1 The Foundations of Astrobiology

Before embarking on our voyage into the vast and multifaceted domain of astrobiology, every budding practitioner and student of this field should be equipped with a few basic tools to venture forth on this grand journey. First, it is vital to grasp some foundational terms that crop up often, such as ‘life’, ‘habitability’, and ‘astrobiology’ itself. Although these concepts may seem self-evident at first glimpse, all of them are inherently complex in actuality, and have accordingly attracted intense debate since at least the twentieth century, sometimes even commencing centuries and millennia earlier. In fact, a universal definition of life remains elusive to this day, and philosophers and scientists continue to debate this matter.

Second, it is important (arguably even essential) to understand the historical development and growth of astrobiology. Pursuing this historical path is valuable for a minimum of two reasons, though it may appear to deviate from the scientific goal(s) of this textbook. For starters, it will help us comprehend and appreciate how, on the one hand, astrobiology has ancient roots and, on the other hand, it is a remarkably young and dynamic discipline. Next, from a broader perspective, the history of science can expand our horizons, and consequently enable us to gain a better picture of where this field might be headed towards in the turbulent twenty-first century.

Thus, the purpose of this chapter is to fulfil the aforementioned objectives. In the first part, we provide working definitions for some fundamental terms encountered in astrobiology, after which we explore an abbreviated history of astrobiology in the second part.

1.1 Key concepts and definitions

In this section, we will carefully examine some of the central concepts that are often encountered in astrobiology.

1.1.1 Astrobiology and exobiology

Since this book deals with astrobiology, it is natural to start with posing the question: *what is astrobiology?* This sweeping question can be broken up into additional segments: what answers are astrobiologists seeking? What targets do they strive to survey? What are some of the disciplines that astrobiology draws on? To answer this plethora of questions, the working definition of astrobiology delineated by the NASA Astrobiology Institute (NAI) is reproduced in its entirety,[1] because it is fairly succinct, yet thorough.

---

[1] https://astrobiology.nasa.gov/nai/about/index.html.

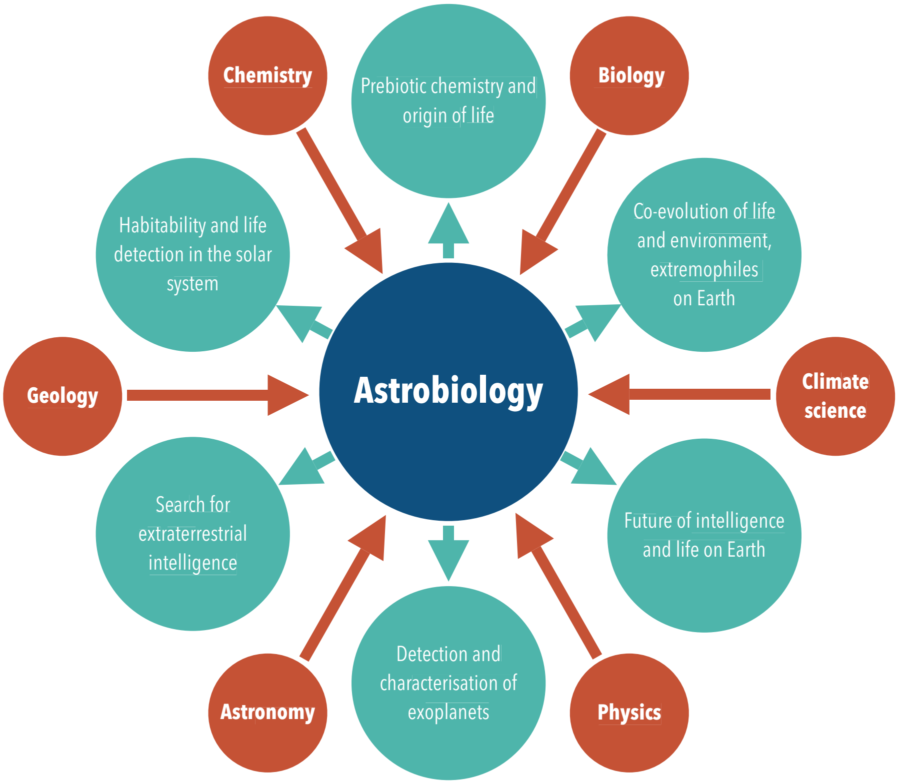

Figure 1.1 The major fields employed in astrobiology (small dark circles) to analyse its key areas of research (large light circles).

> Astrobiology is the study of the origins, evolution, distribution, and future of life in the Universe. This interdisciplinary field requires a comprehensive, integrated understanding of biological, geological, planetary, and cