not carry over to exolanguage – for two reasons.

First, while it seems to be a reasonable assumption that the minimum number of lexical entries in any exolanguage cannot be lower than in human language – on the assumption that we would be dealing with a civilization at least as advanced as our own – we have no way to guess the size of the primary signal repertoire. The human repertoire of fewer than 100 distinctive elements is determined by the structure of the human speech apparatus, sensitivity of the auditory system, and limitations of human perception. Communication via other modalities – e.g., absolute perception of colour frequencies – could conceivably be based on a much higher number of primitive elements. That said, it remains unlikely that the number would be a great as the minimum necessary number of signs.

Second, an exolanguage might allow free and productive combination of a small number of signs into extremely complex compounds. Consider Weilgart's (1962) highly speculative artificial language *aUI*: "the Language of Space"; where the distinction between morpheme and phoneme has been erased. Every primitive element – Weilgart designed 31 symbols – contributes not only form but (albeit rather abstract) meaning. The system is designed to avoid homophony: every combination of discrete elements is simultaneously a combination of meanings.

Double articulation may represent a uniquely human solution to uniquely human problems of communication. Nevertheless – and regardless of the cognitive resources at the disposal of other beings – double articulation remains a general tool for maximizing signal economy, allowing the number of morphemes to be on an entirely different scale from the number of phonemes. I will use the term *phonemes* here, even though the medium may be something other than sound. Even if a communication medium such as absolute colour perception could generate *thousands* of minimally distinct phonemes, double articulation could use these to create *millions* of minimal morphemes. Double articulation is simply the most economical strategy for increasing the number of morphemes beyond the number of phonemes. It is therefore likely to be found in any complex communication system, on the grounds that the null hypothesis should be that any species should make maximal use of the cognitive resources at its disposal.

In summary, recurring patterns could result from a) mere reoccurrence (of a concept), b) hierarchical syntactic structure, c) double articulation, or d) meta-linguistic signaling of a deliberate, structured message as opposed to random noise. Any alien civilization that attempts communication using a closed system of pre-defined codes makes it impossible even begin to decode their messages. Any speculation on the structure of their exolanguage would be irrelevant. Any civilization advanced enough to attempt such communication must surely be aware of this.

Recursion and binarity

What kind of language structure should one expect? Human languages show a strong tendency for clauses to be built from binary structures. Second-position phenomena such as Wackernagel clitics and verb-second constructions are not uncommon: clausal elements preceded by one – and only one – element, which can be of any category. There are *no* attested instances of strict third-position phenomena where an element occurs obligatorily in third position (preceded by exactly two elements which can be of any category). It seems that, where ordering restrictions are defined in terms of positions rather than discrete categories, the number two is the one that matters.

Indeed, the common assumption in most contemporary models of generative syntax is that all syntactic structure is necessarily binary. Apparent exceptions – in particular, ditransitive verbs – have been argued to be implicitly binary (e.g. Larson, 1988).

Must the same pattern carry over to exolanguages, or is it an accident of the human lifeworld? We all are aware of binarity as a common feature of everyday life. Could linguistic binarity be nothing more than another instance of this? Could we predict a corresponding universal principle somewhere else of trinarity or – given Arbib's (2011, *this volume*) example of an alien octopus-like lifeform – even "octonarity"?

To answer this, consider how linguistic structure is built. In generative theory, one of the most basic properties attributed to human grammar is recursion: the ability of a structure to incorporate within itself a self-similar structure – in theory allowing potentially infinite embedding. Hauser, Chomsky, and Fitch (2002) argue that recursion is *the* defining characteristic of human language: given recursion, all else follows.

How does recursion apply to binarity? Assume that the building of the simplest structures involves joining two elements A and B. How does one go further? One possibility would be to create an entirely new operation that combines three – rather than two – elements, then one that combines four rather than two or three, and so on.[7] Each move requires a new operation. By contrast, one could simply apply a generalized recursive operation that takes any two complex elements and treats them as though they were atomic: each level would never be more than binary, but the entire structure could become limitlessly complex. Binarity may not be a universal property of language (Holmer, 2011), but the combination of binarity and recursion *is* an optimally efficient structure-building strategy.[8]

Some have recently argued (Everett, 2005; Evans and Levinson, 2009, and sources cited therein) that recursion is not universal in human language. Their target, however, is a narrow one: the embedding of a particular linguistic category within a self-similar structure, such as a clause within a clause. Recursive binarity is a much broader principle; it is hard to imagine language structure that does not in some way incorporate this feature, in the sense that it does not allow complex elements to be embedded as constituents in other complex elements. Understood this way, recursion is not an innate property of human language (as proposed in generative grammar) but a matter of economy (like double articulation). If we assume the need to build grammatical structures of a given complexity, the simplest possible mechanism is the combination of binary branching with recursion (i.e. the recursive application of binary branching throughout the structure). Given this, we propose that an articulated hierarchical structure of some form or another should also be found in exolanguage.

Non-configurational languages might seem to pose a counterexample (Austin and Bresnan, 1996). In such languages, one cannot divide a clause meaningfully into smaller contiguous units: i.e., these languages lack phrases in the traditional sense, and word order is entirely free. However, what these languages show is not a lack of syntax but the use of other strategies to express it. One obvious alternative to linear ordering is morphology, which certain languages take to extremes in indicating constituency.

Take the Australian language Kayardild (Evans, 1995): every word that belongs syntactically to a given phrase is assigned the morphology corresponding to the function of that phrase. Words that are deeply embedded structurally – that would be deeply embedded linearly in a configurational language – are marked by several levels of affixation. A single word can show several overt levels of case marking, tense, etc.

1 *maku-ntha* *yalawu-jarra-ntha* *yakuri-naa-ntha*
woman-COMP catch-PST-COMP fish-TNS[9]-COMP
dangka-karra-nguni-naa-ntha *mijil-nguni-naa-ntha*
man-GEN-INSTR-TNS-COMP net-INSTR-TNS-COMP
"I saw that the woman must have caught fish with the man's net."
(Evans, 1995, p. 115)

This is an extreme example; but it illustrates how constituency is not logically dependent on linear ordering.[10] In several other Australian languages, including those with less morphological marking than Kayardild, word order is notoriously free. Constituency can be seen as a possible functional universal, grounded in economy, that can be realized through linear ordering, morphological marking, or both.

Speech acts

What is language used for? In linguistics, utterances traditionally can be divided into different speech-act or *illocutionary* types. Three of these – stating facts, requesting information, and issuing instructions – are more or less universal in human language, differentiated by appropriate grammatical constructions: declarative clauses for stating facts, interrogative clauses for requesting information, and imperative clauses for issuing instructions. The different syntactic constructions can be distinguished by means of word order (2a, 2b) or morphology (2c), as illustrated by the following German examples.

2 a.

Du	*is-st*	*deine*	*Wurst.*
2s	eat-2s	2s.POS	sausage

"You are eating your sausage."

b.

Is-st	*du*	*deine*	*Wurst?*
eat-2s	2s	2s.POS	sausage

"Are you eating your sausage?"

c.

Iss	*deine*	*Wurst!*
eat.IMP	2s.POS	sausage

"Eat your sausage!"

The functional differerence between these three speech-act types and the concomitant difference between syntactic clause types is a good candidate for a universal principle of communcation systems.

One might object that, on a more abstract level, these three types amount to a single speech-act type: issuing instructions. A request for information is an implicit instruction to the interlocutor to supply that information. Similarly, a statement is an implicit instruction to the interlocutor to be aware of the fact stated and to acknowledge this awareness in some way: e.g., by nodding. However, this objection is not really a problem, since it can be argued that on this level, *all* human behaviour can be interpreted as some manipulation of the environment. That the three illocutionary types interrelate does not mean that they cannot be distinguished.

Another possible objection might be that humans tend to express instructions or commands indirectly whenever possible, by cloaking them in statements of fact (3a) or in questions (3b).

3 a. I would be very grateful if you would lend me the book.
  b. Could you lend me your book?

Clearly, this is a conflict-avoidance mechanism: one that is probably necessary for survival in a complex society comprising individuals, genetically designed for life in small clan groups (Dunbar, 1992), who are forced – as a matter of historical contingency – to live in much larger ones. This accords with other human conflict-avoidance mechanisms, such as politeness. What all these mechanisms do is bridge the mismatch between modern human demography and our genetic heritage. Once the species has adapted biologically to the modern lifeworld, politeness may no longer be required for human communication. We have no way of knowing how many generations such adaptation would require, though we might guess that politeness is a property of fairly young civilizations. The period of time requiring these conflict-avoidance mechanisms should presumably be short compared to the civilization's lifespan. One might therefore speculate that such mechanisms should be rare among exocivilizations.

In the light of this, it is very likely that these three illocutionary types (stating facts, requesting information, and issuing instructions) may all be universally necessary functions of language (cf. however the discussion of information searches and interrogatives in the following section for a potential counterargument). We should not forget, however, that human societies employ other speech-act types as well, such as performatives: e.g. *I hereby pronounce you man and wife* or *I promise to buy you a beer*. These are typically the product of the structure of human societies.

One should therefore allow the possibility that exolanguages may possess speech-act types we cannot even imagine, but which are the product of the societies in which they are used. Such speech acts would presumably not be relevant to us. Consider an alien visitor to Earth, whom we would hardly address with any performative speech acts; neither would we include any performatives in our outgoing interstellar messages.

Further, the most basic function of human language lies elsewhere, in something so basic that it is often forgotten: establishing and maintaining social relations. Humans primarily communicate, not to transfer or request information, nor to issue instructions, but to acknowledge awareness of the other as interlocutor: to greet one another. One reason why this function is not central to linguistic typology is that there is no obvious clause type which corresponds to it. Within any given language, one finds an entire spectrum of constructions that can express greetings, from formulaic expressions lacking internal structure (*Hello!*) to interrogative (*How are you?*), imperative (*Have a nice day!*) and declarative clauses (*I'm pleased to meet you*).

If this is important in human communication, it is even more so in interstellar contact. Even assuming that exolanguages have speech-act types corresponding to declaratives, interrogatives, and imperatives, one should hardly expect a message from an alien civilization to encode either an instruction or request for information, given the necessary timescale involved in interstellar communication. Any message we might expect to receive from another civilization might well have the secondary purpose of stating facts about the civilization that sent it, but the most important function would be expressing some kind of greeting, establishing contact. If this is encoded in any idiosyncratic construction which is not typical of how the grammar of the language otherwise works (as is often the case with human languages), any message we receive may in any case be quite unrepresentative of the exolanguage involved.

For the remainder of this paper, I will assume that any exolanguages will somehow be commensurable with human language: their functions, at least in part, the same. This is not an entirely unproblematic assumption, but I see no way of progressing without it.

Interrogatives

Interrogatives are the grammatical encoding of information requests. One might consider them a special instance of requests in general. However, one reason why they are central to human society is the following: Much of human culture depends on shared knowledge; yet our only access to knowledge as individuals is personal experience. Being able to harness the personal experience of others becomes an important source of knowledge; requesting such knowledge becomes an important social tool.

Our interaction with other sources of information, such as encyclopedias or Google, does not normally involve the use of interrogatives. Instead, one simply looks up e.g. a relevant reference term and accesses the information directly. Given this behaviour, one might imagine a society where information is accessed directly, without social interaction and where, given time for linguistic structures to adapt, interrogatives may no longer be a central feature of communication or may even disappear altogether. However, for now I will simply assume their existence.

Polar interrogatives

A polar (yes/no) interrogative involves, in the simplest case, a full proposition, uttered together with some signal requesting a comment on its truth value. Thus, the interrogative *Did you feed the cat?* corresponds to the statement *You fed the cat* together with a request for confirmation or denial. Sometimes this request is implicit and only revealed by context.

Most times, however, the signal is overt: e.g., a marker of uncertainty expressed by intonation. In many human languages, polar interrogatives and declaratives differ only by intonation patterns. Typically, the interrogative pattern involves a rise in intonation towards the end of the utterance, though the converse is also known to occur.[11]

4	a.	*Juan*	*ha*	*comprado*	*un*	*coche.*
		Juan	has	bought	1	car
		"John has bought a car."				
	b.	¿Juan	ha	comprado	un	coche?
		Juan	has	bought	1	car
		"Has John bought a car?"				

In the Austroasiatic language Kammu spoken in northern Laos, polar interrogatives can be identified either by intonation alone or by insertion of the particle *béc*, signaling uncertainty (5a). This particle can be translated as "perhaps" (5b), leading to grammatical – but hardly functional – ambiguity when translated into English (5c).

5	a.	*kə̀ə*	*cə̀ə*	*rɔ̀ɔt*	*béc*						
		3S.M	IRR	come	Q						
		"Is he coming?"									
	b.	*mə̀h*	*yʌ̀ʌ*	*ò*	*pɨ̀p*	*pɛ̀*	*béc*	*ò*	*pə́ə*	*pɨ̀an*	*ká*
		be	for	1S	meet	goat	Q	1S	NEG	get	fish
		"Maybe it's because I met a goat[12] that I can't get any fish."									
	c.	*kùuñ*	*béc,*	*pə́ə*	*kùuñ*	*béc*					
		see	Q	NEG	see	Q					
		"Maybe I'll see it; maybe not." ("Will I see it; won't I see it?")									

While some languages use such strategies, most display an overt segmental marker that is unambiguously interrogative. Presumably this reflects the importance to human communication of being able to identify interrogatives unambiguously. Overt markers can be of various types. One simple strategy is to present the proposition – or parts of it – twice, once in negated form; this gives the interlocutor the option to select the correct alternative. This is illustrated for Chinese in (6).

6	*Ni*	*yao*	*bu*	*yao*	*he*	*cha?*
	2S.	want	NEG	want	drink	tea

“Do you want to drink tea?”

The interrogative marker can also be a separate particle that, depending on the language, can be inserted in various positions: clause-initial, clause-final, or some other fixed position. Some languages use less transparent mechanisms such as reversing the order of the first two clause elements, as is the case in German and Swedish and, to a certain extent, English; but this is extremely unusual.

Given the extreme variation in human language, one would be hard pressed to guess what interrogative constructions to expect in an exolanguage. Among other possibilities, the propositional content may be expressed in one medium (e.g., as the morphemes or words in the Spanish example above), the interrogativity in another (e.g., as intonation in the same example). Two entirely different modalities might also be involved: e.g., acoustic and visual.[13]

Content interrogatives

The other major interrogative type in human language is the content interrogative: the *wh*-question, a proposition containing a variable such as *who*, *what*, etc. The proposition is semantically a statement; the illocutionary force of the question addresses the variable’s identity. Thus, an English wh-question like *What did you buy?* can be rephrased as (7).[14]

7	Proposition:	You bought X.
	Request for information:	Identity of X?

Most human languages collapse this to a single clause, the simplest version of which includes a question word as a variable – as e.g. in Chinese (8a). The presence of the variable is sufficient indication that one is dealing with a question. The illocutionary force of the question can be made more salient by *wh*-movement, placing the variable in a predefined position: usually clause-initial, as in English (8b).

8 a. *Ni* *mai-le* *shemme?*
2S buy-PRF what
"What did you buy?"
b. *What* did you buy?

Some languages more or less preserve a double structure, where the question of the variable's identity is the primary construction, and the proposition is expressed by a secondary construction embedded within a nominalization. This is the case in Austronesian languages, represented here by two aboriginal languages from Taiwan: Puyuma (9a) and Seediq (9b).

9 a. *A* *manay* *nu=tr<in>ima-an?*
NOM what 2S=<PST>buy=LF
"What is it that you bought?"

b. *Maanu* *ka* *b<n>rig-an=su?*
what NOM <PST>buy-LF=2S
"What is it that you bought?"

How is this relevant to exolanguage? All the examples given here involve single question variables. While it is possible to make multiple content questions such as *Who bought what?*, human languages are not specifically adapted to this,[15] and such constructions easily create ambiguity: e.g., when the variable doubles as an indefinite pronoun, as in Chinese (10).

10 *Shei* *mai-le* *shemme?*
who buy-PRF what / something
"Who bought what?" / "Who bought something?"

Clearly, human communication assigns greater value to the salience and clarity of a single information request than the capacity to raise several question variables simultaneously. This suggests a possible dimension for variation: assuming that content questions exist in a given exolanguage, one has no a priori reason to expect them to be restricted to a single variable at a time. Regardless of the mechanisms involved, a system capable of dealing with multiple variables would not only make multiple interrogativity possible (11a), but allow other constructions impossible to express in English (11b).

11 a. · *X* bought *Y* in the bookshop.
· Supply values for *X* and *Y*.

b. *What did Bill see the man that John gave _ to?
· Bill saw *X*.
· *X* is a man.
· John gave *Y* to *X*.
· Supply value for *Y*.

In brief, the human tendency to use content interrogatives with single variables more likely derives from human communicative needs and limitations on human cognition than from universal structural requirements. We should therefore not expect this property to carry over to exolanguage.

Grammatical categories

What will the basic components of exolanguage syntax be? Can they be grouped in ways reminiscent of word classes – parts of speech – found in human language: i.e., based on categories of morphosyntactic behaviour: linear position, inflection, etc.; and categories of meaning: objects, actions, etc.? Do categories differ in morphosyntactic behaviour? If they do, do the categories which can be distinguished on morphosyntactic grounds correspond in any way to natural categories which can also be distinguished semantically? If so, which semantic distinctions are relevant?

To answer these questions, one must examine the variation in human language. It seems clear that human language, used by a single species in a single lifeworld – and which might have developed from a single protolanguage[16] – represents the absolute minimum imaginable linguistic variation. Anything not universal in human language can hardly be assumed for exolanguage.

How much *is* universal in human language? One can argue that the meanings are (more or less) universal: all human languages have ways to refer to objects; to happenings; and to properties such as colour, size, quality, manner of action, frequency, etc. However, the mapping of these meaning types onto grammatical categories is far from universal. In various east and southeast Asian languages including Chinese, typical adjectival meanings are expressed by something that behaves like a verb.[17] In the Austronesian aboriginal languages of Taiwan, adverbial meanings are expressed by what behaves like a verb in every respect, morphologically and syntactically: they inflect for tense, mood, and voice like any other verb (Holmer, 2010). In Finnish, negation is a verb; in German, it is an adverb; in Estonian, it forms a category of its own with some properties of both verb and adverb.

Furthermore, European languages do not represent any maximal repertoire of word classes, from which other languages differ by reduction. A survey of the languages of the world reveals many examples of word classes not found in European languages: e.g., expressives and ideophones in southeast Asian and sub-Saharan African languages; classifiers in most east Asian languages: Chinese, Thai, etc. When it comes to word classes, variation is the norm.

Does one find *no* universal word-class distinctions? One frequently offered candidate is the noun-verb distinction; however, recent research suggests that even this distinction is not universal. That said, what one concludes depends on what one means by these categories. All human languages have ways of referring to objects and to actions or events. One must assume that any exolanguage possesses analogous mechanisms. However, from a linguistic point of view, nouns and verbs are distinguished not by the meanings they represent, but by their grammatical behaviour. In this latter sense, Native American languages of the Pacific Northwest are sometimes claimed to lack a noun-verb distinction (Evans and Levinson, 2009). In Nuuchahnulth, both nouns (words with nominal meanings) and verbs (words expressing actions) can behave syntactically as *either* nouns or verbs: e.g., by having definite articles attached or showing verb morphology. This is illustrated in (12a,b) by examples from Tallerman's (2009, p. 469) response article to (Evans and Levinson, 2009). This property does not extend to personal names: names in Nuuchahnulth cannot show verb morphology (12c).

12 a. *mamuuk-maa* *quuʔas-ʔi*
work-3S:INDIC man-the
"The man is working."

b. *quuʔas-maa* *mamuuk-ʔi*
man-3S:INDIC work-the
"The working one is a man."

c. **Jack-maa*
Jack-3S:INDIC
("He is Jack.")

Thus, Nuuchahnulth does indeed have word-class distinctions; but the noun-verb distinction does not seem to be one of them. Instead, one finds a name - non-name distinction. Names can be seen as the extreme point on a REFERENT ↔ PREDICATE scale: the diametric opposite of verbs.

All human languages distinguish referents from predicates. Even in Nuuchahnulth, one element – the prototypically verbal meaning in (12a) and the prototypically nominal meaning in (12b) – bears verb morphology and is interpreted predicatively; the other is interpreted as the subject referent. The ungrammaticality of inflecting names as verbs reflects the prototypically referential nature of names, which normally do not serve as predicates – though the English translation of (12c) shows that this is not an absolute restriction. As in English, Nuuchahnulth also displays a word class distinction along the REFERENT ↔ PREDICATE scale. However, while English draws the line between nouns and verbs, Nuuchahnulth draws it closer to the REFERENT extreme, between names and the rest.

So although the referent-predicate distinction generally corresponds to the noun-verb distinction, it is logically independent. The referent-predicate distinction may well be universal in human language. We should not be surprised to find, in any given exolanguage, some mechanism for directing attention to a referent or referents and some mechanism for addressing situations involving the referent(s). At the same time, it is not necessarily the case that this distinction requires a word-class distinction in any morphosyntactically recognizable sense.

Braine (1963) proposed that the earliest stages of language acquisition involve two classes of lexical items: content words (Braine's *open class*) and function words (Braine's *pivot class*). In his model, the earliest syntactic rule at the stage of two-word formations allows any two-word combination *except* linking two members of the pivot class: in other words, any syntactically well-formed construction must involve at least one content word.

The distinction between content words and function words is another plausible candidate for universality. Applied to morphemes rather than words, it *is* universal to human language: all languages distinguish open classes of morphemes with lexical meanings from closed classes of functional morphemes. This is likely to be part of any exolanguage, on the grounds that functional morphemes are the universal mortar binding the lexical bricks of language. At the same time, the degree of abstraction characterizing the functional meanings makes such a universal claim almost trivial. Further, given that the precise boundary between functional and lexical elements varies between languages,[18] it would be impossible to predict where this boundary would be drawn in exolanguage. Finally, these dimensions – of lexical meaning and syntactic function – might conceivably be expressed in different modalities (e.g., as is the case in certain languages in central Africa, where the lexical meaning is expressed by the segmental material – consonants and vowels – while tense-marking is expressed by tone).

In an exolanguage, word-class categorization may be quite different from that found in any human language; it may make distinctions unknown in human languages. The behaviour of word-class categorization is arguably dependent on the cognitive properties of the life-forms involved. To illustrate how human cognition can affect word-class categorization, I will look at one word class: personal pronouns.

Pronouns

The purpose of pronominal systems is to keep track of referents. In the simplest case, a pronoun's inclusion in a clause serves two functions: economy and reference. Pronoun use is economical by allowing one to replace a noun phrase with a single, usually shorter, proform. It is referentially useful by allowing one to indicate that the referent is already present in the discourse – and not simply another instance of the category indicated by the NP.[19]

Pronominal systems are generally limited. They are used to identify speaker (*I*), listener (*thou*), and other parties (*she, he*), and groups including these individuals (*we, they*). They can indicate number: e.g., singular vs. plural (or dual); and gender (*she/he/it* in English). They can indicate social status, or the familiarity between speaker and listener as in various familiar, polite, and honorific forms. All these properties reflect the society in which they are used: hierarchical, egalitarian, etc. They have one thing in common: tracking a limited number of referents in any given discourse.

Assume that the need to limit the number of referents is a function of limitations on human cognition. Assume further that an alien species may not be subject to the same cognitive limitations – or not to the same degree. What consequences might this have? In particular, what are the alternatives to deictic pronominal reference?

One possible candidate is absolute reference, where every individual has a unique name that is always used in full, never replaced by a proform. While such systems are conceivable, they have functional disadvantages: inefficiency and memory load aside (which might or might not be an issue, depending on the cognitive capacities of the species), the most serious problem would be the difficulty of avoiding deictics in other circumstances. Even if the culture had an absolute reference system for individuals, it could not possibly have one for all other objects in its lifeworld (and even less so for objects outside its traditional lifeworld, e.g., off-planet). Here, at least, deictics – specifically, demonstratives – are probably necessary. If one must assume demonstratives for inanimate objects, it would seem extremely hard to avoid generalizing this to animate ones – creating what is, in effect, a pronominal system. Japanese and Basque are just two of countless human languages that have developed third person pronouns from demonstratives.

One can therefore probably conclude that any alternative to human pronominal systems cannot involve absolute reference. A more likely alternative is a system that, like human pronominal systems, uses deictic reference, but permits a greater number of referents in any given discourse than does human language. How might this work?

Consider a passage from the Basque novel *Babilonia* by Joan Mari Irigoien (1990), compared to an English translation. All details are irrelevant except the pronominal system. The English translation is hard to follow: sometimes the pronoun *he* refers to Zipriano, sometimes to Demetrio, and sometimes to Bernardo. The context of the story makes it difficult to work out the pronominal reference without use of indices.

> Zipriano$_i$ had two brothers, Demetrio$_j$ and Bernardo$_k$, and both had gone to the war, but although it didn't worry (him$_i$) that Demetrio$_j$ was at war, because **he**$_j$ [hura] was almost as strong as **he**$_i$ [bera], (he) really did worry about Bernardo$_k$, because **he**$_i$ [bera] could not imagine how **he**$_k$ [hura], that weak boy, could go to the hills to fire guns, and not **he**$_i$ [bera], the "Strongman of Babilonia".

These problems do not arise in the Basque original. The pronoun *bera* refers consistently to Zipriano: the Topic referent of the entire passage. The pronoun *hura* refers to the local embedded subject Demetrio in the first section, and Bernardo in the second section (the change in reference being marked by the overt mention of the name Bernardo).

> *Ziprianok*$_i$ bi anaia zituen, *Demetrio*$_j$ eta *Bernardo*$_k$, eta biak ziren gerrara joanak, baina *Demetrio*$_j$ gerran zebilela jakiteak kezkatzen ez bazuen$_i$[20] ere, **hura**$_j$ ia **bera**$_i$ bezain kementsua zelako, *Bernardorenak*$_k$ bai, guztiro kezkatzen zuen$_i$, **berari**$_i$ ez bait zitzaion buruan sartzen nola mutil makal **hura**$_k$ mendira joan zitekeen tiroak botatzera, eta ez **bera**$_i$, "Babiloniako Indartsua".

The Basque system thus goes one step further than the English one – unless one uses phrases like *the former* or *the latter* to make up for this weakness. Basque represents an instance of a system in which the number of referents tracked by the pronominal system can be marginally increased when compared with English. Similar results are obtained with the obviative systems found in some native American languages as well as Chamorro in Guam; see (Aissen, 1997) for an excellent overview of the phenomenon. That said, human languages seem incapable of going much further.

An exolanguage could possibly exploit this dimension to increase the number of available referents: either by extending reference back through the discourse using such deictics as “the third-but-last referent mentioned”, etc.; or by making reference to conventionalized classifications analogous to gender, social function, or social status – however these might be construed in the society involved. The possibilities for variation are virtually endless. Meanwhile, the point remains that any language – human or exo- – requires a deictic reference system of some kind to be efficient.

Semantics

The semantic makeup of a language is the part of the language most dependent on its speakers’ culture and lifeworld. Typically, any semantic generalizations that can be made for human language – such as patterns of metaphor – can be derived from shared experience of a distinctly human lifeworld. One cannot point at any semantic features of language that can be considered universally necessary for communication, without taking the lifeworld into account.

Rather, the semantics of any language reflects the semantic needs of the culture using the language, including anything of relevance to that culture. Without knowledge of that culture, what to include is anybody’s guess. Of course, one can suggest some trivially obvious candidates for inclusion: some reference to physical motion, to perception, to activities that can be performed by the agents in question. All of this is necessarily vague: we have no idea what lifeworlds may be involved. Perhaps the likeliest candidate for inclusion is reference to the agents’ wellbeing or survival.

To this one might well add spatial semantics: some means of talking about spaces. This is one domain where interesting contrasts might be found, which could be predictable to some extent from the life-world of the culture. One finds an almost universal distinction in human language between *here* (close to speaker) and *there* (not close to speaker). This is often extended to include further degrees of distance: position relative to the position of the interlocutor, varying degrees of distance from either speaker or interlocutor – as well as position within or beyond the range of visibility, or vertical position relative to the speaker or interlocutor.

An alternative making no use of deixis is, once again, absolute reference: in this case, a system of absolute coordinates. Such a system is potentially less cumbersome than one for individuals or objects. While a system of absolute reference for objects must necessarily be open, a coordinate system could easily be closed. Indeed, for communication across large distances, involving many interlocutors using radio or similar technology, a coordinate system would be more useful than human deictic systems. The latter are suited to communication across short distances, where speaker and interlocutor have a more-or-less shared spatial point of view.

Another potential dimension for variation presents itself: as a species, humans are largely surface bound, with great freedom of motion in two dimensions but extremely restricted motion vertically. This is reflected in spatial semantics. All human languages possess concepts referring to the vertical dimension: all distinguish *upwards* from *downwards* and possess corresponding verbs such as *rise* and *fall*. Meanwhile, they do not universally distinguish directions of horizontal motion: some languages do, but the distinctions are entirely language-specific. Horizontal motion verbs focus on the *manner* of movement: e.g., walking, running, crawling, or floating. Even when we refer to the motions of species not as vertically restricted as ourselves – e.g., fish swimming or birds flying – the standard visualization is motion in the horizontal plane. A hawk diving to catch a rabbit would hardly be described as flying in the typical sense. Clearly, our categorization of spatial semantics and motion verbs is determined by our lifeworld.

To see this, we need only consider Arbib's (2011, *this volume*) octoplus: an aquatic species whose lifeworld makes the concepts of up and down much less relevant than they are to us. If we go further and imagine an entirely interplanetary-space-based species whose life-world is characterized by weightlessness, such concepts should be entirely meaningless, unless they survive as linguistic relics. Extrapolating from life as we know it, it seems reasonable (although admittedly not logically necessary) to assume that all life – and therefore all civilization – must have originated on the surface of a planet. Likewise, given that some kind of communication system is a prerequisite for civilization as we know it, we also assume that language or other communication system must have evolved before the species left the planetary surface to move into space. On the planetary surface, the vertical dimension would have been extremely important – perhaps especially when the civilization was entering its space age, given the enormous amounts of energy required to exit the gravity well of a planet. Therefore, it is likely that planet-based species would minimally need to possess spatial deictics expressing concepts of up and down: i.e., towards or away from the planetary centre.

If, however, a civilization has left the surface of any planet for good – assuming this scenario is even possible – concepts referring to the vertical dimension would lose their value. Once the language had adapted to the new lifeworld, such concepts would serve no function and either disappear or be adopted for other purposes. The only way they might remain useful is when taking another referent, such as the primary star, as a base point. Even then, it seems unlikely that they would be as central to lifeworld experience as up and down are for us.

Conclusion

What we have seen in the preceding sections is that human linguistic diversity is the norm rather than the exception. We have no reason to expect to to be able to predict properties of exolanguages when we cannot predict those properties for our own species. That said, within diversity one can find plausible dimensions for predictable variation between human language and exolanguage. How far we can extrapolate along these dimensions is impossible to guess in advance.

Human languages appear to possess a number of universal properties: e.g., double articulation, pronominal systems, deixis, horizontal-vertical asymmetries in spatial semantics, indirectness as a means of conflict avoidance. Some of these are crucially dependent on limitations of human cognition, or on specific properties of human experience of the distinctly human lifeworld: e.g., deictic systems and horizontal-vertical asymmetries. Such properties cannot be expected to carry over to exolanguage – though if one can identify how they relate to their underlying causes, one might gain insight into how alien lifeworlds and cognitive capacities should shape the development of alien language and be able to predict which linguistic properties to expect, given various sets of assumptions. Other properties of human language derive from what appear to be universally optimal solutions for communication. These *can* be expected to carry over to exolanguage.

In the end, making educated guesses about hypothetical exolanguages requires understanding that some universal properties of human language are functionally necessary – or at least optimal – for language to serve its communicative function; others will prove specifically human. The crucial task for modern linguistic theory is recognizing which is which.

Acknowledgments

I gratefully acknowledge the support, in writing this paper, of the Bank of Sweden Tercentenary Foundation and the Center for Cognitive Semiotics at Lund University. The fieldwork was generously made possible by a 2009 Faculty Research Grant for Taiwan Studies from the Ministry of Education of the Republic of China (Taiwan).

Notes

[1] Here we instead refer to valuable suggestions in Dunér (2011).
[2] A typical example is Burroughs' story *A princess of Mars*, published in 1917.
[3] Cf. Arbib's (2011) inspiring presentation at the symposium *Interstellar Communication: Semiotic, Linguistic and Cognitive Approaches*, published in this volume.
[4] Note, however, that certain frequencies are more affected by the medium than others, e.g. high frequencies carry poorly in liquids.
[5] The Klingon language designed by Marc Okrand for the science fiction series *Star Trek* is intended to be extremely exotic from a human point of view. However, due to the humanoid nature of the Klingons, it is improbably human, being no more exotic than some native American languages by which Okrand was inspired, and in some senses notably less exotic than some other existing human languages mentioned in the present paper (e.g. !Xóõ or Kayardild).
[6] Some natural human languages do indeed have a very developed system of lexical tones (one Hmong-Mien language, Black Miao, is indeed reported to have distinct 5 level tones, cf. Goddard 2005, 36). However, in all human language, tone is used as one dimension for creating phonetic contrasts, not as the only dimension. In contrast, in the science fiction film *Close Encounters of the Third Kind*, a recurring theme is the *re-mi-do-do-sol* snippet, an unlikely tribute to the Western musical tradition, which appears to function as something between a callsign and a greeting.
[7] Unless a generalized process of iteration is developed, in which case we are not really talking about structure in any meaningful sense.
[8] Krifka (2006; 2007) argues that the topic-comment structure in human language may have its origin in the left-right asymmetry in bimanual coordination – one hand holds an object while the other hand manipulates it. This suggestion is supported by similar facts in both sign language and gesture (Engberg-Pedersen, 2011). However, this is not necessarily a counterargument againt binarity as a universally optimal strategy for building structure, but rather outlines a species-specific human way of making use of binarity for communicative purposes.
[9] This case is actually a form of the ablative case, but its function here is to reflect the fact that the verb is in past tense, hence we have glossed it as TNS for "tense" here.

[10] Interestingly enough, however, although the function of each word is fully recoverable from the morphology, constituents in Kayardild still tend to be linearly contiguous (op.cit., p. 234). Thus, Kayardild does not actually use this mechanism to replace linear contiguity, simply to supplement it.
[11] In the Austronesian language Puyuma, spoken in southern Taiwan, the situation is almost reversed. In Puyuma, declaratives are recognizable by a final high pitch not found in interrogatives. Thus *Tremakaw dra belbel i Sigumulri* [stole ACC bananas NOM Sigimulri] is linearly ambiguous: with a high pitch on the final syllable *–lri* it means the declarative: *Sigimulri stole bananas*, while with a low pitch on the final syllable it represents the corresponding polar question: *Did Sigimulri steal bananas?*. For more on this phenomenon cf. Teng (2008, p. 211).
[12] In traditional Kammu culture, goats are viewed as bringing bad luck.
[13] Although, if a message is encoded into a single medium, such as radio, we would probably not notice this.
[14] It could also be expressed using formulae from predicate logic. However, this would not contribute any more clarity for our present purposes, so a simple rephrasing using English can suffice here.
[15] Here Bulgarian, where both question words can appear together in the initial interrogative position, is an exception.
[16] This represents the monogenetic view of language origin, which constrasts with the polygenetic view, which claims that language may have developed independently in various human populations. We will not address this issue here.
[17] The lack of distinctive verb morphology makes this statement trivial, however.
[18] E.g. the concept of "want" or "wish" is expressed as a lexical verb in English, but as a derivational prefix *mk–* in the Austronesian language Seediq, spoken in Taiwan, cf. Holmer (1996, pp. 49 ff).
[19] E.g. the example *Bill bought bacon. I fried bacon. We ate bacon.* is not only uneconomical. It is also unclear whether or not the three clauses actually refer to the same bacon, or if they are three independent instances (pragmatic considerations normally force us to view them as non-coreferent, precisely because the noun is used instead of the pronoun).
[20] In some cases here we also indicate reference indexes on verbal forms, since they incorporate agreement morphemes.

References

Aissen, J. 1997. "On the syntax of obviation." *Language* 73:705–750.

Arbib, M. 2011. "If space aliens were like octopuses, what would we 'say' to them?" Paper presented at the symposium on *Interstellar Communication: Semiotic, Linguistic, and Cognitive Approaches*, Pufendorf Institute for Advanced Studies, Lund. Published in this volume.

Austin, P. and Bresnan, J. 1996. "Non-configurationality in Australian aboriginal languages." *Natural Language and Linguistic Theory* 14:215–268.

Beam Piper, H. 1956. "Omnilingual." *Astounding Science Fiction*. February 1957.

Braine, M. 1963. "The ontogeny of English phrase structure: The first phase." *Language* 39:1–13.

Burroughs, E. R. 1917. *A Princess of Mars*. Chicago IL: MacClurg.

Clarke, A. C. 1973. *Rendezvous with Rama*. Orlando FL: Harcourt Brace.

Dunbar, R. 1992. "Neocortex size as a constraint on group size in primates." *Journal of Human Evolution* 22:469–493.

Dunér, D. 2011. "Cognitive foundations of interstellar communication." In *Communication with Extraterrestrial Intelligence*, edited by D. Vakoch, 449–468. Albany NY: SUNY Press.

Elgin, S. 1984. *Native tongue*. New York: DAW Books.

Engberg-Pedersen, E. 2011. "Cognitive foundations of topic-comment and foreground-background structures: Evidence from sign languages, cospeech gesture and homesign." *Cognitive Linguistics* 22–4:691–718.

Evans, N. 2005. *A Grammar of Kayardild: With Historical-Comparative Notes on Tangkic*. Berlin: Mouton de Gruyter.

Evans, N. and Levinson, S. 2009. "The myth of language universals: Language diversity and its importance for cognitive science." *Behavioral and Brain Sciences* 32:429–492.

Everett, D. L. 2005. "Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language." *Current Anthropology* 46:621–46.

Goddard, C. 2005. *The Languages of East- and Southeast Asia: An Introduction*. Oxford: Oxford University Press.

Hauser, M., N. Chomsky, and W. Tecumseh Fitch. 2002. "The faculty of language: What is it, who has it, and how did it evolve?" *Science* 298:1569–1579.

Hjelmslev, L. 1961. *Prolegomena to a Theory of Language*. Baltimore MD: Waverly Press.

Hockett, C. 1966. "The problem of universals in language." In *Universals of Language*, edited by J. Greenberg, 1–29. Cambridge MA: MIT Press.

Holmer, A. 1996. *A Parametric Grammar of Seediq*. Lund: Lund University Press.

—. 2010. "Seediq adverbial verbs – a review of the evidence." In *Austronesian and Theoretical Linguistics*, edited by R. Mercado, E. Potsdam, and L. Travis, 163–182. Amsterdam: John Benjamins.

—. 2011. “Deconstructing generativism.” Paper presented at *Grammar in Focus*, February 2011. Lund: Lund University.

Irigoien, J. M. 1990. *Babilonia*. Donostia, Basque Country: Elkar.

Krifka, M. 2006. “The origin of topic/comment structure, of predication, and of focusation in asymmetric bimanual coordination.” Paper presented at *Nascent Language Conference*, Rockefeller Foundation, Bellagio, Italy, October 4–8, 2006.

—. 2007. “Functional similarities between bimanual coordination and topic/comment structure.” *Interdisciplinary Studies on Information Structure* (ISIS): Working Papers of SFB 632, 08, 61–96.

Larson, R. K. 1988. “On the double object construction.” *Linguistic Inquiry* 19:335–391.

Tallerman, M. 2009. “If language is a jungle, why are we all cultivating the same plot? Comment to Evans & Levinson.” In Evans and Levinson 2009.

Teng, S. F.-C. 2008. *A Reference Grammar of Puyuma, an Austronesian Language of Taiwan*. Canberra: Pacific Linguistics.

Thompson, E. 2007. *Mind in Life: Biology, Phenomenology and the Sciences of Mind*. Cambridge MA: Harvard University Press.

Traill, A. 1986. “Phonetic and phonological studies of !Xóõ Bushman.” *Quellen zur Khoisan-Forschung* 1. Amsterdam: John Benjamins.

Warner, R. 1984. “Exolinguistics: the state of the art.” *OMNI* 6.11:93.

Weilgart, J. 1962. *aUI: The Language of Space*. Decorah: Cosmic Communication Co.

Zahavi, D. 2001. “Beyond empathy. Phenomenological approaches to intersubjectivity.” In *Between Ourselves: Second-Person Issues in the Study of Consciousness*, edited by E. Thompson, 151–167. Thorverton: Imprint Academic.

CHAPTER EIGHT

PREPARATIONS FOR DISCUSSING CONSTRUCTIVISM WITH A MARTIAN (THE SECOND COMING)

GÖRAN SONESSON

The use of pictures, such as the Pioneer plaque, in preparing messages for intelligent beings on other planets can easily be seen to overrate the naturality with which pictures present the world of our experience. However, there is a second-order naivety that distinguishes those who pinpoint the conventionality of pictures and is more seldom observed. Indeed, in the psychology of perception, constructivism, which supposes all our perception to rely on social constructs, has long been shown to be untenable – notably by James Gibson, who instead suggested that some general ecological principles must be taken into account. In parallel fashion, I have argued that the critique of iconicity should be supplanted by a semiotic ecology: the general principle accounting for the fact that pictures are perceived as signs by human beings. This means that, if we are to fashion messages understandable to a Martian – or, to put the issue more seriously, to some intelligent extraterrestrial and, most probably, extrasolar being – we do not need to find out in what society he lives: but we have to go far beyond the anthropological universals determining human ecology, which may well be a much more formidable task.

Of apes and men: Recognizing the message as such

Unlike the scholars at the SETI institute, I am not interested in communication with extraterrestrial intelligence (which the researchers at SETI call CETI) in itself. Rather, for me, thinking about communication with extraterrestrials is a *test case* (imaginary so far) for the constraints imposed on semiosis. In fact, traditionally there are two or three ways of investigating the constraints on the *specificity* of the (human) semiotic function: studying child development; scrutinizing the capacities of apes, monkeys, and other animals; and analysing cultures whose members are not familiar with some kinds of semiotic resources ("primitive" cultures) such as, most classically, pictures. Indeed, the title of the present chapter is a paraphrase of an article written many years ago by the primatologist David Premack, called "Preparations for discussing behaviourism with chimpanzee" (Premack, and Schwatz, 1966). While Premack must have supposed behaviourism to work if he ever were to engage in such a discussion, my own contention is that constructivism is not what is needed to have a conversation with an extraterrestrial. Constructivism, in the relevant sense, is a particular conception about how not only pictures, but all phenomena of the common-sense world, are perceived. It is opposed to the *direct registration theory* of James Gibson. "Ecological psychology" is a better term: perception is only direct for those embedded in an environment of shared presuppositions. Cros' and Niemans' codings, which I will discuss below, are clearly very indirect constructs.

At the heart of the distinction between constructivism and other views of perception – as well as the problem of communicating with extraterrestrials – is what Douglas Vakoch (1999) has called "the incommensurability problem", which may be paraphrased as follows: the models constructed by scientists on Earth vary considerably, in part because of their different social and historical backgrounds; so it would be surprising if such a variability were not augmented by the scientists having come from different planets, in which case biology may also be different. This issue not only is relevant to scientific models but applies to the transmission of any kind of messages. Indeed, in my own version of the communication model (Figure 1), which – deriving its inspiration from the Prague school of semiotics – takes into account the active construal of the message on the part of the receiver, the pool of knowledge, including norms, abductions, and sign systems held in common by the protagonists of the communication process is – following the parallel suggestions of Lotman and Moles – supposed to overlap only in part at the beginning of the process (Sonesson, 1999).

If the act of communication may still succeed, this must either be because the sender takes pains to adapt his pool of knowledge to that of the receiver, or because the receiver does so with respect to the knowledge of the sender – or some combination of both approaches. In the first case, we have what the Tartu school calls a *receiver-culture*; it is, as I have formulated elsewhere, a culture in which it is felt to be the task of the sender to recover the norms and interpretations characteristic of the receiver. The classical case is the pedagogical situation. In the case of a *sender-culture* on the other hand, the receiver is assigned the task of recuperating the part of the pool of knowledge peculiar to the sender that does not overlap with his own. High art, as well as mystery cults, are of this kind (Sonesson, 1999). Hermeneutics, as a science with practical goals, was developed for the latter situation. Philosophical hermeneutics, on the other hand, often envisions some kind of combination of the two processes: a "fusion of horizons" in Gadamer's (1960) famous phrase. The incommensurability problem, in its extreme forms, suggests the opposite case: the overlap between the two initial pools of knowledge approaches zero.

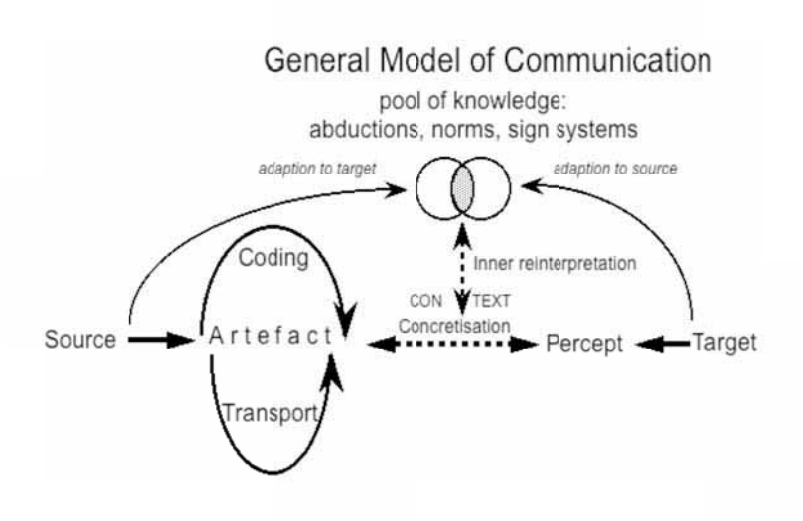


Figure 1. General model of communication.

It is important to recognize that, in a situation of communication, the first problem is not to find out *what* the message means: it is to realise that there *is* a message. That is, it involves the recognition of the message *as such* – *as* a message, rather than a message about something in particular. Even those theories of communication that insist on the act by means of which meaning is produced and conveyed – on the enunciation rather than the utterance – are not very clear about this issue. *Speech act theory* (Austin, Grice, Searle, etc.) separates the content of the message ("locution") from how it is to be taken ("illocution") and even the effect it may have or not have depending on circumstances ("perlocution"); but it is very vague about uptake: the necessity for the message to be attended to as such. In Jakobson's (1960) model of communication, one of the functions

– called the *phatic function* – is supposed to assure that the message gets through; but Jakobson has very little to say about the way this is brought about, apart from giving the commonplace example of checking whether the telephone line is open by saying “hello”. Luis Prieto (1966) has been much more insistent on the difference between the message (“indice”) and the information that somebody is sending a message (“indication notificative”; cf. Sonesson 2012). But somehow the essential question gets lost in the discussion of intentions. After all, even an unintended message has to be recognized as such.

The first incommensurability problem thus concerns the recognition of the message *as* a message. Such recognition requires us to share some common presumptions about the shape of possible messages. This is nicely illustrated by the examples quoted by Vakoch (1999; 2003) of messages that a mathematician in the 1820s suggested could be formed by clearing massive stretches of Siberian forest to produce geometrical figures; and which others hoped to obtain by digging geometrically arranged channels in the Sahara to be filled with kerosene and set aflame during the night. Even if the Martians or inhabitants of the Moon could see these shapes and recognize them for what we think they are, they would only learn anything about us to the extent that they understood that these are messages send by us – and, even more fundamentally, messages, period.

John R. Searle (1969) claims we can only see patterns in the desert sand as writing *if* we suppose somebody intended that we should understand somebody had the intention… etc. But the opposite is of course true: it is only *because* we see something as being (typically) writing that we suppose somebody had the intention… etc. If it is impossible that somebody was around, then, miraculously, God, some ghost, or ET must have been doing the writing (Sonesson, 1978). The astronomer Richard Hoagland says he has discovered in pictures taken of the planet Mars a sculpture of a monkey’s head, together with some other strange constructions, which must be traces of an ancient Martian civilization. For obvious reasons, other astronomers think this is as absurd as affirming that the man in the moon has been painted by intelligent beings (cf. Wikipedia contributors, 2013). However, what Hoagland presents us with is actually an iconical sign of another putatively iconical sign: i.e., a photograph of the monkey’s head. If his claim were borne out by direct observation, then we would have to admit that von Däniken’s space gods – with their superior technological resources – had landed on Mars and edified the monkey’s head, just so as to bewilder us (Däniken, 1973). This is parallel to a case considered by Arnheim: a prototypical picture should possess configurational and other holistic properties not found in ink blots that, in their natural state, are all too irregular and, in their Roschach version, too symmetrical (Sonesson, 1989, pp. 254 f.; 2012; Arnheim, 1966, pp. 93 ff.).

In the end, what is needed are criteria for some shape being a message. One such criterion is no doubt *ruleboundedness*: regularity, repetition, etc.: that is, symbolicity in the Peircean sense. Simple examples are the cleared stretches of Siberian forest producing geometrical figures and the geometrically arranged channels in the Sahara lighted with kerosene. Interestingly, as we shall see, this is also what is found in some more complex proposals for extraterrestial messages – e.g., Cros' and the Niemans' schemes, as well as Drake's later proposal – in the first case, the same number for each line; in the second, "551 = 19 x 29" – though the same clue has to do service a second time as a signifier of "mathematicalness". Another such criterion is similarity: that is, iconicity (but this may lead to projection, as in Hoagland's monkey face and van Däniken's wrist watches and helmets). Indexicality, on the other hand, as found e.g. in traces, could easily suggest no intention to communicate: that is, messages involuntarily produced.

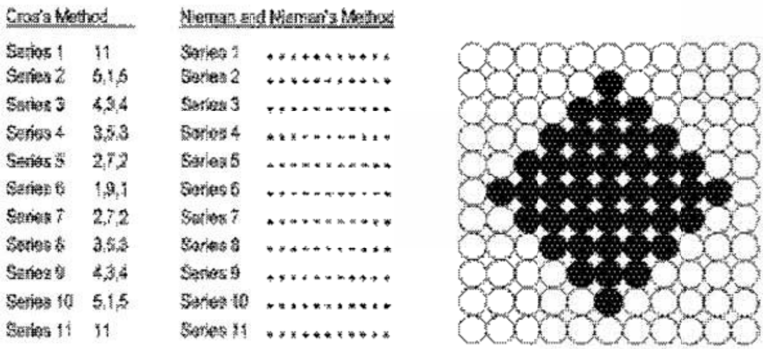
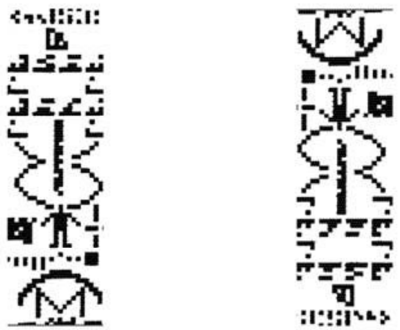


Figure 2. Cros' and the Niemans' method and the resulting picture according to Vakoch 1999 (reproduced with the permission of the author).

Interstellar communication projects into space problems long faced by archaeology in time. Thus, archaeologists are wont to ask: is the Berekhat Ram figure – an object dated to between 250,000–280,000 BP – the likeness of a woman? Before that another question must be posed: do the traces of abrasion left on it show regularity in a fashion suggesting "anthropogenic" movements? Although it has never been claimed to be a picture, Marschack's "calendar" – if it were indeed a calendar: i.e., another kind of artefact with a cultural imprint – would have to evince some kind of regularity in the very way its traces are disposed (Sonesson, 1994). Thus, we recognize the same interplay of iconicity, indexicality, and symbolicity as in interstellar messages. This brings us to the critique of iconicity.

Disquisitions on Bierman's key

Writing the history of interstellar messages, Vakoch (1999; 2001; 2003) tells about some ingenious ways of constructing messages invented by Charles Cros in 1869, by the Niemans in 1920, and by Drake in the 1960s, the ideas of whom are quoted from Vakoch in the following (Figure 2). Cros suggested that several series of numbers should be sent out into space, each one having the same final sum. When the numbers were translated into strings of beads of two different colours and these strings were aligned one over the other, a figure would appear. According to the Niemans' proposal, dots and dashes would be used instead – again corresponding to beads of different colours, with the dots and dashes of each string making up the same sum. Drake's proposal is of the same general kind but more complex: the message sent consists of 551 bits of information, the only factors of which are 19 and 29. When these numbers are taken to be the length and width of the message, the result is a pixelated pattern, which could be interpreted to be a stocky biped placed beside the star and nine planets of our solar system, along with an oxygen and a carbon atom, with their electrons (Figure 3). The result of the reconstruction may be said to be of the same general kind as the better-known Pioneer plaque.



3a 3b

Figure 3. The result of constructing Drake's pictures (3a) according to Vakoch 1999 (reproduced with the permission of the author); and the same picture turned upside down (3b).

Of course, the idea is that, if these extraterrestrial beings are intelligent, they will be familiar with the same mathematics we are, and they will know the same chemistry (and also, as I will insist below, they will represent them the same way). Even granted that, however, these proposals beg the question: why would these hypothetical extraterrestrial scientists believe in the first place that these are messages – which is the primary requisite for setting out to reconstruct them. The only thing that may make such a scenario even remotely plausible is if ordinary perception is already a construction, as the constructionists maintain. But there is no reason to think so.

Although these codings are much more complicated, they remind me of a parable constructed by Arthur Bierman (1963, p. 249) with the purpose of proving the impossibility of iconicity. This story, I submit, is instructive in a different way than Bierman intended. A man receives by mail a parcel, which contains something the man takes to be a blueprint. Using pieces of metal, he sets about constructing a machine according to the blueprint; but when he switches it on, he is electrocuted. The next morning, his widow receives a letter, explaining that the figures marked on the paper must be cut out and put together, to obtain a paper machine. But is the moral of this story really that there are no iconic signs?

I think not. Like all activities taking place in the *Lifeworld*, i.e. the world taken for granted, the interpretation of pictures depends on certain things being taken for granted – but not necessarily on any particular conventions. "Normal" conditions are thought to obtain. When a sign differs from what might be expected, it is indeed necessary to have it "anchored" – to use the classical Barthesian term. When opening the parcel, the man will note a number of things: it contains iconical signs, rather than writing or scribbles, etc.; the particular style of the pictures connotes "blueprint"; the shapes given to the figures suggest that they depict machine parts. These observations determine the use to which the man puts the parcel: since it appears to be a blueprint, he sets about constructing something; since the shapes of the pictures suggest machine parts, and since machine parts are usually made of some sort of metal, he makes his construction out of metal. Apparently, there must also be some sign – probably iconic or indexical – that tells the man how to relate the pieces to each other. But Bierman has been pulling one over on the man. What seems to be a blueprint is really a cut-out sheet; instead of being pictures, the figures are identity signs; and what seem to be their borders are really indexical signs for where one must cut.

Interestingly, while instructions would be needed to discover that the sheet of paper could be seen as an identity sign – and thus a secondary iconical sign, as we shall see below – none was necessary for the man to take it as a picture. If the sheet, considered as an expression, is ambiguous between two readings, then one of them – which happens to be incorrect here – would seem to suggest itself more readily. There is no hint in the story that the man put the pieces together incorrectly: thus, something was apparently read off from the picture iconically (and indexically). In this sense, Bierman's parable presupposes the truth of the very thesis it is supposed to disprove: that similarity, as such, can explain depiction. Not depiction, but the function of depiction, is at issue (Sonesson, 1989, pp. 220 ff.; 1998; 2000; 2001).

In our case, however, incommensurability is much greater. We have no reason to suppose the sender and the receiver of an interstellar message to share such understandings as permit the man in Bierman's story to make an interpretation, even if it happens to be the wrong one. Here it is true, in a much more acute sense, that normal conditions do *not* obtain. In fact, if depiction, on the face of it, stands at the beginning of Bierman's story, it only emerges as a result at the end of the coded messages aimed at extraterrestrials. This is, I think, a decisive difference.

Both Vakoch (1999) and Arbib (1979) locate the problem of the Drake message on the depiction end: the extraterrestrials would not be able to interpret it – they contend – if trying to read it upside down, with the legs of the biped pointing skywards. Arbib even proposes a possible – but obviously erroneous – interpretation of the inverted image. The philosopher Edmund Husserl (1980) long ago encountered the same problem, without having to take the perceptual habits of extraterrestrials into account: he suggested that pictures were essentially non-arbitrary, but that a convention was needed for telling us what was up and down. In a rejoinder to Husserl, I long ago refuted the last part of this affirmation: it is sufficient to turn a picture slowly around; at some point, the configuration giving rise to a depiction will emerge of itself. This is nicely illustrated by the comic strip *The Upside Downs*, created by G. Verbeck in 1903: at the end of each strip, you have to turn the entire strip on its head to follow the rest of the story. Thus, each drawing has a double interpretation, in which what was a hat may, after inversion, appear as a skirt, and so on. When you turn the figure around, not only does a new configuration (*Gestalt*) appear at some given point – but also a new representation. At least, so it is for human beings (Sonesson, 1989, pp. 220 ff.; 1998; 2000; 2001).

If extraterrestrials are like human beings, they will certainly not have more problems finding what is upside down in the picture than perceiving the picture as such. Nothing permits us to conclude, however, that extraterrestrials share the ecological world characteristic of human beings.

But this ignores the primary problem, which is anterior to the depiction: why should the extraterrestrials think there is any message at all?

The construction of the world – and its pictures

As we have seen, the pictures making up the blueprint are really the givens in Bierman's story: it is the machine that is constructed in their image (or, as is happens, out of them). Bierman's formal arguments rather go to prove that pictures as such are constructs of our perception. As is well known, Nelson Goodman (1968) later on gave more famous formulation to those same arguments. The messages conceived by Cros, the Niemans, and Drake are really better illustrations of this constructionist theory of picture perception than Bierman's story. Indeed, if human beings really have to construct each picture before perceiving it, then we may imagine the extraterrestrials doing the same once they have the proper instructions.

There is a parallel between the extant conceptions of picture perception and the psychological theories about our perception of the world. Three schools of perceptual psychology are commonly distinguished. The most venerable is known as constructivism and goes back to Helmholz; but has, in recent times, most famously been represented by Gregory, who claims that impoverished stimuli can only give rise to percepts thanks to inferences or hypotheses. Gregory conceives these to be social constructs. As for the brand-new version of constructivism proposed by Hoffman (2004), it seems to abandon all tenets of the classical tradition and is hard to distinguish from ecological psychology, even though its formulation is very much steeped in constructionist language. It also seems to embrace the nativism otherwise characteristic of *Gestalt* psychology. According to the second school, *Gestalt* psychology – represented by Köhler, Koffka, Arnheim, etc. – innate mechanisms organise perception (based again on impoverished data) into configurations. Ecological psychology, which originates with the work of James Gibson, has been pursued (more or less adjusted) by Reed, Neisser, Kennedy, Hochberg, etc. According to it, the principles of "ecological physics" explain how percepts emerge from stimuli. Thus, it supposes human perception to be a function of the human ecological niche or *Umwelt*: that is – in phenomenological terms – of our *lifeworld* (Sonesson, 1989, pp. 255 ff.).

Psychologists have maintained that all three theories are descriptively inadequate: constructivism because no criteria have been proposed for when a hypothesis is confirmed; *Gestalt* psychology, because its laws are mysterious; and Gibsonianism, because no list of the invariants picked up from the environment can at present be given. Many psychologists – such as Neisser and Hochberg – clearly think some combination of constructivism and ecological psychology would be more to the point. In our terms, ecological psychology may account for the general presuppositions of the human *Lifeworld*, while constructivism takes care of the things taken for granted in the various sociocultural *lifeworlds* (Sonesson, 1993, pp. 352 ff.).

Only ecological psychology seems to have anything relevant to say about pictures. The paradox of perceptual psychology is that ecological psychology is alone in attending to the difference between perceiving the real world and those signs of it called pictures. *Gestalt* psychology and constructivism often use pictorial examples – configurations and illusions, respectively – to illustrate real-world perception. Against this, Gibson (1978) claimed that no conclusion about the real world can be derived from pictorial examples. Although he never says so in as many words, Gibson clearly supposes the picture to be a sign. Indeed, he talks about pictures having referential meaning – as opposed to the "affordances" of perception. All animals can understand the meaning of surfaces. But according to Gibson, only human beings can interpret markings on a surface: that is, have indirect perceptions. If this is taken to mean seeing the markings as markings – and not as the real thing – Gibson may be right. The jury is still out on some primates having this capacity as well (Sonesson, 2009).

To see the picture as a picture clearly requires the capacity to perceive wholes (*Gestalts*) as such; to take contours to be equivalent to the sides of objects; and to accept the 2D surface as surrogate for a 3D world. The picture supposes a *similarity* on the background of a fundamental *difference*. But the problem may well be to see the difference rather than the similarity. Gibson (1978) observes that – besides conveying the invariants for the layout of the pictured surfaces – the picture must contain the invariants of the surface that is doing the picturing: those of the sheet of paper, the canvas, etc., as well as those of the frame, the glass, and so on. The difficulty, clearly, consists in seeing, at the same time, both the surface and the thing depicted (Sonesson, 1989, p. 251). There are indications that neither animals nor small children perceive the difference. Studies of picture perception in animals from doves to apes and in children do not distinguish the perception of the picture as a picture from the identification of another member of the category. But studies of self-recognition in mirrors – if, in spite of Eco (1999) the mirror is seen as a sign – indicate that apes, but not monkeys, can make the distinction.

According to Peirce, the sign – i.e., the expression – is something that "stands for that object not in all respects, but in reference to a sort of idea, which I sometimes called the ground of the representation" (Peirce, CP

2:228). In this sense, the "ground" is the point of view from which the expression represents the content. Elsewhere, it is said that the ground is an "abstraction": for instance "the blackness of two black things" (Peirce, CP 1:293). In my interpretation, the ground is the moment in which iconicity becomes a relation: that is, in Peircean terms, a kind of Secondness. Like indexicality, which is already in itself a relation, it must be combined with the semiotic function to be turned into a sign. But – as we shall see – there are two ways in which this may happen to iconicity. Either similarity is a prerequisite for the sign, or the reverse.

Primary Iconicity in the Human Lifeworld

The relative part played in a sign by iconicity and conventionality may be used to distinguish *primary* from *secondary iconicity*. A primary iconic sign is a sign if the perception of a similarity between an expression *E* and a content *C* is at least a partial reason for *E* being taken as the expression of a sign, the content of which is *C*. That is, iconicity is really the motivation (the ground) – or rather, one of the motivations – for positing the sign function. A secondary iconic sign on the other hand is a sign if our knowledge that *E* is the expression of a sign – the content of which is *C*, in some particular system of interpretation – is at least a partial reason for perceiving the similarity of *E* and *C*. Here, then, it is the sign relation that partially motivates the relationship of iconicity.

That pictures are instances of primary iconicity is shown by the child's capacity for interpreting pictures when first confronted with them at 19 months of age, as demonstrated in Hochberg's famous experiment; as well as by the ease with which pictures are employed by populations whose own culture ignores them – at least, so long as the culture in question is within the bounds of our own Earth. On the other hand, we *do* have to learn that – in certain situations, and according to particular conventions – objects that are normally used for what they are become signs – of themselves, of some of their properties, or of the class of which they form part: a car at a car exhibition, a stone axe in the museum showcase, a tin can in a shop window, an emperor's impersonator when the emperor is away, and a urinal at an art exhibition (if it happens to be Duchamp's "Fountain"). When Man Ray makes a picture of a billiard table, we need no convention to recognise what it depicts. However, if Sherrie Levine's (real, three-dimensional) billiard table is to represent Man Ray's picture, there must be a label inverting the hierarchy of prominence of the *Lifeworld*. This shows that – among the properties determining the probability of an object functioning as the expression of an iconic sign – is to be found three-dimensionality, rather than the opposite.

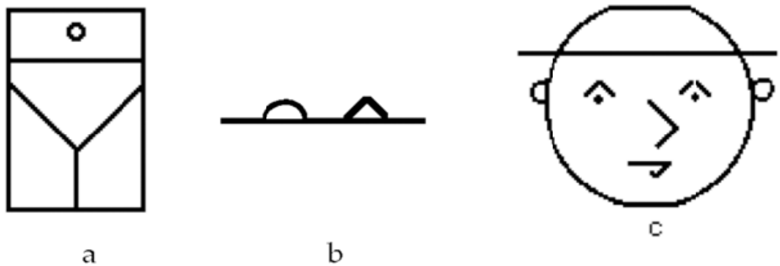


Figure 4. Two droodles and a picture which can be read as a droodle: a) Olive dropping into Martini glass or Close-up of girl in scanty bathing suit (from Arnheim as adapted in Sonesson 1989). b) Carraci's key (Mason behind wall); c) face or jar (inspired by Hermerén, 1983, p. 101).

One kind of picture is really a limiting case: the "droodle", a picture that needs a key such as Carraci's mason behind a wall (Figure 4b). In one droodle, which I borrow from Arnheim (1969, pp. 92 f.), ambiguity is noted immediately in the title: "Olive dropping into martini glass or Close–up of girl in scanty bathing suit" (Figure 4a). While both scenes are possible to discover in the drawing, both are clearly underdetermined by it. There are two ways in which we can try to avoid such ambiguity. One is to fill in the details: in particular, the characteristic differences between an olive and a navel, the air and a pair of thighs, etc. At some point the droodle will then turn into a genuine picture. The other possibility – the only one considered by constructionists – is to introduce an explicit convention such as Carraci's key. According to Hermerén (1983, p. 101) it is only because of "the limitations of human imagination" that we see Figure 4c as a human face; for it can equally be perceived as "a jar from above, with some pebbles and broken matches on the bottom, and a stick placed across the opening". All depends on what is meant by the limits of human imagination: *Gestalt* principles, the face as a privileged perceptual object, etc., all conspire to make one of the readings determinate. While it is possible to find the elements that Hermerén suggests should be in the picture, it is impossible to see the interpretation as a whole without being disturbed by the other reading. Thus, it seems that, when an expression has similarities to different contents or referents, one may be favoured because of properties of the expression itself and is not overwritten by convention. There may be objects like the human face that, because of the particular facts of human embodiment, are more easily recognizable than others.

At Home in the Terrestrial Lifeworld

If our capacity to experience pictures directly – as opposed to secondary iconical signs – depends on the particular *Lifeworld* we are inhabiting: that is, on the ecology typical of human beings as it has evolved on Earth; then there is every reason to suspect that extraterrestrial beings, however intelligent, would not share this capacity with us. What are for us primary iconical signs would be secondary to them. While we function according to ecological psychology, they would have to follow the precepts of constructivism. No doubt there would be other phenomena that would be primary iconicities to them, but that we can only hope to interpret, if ever, according to the regime of secondary iconicity.

In the case of the biped in Drake's picture – once it is reconstructed as a picture – or the more explicit man and woman of the Pioneer plaque, the problem is not so much that the characteristic body shape of human beings must be recognized. Even in a normal picture, we can only recognize objects of the world with which we are already familiar – at least with their general type. Thus, if the extraterrestrials have different body shapes from ours and have never seen human beings, they obviously cannot recognize the human shape. The more general issue involves the possible embodiement of signs themselves. As I noted above, the faculty to interpret pictures at least presupposes the ability to perceive wholes as such, to take contours to be equivalent to the sides of objects, and to accept 2D forms as stand-ins for 3D objects. There is no particular reason to suppose this forms part of the ecology of extraterrestrial beings.

In a more general sense, these observations are also valid for markings, on a surface, that are *not* pictures. If our ability to interpret pictures is part of our competence as inhabitants of the human *Lifeworld*, then all other sign systems may well be dependent on the same particular ecological niche. Suppose that those people are right who think that our conception of mathematics, as well as our contemporary theories of physics, astronomy, and chemistry, must be known to extraterrestrial beings – either because they accept the same theories, or have entertained them at some earlier stage of their development (as we would recognize Newtonian physics with other intelligent beings). This would only be relevant to the content side of the sign. Even in the case of the natural sciences, the expression side of the signs is wholly within the limits of our human *Lifeworld*. Suppose that the extraterrestrials are well aware of hydrogen transitions, pulsars, and the layout of our solar system. It is still highly improbable for them to use the same surface markings to convey them to others as we would. Their *Lifeworld* would, most certainly, predispose them differently.

It is still possible that iconicity – in a wider sense than pictorality – may be of some help. Peirce pointed out that iconic signs convey more information than is contained in them; thus, "with two photographs you can make a map". This property – which Greenlee called "exhibitive import" (Greenlee, 1973) – depends on our knowledge of the *Lifeworld*. Because of our familiarity with the layout of the *Lifeworld*, we are able to fill in the blanks in the representation. We can "see in" what we know should be there. If the extraterrestrials live in a different *Lifeworld* – as they most certainly do – they should be unable to derive any help from exhibitive import (cf. Sonesson, 1989, pp. 302 ff.).

Perhaps there is another type of iconic surplus: something we might call "introversive semiosis", echoing a term used by Jakobson for signs referring to other signs rather than to the world. Peirce's favoured examples of iconicity were mathematical expressions. Jakobson discovered an iconicity in grammar. Such projections of the selection axis onto the axes of combinations – using Jakobson's phrase – is reminiscent of those messages with a regular structure that Arbib (1979) suggests should be used in communicating with extraterrestrials. More importantly, perhaps, what is needed are expressions that mirror the system character of the system. This might be feasible if there is what Deacon (2003) has called *semiotic constraints*: generalities of all conceivable semiotic systems. Of course, in keeping with earlier philosophers such as Husserl, Deacon is generalizing from the case of logic and mathematics to the less tightly organized system of verbal language (Sonesson, 2006; 2007). Are those generalizations really justified? Even if logic and mathematics are universal, do they not need to take on a special embodiment to be conveyed to us – one that is different from what any extraterrestrial may need?

In conclusion, we have seen that, although pictures are based on primary iconicity, this iconicity is relative to the peculiar human *Lifeworld*, which is most probably not shared by intelligent extraterrestrial beings; and that, although some more abstract kinds of iconicity may stand a greater chance of giving rise to universal messages, the very act of conveying them may require a form which is peculiar to the human *Lifeworld*.

Acknowledgments

An earlier draft of this paper was presented in the workshop headed by Douglas Vakoch of the SETI institute, organized during the conference of the International Association for Semiotic Studies in Lyon 2004. The present version was part of the symposium organised by David Dunér at the Pufendorf Institute for Advanced Studies at Lund University, on May 5, 2011. The author wishes to thank the participants in these events for stimulating discussion.

References

Arbib, M. 1979. "Minds and millennia: The psychology of interstellar communication." *Cosmic Search* 1:3. http://web.archive.org/web/radobs/vol1no3/minds.htm

Arnheim, R. 1966. *Towards a Psychology of Art.* London: Faber & Faber.

—. 1969. *Visual Thinking*. Berkeley CA: University of California Press.

Däniken, E. von. 1973. *Meine Welt in Bildern*. Düsseldort & Wien: Econ Verlag.

Deacon, T. 2003. "Universal grammar and semiotic constraints." In *Language Evolution*, edited by M. Christiansen, and S. Kirby, 111–139. Oxford: Oxford University Press.

Eco, U. 1999. *Kant and the Platypus: Essays on Language and Cognition.* London: Secker & Warburg.

Gadamer, H.-G. 1960. *Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*. Tübingen: Mohr.

Gibson, J., 1978. "The ecological approach to visual perception of pictures." *Leonardo* 4:2:227–235.

Goodman, N. 1968. *Languages of Art*. Oxford: Oxford University Press.

Greenlee, D. 1973. *Peirce's Concept of Sign*. The Hague: Mouton.

Hermerén, G. 1983. *Aspects of Aesthetics.* Lund: Gleerups.

Hoffman, D. 2004. "Images as interstellar messages." In *Between Worlds: The Art and Science of Interstellar Message Composition*, edited by D. Vakoch. Cambridge MA: MIT Press.

Husserl, E. 1980. *Phantasie, Bildbewusstsein, Erinnerung.* Husserliana XXIII. The Hague: Nijhoff.

Jakobson, R. 1960. "Linguistics and poetics." In *Style in Language*, edited by T. Sebeok, 350–377. Cambridge MA: MIT Press.

Peirce, C. 1931–1935. *Collected Papers*. Cambridge MA: Harvard University Press (quoted as CP).

Prieto, L. J. 1966. *Messages et signaux*. Paris: Presses universitaires de France.

Premack, D., and A. Schwatz. 1966. "Preparations for discussing behaviorism with chimpanzee." In *The Genesis of Language: A Psycholinguistic Approach. Proceedings*, edited by F. Smith, and G. A. Miller, 295–335. Cambridge MA: MIT Press.

Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sonesson, G. 1978. *Tecken och handling*. Lund: Doxa.

—. 1989. *Pictorial Concepts*. Lund: Aris/Lund University Press.

—. 1993. "Pictorial semiotics, Gestalt psychology, and the ecology of perception." *Semiotica* 99(3/4):319–399.

—. 1994. "Prolegomena to the semiotic analysis of prehistoric visual displays." *Semiotica* 100(3):267–332.

—. 1998. "That there are many kinds of pictorial signs." *Visio* 3(1):33–54.

—. 1999. "The life of signs in society – and out of it." In *Trudy po znakyvym sistemam – Sign Systems Studies* 26, edited by P. Torop, M. Lotman, K. Kull, and Ü. Pätli, 88–127. Tartu: Tartu University Press.

—. 2000. "Iconicity in the ecology of semiosis." In *Iconicity – A Fundamental Problem in Semiotics*, edited by T. Johanson, M. Skov, and B. Brogaard, 59–80. Aarhus: NSU Press.

—. 2001. "From Semiosis to Ecology." *Visio* 6(2–3):85–110: Cultural cognition and space cognition/Cognition culturelle et cognition spatiale, edited by A. Quinn.

—. 2006. "The meaning of meaning in biology and cognitive science: A semiotic reconstruction." In *Trudy po znakyvym sistemam – Sign Systems Studies* 34(1):135–214.

—. 2007. "From the meaning of embodiment to the embodiment of meaning." In *Body, Language and Mind. Vol. 1: Embodiment*, edited by T. Zimke, J. Zlatev, and R. Frank, 85–128. Berlin: Mouton.

—. 2009. "New considerations on the proper study of man – and, marginally, some other animals." *Cognitive Semiotics* 4:133–168, Spring 2009: Anthroposemiotics vs. Biosemiotics, edited by G. Sonesson, and J. Zlatev.

—. 2012. "The foundation of cognitive semiotics in the phenomenology of signs and meanings." *Intellectica* 58(2):207–239.

Vakoch, D. 1999. "The view from a distant star." *Mercury*, March/April, 26–39. http://www.astrosociety.org/pubs/mercury/9902/vakoch.html

—. 2001. "Messages from ET." http://www.space.com/searchforlife/vakoch_messages_010706.html

—. 2003. "Messages from the moon, to Mars, to the stars." http://seti.org/seti/our_projects/interstellar_messages/overview/history/messages.html

Wikipedia contributors, 2013. "Richard C. Hoagland." *Wikipedia: The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Richard_C._Hoagland&oldid=540691754 (accessed February 26, 2013).

CHAPTER NINE

THE EVOLUTION OF LINCOS: A LANGUAGE FOR COSMIC INTERPRETATION

MARIA G. FIRNEIS AND JOHANNES J. LEITNER

Introduction

In the 1820s, Carl Gauss in Germany and Joseph Johann Littrow in Austria proposed ideas on how to communicate with non-terrestrial species. Unfortunately, they never published their proposals; they only discussed them in public talks. Gauss is said to have proposed a huge Pythagorean rectangular triangle of sides three and four units in length (the units were undefined; one unit length could be about 100 miles) and a hypotenuse of five units, with proportional squares arising from the three sides, to be constructed in the plains of Siberia. The intention was that it should be visible to potential observers on the Moon (Raulin-Cerceau, 2010, see Figure 1).

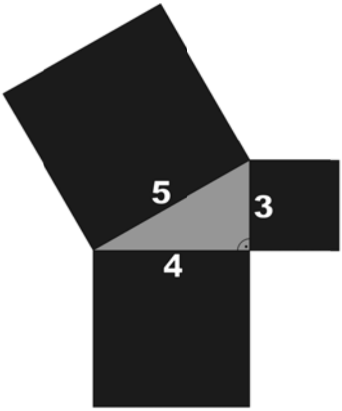


Figure 1. The huge Pythagorean rectilinear triangle proposed by Gauss to be constructed in the plains of Siberia.

Littrow proposed a similar triangle, to be sited in the Sahara, constructed as gigantic channels filled with water and topped with kerosene. The kerosene would be lit at night as a fiery signal for observers on the Moon. Sometimes his idea is presented differently: the channels as set of huge, 30 km-diameter concentric rings. Again, his intention was something discernible as an artificial structure hinting of our existence to any would-be observers. Raulin-Cerceau (2010) gives an outstanding summary of other early ideas, centering on French proposals; Dick (1996) offers a likewise outstanding overview of Twentieth Century suggestions.

We have the English biologist Lancelot Hogben to thank for offering a starting point for a language to ease communications. In 1954, he created Astraglossa as a means of communication via short and long radio pulses. These pulses have to be applied when simpler means such as direct transmissions of pictures are not available. In case a "Rosetta stone" offering identical texts in languages such as English and Russian also is not available, pulses are a reasonable tool. He based his proposal on earlier radioglyphic schemes. A radioglyph is a picture intended for interstellar communication sent via radio waves. Morrison (1962) elaborated on Astraglossa. Short pulses represented numbers, long pulses represented such operations as addition or subtraction (Hogben, 1963). The following example from Hogben's book (p. 127) illustrates the basic concept:

1 + 2 + 3 = 6
(i) 1 . . F_a . .1 . 1 . .F_a . . 1 . 1 . 1 . . F_b . . 1 . 1 . 1 . 1 . 1 . 1
(ii) 1 . . F_a . .1 . 1 . .F_a . . 1 . 1 . 1 . . F_b . . F_s . . F_b . . 1 . 1 . 1 . 1 . 1 . 1

F_a signifies addition and F_b identity. F_s introduces the symbol for six.

LINCOS: Background and concept

LINCOS was created as a playground for these nascent ideas. As an abbreviation of "*lingua cosmica*", it is a so-called constructed or artificial language. *Contra* naturally evolved languages, such languages are devised efficiently for grammar, vocabulary, and individual phonology. Sometimes they are created to facilitate human communication (e.g. Esperanto), sometimes to bring fiction to life (Klingon in Star Trek). Artifical languages have given us the science of *astrolinguistics* (Ollongren, 2008).

Hans Freudenthal conceived LINCOS as a means of communication with intelligent aliens. Born 17 September 1905 in Luckenwalde, Germany, died 13 October 1990 in Utrecht, the Netherlands, Freudenthal was for many years an outstanding mathematician at the University of

Utrecht. A specialist in topology and Lie groups, he intended to create a self-explanatory language that could be taught using a large number of examples formulated mathematically. It assumed a common knowledge between human and alien, which for him meant mathematics – if only because the aliens would need to think in ways somehow similar to our own if they were to understand our communicative intentions. He further assumed they would have access to powerful means of communication at long wavelengths, which are as free from cosmic disturbance as is technically possible. One strategy could focus around decimetric radio frequencies, the other around optical frequencies observable by large optical telescopes as used by "Optical SETI" (Ross and Kingsley, 2011).

Freudenthal constructed LINCOS as a sort of a *Gedankenexperiment*, with a 200-or-so-page manual (*LINCOS: Design of a Language for Cosmic Intercourse*, 1960) offering a step-by-step description of the language.

Description of Classic LINCOS

The original text is divided into such topics as mathematics, time, behaviour and space, and motion and mass. Freudenthal planned a second volume addressing social aspects of humanity but never got it written. Evidently, Freudenthal lost interest in the subject due to mounting conceptual difficulties when applying LINCOS for other than mathematical contents due to the potential different sociological aspects of alien receivers (see also Dick, 1996).

Table 1 displays a sample of the mathematical code. The signals are coded *X* and *O* for easy legibility; *ver* (Latin *verum*) means "true" and *fal* (Latin *falsum*) means "false".

Integers, rational numbers, real numbers, complex numbers, and prime numbers (Int, Rat, Rea, Com, and Pri) are introduced one after another. Certain essential constants (e, i, π) are also defined. Combination of terms allows for remarkable complexity: LINCOS is even able to express Einstein's theory of special relativity (see Figure 2).

Bassi (1992) offers a concise review of Freudenthal's book. Bassi suggests that LINCOS be broadcast via radio signals of varying duration and wavelength, a possibility not mentioned anywhere by Freudenthal. Freudenthal's focus lay with claiming that it should be optimally systematic in the sense that syntactic and semantic categories should be marked phonetically (see Bassi, 1992).

Fuchs (1973) used Morse code with signals of short (piep) and long (pieep) duration; an interrupt indicated the "equals sign". Binary notation can be used instead of decimal notation. For the decimal code an example is:

3 = 3 results in
piep - piep - piep -- pieep -- piep - piep – piep

In similar fashion, binary notation can be used to show the same example:

0101010000111000010101
with 000 as “break” and 0111 as “=”.

For chatting with extraterrestrials, Freudenthal (1960) suggests the following dialogue (Table 2), where “ben” and “mal” indicating approval and disapproval respectively, and “tan” (Latin *tan=tamen,* “nevertheless”) for “but”, are added to the earlier “ver” and “fal”.

Mathematical Code	**LINCOS Code**
1 = 1 true	X O X ϵ ver
1 < 2 true	X OO XX ϵ ver
2 = 3 false	XX O XXX ϵ fal
2 > 1 true	XX OOO X ϵ ver
(1 < 2) and (3 > 1) true	X OO XX A XXX OOO X ϵ ver
(1 = 2) and (1 = 1) false	X O XX A X O X ϵ fal
(1 = 2) or (1 < 2) true	X O XX AA X OO XX ϵ ver
(1 + 1 = 2) true	X P X O XX ϵ ver
(1 + 2 = 4) or (1 > 2) false	X P XX O XXXX AA X OOO XX ϵ fal

Table 1. Mathematical code versus LINCOS code.

CH. IV] SPACE, MOTION, MASS 205

Erg ⁝ Cca t_1 . X Vul · $A \in$ Sca Cmt 1 .
→ ˙ Ncs : Cca t_1 . X Ced ·
Nnc Y Vul . $A \in$ Sca Cmt $\sqrt{1 - v^{10}/c^{10}}$, :
Qia , · Pon , $w \in$ Und Ath .
∧ ⁝ ∨ ⌜t . t'⌝ ˙ $t \frown t'$. = Cca . t_1 · ∧ : t Usq t' ·
Apu Y . Usd . Apu z_N · w Mov ⁝
∧ ⁝ ∧n ⁝ $n+1$. ∈ Num · ∧ . $n \leq N$. → ⁝ $w_n \in$ Und Ath .
∧ ⁝ ∨ ⌜t . t'⌝ ˙ $t \frown t'$ = Cca t_1 · ∧ : t Usq t' ·
Apu z_n . Usd . Apu Y · w_n Mov ⁝
∧ : ∧t · tw Apu z_n . ↔ . tw_n Apu z_n :
∧ ⁝ ∧n ˙ z_n Par A . → ⁝ u_n = . w_n Apu z_0 · ∧ · u'_n = . w_n Apu Y ⁝
∧ : U Uni . u_1 . u_2 . Etc : ∧ : U' Uni · u_1' . u_2' . Etc , .
Erg ⁝ Ncs · Cca t_1 . X Ced . $U \in$ Sca Cmt $10\,c^{-1}$:
Erg : Ncs · Cca t_1 . Y Ced . $U' \in$ Sca Cmt $10\,c^{-1}\sqrt{(c+v)/(c-v)}$ ˙
∧ ⁝ Cca t_1 . Y Inq Y : $\beta \in$ Pos . ∧ . $A \in$ Sca Cmt β .
∧ · w Apu z_n . t_n Pst . w Apu Y · ∧ · [1]
w_n Apu Y . t'_n Pst . w Apu Y ˙
→ ⁝ Ncs ˙ Cca t_1 . Y Ced : $t'_n = 10\,t_n$.
∧ . Cmt $\beta = (t_{n+1} - t_n)(c - v)$. ∧ . Cmt $10\beta = (t'_{n+1} - t'_n)(c - v)$ ·
Erg . Cmt $10\beta = 10\,c^{-1}\sqrt{(c+v)/(c-v)}\,(c - v)$ · Erg . β
$= \sqrt{1 - v^{10}/c^{10}}$ ∗

A short proof of Lorentz contraction.

4 28 6. ∗ Hb Inq Ha !
Pon , . $X \frown Y \frown Z \in$ Hom . ∧ · δ . Mul > . Sec 1 ·
∧ , ∨ ⌜a . b . c⌝ ⁝ ∧t ⁝ $t_1 - \delta \leq t \leq t_1 + \delta$. → ⁝ Cel X . t · = a :
∧ : Cel Y . t · = b : ∧ : Cel Z . t · = c : ∧ : Cel Rel $X\,Y$. t · = x :
∧ : Cel Rel $Y\,Z$. t · = y : ∧ : Cel $X\,Z$. t · = z .
∧ ˙ t Loc Y . ⊂ : Cvx · t Loc X . ∪ . t Loc Z , ·
Erg . $z = (x + y)/(1 + xy/c^{10})$, :
Qia , · Pon , $W \subset$ Und Ath . ∧ . $\alpha \frown \beta \frown \gamma \in$ Pos .
∧ . $U \in$ Sca Sec α . ∧ . $U' \in$ Sca Sec β . ∧ . $U'' \in$ Sca Sec γ .
∧ ⁝ ∧ ⌜u . u' . u''⌝ ⁝ u Par U . ∧ . u' Par U' . ∧ . u'' Par U'' ·
↔ ⁝ ∨ ⌜w . t . t''⌝ ˙ $w \in W$. ∧ : t Usq t'' · Apu X . Usd .
Apu Z · w Mov :
∧ · u = . w Apu X · ∧ · u' = . w Apu Y · ∧ · u'' = w Apu Z , .
Erg ⁝ Cca t_1 . X Vul · $U \in$ Sca Sec 1 :
→ ⁝ Ncs ˙ Cca t_1 . X Ced : Nnc Y Vul . $U' \in$ Sca Sec $\sqrt{(c+x)/(c-x)}$ ·
∧ · Nnc Z Vul . $U'' \in$ Sca Sec $\sqrt{(c+z)/(c-z)}$ ⁝
∧ ⁝ Cca t_1 . Y Vul · $U' \in$ Sca Sec 1 :

[1] Inadvertently I introduced a t_n, which for $n = 1$ does not mean the explicit t_1.

Figure 2. An extract from Freudenthal's book illustrating the formulation of Einstein's special relativity in LINCOS code.

Dialogue	**LINCOS**
H_a asks H_b: what X satisfies 2x = 5?	H_a Inq H_b · ?x 2x = 5
H_b asks H_a: 5/2	H_b Inq H_a · 5/2
H_a says to H_b: good	H_a Inq H_b · ben
H_a says to H_b: what x satisfies 4x = 10?	H_a Inq H_b · ?x 4x = 10
H_b says to H_a: 10/4	H_b Inq H_a · 10/4
H_a says to H_b: bad	H_a Inq H_b · mal
H_b says to H_a: 4 × 10/4 = 10 true	H_b Inq H_a · 4 × 10/4 ∈ ver
H_a says to H_b: true, but bad: x = 5/2	H_a Inq H_b · ver tan mal x = 5/2

Table 2. Examples for an intergalactic dialogue based on Freudenthal.

Bassi (1992) readily admits that LINCOS is not a perfect creation: e.g., no syntactic and semantic classification has been developed, while questions starting with "why" or "whether" are "indistinguishable from affirmative statements" (Bassi, 1992, p. 7). Two different words are used for "but". In (1974), Freudenthal acknowledged that the optimal level of formalization in the language had not been achieved.

Cosmic Call

In 1999 and again in 2003, LINCOS was used as the basis for an enterprise called *Cosmic Call*. Dumas (2007; Dutil and Dumas, 2001) describes how a message consisting of a scientific part plus a part created by the interested public was conceived. The public message contained music as well as fifty personal messages.

Using mathematics, a frame of pixels was set around images to overcome the difficulties confronted in the Arecibo/Drake message of 1974 (NAIC staff, 1975) – where, due to interstellar noise, a one-bit error was sufficient to collapse the whole information structure. The 70-meter dish of the Evapatoria, Ukraine, deep-space radar telescope was used for the 6 July 2003 transmission, with a 150 kW transmitter on 6 cm wavelength. Financing was provided by the US (Texas)-based company Team Encounter. Unlike the 1999 transmission, which was divided into 23 frames, the 2003 transmission consisted of one single frame transmitted over and over again (Braastad and Zaitsev, 2003).

Symbols for the numbers 1 through 9 were created using a 5×7 bitmap. Then, as in the original LINCOS, symbols were created for mathematical operations, along with pi and delta and various physical units: kilogram, meter, Kelvin, year, etc. Names were given to important chemical elements. Physical concepts such as proton, Planck constant, cosmological constant, etc.; biological concepts such as adenosine, cell, male, people, etc.; astronomical names for the planets, Sun, and universe; and other concepts like ocean, sky, target, age, etc. followed.

The stage set, a message was assembled including the contents of the earlier Arecibo message, information on the Solar System, a description of DNA, and a Fuller-map representation of Earth's continents. Five stars were targeted (see Table 3), chosen for their position near the galactic plane (Braastad and Zaitsev, 2003) in a region where interstellar scintillation is minimal: between 60° and 90° longitude of galactic centre.

Star (Constellation)	VMag	SpType	RA J2000	DE J2000
HIP 26335 (in Orion)	8.78	K7	05 36 30. 99	+ 11 19 40.3
HIP 43587 (55 Cnc) (in Cancer)	5.96	G8V	08 52 35.81	+ 28 19 51.0
HIP 4872 (in Cassiopeia)	9.56	K5V	01 02 38.87	+62 20 42.20
HIP 53721 (47 UMa) (in Ursa Major)	5.03	G0V	10 59 27.974	+ 40 25 49.92
HIP 7918 (in Andromeda)	4.96	G2V	01 41 47.143	+ 42 36 48.13

Table 3. Target stars from the 2003 Cosmic Call transmission (Braastad and Zaitsev, 2003). Data extracted from http://vizier.u-strasbg.fr.

A New Lingua Cosmica

Beginning in 2000, Alexander Ollongren (2005) started developing a new proposal based on formal lambda calculus and the calculus of constructions, including inductive definitions. The lambda calculus was originally developed by Whitehead and Russell (1910; 1912; 1913) and popularized by Kleene (1935) and Church (1936). It is a formal language for investigating functions or maps. It describes the definitions of functions and formal parameters as well as the input of actual parameters. It provides a successful characterization of computability (Barendregt, 1984). The theory behind it is analogous to that of number theory.

Kleene (1936) proved that lambda computability is equivalent to Gödel-Herbrand recursiveness; while Turing (1937) showed that Turing-machine computability is equivalent to lambda definability. In the 1970s it was shown that logical proofs could be formulated as lambda representations (Barendregt, 1984). The lambda calculus laid the foundations for programming languages like LISP by providing their operational semantics.

In 2010, Ollongren expanded the LINCOS system to include logic terms he called entities, which, by their concrete declaration, may be variables, hypotheses, or definitions, including inductive definitions. Global variables and constants are introduced via the definition of maps expressed by typed λ abstraction (Barendregt, 1992). Logical properties are expressed via the binary connectives $\wedge$ (“and”) and $\vee$ (“or”) and the unary operator ¬ (“negation”).

Through constructive logic, the members of sets with special properties are expressed as unparameterized maps, allowing a general λ abstraction to verify a test hypothesis: i..e., λ abstractions in any given LINCOS text are representations of reality modes. Nesting of λ abstractions may occur. The scope of local variables is restricted, so that they may occur elsewhere in the text. Global variables form a superior body of λ abstractions and therefore cannot be redefined. One further token, the semicolon, serves as a mapping operator.

The Aristotelian syllogism – called Barbara by medieval scholars – will serve as illustration. It may be read: <u>*all*</u> *astronomers are human, and* <u>*all*</u> *stargazers are astronomers; therefore,* <u>*all*</u> *stargazers are human* (Ollongren, 2005). In LINCOS it looks like this:

FACT Barbara
(All x : D)
((astronomers : x) → (human : x)) □ ((star gazers : x) → (astronomers : x))
→ (star gazers : x) → (human : x)

Base *D* represents the general aspect of the problem whose properties are "astronomers", "human", and "stargazers".

As Ollongren (2010) allows, questions remain whether an alien culture will be able to interpret all the symbols contained in an interstellar message, "where the tokens serve as links between abstract and concrete signatures" (Ollongren, 2010, p. 1442). In a later paper, Ollongren (2011) elaborates a logical arbiter function that, via parallel processes, interprets a dynamic process in a linguistic system. As illustration, he translates the opening scene of Hamlet, where the ghost appears, into *Lingua Cosmica*. No reference to time sequences is necessary; proper state vectors model the necessary interruptions using eight parallel threads.

Turning things around, Elliot (2010) offers a scenario of what might happen if an interstellar message would be received on Earth, based on recognizable mathematical/logical structures. Using Zipf's Law, phrase-like chunks hinting at words or word sequences – where the frequency of a word is inversely proportional to its rank in a frequency table – would allow cryptoanalysis to reveal the underlying language-like structure in the same way that the linguistic signals of human languages reveal lexical and grammatical patterns: so e.g. correlation profiles between pairs of words have been investigated for various human languages. Such an analysis allows the possibility clearly to distinguish language from interstellar noise: e.g., a pulsar or the random noise generated by interstellar hydrogen.

Maccone's (2010) statistical approach to the Drake equation suggests that the probability of finding intelligent extraterrestrials within a one-sigma distance should be within 1361 to 3979 ly. If so, this large distance would prohibit any communication.

Conclusion

We have provided an overview of the various attempts to convey at least the symbolic fundamentals of communication to an unknown alien intelligence. The updated version of LINCOS – NEW LINCOS – is meant to support auto-interpretation. Auto-interpretation in this sense considers "the case of elementary inductive definitions of terms, and the use of elimination to deduce facts from premises" (Ollongren, 2005, p. 1). NEW LINCOS has room for further improvement. Human beings are on the verge of showing off their existence to a probable interstellar community.

Acknowledgments

The authors acknowledge the financial support of the Research Platform on ExoLife at the University of Vienna (FPF-234) and of their partner Key Consult.

References

Barendregt, H. 1984. *The Lambda Calculus: Its Syntax and Semantics*. Amsterdam: North Holland Publ.

—. 1992. "Lambda Calculi with Types." In *Handbook of Logic in Computer Science* 2, edited by S. Abramsky, D. M. Gabbay, T. S. E. Maibaum, 117–309. Oxford: Oxford University Press.

Bassi, B. 1992. "Were it Perfect, Would it Work Better? Survey of a Language for Cosmic Intercourse." http://www.brunobassi.it/scritti/lincos.html.

Braastad, R., and A. Zaitsev. 2003. "Synthesis and Transmission of Cosmic Call 2003 Interstellar Radio Message." http://www.cplire.ru/html/ra&sr/irm/CosmicCall-2003/index.html.

Church, A. 1936. "An unsolvable problem of elementary number theory." *Am J Math* 58:345–363.

Dick, S. J. 1996. *The Biological Universe: The Twentieth-Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge: Cambridge University Press.

Dumas, S. 2007. "The 1999 and 2003 messages explained."

http://www3.sympatico.ca/stephane_dumas/CETI/messages.pdf.
Dutil, Y., and S. Dumas. 2001. "Error Correction Scheme in Active SETI." http://www3.sympatico.ca/stephane_dumas/CETI/dutil_dumas.pdf.
Elliot, J. R. 2010. "Detecting the Signature of Intelligent Life." *Acta Astronaut* 67:1419–1426.
Freudenthal, H. 1960. *LINCOS: Design of a Language for Cosmic Intercourse*. Amsterdam: North-Holland Publ.
—. 1974. "Cosmic Language." *Current Trends in Linguistics* 12:1019–1042.
Fuchs, W. 1973. *Leben unter fernen Sonnen*. München: Droemer Knaur.
Hogben, L. 1963. *Science in Authority*. New York: Norton.
Kleene, S. 1935. "A theory of positive integers in formal logic." *Am J Math* 57:153–173, 219–244.
—. 1936. "λ-definability and recursiveness." *Duke Math J* 2:340–353.
Morrison, P. 1962. "Interstellar Communication." *Bull Philo Soc Wash* 16:78.
Maccone, S. 2010. "The Statistical Drake Equation." *Acta Astronaut* 67:1366–1383.
NAIC staff. 1975. "The Arecibo Message of November, 1974." *Icarus* 26:462–466.
Ollongren, A. 2005. "On Auto-Interpretation in LINCOS." http://www.liacs.nl/~ollongre/auto-Interpretation%20LINCOS.html.
—. 2008. "Astrolinguistics: A Guide for Calling E.T." http://www.alexanderollongren.nl/ASTROLINGUISTICS,%20A%20GUIDE%20FOR%20CALLING%20E%20T.htm.
—. 2010. "On the signature of LINCOS." *Acta Astronaut* 67:1440–1442.
—. 2011. "Processes in Lingua Cosmica." *Acta Astronaut*, doi:10.1016/j.actaastro.2011.09.017.
Raulin-Cerceau, F. 2010. "The Pioneers of Interplanetary Communication: From Gauss to Tesla." *Acta Astronaut* 67:1391–1398.
Ross, M., and S. Stuart. 2011. *Optical SETI: Moving Toward the Light*. Berlin Heidelberg: Springer.
Turing, A. M. 1937. "Computability of λ-definability." *J Symbolic Logic* 2:153–163.
Whitehead, A., and B. Russell. 1910. *Principia Mathematica*, Vol. 1, Cambridge: Cambridge University Press.
Whitehead, A., and B. Russell. 1912. *Principia Mathematica*, Vol. 2, Cambridge: Cambridge University Press.
Whitehead, A., and B. Russell. 1913. *Principia Mathematica*, Vol. 3, Cambridge: Cambridge University Press.

PART III:

CULTURE

CHAPTER TEN

THE DEBATE OVER PANSPERMIA: THE CASE OF THE FRENCH BOTANISTS AND PLANT PHYSIOLOGISTS AT THE BEGINNING OF THE TWENTIETH CENTURY

STÉPHANE TIRARD

Introduction

One of the first presentations of panspermia theory occurred in Edinburgh in 1871 in William Thomson's presidential address to the Congress of the British Association for the Advancement of Science. In strong opposition to Darwin's theory, Thomson claimed that life began on Earth because of germs coming from space via meteorites (Thomson, 1872). With this, Thomson opened a new line of thought about the origins of life on Earth. In 1884, Hermann von Helmholtz joined him. A few years later, the Swedish scientist Svante Arrhenius presented his panspermia theory in his famous book *Worlds in the Making: The Evolution of the Universe* (1903). He claimed that every new planet in the universe receives some microscopic germs, which travel across space because of radiation pressure (Dick, 1996; Fry, 2000; Tirard, 2006).

This paper studies the complex context of the origins of life and of French botany and plant physiology, focusing on such influential scientists of the late Nineteenth and early Twentieth Century as Philippe Van Thieghem, Gaston Bonnier, Julien Constantin, and Paul Becquerel. Its goal is to analyse the various conceptions of the origins of life within this small community of close colleagues. I present their positions on the origins of life and on panspermia in particular as progressively enlightened by their theoretical positions on the evolution of living beings – they were neo-Lamarckians or neo-Darwinians – and their persistence under extreme conditions.

The French botanists and plant physiologists in relation to panspermia

None of the botanists or plant physiologists I wish to discuss were specialists on the origins of life. However, they were very interested in the topic, and they exposed their views clearly. First among these is Philippe Van Tieghem (1839–1914), a professor at the *Muséum d'histoire naturelle* in Paris, where he occupied the chair of *Botanique, Organographie et Physiologie végétale* from 1879 to 1914. In his *Traité de Botanique* (1891), writing about the origins of life on Earth, he claimed:

> The vegetation of earth is only a very little part of the vegetation of the Universe… The vegetation of earth has had a beginning and it will have an end, but the vegetation of the Universe is eternal as the Universe itself.

One notes in this a characteristic aspect of the panspermia hypothesis: Van Tieghem does not explain the beginning of life or the Universe; instead, he prefers to refer to eternity.

Among the defenders of panspermia was the botanist Gaston Bonnier, who was very close to Van Tieghem, his thesis advisor. Bonnier was a very important French botanist: the creator and director of the *Revue générale de Botanique.* He is well known for the numerous editions of his *Flore de France* and his experiments on plant habitat (Tirard, 2003).

In his book, *Le Monde végétal* (1907), Bonnier devoted one chapter to the appearance of life: particularly of life on Earth; and to the notion of spontaneous generation. Upon giving a historical review of that notion from the work of John Tuberville Needham to the debate between Louis Pasteur and Félix Pouchet, he insisted on the need to give up the idea. At the same time, he contested the possibility of creating life in the laboratory. He rejected all the experiments of his colleagues Raphaël Dubois and Stéphane Leduc, who tried to synthesize structures resembling life.

Bonnier did not accept also that the origins of life on Earth rested on a long evolution of mineral matter into living matter. He finally settled (1907) on William Thomson's panspermia theory as the best solution. He claimed that meteorites carried life from one planet to another; only their surfaces were sterilized by atmospheric passage. He suggested examining the interior of carbonic meteorites for microbes, which, he thought, the only type of life they could carry. For him, the Earth provided an ideal environment in which such travellers might flourish. He did not concern himself with the problem of the origins of life; in his opinion, the eternity of life was a perfectly acceptable conclusion. The timelessness of the universe, matter, and life was central to both his and Van Tieghem's conception of panspermia.

Among French opponents to the panspermia hypothesis, first and foremost is the botanist and plant physiologist Paul Becquerel (1879–1955), a student of Bonnier's. He was the biologist of the Becquerel family: a family of famous physicists. He was a keen experimenter; at the beginning of the Twentieth Century, he carried out (2010) experiments on the persistence of bacteria, fungus spores, and seeds under extreme conditions. Becquerel concluded that panspermia was inconceivable.

Next in line is the botanist Julien Costantin (1857–1936), a student of Van Tieghem's. Costantin was professor at the Paris *Muséum d'Histoire Naturelle,* holding the chair of culture from 1901 to 1919. In 1923, he intervened in the debate over the origins of life in his book *Origine de la vie sur le globe* ("The Origins of Life on Earth"). Following an introductory chapter on geology in relation to the origins of life, he firmly rejected panspermia in Chapter Two. After a short historical review, he focused on Arrhenius's version of the theory. Following Marshall Ward, Emile Roux, and Paul Becquerel, he concluded that germs could not survive ultraviolet light exposure in space. Considering this possibility eliminated, he proceeded to develop his thoughts on the evolution of living matter.

For these French botanists and plant physiologists, the origins of life came down to two fundamentally contrasting alternatives. On the one hand, panspermia postulated the eternity of life, the universe, and matter, inspired by the seeming impossibility of evolution from inert to living matter. On the other, evolutionary abiogenesis (Kamminga, 1988) denied that very impossibility at the same time as rejecting spontaneous generation. For scientists like Costantin, the progressive evolution of matter leading to the most primitive living beings seemed plausible.

The debate over panspermia in the context of conceptions of evolution

All of the scientists I have presented were evolutionists. They formulated their divergent theories on the origins of life with that common assumption in mind. At the same time, they had differing understandings of evolution. It is useful to place their debates over the origins of life in the context of the debates between neo-Lamarckism and neo-Darwinism.

For Lamarck, the appearance of life from non-life was an ongoing process, currently active in nature and present at the beginning of each new species (Corsi *et al.*, 2006). For Darwin, this process was finished, having occurred only at the beginning of life. Life *itself* prevented life from emerging again (Tirard, 2005; Peretó, Bada and Lazcano, 2009). He restated this hypothesis in his famous 1871 letter to Hooker:

> It is often said that all the conditions for the first production of a living organism are now present, which could have been present. But if (and oh what a big if) we could conceive in some warm little pond with all sort of ammonia and phosphoric salts, – light, heat, electricity &c. present, that a protein compound was chemically formed, ready to undergo still more complex changes, at the present day such matter would be instantly devoured, or absorbed, which would not have been the case before living creatures were formed (Calvin, 1969).

Most of the debate among the French scientists occurred after August Weismann's revision of Darwinism. Meanwhile, the neo-Lamarckian influence in France at the end of the Nineteenth and the beginning of the Twentieth Century was quite strong (Loison, 2010). Van Tieghem was both a neo-Darwinian and a panspermist. However, his two students, Bonnier and Costantin, were both neo-Lamarckians. Bonnier embraced panspermia and Costantin rejected it.

Bonnier was a transformist who, over the course of his career, struggled endlessly against creationism. He used neo-Lamarckian theory to explain the transformation of species even as his opposition to evolutionary abiogenesis revealed the limits imposed by his Lamarckian viewpoint.

In *Le Monde végétal*, he emphasized the complexity of the various parts of a cell: particularly the protoplasm and nucleus. Protoplasm reveals what he called "cumulated characters", "heredity", and "complicated destiny". He thought the last of these the most important. His neo-Lamarckian viewpoint predisposed him toward a form of finalism, reflected in his conception of life. He wrote: "it is not more difficult to create straightaway an elephant than to create a bit of living matter." Given his transformist principles, he felt unable to explain how living matter possessing this "complicated destiny" could arise by simple evolution of matter. The problem was not only chemical, but philosophical. "To create living substance! How to hope for it in the current state of science?… When the man will solve this problem, he will be more creator than the Creator, stronger than the entire Nature, more powerful than the infinite Universe" (Bonnier, 1907, pp. 383–384). The eternity of life appears as a solution to the question of its creation in all its complexity.

In mentioning a creator, Bonnier was not expressing religious convictions but only highlighting the extreme difficulty in imagining how such evolution of matter could be possible. Present-day nature offered no help. Although species are subject to transformation, the essential complexity of living matter remains unchanged – thus posing an obstacle to how it could ever get started. His conclusion accorded with the results of his work on the physiology of seeds and bulbs, which revealed no interruption in the vital process. Panspermia offered the solution.

Costantin, of course, came to the opposite conclusion. He was likewise a committed neo-Lamarckian. He claimed that life emerged from an evolutionary synthesis of organic inert substances. For him, the advent of life on Earth was only the first stage in the transformation of matter, which continued in the development of various animal and plant species. The same year that Alexander I. Oparin gave his famous lecture – published the following year (Oparin, 1924)[1] – Costantin offered his own thorough analysis of the origins of life (1923). After criticizing the panspermia hypothesis, he described the possible evolution of living matter out of simple chemical elements. He concluded that, on ancient Earth, the elements hydrogen, carbon, and oxygen gave rise to carbon dioxide (CO_2) and water (H_2O). Living matter emerged from the interaction between these molecules and nitrogen coming from the Earth's crust. Under the influence of solar energy, the first "green" matter was produced.

Costantin's proposal is very similar to Herbert Spencer's, but presented more simply. Costantin certainly shows the British philosopher's influence – notably that of Spencer's *Principles of Biology* (1868), in which he describes a possible mechanistic, chemical evolutionary process.

Meanwhile, working in the early 1900s, Becquerel did not expound his ideas on evolutionary theories but focused on presenting his experimental results. He waited until 1924 to offer a neo-Darwinian account of the history of life on Earth. However, one can find early signs of his thinking: e.g., in a paper from 1910, where he offers an intriguing answer to the question of how life should be defined. He thought that his own work on latent life confirmed the definition of life offered by Claude Bernard, referring to Bernard's (1878) *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*. On Becquerel's reading of Bernard, it was no longer possible to claim life as a principle or a mysterious directive independent from the determinism of natural phenomena.

All of these scientists shared an interest in latent life and the persistence of life under extreme conditions

All of these scientists were specialists in the issues of latent life and the relationship between organisms and their physical environments – particularly under extreme conditions (Tirard, 2010). The connection to panspermia is obvious, given its interest in the capacity of organisms to resist extreme conditions in space.

Van Tieghem and Bonnier (1880) focused on the problem of latent life in the context of the physiological activity of seeds and bulbs, which, they claimed, maintained – under extreme conditions – a low level of physiological activity: i.e., they were alive. Costantin did not publish specifically on latent life. However, in *Les végétaux et les milieux cosmiques* (1898), he discusses the relationship between plants and their physical environment:[2] specifically, the conditions of heating, lighting, gravity, and aquatic environment. The book reflects his neo-Lamarckian convictions. He tries to show how environmental conditions influence the transformation of organisms, strongly defending the inheritance of acquired characteristics. As he wrote in (1923), the problem for panspermia is not the continuity or complexity of life but the lack of resistance of living matter to cosmic radiation.

During the 1910s, Becquerel made himself a specialist on the persistence of life under extreme conditions. Preparing his doctoral thesis under Bonnier's supervision, he carried out numerous experiments on seeds, fungus spores, and bacteria, exposing them to varying extreme conditions. He observed that, after the desiccation of fungus spores in a vacuum, their lifeforce was suspended: i.e., strict continuity of vitality was not needed; life could be interrupted.

Becquerel's experimental work connected closely to his interest in panspermia: specifically, its hypothesis on the cosmic origins of life. He considered whether low temperatures and vacuum could moderate the lethal effects of ultraviolet radiation, looking specifically at spores of *Aspergillus niger*, *Sterigmatocystis niger*, mucor, Cramant yeasts, and various sporulated bacteria. He concluded (1910) that, even if these organisms could survive the cosmic extremes of pressure and temperature, they could not survive exposure to ultraviolet radiation. The interplanetary environment was necessarily sterile, and panspermia was disproved. Such French biologists as Charles Richet, Raphaël Dubois, Yves Delage, and Edmond Perrier immediately embraced his results, making them a turning point in the tide of opinion against panspermia.

For the rest of his career – notably at the University of Poitiers – Becquerel remained interested in the panspermia debate, publishing two other important texts (1911; 1924) arguing against the panspermia hypothesis, based on the same evidence on persistence under extreme conditions.

Conclusion

At the end of the Nineteenth and the beginning of the Twentieth Century, French botanists and plant physiologists provided animated debates over the origins of life, the hypothesis of panspermia, the alternative perspectives of neo-Lamarckism and neo-Darwinism, and the resistance of organisms to extreme conditions. Out of this debate, they developed their varying conceptions of the nature of life and its origins – accepting the panspermia hypothesis or rejecting it.

Acknowledgements

I gratefully acknowledge support from PIR-CNRS: *Origines de la vie et environnements planétaires*, 2010.

Notes

[1] In the text, Oparin criticizes panspermia, describes conditions on ancient Earth, and offers a scenario of evolution from mineral to organic and living matter.
[2] Earth was included in the cosmic environment for Costantin; he studied the terrestrial environment from that perspective.

References

Arrhenius, S. 1903. *Worlds in the Making: The Evolution of the Universe.* London: Harpers and Brothers Publishers.

Becquerel, P. 1910. “L’action abiotique de l’ultraviolet et l’hypothèse de l’origine cosmique de la vie.” *Comptes rendus hebdomadaires des séances de l’Académie des Sciences*, 151:86–88.

—. 1911. “La panspermie interastrale devant les faits.” *Revue scientifique* 18 février 1911: 200–206.

—. 1924. “La vie terrestre provient-elle d’un autre monde?” *L’Astronomie, Bulletin de la société astronomique de France*, octobre: 393–417.

Bernard, C. 1878. *Leçons sur les phénomènes de la vie communs aux animaux et végétaux.* Paris: Baillère.

Bonnier, G. 1907. *Le Monde végétal.* Paris: Flammarion.
Calvin, M. 1969. *Chemical Evolution: Molecular Evolution towards the Origin of Living Systems on the Earth and Elsewhere.* Oxford: Clarendon Press.
Corsi, P., J. Gayon, G. Gohau, and S. Tirard. 2006. *Lamarck: Philosophe de la nature.* Paris: Presses Universitaires de France.
Costantin, J. 1898. *Les végétaux et les milieux cosmiques.* Paris: Félix Alcan.
—. 1923. *Origine de la vie sur le globe.* Paris: Ernest Flammarion.
Dick, S. J. 1996. *The Biological Universe: The Twentieth-century Extraterrestrial Life Debate and the Limits of Science.* Cambridge: Cambridge University Press.
Fry, I. 2000. *The Emergence of Life on Earth: A Historical and Scientific Overview.* New Brunswick NJ: Rutgers University Press.
Kamminga, H. 1988. "Historical Perspective: The Problem of the Origin of Life in the Context of Developments in Biology." *Origins of Life and the Evolution of the Biosphere* 18:1–11.
Loison, L. 2010. *Qu'est-ce que le néolamarckisme? Les biologistes français et la question de l'évolution des espèces, 1870–1940.* Paris: Vuibert.
Oparin A. I. 1924. "The Origin of Life." Translation Ann Synge in John Desmond Bernal. 1967. *The Origin of Life.* London: Weidenfeld and Nicholson.
Peretó, J., J. L. Bada, and A. Lazcano. 2009. "Charles Darwin and the Origins of Life." *Origins of Life and Evolution of Biospheres* 39:395–406.
Spencer, H. 1868. *The Principles of Biology.* London: Williams and Norgate.
Thomson, W. 1872. "Address of Sir William Thomson (President), 1871." *British Association for the Advancement of Science, Edinburgh, Report.* London: John Murray.
Tirard, S., 2003. "Gaston Bonnier (1853–1922): un botaniste lamarckien." *Bulletin d'histoire et d'épistémologie des sciences de la vie* 10(2):57–186 [Printed in 2004].
—. 2005. *L'histoire du commencement de la vie à la fin du XIX*e *siècle Cahiers François Viète*, 9:105–118. Second edition. 2008. *La pré-histoire de la vie Hors-série Sciences et avenir: L'univers est-il sans histoire?* 146:30–34.
—. 2006. "William Thomson (Kelvin) Histoire physique de la Terre et histoire de la vie." In *Pour comprendre le XIXe siècle: Histoire et philosophie des sciences à la fin du siècle*, edited by J.-C.-Pont, L.

Frieland, F. Padovani, and L. Slavinskaia, 297–306. Genève: Olski, Collection Bibliothèque d'histoire des sciences.

—. 2010. *Histoire de la vie latente: Des animaux ressuscitants du XVIIIe siècle aux embryons congelés du XXe siècle*. Paris: Adapt-Vuibert.

Van Tieghem, P., and G. Bonnier. 1880. "Recherches sur la vie ralentie et sur la vie latente." *Bulletin de la société botanique de France* 27:83–88, 116–122.

Van Tieghem, P. 1891. *Traité de Botanique.* 2 volumes, second edition. Paris: Librairie F. Savy.

CHAPTER ELEVEN

EXTRATERRESTRIAL LIFE FOR THE MASSES: KNUT LUNDMARK AND THE BOUNDARIES OF SCIENCE

GUSTAV HOLMBERG

Introduction

While astrobiology today is an established part of modern science – there are academic journals, conferences, an IAU commission, PhD programmes devoted to astrobiology – that has not always been the case. The search for extraterrestrial life has at times been defined as extraneous to science: something that did not really belong inside the scientific enterprise. There have been attempts to wall it off from science proper when, at various stages in the history of astrobiology, observatory directors have turned down proposals for radio telescope time, politicians have argued that it was a waste of taxpayers' money, and distinguished biologists have said that it was "a science without a subject" threatening to siphon off funding from more established parts of biology. It has been – to quote a leading historian of the field – an activity "at the limits of science" (Dick, 1996); and sometimes that limit has been drawn in such a way that it was excluded from science. However, the scientific search for and discussion of extraterrestrial life – variously labelled astrobiology, exobiology or bioastronomy – *has* managed to develop and become a part of science. How this came about exemplifies how an ensemble of processes, powers, and people together can shape the boundaries of science (Gieryn, 1999).

To succeed, proponents of the scientific search for life on other planets have drawn on a variety of resources. They have been successful in getting funding from agencies such as NASA, which was important at a time when the pioneers of astrobiology had trouble securing funding from agencies aimed at more established disciplines. Astrobiology-oriented astronomers made vital contributions to topics that were seen to be of high societal relevance, such as the nuclear winter debate and the growing awareness of climate change – proving that the field was not only *l'art pour l'art* pure science pursued for philosophical reasons, but could produce research results with an impact outside of the field. Prominent and well-established scientists – some of them Nobel Prize winners – gave credibility to the field, leveraging their status behind a bid for exobiology and bioastronomy belonging inside the boundary demarcating science from nonscience (Badash, 2009; Strick, 2004). Another resource – policy-wise but sometimes also economically, used to occasional great effect by the astrobiologists – came out of very successful public engagement with science-related activities pursued in a range of contexts: astrobiology has had no trouble making people outside academe or the traditional channels of research policy interested in the field. Organizations such as The Planetary Society, wealthy individuals such as Microsoft co-founder Paul Allen and filmmaker Steven Spielberg, and millions of participants in the pioneering citizen-science project Seti@home: all have been supporters in one way or another of exobiology, an interest fuelled by popular science in the media. The relationship between science and public can have an impact on the workings of science. To gain an understanding of the conditions shaping the scientific search for life in the universe, one ought not to look exclusively on the workings of academic peer review, technological developments that have given new tools for the astrobiologist, traditional funding mechanisms, research policy priority-setting processes, or the dynamics of academic organisation. The field has had a number of prominent spokespersons; their appearances in a variety of media have had an effect on the science of exobiology.

The popularization of exobiology has a long history. Studies of the dynamics of the discipline during its formation need to take into account this perspective of public engagement with science. They can take their cue from recent work in history of science studying publishing, media, and popularization as important parts of scientific practice: specifically, a growing body of historical work that identifies publishers, printers, booksellers, television producers, etc. as significant actors and factors in charting the historical development of science. In this approach to history of science, communication is not something that merely follows once knowledge has been produced; it is conceptualised as an integral part of scientific practice (Secord, 2004; 2000; Topham, 2004; 2009).

Astrobiology news and popularization feature regularly in today's popular science; but, as said, they come with a history: scientists working on extraterrestrial life have routinely published their findings in public outlets. One could argue that, before the discipline began to form – before the arrival of today's academic journals, conferences, graduate schools, and funding programmes for astrobiology – such popular publications functioned as the primary platforms for the development of ideas, concepts, and theories on the scientific search for life in the universe, there being so few professional and academic communication channels available.

This chapter is about the search for extraterrestrial life during its pre-disciplinary era, particularly the 1920s and 1930s. It will focus on Swedish astronomer Knut Lundmark's (1889–1958) work on extraterrestrial life, including the ways he used non-academic publications to popularize and develop his ideas. Lundmark's defence of the panspermia hypothesis – against the criticism levelled by Paul Becquerel – came in what would normally be termed a popular publication, not in an academic journal. I analyse the contents of Lundmark's publications and their reception among professional astronomers. Also, I draw on unpublished sources, such as letters between Lundmark and his publishers preserved in the Knut Lundmark Collection at the Lund University library.

Knut Lundmark

In the spring of 1925, Knut Lundmark received a proposal from Uppsala publisher J.A. Lindblad: would he be interested in writing a popular work for entering in a prize competition? He *was* interested. The editor at Lindblad's proposed the title "The Depths of Space"; Lundmark's counter-proposal, in mid-August, was *Världsrymdens liv: "*Life in the Universe".

The editor was sceptical at first. The subject lacked concrete focus, he wrote in his letter to Lundmark at Greenwich Observatory, where Lundmark was a guest researcher. Lundmark dashed off a quick response and succeeded in persuading the editor, who by the end of August was in agreement. Lundmark delivered the manuscript in December of the same year; in January, he was awarded one of four prizes in the competition – the other three were for books on the nature of light, the nervous system in animals, and syphilis. Lundmark received 1,250 Swedish kronor, a non-trivial sum at the time, and, of course, a publishing contract with royalties. Three thousand copies of *Världsrymdens liv* were printed in 1926. Apparently the topic *was* suitable for popularization. As Lundmark's career unfolded, he continued to publish on extraterrestrial life in a range of media.

By the mid-1920s, Lundmark's career as an astronomer had both positive and not-so-positive prospects. On the one hand, he was receiving international recognition for his work on galaxies. In 1920 at Uppsala University, he defended his PhD dissertation on the distance to M31, the

Andromeda Nebula as it was called at the time, before it became established that it was not a nearby nebula but in fact a large galaxy, as both Lundmark and Hubble argued (Way, 2013). After that he worked at observatories on the United States' West Coast that were emerging as world centres for observational astronomy; and then at the Greenwich observatory (see e.g. Holmberg, 1999; Kärnfelt, 2009).

At the same time, Lundmark lacked a permanent academic position – those were few and far between in Swedish astronomy – and his economic situation was far from secure. This was one reason why he began to involve himself in popular science. Like many other Swedish scientists at the time, he had accumulated significant debts during his time as a student, especially since he had come to Uppsala University from a very poor background. It is hardly surprising that the publisher's proposal interested him. That said, though the book royalties and the prize money were substantial, they were not the only rewards he sought in writing on extraterrestrial life.

Världsrymdens liv begins with a historical review – typical for Lundmark who, throughout his career, would come time and again back to the history of science and culture. He then proceeds to discuss the planets in the solar system, claiming the possibility of life on Mars. He was no sensationalist; his take on the discovery of extraterrestrial life – understandable in the wake of the Martian canal controversy – is that perhaps all science can do when it comes to such matters is to point to possibilities. This caution is evident elsewhere in the book; Lundmark writes, "the banner of critical thinking must be held high" – perhaps anticipating criticism from more traditionally minded colleagues.

The book rounds off with discussion of the panspermia hypothesis. Lundmark gives generous space to his successful Swedish predecessor in public discussion of astrobiology: Nobel-Prize-winning chemist and physicist Svante Arrhenius. He discusses thoroughly Arrhenius' theory and the criticisms levelled against it. He argues that Becquerel's work on the sterilizing effects of ultraviolet radiation in space ought to be checked independently and further verified before taken as proof against panspermia *à la* Arrhenius; till then, the jury on panspermia was still out. He proceeds to develop further Arrhenius' version of the panspermia hypothesis, using the tools of celestial mechanics to discuss the passage of spores between stars – a possibility for a professional astronomer like Lundmark that was not available to Arrhenius.

Through the second half of the 1920s, Lundmark remained quite positive towards panspermia and towards Arrhenius and his approach. In an article on Arrhenius in *Populär astronomisk tidskrift* – the main popular astronomical journal in Sweden – Lundmark describes the problems caused by Arrhenius' lack of experience in astronomy, but also the virtues of Arrhenius' coming to the field from outside. Not being an astronomer, Arrhenius did not command the technical tools of astronomy with which one might further develop his version of panspermia: e.g., through addition of arguments based on celestial mechanics. On the other hand:

> He could gain significant insights in celestial research without, as the case often is with professional astronomers, becoming confused and bogged down with innumerable conflicting details that make it hard to behold the grander scheme behind the varying phenomena. For a scientist of Arrhenius' personality it can be easier than for astronomers themselves to see where the results are heading (Lundmark, 1927).

In many ways, Lundmark shared this style of personality; but it did not endear him to the academic astronomy establishment. Eventually, he obtained the permanent position he coveted, winning the competition for a Lund professorship in 1929. At the same time, his position in Swedish astronomy became quite peripheral compared to the astronomers working at Uppsala and Stockholm/Saltsjöbaden. He developed an increasingly frosty relationship with certain of his fellow astronomers, such as Bertil Lindblad and his group at the Stockholm observatory, at the same time as finding an increasingly warm relationship with the general public (Kärnfelt, 2009; Holmberg, 1999). As the years went on, he was prolific both as an author and public intellectual, publishing on a large number of subjects, some quite distant from astronomy: history of science, August Strindberg, botany, and so on. He was an astronomer with many faces: a generalist looking in many directions to gain a synthesized image of Man's position in the universe. To his colleagues in professional astronomy, this smacked of amateurism: the negative consequence of too-broad interests and too-speculative outlook. To many others, he became Mr. Astronomy in Sweden, a central public figure whose publishing and public speeches aroused the interest of many in astronomy and the sciences.

Lundmark shared Arrhenius' broad outlook on scientific matters and his positive view of panspermia. In the late 1920s, he was publishing much of his work outside of the professional journals. When Arrhenius died, Lundmark was chosen to oversee completion of the half-finished eighth and final edition of *Världarnas utveckling* ("Worlds in the Making"): Arrhenius' major work in astrobiology. Entering the 1930s, he was well established in the academic world at the same time he continued his non-academic publishing. He was an increasingly public figure, a staple of the Swedish media, an oft-heard voice on radio, a public lecturer who penned articles on a diverse set of topics. He wrote on extraterrestrial life for newspaper and magazine articles. He continued to write for many types of publications, including those where his astronomer colleagues would never publish.

Many of the themes of Lundmark's first book on extraterrestrial life reappeared in later publications. Meanwhile, international publishers showed increasing interest in spreading that early work internationally. Translations were discussed; in 1930, Brockhaus published a German translation. As said, Lundmark edited and contributed original material to the eighth edition of Arrhenius' *Världarnas utveckling*. In 1935, his own *Livets välde* ("The Realm of Life") was published by Bonnier's, one of the leading publishers in Sweden with a long history of publishing in the field of popular astronomy. The initiative came from Bonnier's. In March 1930, on publication of the German edition of *Life in the Universe*, Bonnier's contacted Lundmark wishing to publish another, enlarged and revised book on the topic. Lundmark the junior astronomer of the mid-1920s had published what was a slim volume from a small publisher: Lindblad's. Lundmark the established astronomer and public figure of the 1930s had the printing and marketing resources of the nation's premier publisher. The resulting lavishly illustrated (and expensive) volume was released in late 1935 with good exposure in Bonnier's prestigious Christmas catalogue. Together with the nation's leading authors, historians, and heroes, he had arrived at a kind of cultural Parnassus. He received an advance of 4,500 Swedish kronor for the book – about a third of the yearly income of a Swedish university professor – with the promise of royalty checks in the future.

Boundaries

These books on extraterrestrial life sold as popular works. Reading them, though, one gets the feeling they were about more than just popularization. Surely, that feeling is present not only in this reader but also in the author: for Lundmark, life in the universe *is* a topic fit for scientific discussion. Written by a scientist, these works popularized both facts and methods. Some parts are quite advanced for being popular works, at the same time they remain accessible. Their style is in keeping with his other popular publications. He argued consistently for presenting a diversity of results and hypotheses in popular forums and not just catering to the current scientific consensus (see Kärnfelt, 2004; Holmberg, 1999, pp. 187–190).

Lundmark argued that extraterrestrial life had become a scientifically relevant subject – it fell within the boundaries of science – even as it had earlier mostly been about hunches and flights of fantasy. Advances in astrophysics made it possible to treat the subject scientifically, as commented upon positively by an anonymous reviewer for *Populär astronomisk tidskrift* in 1926. A review in the same journal of *Worlds in the Making* – written in 1929 by Carl Schalén – was much more sceptical of Lundmark's/Arrhenius' position, describing it as too speculative. A similar tone is evident in Bertil Lindblad's 1936 review of *The Realm of Life.* Lindblad – the foremost Swedish astronomer of the time – uses his review in what was a leading popular publication to state what he saw as the problems inherent in any such multidisciplinary undertaking as astrobiology. So, for example, because he was not a biologist, Lundmark clearly errs where he writes on biology: a not-uncommon argument against the scientific study of extraterrestrial life in the history of modern science.

Concluding remarks

Lundmark addresses quite advanced issues in his books on extraterrestrial life: e.g., *Världsrymdens liv* includes a page of dense equations. At the same time, he uses the books to weigh various ideas one against another. At the time, no established arena existed for astrobiology – or, as Lundmark liked to call it, cosmobiology. The openness of the field suited his open style. He intended his books not only to popularize knowledge already gained and secured, but also to provide a laboratory for discussing matters that remained quite open. He used them as a platform to discuss matters of interpretation in measurements and to develop theories and instruments of importance to astrobiology in ways not really possible in the standard astronomical journals where he published his scientific work. His books not only served to popularize – and popular they were, among both readers and publishers – but also played a serious role in scientific discussion. They cast real doubt on the possibility of making any clear-cut distinction between the popular and scientific literature of the 1920s and 1930s on extraterrestrial life. They reveal as well the importance of publishers and book markets when writing the history of astrobiology.

References

In the Knut Lundmark archive of the Lund University library: letters between Lundmark and the Bonnier's, Brockhaus, Lindblad's, and Geber's publishing houses.

Arrhenius, S. 1924–1929. *Världarnas utveckling.* Stockholm: Geber.
Badash, L. 2009. *A Nuclear Winter's Tale: Science and Politics in the 1980s*. Cambridge MA: MIT Press.
Dick, S. J. 1996. *The Biological Universe: The Twentieth-Century Extraterrestrial Life Debate and the Limits of Science*. Cambridge: Cambridge University Press.
Gieryn, T. F. 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago IL: University of Chicago Press.
Holmberg, G. 1999. *Reaching for the Stars: Studies in the History of Swedish Stellar and Nebular Astronomy, 1860–1940*. Lund: History of science and ideas department, Lund University.
Kärnfelt, J. 2004. *Till stjärnorna: Studier i populärastronomins vetenskapshistoria under tidigt svenskt 1900-tal*. Göteborg: Folkuniversitetets akademiska press.
—. 2009. *Allt mellan himmel och jord: Om Knut Lundmark, astronomin och den publika kunskapsbildningen*. Lund: Nordic Academic Press.
Lundmark, K. 1926. *Världsrymdens liv*. Uppsala: Lindblad.
—. 1927. "Arrhenius och kosmologien." *Populär astronomisk tidskrift*, 177–181.
Lundmark, K. 1935. *Livets välde: Till frågan om världarnas beboelighet. 1*. Stockholm: Bonnier.
Secord, J. A. 2004. "Knowledge in transit." *Isis* 95:654–672.
—. 2000. *Victorian Sensation: The Extraordinary Publication, Reception, and Secret Authorship of Vestiges of the Natural History of Creation*. Chicago IL: University of Chicago Press.
Strick, J. E. 2004. "Creating a cosmic discipline: The crystallization and consolidation of exobiology, 1957–1973." *Journal of the History of Biology* 37(1):131–180.
Topham, J. R. 2004. "A view from the industrial age." *Isis* 95:431–442.
—. 2009. "Introduction." *Isis* 100(2):310–318.
Way, M. J. 2013. "Dismantling Hubble's legacy?" In *Origins of the Expanding Universe: 1912–1932*, edited by M. J. Way, and D. Hunter. San Francisco CA: Astronomical Society of the Pacific.

CHAPTER TWELVE

THE ORGANIZATION OF ASTROBIOLOGY IN A SOCIOLOGICAL CONTEXT

URSZULA K. CZYŻEWSKA

Introduction

A variety of scientific studies, including the astrobiological, can be described in an almost infinite number of ways, from different perspectives. Astrobiological studies – especially in their institutionalized form, with their own history – can be described as a special structure that changes over time. This structure consists of scientific institutions, research centres, laboratories, and a variety of national and international publications.

This paper not only synthesizes information for a historical approach to the organization of astrobiology, but also offers a new heuristic framework for sociology of science: one that does not just study the importance of scientific final products or a given knowledge system e.g. theories, laws, patents, invention, discoveries; neither does it investigate their contents directly, as one might expect. It focuses on defining the nature of scientific ideas and their relations to other ideas and to various institutional and personality factors (see e.g. Barber, 1968), along with the implications: ideological, philosophical, aesthetic, religious. The sociology of science deals with scientists as human beings – the so-called *human factor* – along with their activities including research and development, their connection with other groups from the wider scientific community, and their inevitable and sometimes complex relationships to society.

It is reasonable to ask how, if at all, scientometric data reflects the organization of one or another branch of knowledge, as well as whether the outcomes of such relations are relevant for sociology of science. What does the data have to say about various institutional and personality factors? I focus carefully on describing the structure within which astrobiological studies take place and are popularized. To support my theoretical considerations, I also present a general scientometric analysis.

A few remarks on sociology of science

As a field, sociology of science closely relates to sociology of scientific knowledge. Both terms were popularized in the second part of the Twentieth Century by American sociologist Robert K. Merton (1910–2003).[1] Sociology of science deals with the social conditions and effects of science, along with the structures and processes of scientific activity (Barber, 1968; Merton, 1973; Ben-David and Sullivan, 1975; Leydesdorf 1992). The field has long been neglected both by natural (empirical) and non-empirical scientists. Another American sociologist, Edward Shils (1910–1995) counted the study of science/scientific institutions among the major undeveloped areas of sociological inquiry (Barber, 1952). Although the interaction between science, scientists, and society has been the subject of occasional interest by scholars, it remains underexplored. Some thoughts on the organization of astrobiology in light of sociology of science would greatly add to the discussion.

Barber (1952, p. 26) describes science as "a special kind of thought and behaviour which is realized in different ways and degrees in different historical societies", thereby emphasizing the sociological component of science. Considering astrobiology in a sociological context gives rise to a whole set of questions. How did it emerge and become institutionalized? How is it maintained and controlled? How is it organized? What determines the changes in its scientific organization, and how do those changes affect research? (Merton, 1968; 1973; Ben-David and Sullivan, 1975; Shapin, 1995). Further questions relate to the stages in astrobiology's development; the organizational structure of the science within which it falls, including processes of scientific discovery; its value systems, instrumental needs, and politics; its dialogue with religion; economic factors including financial support; astrobiology courses designed for different levels of the education system; the motivations of, rewards for, and communications among its researchers; and society's awareness of it (Barber, 1952; 1968; Merton, 1973; Jakosky, 2000; Lemarchand, 2000). Such themes are essential for understanding the social aspects of any science, including astrosociology – defined by Jim Pass (2004, *online*) as the study of "all social forces, organized activities, objectives and goals, and social behaviours directly or indirectly related to spaceflight and exploration or any of the space sciences." Though astrosociology has been broadly criticized, many appreciate its attempt to discern the relationship between society and universe.

Aware of the many diverse motifs connected with this subject, I chose to focus on certain aspects of astrobiology's sociological context: its institutional organization and the communication among its scientists.

Astrobiology's organization

To understand astrobiology's organization requires understanding its institutional origins in NASA's Astrobiology Institute (NAI) at Ames Research Center. Astrobiology emerged in the second half of the Twentieth Century focused initially on Project Phoenix, which looked for incoming radio signals. Around 2004, NASA resumed study of the nature and origins of life: terrestrial and extraterrestrial. The research focused on life in the universe – not specifically on intelligent life (Basalla, 2006). Instead of asking, "where are they?"[2] astrobiologists asked: how does life begin and evolve? Does life exist elsewhere in the universe?, What is the future of life on Earth and beyond? (Lemarchand, 2000; Des Marias *et al.*, 2008). The change of focus did not mean that the search for extraterrestrial intelligence had stopped or even diminished, the search for which was revived and incorporated into NAI studies in the late 1990s, albeit not as the dominant venture it had been in the past.[3] A key turning point came at a press conference given at Ames by NASA Administrator Daniel Goldin 19 May 1996, at which he officially named Ames NASA's Lead Center in Astrobiology, with the mission of exploring collaborative possibilities for an Astrobiology Institute (NAI History, 2008). The resulting NAI was placed under the directorship of Baruch Blumberg (1925–2011), winner of the 1976 Nobel Prize in physiology and medicine. George Basalla writes:

> Blumberg directed a "virtual" Institute because its teams of researchers, scattered around the world in universities and laboratories, communicated with one another electronically. On occasion they assembled for more traditional scientific meetings (2006, p. 172).

This represented a tremendous shift in traditional thinking about life and the way it should be studied, as described in *The NASA Astrobiology Roadmap*. It led to the formation of an international, highly multidisciplinary group of researchers, including geologists, chemists, oceanographers, planetary scientists, molecular biologists, virologists, zoologists, and palaeontologists: each discipline contributing to the overall understanding (Cockell, 2002). Most were recruited at conferences or workshops. Henry Gee writes (2001, p. 1079):

> In the age of increased narrowness of research goals, it is invigorating to lift one's eyes to the stars and consider life in its broadest sense. One consequence of astrobiology will be to deepen our understanding of our own place in the Universe, our uniqueness and our potential.

Even economic theory has its role, weighing costs and benefits; while sociology is needed to understand public perspectives and risk perception. For nearly fifteen years, the institute has flourished as its researchers have studied terrestrial extremophiles, looked for signs of life in examined meteorites and soil samples from Mars, and detected possible biological activity in the atmosphere of one of the many discovered exoplanets.

Effective communication among researchers is an indispensable part of scientific activity. The development of an organization to the field – the assembly of a group of scientists to achieve a common scientific aim – requires collaboration among research centres, governmental space agencies (NASA, ESA, JAXA), and societies (the European Astrobiology Network Association, the Nordic Astrobiology Network, the Finnish Astrobiology Network, The Swedish Astrobiology Network, the Centre of Advanced Studies in Astrobiology and Related Topics, etc.) – all of whom play an essential role in not just conducting but popularizing astrobiological research. I wish to focus on the NAI.

The NAI is one of four components of the NASA Astrobiology Program.[4] It represents a partnership between NASA, universities, and other research agencies to conduct multidisciplinary research, to train young scientists, and to promote to the public the adventure that is the study of the living universe. It was founded in 1998 with eleven institutions, including Arizona State University, the Massachusetts Institute of Technology, Michigan State University, Pennsylvania State University, University of California, University of Colorado, University of Rhode Island, University of Washington, Harvard University, the Marine Biological Laboratory (Woods Hole), along with NASA team centers, the Ames Research Center, the Johnson Space Center, the Jet Propulsion Laboratory, the Scripps Research Institute, and the Carnegie Institution for Science (NAI History, 2008).

The NAI has developed a global partnership program including among others the Australian Centre for Astrobiology, the *Centro de Astrobiología*, the Astrobiology Society of Britain, the *Société Française d'Exobiologie*, the European Exo-/Astrobiology Network Association, the Russian Astrobiology Centre, the *Instituto de Astrobiología Colombia*, and the Nordic Network of Astrobiology Graduate Schools. Foreign agencies may become either associate or affiliate members. Associate membership requires a formal agreement between NASA and the organization's national government; affiliate membership does not. Membership brings such benefits as participation in NAI meetings and workshops, access to its postdoctoral fellowship program, and opportunities for exchange of researchers and students (International Partners 2009).

Among the important creations of the early modern period in science were the creation of national and international societies and professional publications including journals and books, allowing speedier and broader communication of scientific work. Naturally, this includes astrobiology, whose contemporary communications make use of the Internet, including websites, databases, and webzines[5], along with traditional printed materials. The primary method for presenting research results is through conferences such as the biennial Astrobiology Science Conference (AbSciCon), where scientists are encouraged to push the boundaries and bridge the barriers between disciplines.

Thousands of books address astrobiology but many of them do so only in a narrow way as part of considering other disciplines. This makes it difficult to determine which books in the field are most significant. Meanwhile, scientometric data shows that the most representative journals of the field are *Astrobiology* and the *International Journal of Astrobiology*. The overall number of listed journals is one way to measure the maturity of a field (see e.g. Garfield, 1972; Woolf, 2006; Chang *et al.*, 2011). A journal's ranking rises when it becomes listed on the ISI journals list and has a high impact factor.

From its beginnings, the field has made a concerted effort to emphasize the importance of education: training the next generation of astrobiologists and keeping society aware of results. Many universities have at least one course in astrobiology. Many of those have degree programs. US undergraduate programs can be found at Rensselaer Polytechnic Institute's Center for Astrobiology, South Carolina State University, the University of Rhode Island, and Penn State University's Astrobiology Research Center; postgraduate programs can be found at UCLA's IGPP Center for Astrobiology, the University of Washington, Penn State University's Astrobiology Research Center, and University of Arizona's Center for Astrobiology. Stanford University offers an online course. Students specialize either in either biology or astronomy while approaching astrobiology through specially designed curricula, textbooks,[6] workshops, science camps, and training classes (Woolf, 2006; Wells *et al.*, 2007).

As a panelist at the 2011 Origins conference, Gerda Horneck presented an example of interdisciplinary astrobiology studies from Germany. She stressed the importance of educating the very youngest students, starting from kindergarten (2011, *online*). She described children eagerly participating in such exercises as baking of chocolate sprinkle cookies, where the sprinkles represent the seeds of life.

General scientometric overview

Scientometrics analyses science through measuring the impact of scientific publications by qualitative, quantitative, and computational methods. It relates closely to such fields as history of science and technology, philosophy of science, and sociology of scientific knowledge. Modern scientometrics is based mostly on the work of Derek J. de Solla Price (1922–1983) and Eugene Garfield (1925–present), who founded the Institute for Scientific Information (ISI) (see Garfield, 1972).

Scientometrics' history can be traced back to the beginning of the Nineteenth Century, in such fields as law and psychology. The growth of databases in the 1960s–1970s made widespread citation analysis possible. The Twenty-first Century has brought a proliferation of databases, author-focused indices, and an increasingly important Web (*Research Trends*, 2007a). Scientometrics researches article output, citation comparisons between disciplines, and geographical distribution of publications. It has become central to monitoring progress in the scientific performance of research groups, departments, universities, and other institutions (Huang *et al.*, 2011; Geraci and Esposti, 2011; Chang *et al.*, 2011). It has been used by governments and other funding bodies to determine allocation of research funds (*Research Trends*, 2007b).

Various methods are available for data gathering; most are based on bibliometric approaches. Commercial and freeware tools include ISI Web of Knowledge™, Scopus™, and Google Scholar™. I will focus on basic, quantifiable *research assessment measures* (RAM)[7] within the Thomson Reuters ISI Web of Science database: an online version of the index that allows subscribers to search across numerous citation databases; including such sub-databases as SCI-EXPANDED, SSCI, A&HCI, CPCI-S, and CPCI-SSH. I will use it to evaluate the following factors relative to astrobiology: institutions; funding agencies; and document types, titles, languages, year of publication, subject area, authors, and geographical distribution. Chang *et al.* (2011, p. 22) write: "ISI would seem to establish the 'gold standard' database for purposes of generating RAM for journals and individual researchers in a wide range of disciplines in the sciences and social sciences for an extended period."

The word "astrobiology" appears 1,290 times[8] within the publications surveyed: in titles, as elements of key terms, or in abstracts. The minimum record count (threshold) was set at two. Related terms produced the following results: “exobiology” (393 mentions), “bioastronomy” (49), “biocosmology” (2), and “cosmobiology” (13). A total of 351 institutions are represented. Table 1 presents the top ten. Not surprisingly, NASA holds first rank.

The top ten funding agencies by reference are NASA (4.884%), NSF (1.705%), ESA (0.853%), NASA Astrobiology Institute (0.775%), CNES (0.620%), National Science Foundation (0.620%), COMAT Aerospace Toulouse France (0.388%), National Aeronautics and Space Administration (0.338%), ILEWG (0.310%), International Lunar Exploration Working Group ILEWG (0.310%). The top ten document types are article (56.899%), proceedings paper (32.016%), meeting abstract (6.899%), review (5.581%), editorial (4.109%), news item (0.930%), letter (0.775%), book review (0.698%), book chapter (0.310%), and correction (0.233%).

INSTITUTION	1	NASA
	2	CalTech
	3	University of Arizona
	4	SETI Institute
	5	CNRS
	6	University of Colorado
	7	Open University
	8	University of Washington
	9	Harvard-Smithsonian Center for Astrophysics
	10	CSIC

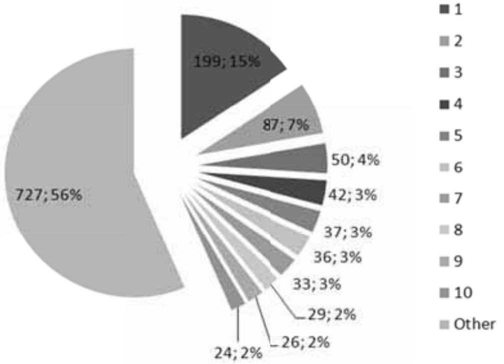


Table 1. Scientometric data by institution referenced.

The term “astrobiology” appears in a total of 128 journals. Table 2 shows the top ten. The more issues a journal has, the higher in general its percentage.

SOURCE TITLES		
	1	Astrobiology
	2	Astrophysical Journal
	3	Proceedings of the Society of Photo Optical Instrumentation SPIE
	4	International Journal of Astrobiology
	5	Icarus
	6	Astronomy Astrophysics
	7	IAU Symposium Proceedings Series
	8	ESA Special Publications
	9	Advances in Space Research
	10	Origins of Life and Evolution of Biospheres

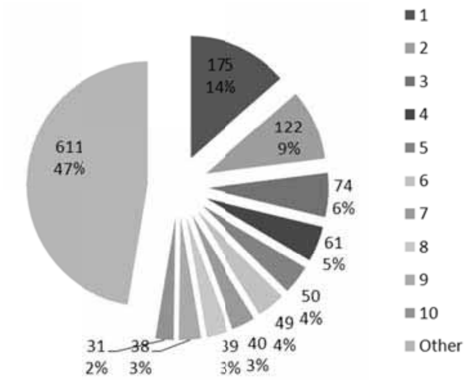


Table 2. Scientometric data by journal title.

Given the search terms, only two languages are represented in the data: English (99.690%) and Spanish (0.233%). A total of three articles were published in the Spanish-language journals *Revista Mexicana de Ciencias Geologicas* and *Acta Bioquimica Latinoamericana*.

The data covers the period 1998–2012. A (very) few articles were published in 1975, 1983, and 1997, but the minimum record count was not reached. Over the past decade, percentage of overall publications by year are 5.504% (2002), 9.070 (2004), 7.907% (2005), 9.535% (2006), 11.085% (2007), 9.457% (2008), 7.519% (2009), 11.628% (2010), 11.550% (2011), and 4.651% (2012). Note that 2010 was the peak year for publications. I do not consider publications from 1998, 1999, 2000, 2001, or 2003.

The top ten subject areas represented are astronomy astrophysics (61.4550%), geology (25.814%), life sciences biomedicine: other topics (23.178%), optics (9.225%), engineering (7.519%), science technology: other topics (5.349%), geochemistry geophysics (3.643%), meteorology atmospheric sciences (3.643%), chemistry (3.256%), and physics (3.178%)

The top nine authors, after "anonymous" (4.341%)[9], are R.B. Hoover (2.171%), P. Ehrenfreund (2.016%), G. Horneck (1.628%), H.G.M. Edwards (1.473%), F. Raulin (1.473%), C.S. Cockell (1.395%), A. Brack (1.163%), R.I. Kaiser (1.163%), and J. Chela-Flores (1.085%).

Table 3 shows the top ten countries represented. Given the history of astrobiological studies, it is not surprising that the USA is on top.

COUNTRIES		
	1	USA
	2	UK
	3	France
	4	Germany
	5	Spain
	6	Netherlands
	7	Italy
	8	Japan
	9	Canada
	10	Austria

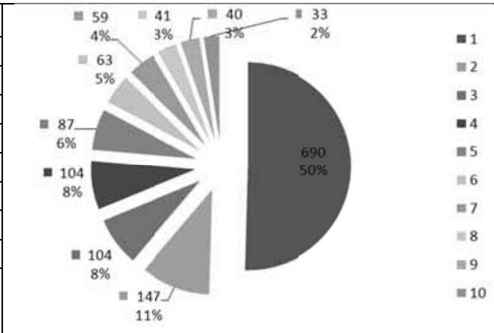


Table 3. Scientometric data by countries represented.

Discussion

Scientometric analysis offers basic insights into the organization of science, but it is not error free. Indeed, it is constantly criticised, and some reject it outright as a source of reliable data. Closer examination of the results from the ISI TR data reveals grounds for suspicion. Consider the potential for confusion between full names and acronyms for universities, funding agencies, etc.: e.g., the International Lunar Exploration Working Group vs. ILEWG, the National Science Foundation vs. NSF, the National Aeronautics and Space Administration vs. NASA; or the potential for confusion with second authors' names or use only initials; or the potential for confusion between United Kingdom, UK, and England. All of these have potentially serious consequences for the reliability of results and the validity of conclusions.

At the same time, scientometric research *is* important for understanding the organisation and practice of astrobiology, even though it cannot be considered as conclusive argument when methodological and sociological characteristic of astrobiology is concerned. It offers insights into where astrobiologists come from, where they publish, what they write about (clearly, astronomy and astrophysics are key subject areas, even while other publications reflect the tremendously multidisciplinary nature of astrobiology), and so on. The articles they publish reveal a wide range of objectives, from theoretical to classificational, structural, and synthetic goals (Zeidler, 2010).

Conclusions

The chapter discusses astrobiology's organisation and institutional origins in a sociological context, accompanied by general scientometric analysis. It described the activities by which researchers and their international partners exchange ideas, and data and help to train young astrobiologists.

The research reported here affords a number of conclusions. First, the organization of any scientific discipline can be expressed within sociology of science, which outlines the scientific activity and achievements of the relevant research group and those associated structures, including institutions requiring separate, in-depth study. Astrobiology is a hierarchical, institutionalized science, open to international cooperation. In the case of astrobiology, the leader in the field is the NAI. Scientometric analysis of RAM ISI data describes the quantitative characteristics of the field. The results must be taken cautiously and interpreted narrowly; but, at the same time, they allow identification of research trends.

Acknowledgments

Special thanks to J. Herda, PhD, at the Institute for Philosophy of Nature and the Natural Sciences, Department for Philosophy of Biology, KUL, for her helpful suggestions on this paper. This work was financed by Ministry of Science and Higher Education funds for science 2010–2012 as research project NN101253239.

Notes

[1] "…The sociology of knowledge is defined as the part of sociology that studies the nature of and relations between different types of idea systems, on the one side, and the relations between these idea systems and a variety of institutional (or social-structural) and personality factors, on the other, then the sociology of science is one part of the sociology of knowledge" (Barber, 1968, p. 92).

[2] The early years of NASA's interest in SETI coincided with the failure of the 1976 Viking landers to find evidence of life on Mars; in the aftermath of which the University of Maryland hosted a 1979 symposium entitled "Where Are They?", revisiting Enrico Fermi's question on the aliens' whereabouts (Basalla, 2006).

[3] In the summer of 2003, NASA formally named the scientists of the SETI Institute to one of twelve new teams in its Astrobiology Institute.

[4] The other components are Exobiology and Evolutionary Biology (EXO), established in 1965 to support research into the origins of life, with focus on the potential for life to exist elsewhere in the universe; Astrobiology Science and Technology Instrument Development (ASTID), established in 1998 to support development of instrument prototypes; and Astrobiology Science and Technology for Exploring Planets (ASTEP), established in 2001 to support explorations of extreme environments so as to develop techniques for astrobiological exploration of other worlds in our solar system (Our Research 2007; ASTEP Projects 2008, Projects 2011a, Projects 2011b).

[5] See e.g. *Astrobiology Magazine*: http://www.astrobio.net/.

[6] As of 2007, five undergraduate textbooks in astrobiology had been published in the US: four intended for non-science students and only one for science majors. At the high school level, 2003 and 2005 each saw the development of full-year astrobiology curricula (Wells *et al.*, 2007).

[7] Various ISI RAM analyses, some calculated annually and others daily, are presented in (Chang *et al.*, 2011). These include the classic *two-year impact factor* (2YIF), *five-year impact factor* (5YIF), immediacy (or *zero-year impact factor*: 0YIF), Eigenfactor, article influence, *citation performance per paper online* (C3PO), h-index, Zinfluence, *papers ignored by even the authors* (PI-BETA), and *impact factor inflation* (IFI), along with three new RAM: *historical self-citation threshold approval rating* (H-STAR), *two-year self-citation threshold approval rating* (2Y-STAR), and *cited article influence* (CAI).

[8] The results were re-calculated 31 July 2012.

[9] Authorship in these cases is anonymous because the publications were a part of a journal special issue or conference proceedings mainly containing article abstracts.

References

"ASTEP Projects." 2008. http://astrobiology.nasa.gov/astep /projects/ (accessed June 14, 2011).

Basalla, G. 2006. *Civilized Life in the Universe: Scientists on Intelligent Extraterrestrials*. Oxford: Oxford University Press.

Barber, B. 1952. *Science and the Social Order*. New York: Collier Books.

—. 1968. "The sociology of science." In *International Encyclopedia of the Social Sciences*, edited by D. L. Sills, 13:92–100. New York: The Macmillan Co & The Free Press.

Ben-David, J., and T. A. Sullivan. 1975. "Sociology of science." *Annual Review of Sociology* 1:203–222.

Chang, C. L., M. McAleer, and L. Oxley. 2011. "What makes a great journal great in the sciences? Which came first, the chicken or the egg?" *Scientometrics* 87:17–40.

Cockell, C. 2002. "Astrobiology – a new opportunity for interdisciplinary thinking." *Space Policy* 18:263–266.

Des Marias, D. J., J. A. Nuth III, L. J. Allamandola, A. P. Boss, J. D. Farmer, T. M. Hoehler, B. M. Jakosky, V. S. Meadows, A. Pohorille, B. Runnegar, and A. M. Spormann. 2008. "The NASA astrobiology roadmap." *Astrobiology* 8(4):715–730.

Garfield, E. 1972. "Citation analysis as a tool in journal evaluation." *Science* 178(4060):471–479.

Gee, H. 2001. "Nature insight: Astrobiology." *Nature* 409:1079.

Geraci, M., and M. D. Esposti. 2011. "Where do Italian universities stand? An in-depth statistical analysis of national and international rankings." *Scientometrics* 87:667–681.

Huang, C., A. Notten, and N. Rasters. 2011. "Nanoscience and technology publications and patents: A review of social science studies and search strategies." *The Journal of Technology Transfer* 36:145–172.

"International partners." 2009. http://astrobiology.nasa.gov/nai/international-partners/ (accessed September 09, 2011).

ISI Web of Science. "Essential science indicators." *Thomson Reuters ISI*. http://thomsonreuters.com/products_services/science/science_products/a-z/essential_scien ce_indicators/ (accessed September 23, 2011).

Jakosky, B. 2000. "Philosophical aspects of astrobiology: A new era in bioastronomy." In G. Lemarchand, and K. Meech (ed.), *Astronomical Society of the Pacific Conference Series* 213:661–666.

Lemarchand, G. A. 2000. "A new era in the search for life in the universe." *Astronomical Society of the Pacific Conference Series* 213:7–18.

Leydesdorf, L. 1992. "The knowledge content of science and the sociology of scientific knowledge." *Journal for General Philosophy of Science* 23:241–263.

Merton, R. K. 1968. *Social Theory and Social Structure*. New York: The Free Press.

—. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: The University of Chicago Press.

"NAI history." 2008. http://astrobiology.nasa.gov/nai/about/history/ (accessed September 18, 2011).

"Our research." 2007. http://astrobiology.nasa.gov/nai/research/ (accessed June 14, 2011).

"Outreach and education panel." 2011. http://www.exobiologie.fr/origins2011/slides/ (accessed August 15, 2011).

Pass, J. 2004. "Understanding space and society from a grounded perspective…" http://.astrosociology.com

—. 2010. "The astrosociological implications of astrobiology." In G. A. Robertson (ed.), *Space, Propulsion & Energy Sciences International Forum-SPESIF*. American Institute of Physics, 402–417.

"Projects" 2011a. http://astrobiology.nasa.gov/astid/projects/ (accessed June 14, 2011).

—. 2011b. http://astrobiology. nasa.gov/exobiology/ projects/ (accessed June 14, 2011).

Research Trends. 2007a. "Scientometrics from past to present." *Research Trends* 1:2.

—. 2007b. "Scientometrics from past to present: part two." *Research Trends* 2:2.

Shapin, S. 1995. "Here and everywhere: Sociology of scientific knowledge." *Annual Review of Sociology* 21:289–321.

"The Stanford astrobiology course." *Programs in the United States*. http://www.stanford.edu/group/astrobiology/cgi-bin/?page_id=663 (accessed October 20, 2011).

Wells, L., J. Armstrong, and J. A. Huber. 2007. "Disciplinary aspirations and educational opportunities." In *Planets and Life: The Emerging Science* of Astrobiology, edited by W. T. Sullivan III, and J. A. Baross, 548–557. Cambridge: Cambridge University Press.

Woolf, N. 2006. "Why is astrobiology important for the USA and NASA?" http://www3.geosc.psu.edu/astrobiominor/Why Astrobiology.pdf

Wynn-Williams, D. D. 2001. "Why astrobiology needs collaboration?" *Astronomy and Geophysics* 42(5):5.20–5.21.

Zeidler, P. 2010. "Co, w jakim celu i w jaki sposób badają chemicy? Współczesne kontrowersje wokół przedmiotu i specyfiki metodologicznej praktyki badawczej chemii" ["What, for what purpose and how examine the chemists? Contemporary controversy concerning the subject and the specific methodological practices of chemistry research"]. In *Filozofia przyrody współcześnie* [The Philosophy of Nature Today], edited by W. M. Kuszyk-Bytniewska, and A. Łukasik, 111–130. Kraków: Universitas.

CHAPTER THIRTEEN

ASTROBIOLOGY AND THEOLOGY: UNEASY PARTNERS?

CHRISTOPHER C. KNIGHT

Introduction

From the theologian's perspective, the astrobiological project of the early twenty-first Century raises a number of questions. At one level, these arise from the astrobiologist's judgment that the possibility of extraterrestrial life – even self-conscious, intelligent life – is not negligible: a belief that is clearly incompatible with at least some traditional theological positions. The theologian will inevitably ask how this is to be assessed in relation to a whole range of current theological opinions.

At another level, more general questions arise. It is with these I shall begin, since they are relevant even to those who do not find the theological task a meaningful one. These questions are essentially psychological, sociological, and philosophical. Why – theologians ask – do so many people have such a strong interest in astrobiology? Is it simply the natural curiosity that goes with any scientific pursuit; or is something else at work, about which theologians may have insights because of the psychological, sociological, and philosophical tools that have long been part of their intellectual toolkit?

Psychology, sociology and philosophy

From a psychological perspective, the unconscious roots of religious belief are clearly relevant to the question of motivation, especially when astrobiological research is linked to the *search for extraterrestrial intelligence* (SETI). A number of commentators have noted the religious undertones of certain strands of astrobiological speculation, with Karl S. Guthke noting how these sometimes represent a kind of "religion or alternate religion" (Crow, 1986, p. 559) and Frank Tipler (1981) making similar comments in relation to the writings of Frank Drake and Carl Sagan. Indeed, a widespread – if usually unconscious – belief in some kind of "salvation" through extraterrestrial beings was noted more than half a century ago by C.G. Jung in his fascinating book on the early stages of the UFO phenomenon; he argues there that UFO sightings represent the eruption into visionary experience of an archetype "that has always expressed order, deliverance, salvation and wholeness." It is characteristic of our time, he goes on, that "in contrast to previous [religious] expressions the archetype should now take the form of an object, a technological construction, in order to avoid the odiousness of mythological personification" (Jung, 1959, pp. 22–23). From this perspective, the psychology of religion is far from irrelevant to the study of astrobiology, at least as far as its motivation and the possible resultant distortion of its speculations are concerned.

At a sociological level, comparable questions arise. Ever since the publication of Thomas Kuhn's *The Structure of Scientific Revolutions* (1962), philosophers of science have been aware that there are aspects of the practice of scientific communities that are better understood in terms of sociology than in terms of some kind of straightforward logical methodology. Few philosophers of science would now, perhaps, accept the way in which the very rationality of science was sometimes challenged in the immediate aftermath of Kuhn's analysis; nevertheless, many of them would still recognize the validity of aspects of his thinking, especially in fields of enquiry in which theoretical speculation is (as in astrobiology) strikingly underdetermined by data or involves a greater degree of interdisciplinarity than is usual in scientific work.

Indeed, sociologists of religion – who are well aware of the way in which a community's "rational" discourse can sometimes be based on little more than the shifting sands of bold speculation or wishful thinking – might well wonder whether there is occasionally, in parts of the astrobiological community, a kind of discourse comparable to that which sociologists of knowledge have observed in other intellectual communities, and analysed in terms of Alfred Schutz's concept of "recipe knowledge." In such communities, they have noted, it is quite possible for discussion to require little more than that "the individual must be able to reproduce a small number of stock phrases and interpretative schemes, to apply them in "analysis" or "criticism" of new things that come up for discussion, and thereby to authenticate his participation in what has been collectively defined as reality in these circles." As these commentators go on to note, statistically speaking, "the scientific validity of this intellectuals' "recipe knowledge" is roughly random" (Berger *et al.*, 1974, p. 12).

On top of these issues, the theologian can hardly help be aware of the average scientist's lack of historical and philosophical awareness, if only because those scientists who are in the forefront of the so-called "new atheism" are so frequently illiterate in these areas. While historians of science have long since abandoned the simplistic myth of an intrinsic and continuous conflict between science and religious belief, this myth still has a strong influence among members of the scientific community and the wider public (Numbers, 2009). This influence is evident among astrobiologists who still, for example, often ignore complex historical reality in favour of the myth when speaking of events such as the execution of Giordano Bruno.

Moreover, philosophical ignorance among scientific advocates of the new atheism is endemic, and in this they seem to be typical of the wider scientific community. When the philosopher of biology, Michael Ruse (2009), comments that it makes him "ashamed to be an atheist" when he reads Richard Dawkins's atheistic attacks on religion, he is not, perhaps, referring only to Dawkins's naïve philosophical assertions, which have parallels in the kind of logical positivism that few philosophers have tried to defend since the radical reassessment of the philosophy of science inaugurated by Karl Popper and Thomas Kuhn in the mid twentieth century. Nevertheless, part of Ruse's criticism of Dawkins clearly relates to the way in which debates that have occurred within the philosophy of science over the past half century (Knight, 2001, pp. 47–51) seem to have had little or no impact on Dawkins and on many of his scientific colleagues, whose views manifest a very poor understanding of what constitutes meaningful enquiry and of what the metaphysical foundations and implications of the scientific enterprise might be.

Put simply, scientists are wrong if – as they often do – they make the "scientistic" claim that they work simply from observation and without metaphysics. Astrobiologists need to be particularly aware of this error, since their own metaphysical foundation is sometimes far more questionable than that of the average working scientist (which latter has the virtue of being defensible provided that the limitations of the scientific understanding are acknowledged.) As Michael J. Crowe (1986, p. 555) has commented, the point to be stressed in astrobiology "is not that metaphysical ideas must not enter the debate but that authors recognize and candidly admit their presence, rather than claiming that the issues can be analysed solely in scientific categories."

Theological issues

When all these psychological, sociological, and philosophical considerations are taken into account, where does the theologian stand in relation to the astrobiological enterprise? In practice, this will depend very much on the particular theological tradition being defended, and even those defending very similar theological positions may find themselves in violent disagreement.

One might perhaps expect there to have been a gradual acceptance of astrobiological insights by theologians over time, but in fact, if we look at the period from the birth of modern science up to the present time, we can find adherents of opposing viewpoints in every generation. For John Ray – one of the great scientists and Christian apologists of the early period – it was evident that "Every fix'd star … is a Sun or Sun-like Body, and in like manner incircled with a Chorus of Planets moving about it", with each of these planets "in all likelihood furnished with as great variety of corporeal Creatures, animate and inanimate, as the Earth is …" (Ray, 1704, pp. 18–19). Although he himself drew back from speculating about extraterrestrial intelligence, others of his generation were less cautious, and throughout the period of modern science there have been religiously-inclined scientists happy to speculate about, and even insist on, the high probability of, extraterrestrial life and intelligence.

By contrast, theological positions that stress the cosmic implications of events on Earth have for centuries manifested an opposition to this view. Still, for a theologian like Brian Hebblethwaite (1987, p. 168), it is "an implication of the Christian incarnation, properly understood, that there are no other intelligent, personal creatures in God's creation than human beings on earth."[1]

Arguments about extraterrestrial life can, in fact, be found in all generations of religious believers from the birth of modern science to the present day, and when these arguments are studied in detail, historians can find no straightforward correlations between particular theological and astrobiological views. When Michael J. Crowe, for example, studied in depth the arguments of the period from 1750 to 1900, he was struck by the way in which one of most remarkable features of the conflict within religious groups was

> the degree to which it may be characterized as a night fight in which participants could not distinguish friend from foe until close combat commenced. Allies in a dozen conflicts, authors agreeing on a hundred issues, disagreed on extraterrestrial life. Anglicans argued against Anglicans, Catholics against Catholics, materialists against materialists

> […] The fact that pairs as dissimilar as Hume and Wesley, Paine and Whewell, Maunder and Wallace agreed on fundamental issues in the debate, whereas pairs as close in overall philosophies as Whewell and Sedgwick, Buchner and Strauss, Poble and Searle opposed each other, must make historians cautious of claiming correlations (Crowe, 1986, p. 558).

Theology and naturalism

It should perhaps be noted, however, that since the second half of the nineteenth century, the proportion of mainstream Christian theologians who have stressed human uniqueness, in the way that Hebblethwaite still does, has undoubtedly fallen. One of the reasons for this is that there has arisen a new stress on natural processes as the locus of divine action – a position not unknown in the medieval period but largely forgotten until the post-Darwinian controversies.

An early example of this stress is the defence of Darwinian evolution mounted by the Anglican priest, Aubrey Moore, who – well over a century ago – argued that Darwin's view was "infinitely more Christian" than the notion of creation as a series of "special" divine acts. The Darwinian view, he argued, "implies the immanence of God in nature, and the omnipresence of his creative power." Those, he went on, "who oppose the doctrine of evolution in defence of a "continued intervention" of God seem to have failed to notice that a theory of occasional intervention implies as its correlative a theory of ordinary absence" (Moore, 1889, p. 184).

At a theological level, the dialogue between science and theology of the last few decades has been based, in large part, on the kind of view of divine immanence that Moore stressed. Over the last forty years, in particular, this dialogue has been dominated by the work of three "scientist-theologians" – Ian Barbour, Arthur Peacocke and John Polkinghorne – who have argued for the consonance of scientific understanding and religious belief, in part through the stress on natural processes that Moore advocated. Although they offer slightly different perspectives (Polkinghorne, 1996; Barbour, 2012), all three agree with the statements of Peacocke (1993) that "it is chance operating within a lawlike framework that is the basis of the inherent creativity of the natural order, its ability to generate new forms, patters and organizations of matter and energy" (p. 65) so that "to a theist, it is now clear that God creates the world through what we call "chance" operating within the created order, each stage of which constitutes the launching pad for the next" (p. 119).

It is noteworthy, however, that all three of these scientist-theologians have retained a theological view which – while avoiding the notion of the laws of nature being temporarily suspended in "supernatural" divine acts – still assumes that there is sometimes a kind of divine interference with the world (see Knight, 2007, pp. 22–27). Here, the nondeterministic universe revealed by quantum mechanics has led them to speculate about whether quantum phenomena or other features of modern science allow scientifically literate speculation about a "causal joint" which allows God to affect the world "in, with, and under" the laws of nature. The point here is that all three – together with younger followers like Philip Clayton and Robert John Russell – have attempted to reconfigure, rather than simply abandon, the old theological distinction between events due to "special" and "general" modes of divine action (the former being the technical term for God's active "response" to events in the world, the latter referring to naturalistic processes which, from the very beginning, God has "built into" that world through its benign design).

This "causal joint" approach to divine action has, it should be noted, come in for a certain amount of theological criticism in recent years, with Nicholas Saunders going as far as to see it as so lacking in coherence that "contemporary theology is in a crisis." (Saunders, 2002, p. 215). It remains, nevertheless, the predominant view among scientifically-informed theologians, and it has the implication that naturalistically-based statistical arguments cannot be used straightforwardly to indicate whether life or self-conscious intelligence have developed elsewhere in the universe. At least in principle, according to this model, God might have interfered with the probabilities of events to bring about or prevent some particular event during the history of the cosmos's formation.[2] For a theistic astrobiology, this model has the implication that God might, on different potentially inhabitable planets, have acted in such a way as to enhance or diminish the probability of emergence of life or of intelligent, self-conscious life, so as to bring about his desired number of planets inhabited by intelligent beings (whatever that number might be judged to be.) Thus for this model, whatever our estimate of the purely naturalistic probability of such emergence might be, another – more purely theological – factor must also be taken into account, even though in practice there seems to be no consensus among theologians on how this should be assessed.

Convergent evolution and a new approach to divine action

This is not the only kind of theistic astrobiology that is possible, however. Part of the motivation for this theological approach – at least among some of its supporters – seems to be that it gets over the problem of the unpredictability of biological evolution proclaimed by people like George Gaylord Simpson (1964) and popularized more recently by Stephen Jay Gould (1989). Here, however, recent thinking on convergent evolution by people like Simon Conway Morris (2003) suggests that in practice, once life has begun, something essentially humanoid is a likely outcome. In this sense, a fully naturalistic theism seems consonant with scientific perspectives, in that the theistic naturalist can perceive God as creating a world that can subsequently "make itself" naturalistically so as to bring about an intended end: the emergence (perhaps in many parts of the universe) of creatures "in the image and likeness of God." If Morris is right in his biological speculations, therefore, this part of the motivation for the causal joint model of divine action is no longer valid.

There is, however, a further reason for people defending this "causal joint" account of divine action, which is based on a particular, somewhat anthropomorphic view of what it means to speak about God. This view – which stresses the parallels between divine and human action – fails to appreciate, in several respects, either the traditional, medieval philosophical theology of the Christian community or the more subtle and profound aspects of modern philosophical theology.

Wesley Wildman, commenting on one scholar's application of this view to questions about divine action, has seen it as the product of a "personalistic theism … of the distinctively modern kind … a distinctively Protestant deviation from the mainstream Christian view." It is a view, says Wildman, in which God is seen as "a being among beings, whose action in the world is properly subject to the quest for its mechanism, the causal joint that links the divine intentions to the created world." What happened, he goes on to ask rhetorically, "to the classical doctrines of aseity and immutability, the affirmations that God is self-contained and does not change through acting or feeling? What happened to God as the ground of being or being itself, as pure act and first cause? How [do the upholders of this view] deflect the classical intuition that God as a being can be no God at all but merely an idol of the human imagination?" (Wildman, 2006, p. 166).

If this analysis of Wildman's is correct – and I believe it is – then we must recognise that if modern "scientific" atheism is usually philosophically illiterate, then much of the theistic reaction to it, at least in the so-called "dialogue between science and theology," also suffers from a lack of philosophical insight (albeit not at quite as crude a level).

Here, however, it is of interest that two recent accounts of divine action – both taking account of scientific perspectives – reflect something of Wildman's appeal to medieval traditions of philosophical theology, and in so doing offer an interesting way forward. Both, moreover, avoid the prevalent distinction between special and general modes of divine action.

One of these accounts, by Denis Edwards (2010), restates the Thomist view that speaks in terms of primary and secondary causes, though with an interesting twist (which separates it from the Western medieval tradition) in that it does not see "miraculous" events as events with no secondary cause. For Edwards, God – as primary cause – always works through the laws of nature, which are the secondary causes through which all events occur. The miraculous, for him, simply represents an aspect of the laws of nature that is not understood.

I myself have independently developed a very similar view through aspects of the Byzantine philosophical tradition, especially as set out in the seventh century work of Maximos the Confessor (Knight, 2007). Despite its very different starting point, convergence between this second account and that of Edwards' is evident.[3]

For astrobiology, these approaches have the important implication that the probability of extraterrestrial life – as perceived (or at least in principle perceivable) by the sciences – is precisely what God has willed for his creation. There is no extra factor to which theologians have some sort of privileged access. They provide a theoretical background, therefore, to a view some Christians have already developed, which allows them to speculate about other incarnations comparable to that which they see in the person of Jesus.[4] It also provides a background to a speculative extension to other planets of a more pluralistic theological view that I myself have developed in relation to humans on our own planet (Knight, 2010). In this account, I argue that human experience which is interpreted in terms of divine revelation or spiritual enlightenment will be a natural psychological outcome of the kind of evolution towards something essentially humanoid that Simon Conway Morris (2003) has insisted will occur in any biological system once life has begun.

Conclusion: Scientific caution and theological development

Such theological speculation may, however, be scientifically premature. As Simon Conway Morris (2003) has suggested, even if we believe (with him) that almost any planet with life would produce living creatures that we would recognize as parallel in form and function with our own biota, we should also recognise that as yet we have no idea how rare the emergence of life might be, so that we must admit the possibility that we are alone and unique in the cosmos. What is more, even if life is common in the cosmos, a comparable admission about the development of self-conscious intelligence arises both from Stephen Jay Gould's stress on contingency in evolution (which leads him to see human-type intelligence as a sort of freak accident) and from Richard Dawkins's observation that some adaptive features have appeared independently on earth many more times than have others. Dawkins sees this fact as indicating that there are "certain evolutionary pathways which life is "eager" to go down" while other pathways, as he puts it, "have more resistance [so that] evolution repeatedly races down the easy corridors, and just occasionally, and unexpectedly, leaps one of the hard barriers" (Dawkins, 2005, p. 604). As he notes a little later, "true syntactic language seems to be unique to one species, our own" (p. 610) – which seems to imply, from his earlier argument, that in any earth-like biological system, the kind of intelligence assumed by SETI research may have only a very small probability of evolving.

These, though, are simply factors that scientific analysis must take into account. What is important here is that theologians – while they may perceive a number of scientific reasons for scepticism about the more extreme claims about probability that are sometimes made within the astrobiological community – now seem to be moving in a direction in which hard scientific evidence for extraterrestrial life, and even of extraterrestrial intelligence, would deepen rather than challenge theological thought. Such a discovery would simply bring into clearer focus an existing set of questions in the ongoing theological enterprise, based on the theological interpretation of naturalistic explanation that since the time of Darwin has come increasingly to characterize that enterprise.

Notes

[1] One wonders here whether, if extraterrestrial intelligence were discovered, those like Hebblethwaite – who effectively "predict" that extraterrestrial intelligence does not exist – would reinterpret their previous views as "auxiliary hypotheses" which, when falsified, need have little effect on their "core theories." Certainly the view of theological development outlined by Nancey Murphy (1990), on the basis of Imre Lakatos's views on the development of science, suggests that this reinterpretation would be open to them.

[2] In practice, we find that Peacocke sees "special" action as something appropriate only to God's redemptive – as opposed to creative – action, but this division between creative and redemptive action is essentially arbitrary. Polkinghorne here is arguably more coherent, in that he clearly affirms the possibility that God might, if this basic model is correct, "guide" the creative process through occasional "special" actions.

[3] At first sight, admittedly, these approaches may seem very different, in that Edwards seems to abolish the distinction between special and general divine action by subsuming the latter into the former, while I do so by subsuming the former into the latter, proclaiming what I call a “strong theistic naturalism.” A more careful reading of the two approaches indicates, however, that the two views are very close. In both, the laws of nature are not something that God must sometimes set aside or manipulate to bring about his will but they are, so to speak, nothing other than his action itself.

[4] It is significant here, perhaps, that some Christians have already started using a hymn that asks: “Who can tell what other cradle / high above the milky way / still may rock the King of Heaven / on another Christmas day?”

References

Barbour, I. G. 2012. “John Polkinghorne on three scientist-theologians.” In *God and the Scientist: Exploring the Work of John Polkinghorne*, edited by F. Watts, and C. C. Knight. Farnham: Ashgate.

Berger, P., B. Berger, and H. Kellener. 1974. *The Homeless Mind: Modernization and Consciousness*. Harmondsworth: Penguin.

Crowe, M. J. 1986. *The Extraterrestrial Life Debate 1750–1900: The Idea of a Plurality of Worlds from Kant to Lowell*. Cambridge: Cambridge University Press.

Dawkins, R. 2005. *The Ancestors’ Tale: A Pilgrimage to the Dawn of Life*. London: Phoenix.

Edwards, D. 2010. *How God Acts: Creation, Redemption and Special Divine Action*. Minneapolis MN: Fortress.

Gould, S. J. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: W. W. Norton.

Hebblethwaite, B. 1987. *The Incarnation: Collected Essays in Christology*. Cambridge: Cambridge University Press.

Jung, C. G. 1959. *Flying Saucers: A Myth of Things Seen in the Sky*. London: Routledge and Kegan Paul.

Knight, C. C. 2001. *Wrestling With the Divine: Religion, Science, and Revelation*. Minneapolis MN: Fortress

—. 2007. *The God of Nature: Incarnation and Contemporary Science*. Minneapolis MN: Fortress.

—. 2010. “*Homo religiosus*: A theological proposal for a scientific and pluralistic age.” In *Human Identity at the Intersection of Science, Technology and Religion*, edited by N. Murphy, and C. C. Knight. Farnham: Ashgate.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago IL: University of Chicago Press.

Morris, S. C. 2003. *Life's Solution: Inevitable Humans in a Lonely Universe*. Cambridge: Cambridge University Press.
Murphy, N. 1990. *Theology in an Age of Scientific Reasoning*. Ithaca NY: Cornell University Press.
Numbers, R. L., ed. 2009. *Galileo Goes to Jail and Other Myths about Science and Religion*. Cambridge MA: Harvard University Press.
Polkinghorne, J. 1996. *Scientists as Theologians: A Comparison of the Writings of Ian Barbour, Arthur Peacocke and John Polkinghorne*. London: SPCK.
Ray, J. 1704. *The Wisdom of God in the Works of Creation*, fourth edition.
Ruse, M. 2009. "Why I think the new atheists are a bloody disaster." http://blog.beliefnet.com/scienceandthesacred/2009/08/why-i-think-the-new-atheists-are-a-bloody-disaster.html
Saunders, N. 2002. *Divine Action and Modern Science*. Cambridge: Cambridge University Press.
Simpson, G. G. 1964. "The non-prevalence of humanoids." *Science* 143:769–775 (reprinted in shorter form as chapter 13 of his *This View of Life: The World of an Evolutionist*. New York: Harcourt, Brace and World, 1964.)
Tipler, F. 1981. "Additional remarks on extraterrestrial intelligence." *Royal Astronomical Society Quarterly Journal* 22:279–288.
Wildman, W. J. 2006. "Robert John Russell's theology of God's action." In *God's Action in Nature's World: Essays in Honour of Robert John Russell*, edited by T. Peters, and N. Hallanger. Aldershot: Ashgate.

Chapter Fourteen

Some Philosophical and Theological Implications of Modern Astrobiology

Ludwik Kostro

It is obvious that modern astrobiology, being essentially concerned with life, consciousness and intelligence as cosmic phenomena, implies new philosophical and theological ideas. We can even say that astrobiology will trigger new opinions concerning the whole world. These ideas can produce a profound revolution in people's outlook on the whole universe. The authorities of various churches are aware of the problem and are beginning to make their theologies more cosmically informed.

Preliminary historical data concerning human images of the universe

Mankind has changed its mythological, theological, and philosophical image of the universe many times. In Biblical times in Canaan, the image was similar to those created by the Sumerian and Egyptian civilizations, which believed that all things, including gods, originated from immense, eternal waters, themselves constituting a divine entity. This deity was called Mammu by the Sumerians, who regarded it as the Mother of All Things, and Nun (a male deity) by the Egyptians. The divine waters constitute a chaotic, cosmic soup. The primary god, who emerges from the waters, puts everything in order, divides and separates the waters into the upper (heaven), ambient (enclosing Earth as a drifting disc), and lower waters. The whole universe arises from that abyss of waters. The primary god – personification of the forces that put everything in order – must fight the water dragon as the personification of chaos. The house of the primary god is at the top of the universe in the midst of the upper waters; while the world of the dead and departed is deep in the underworld. A similar world-image is to be found in the Bible written in Canaan, which was situated between Mesopotamia and Egypt (Smith, 2001; 2002; Cross, 1997; Armstrong, 1993). The Biblical God – named El, El Elyon, Elohim, and later Yahweh – divides the waters into the upper, ambient, and so on: see the first chapter of Genesis. The Hebrew word *bara*, meaning "to separate", is often also translated as "to create" (see e.g. Grande Lessico, 1988), even though many Biblical scholars stress that it should not be taken to mean "creating out of nothing". Instead, it expresses the divine struggle with the waters and the victory over chaos. By their judgment, the opening line of the Bible – "in the beginning God created the heavens and the earth" – is an erroneous translation made under the influence of a much later doctrine of creation out of nothing. The Hebrew indicates only the beginning of the process of separation. That is why many modern translations, especially the so-called ecumenical ones, read: "when God began to create (= to divide, to separate) the heavens and the earth" ("lorsque Dieu commença la crèation du ciel et de la terre…" in French; "quando Dio comició a crear il cielo e la terra…" in Italian). The erroneous translation contradicts the rest of Genesis 1 and 2 and other texts in the Bible. Genesis 2:1 and Exodus 20:11 suggest that the creation of the heavens and the Earth was not a point in time but a process of dividing up the cosmic soup over a period of six days.

True, in the first chapter of Genesis, the celestial bodies are not considered gods, and the water dragon is not mentioned; but elsewhere in the Bible, God fights this primeval sea monster that represents the forces of chaos: sometimes called Leviathan, sometimes Rahab or simply Dragon (see e.g. Job 9:13, 26:12, Psalms 89:10, Isaiah 51:9). One finds Biblical passages in which the celestial bodies are divine beings; they are Yahweh's hosts: e.g., in Deuteronomy 4:19–20, Yahweh does not permit the Israelites to worship the hosts of heaven – sun, moon, stars, and other celestial bodies – because He has allotted them to others to worship. Yahweh's house is high at the top of the upper waters, while the world of departed, Sheol, is underground.

Elements of the Old Testament image of the universe remain in the New Testament: e.g., the idea that "by God's word the heavens existed and the earth was formed out of water and by water" (2 Peter 3:5). At the same time, the image continued to evolve.

In the Medieval image of the universe, the Earth is at the centre. The house of God remains at the top: see e.g. the picture by Hartmann Schedel *Liber Cronicarum*, Nuremberg 1493; Hell remains below. According to Medieval catechesis, volcanoes prove the existence of Hell.

The Copernican revolution and its philosophical and theological repercussions

It was only when Copernicus discovered that the Earth was not the centre of the universe that the ancient image was overturned. The Copernican principle states that the Earth holds no central, specially favoured position. This developed into the mediocrity principle (our planet and solar system are no greater nor lesser than other planets and systems) and, finally, in modern cosmology, into the principle of homogeneity and the cosmological principle: each indicating different aspects of the same principle.

The Copernican revolution radically changed theological perspectives as well. In Christian theology, Heaven (Paradise) with its God, angels, and saints and Hell with its Satan, demons, and condemned (and, for the Catholics, Purgatory as well[1]) had to take up a new location, beyond our universe, in other dimensions.

Philosophical arguments for the cosmic nature of life, consciousness, and intelligence

Because of developments in astronomy and astrophysics, many scientists are deeply convinced that life, consciousness, and intelligence are all cosmic phenomena, existing not only on Earth but on many Earth-like planets. Astrobiology has become an accepted branch of science.

Modern science has come to the conclusion that nature is homogeneous: one of the results of the Copernican revolution. Before Copernicus presented his new vision of the solar system, a majority of scholars were convinced that the universe was divided into two parts: the perfect super-lunar world and the imperfect sub-lunar world. When, using his telescope, Galileo Galilei discovered spots on the surface of the sun, scholars gradually came to the conclusion that the super-lunar world is imperfect as well. Thus, step by step, the principle of the homogeneity of nature was born. In modern philosophy of science and in logic, this principle plays a critical role, implicitly constituting the basis of modern science: “implicitly” because many scientists are not aware of its influence. Philosophy tries to explicate its role and investigate its epistemological status.

The homogeneity principle states that *the same physical phenomena given the same circumstances run the same way*. In other words, what the laws of nature as described by physics, chemistry, biology, and other sciences admit necessarily happens, under favourable conditions.

The homogeneity principle found its mathematically idealized expression in physics thanks to Emmy Noether (see e.g. Jarczyk, 2010; Chown, 2007). She demonstrated in a precise, mathematical way that the laws of nature are invariant with respect to displacement in space: the same physical laws that govern Warsaw govern Rome. She showed how this principle is connected with the law of conservation of momentum. She further demonstrated that the laws of nature are invariant with respect to displacement in time. The same laws were valid a hundred years ago, are valid now, and will be valid one hundred and more years into the future. She showed how this further principle is connected with the law of conservation of energy. Existing energy can neither increase nor decrease with passage of time. Finally, she demonstrated that the laws of nature are invariant with respect to rotations and showed how this connects with the law of conservation of angular momentum.

In brief, the invariance of the laws of nature with respect to displacement in space and time and to rotation safeguards the conservation of momentum, energy, and angular momentum. That said, Noether's theorems remain mathematical idealisations whose target is the laws of classical physics. Richard Feymann expressed his doubts whether the homogeneity principle applies to quantum mechanics. As the Heisenberg Uncertainty Principle states, (1) uncertainty of position in space relates to uncertainty of momentum, (2) uncertainty of time interval relates to uncertainty of energy, and (3) the uncertainty of rotational angle relates to uncertainty of angular momentum. However, in *Principles of Quantum Mechanics* Ramamurti Shankar has shown, that also the laws of quantum mechanics are invariant with respect to displacement in space and time and to rotation (Shankar, 2004). Thus quantum mechanics is meant to be valid throughout the universe spatially, temporally, and rotationally: ubiquitously. Its laws are respected in all the basic branches of science, from quantum chemistry to molecular biology and far beyond. It became clear that the homogeneity principle is really universal, it is universal for quantum phenomena as well. At the same time, it suggests that the laws of nature are not so deterministically invariant as Emmy Noether believed. Modern science allows the possibility that laws of nature and so-called universal constants *can* evolve on the basis of other reasons as well.

If, on an exo-Earth, favourable conditions are present, then life and consciousness will appear. The requisite key events must be indicated, studied, and incorporated into the homogeneity principle. They constitute the *conditiones sine qua non*. For many philosophers of science, the homogeneity principle must be considered one of the foundations of modern science. It provides the grounds for affirming that life and consciousness are truly *cosmic* phenomena. We can even say that the appearance of life and consciousness, given the required conditions and key events, is a law of nature. Of course, the origins of life and consciousness pose a much more complicated question than a simple matter of physics or chemistry.

Consider two simple experiments: one from physics, one from chemistry. When one sends electrical current (the flow of electricity) in the same direction through two parallel wires, an electromagnetic repulsive force appears between them. However, if one sends the flows of electricity in opposite directions, an attractive force appears instead. When an unbound piece of potassium or sodium comes into contact with water, fire (violet for potassium, yellow for sodium) and great heat result. Should one repeat these experiments in different places and times, under identical conditions, one can assume *with certainty* that they will run the same way. A countless number of experiments over the last three centuries have convinced the scientific community of the merits of the homogeneity principle.

Electromagnetic and chemical processes are central to all living beings. The biological evolution is simply a new kind of chemical evolution, and organisms are sophisticated chemical laboratories. Gerald Francis Joyce writes (1994), "life is a self-sustained chemical system capable of undergoing Darwinian evolution". The evolution of consciousness is likewise connected with electromagnetic and chemical processes. We know that consciousness is intimately connected with neuro-chemical processes in the brains of human beings and other animals. Anaesthetists know which chemical substances to apply to induce unconsciousness and which will induce a return to consciousness.

The epistemological status of the homogeneity principle

Italy's Professor Philippo Selvaggi, who examined my doctoral thesis, was one of the prime defenders of the principle of homogeneity in nature. In *Filosofia delle Scienze* (1953), Selvaggi called it the *fundamental principle of induction* and expressed its different aspects: (1) *the principle of physical causality*, by which the same causes in the same circumstances produce the same effects; (2) *the principle of the constancy of the laws of nature*, by which their corporal nature is determined to one and operates always in the same way, obstacles notwithstanding; (3) *the principle of physical determinism*, by which, once one knows the state of a system and the laws that regulate it, it is always possible to foresee future events produced within the system. In this way, Selvaggi's formulation of the homogeneity principle reduces it to deterministic phenomena concerning also the statistic and probability. At the same time, the ubiquity of *indeterministic* quantum mechanics' own laws shows that they, too, are valid throughout spacetime and cover the rotational phenomena. In my opinion, the Polish logician Leopold Regner's own formulation of the homogeneity principle (1973) is, in this way, better than Selvaggi's, because Selvaggi's cannot be applied to all phenomena. Selvaggi was aware of this, affirming that his *fundamental principle of induction* cannot be considered either the most general physical law or a simple postulate; therefore, he looks for a more general formulation according to which "it constitutes an analytical principle that is obvious by itself and […] is an explicitness of the metaphysical principles of causality and sufficient reason applied to the corporal nature" (Selvaggi, 1953).

Let me now present Leopold Regner's own formulation of the homogeneity principle. According to him (1973):

> The conviction that what happens in determined circumstances W, will happen again exactly and unfailingly everywhere and always where and when there will be circumstances totally similar to W, is based on the recognition of *the principle of homogeneity of nature*. The homogeneity of nature consists on this that the course of a phenomenon does not depend on the circumstances of place (where) and time (when). […] The principle of the homogeneity of nature is not a kind of a major premise that is present in an implicit way in every inductive inference but it is something that can be called *preliminary assumption* (praeambula) of induction. The principle of the homogeneity of nature is not a proved affirmation but it is a certain kind of *postulate* or a presupposition about the properties of the Universe.

In my opinion, the homogeneity principle constitutes a major implicit premise in every inductive inference in scientific practise going back almost four centuries. The results obtained so far have convinced the scientific community that nature must be considered homogenous, and this homogeneity permits human reasoning in the natural sciences to be nearly unfailing. The homogeneity principle provides good – perhaps unassailable – grounds for the conclusion that life, consciousness, and intelligence are all cosmic phenomena. I am convinced that future observations will prove the truth of this conclusion.

The laws of large numbers

Philosophers of science often emphasize the role of the laws regarding large numbers as they apply to the natural sciences. On their basis, we are able to discover regularities in phenomena involving a very large number of elements: e.g., the so-called *half life* of free neutrons. For any large number of free neutrons, half will decay into protons, electrons, and electron anti-neutrinos within ten to fifteen minutes. The laws of large numbers further enable us to calculate the frequency of a large number of events recurring: When we have e.g. a great population of human beings then every 85th human pregnancy, in such a population, will result in twins. Attempts have been made to calculate the frequency of Earth-like planets in the Milky Way. According to the latest estimates by e.g. Jan Skowron and Joseph Catazarite, presented at the 2011 Sagan Workshop at the California Institute of Technology in Pasadena, between one and three percent of Sun-like stars have Earth-like planets.

Nobel Prize winner Christian de Duve has shown that, in evolutionary processes, the great number of trials or opportunities provided plays a decisive role. He observes that chance does not exclude inevitability nor probability exclude certainty. To prove this, he uses the equation:

$$P_n = 1 - (1 - P)^n$$

Where n = number of trials/opportunities, P = probability (n = 1), and P_n = probability after n trials. When the number of trials/opportunities is large enough, P_n approaches 1, and the appearance and evolution of life becomes inevitable. De Duve concludes that life is a cosmic imperative and counters Einstein's statement that "God does not play dice with the world" with "God plays dice with the universe, because he is sure to win".

The observational evidence for extraterrestrial life, consciousness, and intelligence remains to be found. At the same time, the history of science has shown that the human mind reaches truths earlier than does its senses – sometimes many centuries before observational evidence arrives. Twenty years ago, we had no observational evidence of exoplanets. Now we are awash in indirect and direct observational evidence that our galaxy is a great producer of exoplanets. So far, we have detected around 900 certain exoplanets and 3,500 likely candidates. Although our civilization recently became technological, our observational technologies have evolved rapidly, becoming more and more precise and effective. I believe we are not so far from the observational evidence that life, consciousness, and intelligence are cosmic phenomena.

Attempts to make Christian theology more cosmically informed

Various theological arguments – by theologians and scientists alike – have been provided in affirmation that life, consciousness, and intelligence are cosmic phenomena. Some claim that, were the universe deprived of life and consciousness, God's act of creation could not be regarded as the granting of sensible existence but only a toy for God to play with, totally deprived of sense. Only beings with sensibility and consciousness can experience existence as a gift. Beings deprived of sensibility and consciousness are beings for others and never for themselves. If, in the whole universe, only Earth possessed sensible and conscious existence, the rest of the enormously great universe would be but the Creator's toy. In *God's Universe* (Gingerich, 2006), Owen Gingerich quotes the Archbishop of Paris Etienne Tempier who, in 1277, on Pope's John XXI request, enumerated the contemporary heresies: among them, "who affirms that God created life and intelligent beings only on Earth is a heretic, because he limits the creative power of God. God can create life and intelligence where he wants."[2] In *Jesus* (Guitton, 1956), Jean Guitton claims that, were life to exist only on Earth, the Universe would be a plinth too great for the statue: "*Le socle est trop grand pour la statue.*"

At the same time, modern astrobiology poses problems for Christian theology, which is often still geocentric, if not provincial. According to Christian doctrine, Heaven is governed by earthly representatives: Jesus of Nazareth sits on the right hand of God the Father; His mother Mary is considered by Catholicism to be the queen of Heaven; and Saint Peter holds the keys to the Kingdom of God. All are human beings. If our galaxy and others possess habitable planets, then Christian theology must change, taking into account the existence of exo-civilizations.

The Holy See is aware of the problem. In response, Rome encourages the creation of new theological ideas. In an attempt to make their theology more cosmically informed, some theologians introduce the notion of multiple incarnation or return to the ancient adoptive form of Christology.

The notion of multiple incarnation was first proposed by the recognized Jesuit, scientist, and philosopher Pierre Teilhard de Chardin in *God, Church and the Extraterrestrials* (Vigne, 2000). The Polish cosmologist and theologian and Templeton Prize winner Professor Michał Heller (2008) likewise is of the opinion that the Christian notion of incarnation is too absolutized, too focused on uniqueness. The idea of multiple incarnation is that the same person, whom we know as the Son of

God incarnated as Jesus of Nazareth, is incarnated on other worlds in other intelligent beings.

According to adoptive Christology, instead of one divine person incarnated in many forms one has an "adoption" of intelligent beings on every Earth-like planet and so many sons of God. For a long time, adoptive Christology was considered heresy; but now it is accepted even by such prominent Catholic scholars as Alfons Weiser (Weiser, 1974).

The Vatican has organized two conferences on life and intelligence in the universe, in 2005 and 2009. The second conference (6–11 September) welcomed scholars from many prominent scientific institutions and was organized by Pope Benedict XVI. Among the attendees was David Charbonneau, engaged in the research on transiting planets, who debriefed this author on the proceedings (Impey, Lunine, and Funes, 2012).

Finally, I would like to call readers' attention to an interview in the Vatican newspaper *Osservatore Romano* with the director of the Vatican Astronomical Observatory Jose Gabriel Funes under the headline "The extraterrestrial is my brother" (Funes, 2008). Clearly, Father Funes is serious about taking into account the existence of habitable exoplanets with intelligent life.

Conclusions

I have claimed that astrobiology, being concerned with life and consciousness as potentially cosmic phenomena, implies philosophical and theological ideas that can trigger a profound revolution in people's outlook on their world. On the basis of scientific, philosophical, and theological reasoning, we can have the courage to claim that the universe had a second Big Bang: the explosion of life and consciousness everywhere that the required conditions are met. Although we lack concrete observational evidence, nevertheless, the appearance and existence of life, consciousness, and intelligence can be already considered a law of nature. Given the rapid development in our observational technologies, we can hope that, in the coming decades, we will find the looked-for evidence and come to see life, consciousness, and intelligence as common cosmic phenomena. Only then will we be able to state with certainty that the universe is not only physical and chemical but also biological.

Notes

[1] The notion of Purgatory was introduced by Pope Innocent IV on 6 March 1254 in his *Sub Catholicae* (*Encyklopedia Katolicka*, 1979). As such, it remains outside the

belief systems of Orthodox Christians and Protestants (*Enciclopedia del Cristianesimo*, 1997).
[2] Gingerich bases his information on data presented by the work of Stephen J. Dick (1982).

References

Armstrong, K. 1993. *A History of God: The 4000-Year Quest of Judaism, Christianity and Islam.* New York: A. A. Knopf.

Chown, M. 2007. *The Never-Ending Days of Being Dead: Dispatches from the Front of Science*. London: Faber and Faber Limited.

Cross, F. M. 1997. *Canaanite Myth and Hebrew Epic: Essays in the History of Religion of Israel*. Cambridge MA: Harvard University Press.

Dick, S. J. 1982. *Plurality of Worlds: The Origins of the Extraterrestrial Life Debate from Democritus to Kant.* Cambridge: Cambridge University Press.

Encyklopedia Katolicka. Vol. 3. 1979. Edited by R. Lukaszyk, L. Bienkowski, and F. Gryglewicz. Lublin: KUL. (See the entry "czyściec," p. 940.)

Enciclopedia del Cristianesimo. 1997. Edited by G. Ambrosio, with forword by J. Ratzinger. Novara: Istituto De Agostini. (See the entry "purgatorio.")

Funes, J. G. 2008. "The extraterrestrial is my brother." *Osservatore Romano*, Friday, May 23. http://padrefunes.blogspot.com/

Gingerich, O. 2006. *God's Universe*. Cambridge MA: Harvard University Press.

Grande Lessico dell' Antico Testamento. Vol. 1. 1988. Edited by G. J. Botterweg, and H. Ringgren. Brescia: Paideia. (See entry "bara," p. 1572.)

Guitton, J. 1956. *Jėsus*. Paris: Bernard Grasset, p. 295.

Heller, M. 2008. *Tygodnik Powszechny* 23(3074), 8 June 2008.

Impey, C., J. Lunine, and J. Funes, ed. 2012. *Frontiers of Astrobiology*. Cambridge: Cambridge University Press.

Jarczyk, L. 2010. *Wczesny rozwoj Wszechswiata: Model Wielkiego Wybuchu – Big Bang*. Warsaw: Wyd. Naukowo-Techniczne.

Joyce, G. F.1994. *The RNA World: Life before DNA and Protein.* http://en.wikipedia.org/wiki/Life#cite_note-22

Kasting, J. 2010. *How to Find a Habitable Planet*. Princeton NJ: Princeton University Press.

Regner, L. 1973. *Logika*. Cracow: PTT.

Selvaggi, F. 1953. *Filosofia delle scienze*. Rome: Edizioni della Civiltá Cattolica.

Shankar, R. 2004. *Principles of Quantum Mechanics*. Dordrecht: Kluwer Academic Publishers.

Smith, M. S. 2001. *The Origins of Biblical Monotheism: Israel's Polytheistic Background and the Ugaritic Texts*. Oxford: Oxford University Press.

—. 2002. *The Early History of God: Yahweh and Other Deities in Ancient Israel*. Dearborn MI: W. B. Eerdmans Publishing Company.

Vigne, A., ed. 2000. *Dieu, l'Église et les extraterrestres*. Paris: Albin Michel.

Weiser, A. 1974. *Jesus – Gottes Sohn? Antwort auf eine Herausforderung*. Stuttgart: KBW Verlag.

CHAPTER FIFTEEN

ASTROBIOLOGY: THE NEXT REVOLUTION?

JACQUES ARNOULD

Introduction: The street lamp, the drunk, and his keys

Scientific research is always a search in the dark. So it is not surprising that researchers like to use the metaphor of looking for lost keys in the night. Without the beam of a torch or the glow of a lighter to help find the keys, the person looking will have no choice but to start his search beneath the halo of a street lamp, where he has a chance of finding his belongings. This amusing scene is sometimes taken a step further, with the unlucky or careless loser of keys being inebriated: to avoid falling down, he has to keep a firm hold on the lamp post at all times, which of course means the area he can search is even more limited.

Astrobiology is the youngest of scientific disciplines: a bridge between the astronomic and biological sciences. It has seen a surprising yet, in hindsight, foreseeable alliance between the quest for exoplanets, the search for extraterrestrial life, and the study of life's appearance on Earth. It is the heir to age-old questions about the plurality of inhabited worlds. As with any scientific approach, astrobiology requires that the right method and the best conditions be chosen and that the best-suited instrument be built for reaching the defined goal or expected target; but these are not the only challenges it faces. Insofar as possible, it must also avoid the pitfalls illustrated by the amusing image of the drunk looking for his keys by the light of the street lamp: reductionism and dogmatism. It must never forget the immeasurable nature of what it is studying: life itself. The epistemological tools at its disposal – analogy, wonder, and the idea of a Second Genesis– must be aptly and soundly used.

The two pitfalls for astrobiology

In contemporary scientific research, reductionism seems to be requisite. The complexity of reality, the growing diversity in fields of study, and advances in technology are the most obvious reasons and the most acceptable excuses. At the same time, the spirit of research is changing. “Life, now, is more often studied in laboratories,” observes the French biologist and Nobel Prize winner François Jacob (1970, pp. 320–321). “We are no longer trying to define the contours. We simply try to analyse living systems, their structures, their functions and their history.” He does not deny the benefits of methodological reductionism, but he worries about the consequences – in the first place, philosophical; ultimately, scientific – of potential ideological drift. A living organism cannot be reduced to its mere genetic makeup, and the appearance of life forms on a planet cannot be compared to a synthetic experiment in a test tube. To paraphrase Oscar Wilde, life is far too important for reductionist methods alone.

Dogmatism springs from similar origins to reductionism: in particular, from an attitude of self-importance. Like theology and philosophy, scientific research resorts to dogmas – or, if one prefers, such *reference points* as theories and worldviews – that allow researchers to diverge from lighted paths, to step outside the bounds of established knowledge, to venture into dark seas and lands of ignorance – the *terrae incognitae* of explorers – without losing sight of familiar territory and established milestones. There is no scientific research without dogma; but without abandoning dogma, without stepping out of dogma’s bounds, without a break from charted territory – what Thomas Kuhn (1962) referred to as revolutions – there is no scientific discovery. Indeed, a dogmatic attitude is ultimately an obstacle to scientific development.

Founded on a long tradition of philosophical, theological, and scientific debate over the number of inhabited worlds and the question of extraterrestrial life, astrobiology is particularly susceptible to dogmatism: it is much easier to keep a tight grip on the lamp posts of ideologies, firmly anchored in history, that bear the names of the most distinguished thinkers. The focus of astrobiology is among the most complex things the human mind can fathom – life itself – making it all the more essential for freedom to be cultivated.

Life: Necessarily a Gordian knot

The goal here is not to provide a definition of life. Astrobiology has set itself the task of establishing the most accurate definition of life possible.

It must be said that, thus far, the astrobiology community has not reached any consensus on that definition. This will come as no surprise to any historian of science, ideas or philosophy: no civilization, no culture, no school of thinkers has succeeded in this task, and scientists have fared even more poorly than philosophers, poets… even judges! The wisest often content themselves with Saint Augustine's description of time, found in Book XI, chapter XIV, of *Confessions*: "What then is time? If no one asks me, I know what it is. If I wish to explain it to him who asks, I do not know" (Augustine, p. 195); or they take inspiration from René Dubos, who writes (2002, p. 68), "life absolutely cannot be understood as an abstract concept… it is only known as an experience." Both time and life are difficult to confine within the limits of a definition, since the mind that must compose and understand them itself directly experiences them in each moment of its existence. Life is experientially self-evident; it imposes itself on each human being and on human intelligence, knowledge, and science. On the possibility of discovering extraterrestrial life, David Grinspoon (2004, pp. 98–99) echoes Dubos: "perhaps finding extraterrestrial life will be more like falling in love than confirming a specific hypothesis. When it happens, we'll know."

The self-evident nature of life must not overshadow another difficulty, which Thomas Kuhn (1983, p. 176) points out: "the operations and measurements performed by the man of science in his laboratory are not 'given' by the experiment; rather, they are 'acquired with difficulty'. "This is particularly true in the quest for extraterrestrial life: one cannot simply wait for life to manifest itself. Astrobiology proper was born with the start of the space programme, which allowed the exploration of other celestial bodies by automated and manned spaceflights. Today, for all of the thrill of the discovery of exoplanets, Kuhn's observation remains: scientific experience is acquired with difficulty; it is never given.

Acquiring this experience is often laborious and always unfinished. To begin, one requires a vocabulary, a method, a vision, and a scientific paradigm. A scientific approach can never, on its own, create a language of observation. It always relies on what was previously treated as dogma: a well-known scientific theory, a common worldview, ordinary language. The pitfall of dogmatism is always lurking.

Astrobiology has a rich and varied philosophical heritage (see e.g. Dick, 1982; Crowe, 1999; Dick, 2004). Recent developments in astronomy and biology – such as the discovery of exoplanets and extremophiles respectively – generate great interest among the general public; and life in general is a popular topic of discussion, despite the antipathy created by e.g. certain genetic engineering projects. For members of the astrobiology community, all hopes are possible; but they must not forget that, even though life is imposed on experience, it is still a Gordian knot.

According to legend, Gordias was king of Phrygia. The shaft of his cart was fastened with a knot so complex that no one was able to untie it. An oracle predicted that the Asian empire would go to he who succeeded in undoing the knot. In 337 BC, Alexander the Great, determined to conquer Asia, came and sliced the knot with a stroke of his sword then conquered Asia – only to lose it almost immediately. The Gordian knot is probably only ever severed momentarily; it constantly re-forms itself.

Gordian or not, the knot is a beautiful and ancient symbol of life that has nothing linear to it – as chemistry and biology confirm. Rather, it is composed of loops, tangles, and junctions: a combination of imposed inevitability and unexpected complications. Barring capitulation, researchers into life often must sever it or do away with it to solve the mystery. Nevertheless, through twists and turns, life usually manages to evade the sagacity of scholars. Life is here, but it is also elsewhere.

Using analogy wisely

The history of science confirms that, along with an assumption of plenitude, the principle of analogy is the tool most used by those interested in the possibility of extraterrestrial life and the incidence of inhabited worlds. Meanwhile, many pluralist theologies are based on the assumption that only a single creation, at the same time multiple and complete in its diversity, could glorify the omnipotence of the Creator.

Scientific thinking has often called into question the enduring philosophical notions of analogy, correspondence, resemblance (analogy of attribution), relationship, and proportion (analogy of proportionality) that can interconnect two or more words or phrases. A dictionary entry on analogy reads (Ali, 1999, p. 32):

> A simple metaphor, or an authentic expression of a connection, analogy is a form of knowledge which, of course, has not the certainty of demonstration, but nor does it have the narrowness of its fields of application: in alchemy, law, theology, biology and physics, analogy has enabled knowledge to be expanded.

Analogy is polysemous concept sometimes deemed overly subjective, especially when it concerns analogy of proportionality; at the same time, it is commonly associated with the scientific approach. This is not the place to discuss all the different ways analogy can be used. What is important is that analogical knowledge can complement demonstrative knowledge:

filling in areas where the latter is lacking. Of course, one must not confuse the two. Thomistic philosophy uses analogy of attribution to attribute certain qualities to God, but it cannot claim to demonstrate their existence; on the contrary, it must suppose their existence *a priori*. Meanwhile, alchemists, in presenting humanity as the centre of proportional relations between Heaven and Earth, animals and plants, offer a vitalist reading of reality that is not entirely without interest; at the same time, as Gaston Bachelard notes (1983, p. 88), they encourage *fuites de pensée* ("scattered thinking"):

> Need it be said that these analogies promote no research? On the contrary, they lead to scattered thinking; they block the steady curiosity which provides the patience to follow a well-defined order of things. At any time, the proof is transposed. You thought you were doing chemistry in the heart of a vial, but the liver responds instead. You thought you were examining a patient; it is the lining up of stars which affects the diagnosis.

When using analogy, relationships between things are surely more important than properties; but as Bachelard observes, not all relationships are salient. In sum, analogical knowledge connects to other forms of knowledge; it must be distinguished from a modelling or similarity-based approach; caution and precision are needed to determine its boundaries.

The French naturalist Etienne Geoffroy Saint-Hilaire (1818) used the term "analogue" to describe two organs in separate agents that have the same position and the same connections, even if they have different functions: e.g., the arm of a human and the wing of a bird. Today, "homologue" conveys this meaning. Saint-Hilaire's colleague Georges Cuvier (1835–1846) used "analogue" in a different way: to describe organs with the same function, whether or not they are anatomically similar – a usage that biology and epistemology have adopted, using "analogue" specifically to refer to two terms that, in relation to each other, serve the same function as another pair of terms: e.g., to say that an IT network is analogous to the nervous system means that the former is to a country what the latter is to an organism. What this means is that, when discussing analogous conditions between a region on Earth and another planet or other celestial body, one must remember to express the relationship in *four* terms, to give the comparison a context: a region on Earth cannot be declared analogous to a region on Mars *as such*, but only within the context of e.g. testing the scientific instruments that may be used as part of a mission to the Red Planet, or comparing conditions for the existence of organisms, or considering how particular biological systems might function in a Martian environment.

Astrobiology has been compared to theology, not always favourably. Both astrobiology and theology make frequent use of analogy. While the limitations and dangers of analogy must be acknowledged, so must its merits. Philosopher Jacques Bouveresse explains (1999, p. 37): "even if it has been shown almost indisputably that it is based on nothing substantial, an analogy (and at the same time, the means of expression which is based on it) can always maintain at least a certain power of suggestion." This power of suggestion gives researchers a powerful tool to counter the adverse effects of reductionism and dogmatism: for letting go of the street lamp and stepping outside the cone of light, showing strength to accept the unknown and the unknowable.

Allowing the possibility of a Second Genesis

The researchers who introduced the notion of Second Genesis – the emergence of life elsewhere than on Earth – probably knew that they were employing a highly-charged expression. While the word "creation", whose origins are equally Biblical, is used indiscriminately by artists and scientists without causing a stir, the word "genesis" undeniably retains its religious connotations. In certain contemporary societies, questions regarding the origin and evolution of life are met with fear and refusal, particularly in certain religious circles: a worry for both scientists and teachers (see e.g. Arnould, 2007; 2009; Bertka, 2009). This is a genuine challenge that must be addressed carefully by the astrobiology community. Astrobiology research deserves a wider audience; it should be subject to further debate outside the purely scientific world. It would be a shame if Second Genesis hindered this development by calling it radically into question or giving it a faulty image.

At the same time, the perspective "Second Genesis" suggests appears essential to any scientific approach to astrobiology. Religious contexts aside, the concept of genesis implies a process and set of relationships that, *a priori*, are not part of the usual vocabulary; if they were, use of the term would be inappropriate or even usurpative. Accepting and even defending extraterrestrial genesis – the appearance of forms that could be called "living" elsewhere than on Earth; to borrow a philosophical perspective: a non-terrestrial transition from "to be" to "being" – requires breaching the worldviews humans have devised on the basis of their experiences, hopes, and fears and agreeing to question pre-formed answers to the mystery of life. It means waiting for – without being able to expect – an experience such as the philosopher Hans Jonas describes (1987) in *The Concept of God after Auschwitz*. The idea of Second Genesis makes this possible;

since, in the famous words, attributed to Martin Rees: "Absence of evidence is not evidence of absence. "Of course, it could be that Earth *is* alone: the solitary miracle of the biosphere; which raises the question: why us?

Wondering rather than marveling

The question, why us?, can give rise to either of two moods: marvel orwonder. The first is common in writings on natural theology and spirituality based on the common human experience and the spectacles of sky and earth. In his *Etudes de la nature* (1784), Bernardin de Saint-Pierre (died 1814) slips easily into such a tone, at the same time weaving in analogue and a teleological outlook; he risks calling something self-evident that remains to be proven. He presents as proof of the existence of God (the Great Clockmaker or *grand horloger*, see Paley, 1802) the now-famous example of a superior intelligence who decides that melons should have ribs to make them easier to share.

To what extent is discourse on the anthropic cosmological principle overly imbued with a tone of marvel, to the detriment of scientific reason? Marvel risks forming – or even imposing – a *worldview* too particular, too determinate to allow the irruption of anything different, be it new cosmological or biological knowledge.

In setting the tone for answering "why us?" – more generally, in choosing an attitude towards the reality of the cosmic, biological, and astrobiological world – wonder presents a helpful alternative. It is a way of fighting the trivialisation that anesthetises the imagination while avoiding dogmatic behaviours. Philosophers like to draw on wonder as a source for their labours. In his *Metaphysics*, Aristotle writes:

> For it is owing to their wonder that men both now begin and at first began to philosophize; they wondered originally at the obvious difficulties, then advanced little by little and stated difficulties about the greater matters, e.g. about the phenomena of the moon and those of the sun and of the stars, and about the genesis of the universe. And a man who is puzzled and wonders thinks himself ignorant.

Could the same not be said of astrobiologists?

Conclusion: A sea without shores

Speaking 12 September 1962 at Rice University in Houston, John F. Kennedy (1962, p. 669) said, "we set sail on this new sea because there is new knowledge to be gained, and new rights to be won, and they must be won and used for the progress of all people." His words have not lost their prophetic quality: not that the American president could see the future, but he and his advisors understood that, by embarking on their space race, the American and Soviet powers would bring all humanity into a new era. The potential benefits went far beyond the borders of their two countries, their scientists and military, to encompass all human cultures.

Fifty years later, many of the promises have been fulfilled; many others remain to be realised. Astrobiology is one area in which Kennedy's words are still highly relevant. It holds out the promise of a true revolution – not only scientific but also philosophical, theological, and cultural – even though no one can foresee the timeframe involved.

The scientific vessels of today are sailing into a new sea; but it is a sea without shores. Life is the horizon: as difficult for the human mind to arrive at as the origins of the universe. Humanity's destiny is forever to embark on such odysseys.

References

Ali, B. 1999. "Analogie." In *Dictionnaire d'histoire et philosophie des sciences*, edited by D. Lecourt. Paris: Quadrige/PUF.

Arnould, J. 2007. *Dieu versus Darwin: Les créationnistes vont-ils triompher de la science?* Paris: Albin Michel.

—. 2009. *God vs Darwin: Will the Creationists Triumph over Science?* Adelaide: ATF Press.

Augustine. 1955. *The Confessions*. Philadelphia PA: Westminster Press.

Bachelard, G. 1983. *La formation de l'esprit scientifique*. Paris: Vrin.

Bernardin de Saint-Pierre, J.-H. 1784. *Etudes de la nature*, ch. XI, sec. "Harmonies végétales des plantes avec l'homme."

Bertka, C. M., ed. 2009. *Exploring the Origin, Extent and Future of Life: Philosophical, Ethical and Theological Perspectives*. Cambridge: Cambridge University Press.

Bouveresse, J. 1999. *Prodiges et vertiges de l'analogie: De l'abus des belles-lettres dans la pensée*. Paris: Editions Raisons d'agir.

Crowe, M. J. 1999. *The Extraterrestrial Life Debate, 1750–1900*. Mineola NY: Dover Publications.

Cuvier, G. 1835–1846. *Leçons d'anatomie comparée*. Paris: Crochard.

Dick, S. J. 1982. *Plurality of Worlds: The Origins of the Extraterrestrial Life Debate from Democritus to Kant*. Cambridge: Cambridge University Press.

Dick, S. J., and J. E. Strick. 2004. *The Living Universe: NASA and the Development of Astrobiology*. New Brunswick NJ: Rutgers University Press.

Dubos, R. 2002 (1962). *Le flambeau de la vie*. Paris: L'Harmattan.

Geoffroy Saint-Hilaire, E. 1818. *Philosophie anatomique*. Paris: J.-B. Baillière.

Grinspoon, D. 2004. *Lonely Planets: The Natural Philosophy of Alien Life*. New York: Harper Collins.

Jacob, F. 1970. *La logique du vivant*. Paris: Gallimard.

Jonas, H. 1987. "The concept of God after Auschwitz: A Jewish voice." *The Journal of Religion* 67(1):1–13.

Kennedy, J. F. 1962. *Public Papers of the Presidents of the United States*, vol. 1. Washington DC: US Government Printing Office.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago IL: University of Chicago Press; transl., *La structure des révolutions scientifiques*. Paris: Flammarion, 1983.

Paley, W. 2006 (1802). *Natural Theology*. Oxford: Oxford University Press.

CONTRIBUTORS

Michael A. Arbib, University Professor, Fletcher Jones Professor of Computer Science, Professor of Neuroscience, University of Southern California (USA)

Jacques Arnould, PhD, The French Space Agency – CNES, Paris (France)

Urszula Katarzyna Czyżewska, PhD candidate, Department of Philosophy of Biology, John Paul II Catholic University of Lublin (Poland)

David Dunér, Professor, History of science and ideas/Centre for Cognitive Semiotics, Lund University (Sweden)

Maria G. Firneis, Professor, Research Platform: ExoLife/Institute for Astrophysics, University of Vienna (Austria)

Gustav Holmberg, PhD, Senior Lecturer, History of Science and Ideas, Lund University (Sweden)

Arthur Holmer, Associate Professor, General linguistics/Centre for Cognitive Semiotics, Lund University (Sweden)

Christopher C. Knight, PhD, International Society for Science and Religion, Cambridge (UK)

Ludwik Kostro, Professor, Department for Logic, Methodology and Philosophy of Science, University of Gdańsk (Poland)

Johannes J. Leitner, Mag. rer. nat., Research Platform: ExoLife/Institute for Astrophysics, University of Vienna (Austria)

Per Lind, PhD candidate, Philosophy of Religion, Centre for Theology and Religious Studies, Lund University (Sweden)

Mathias Osvath, PhD, Cognitive Science, Lund University (Sweden)

Joel Parthemore, PhD, Centre for Cognitive Semiotics, Lund University (Sweden)

Erik Persson, PhD, Practical Philosophy, Lund University (Sweden)

Jean Schneider, PhD, LUTh – Paris Observatory (France)

Göran Sonesson, Professor, Centre for Cognitive Semiotics, Lund University (Sweden)

Stéphane Tirard, Professor, Centre François Viète d'épistémologie et d'histoire des sciences et des techniques, Université de Nantes (France)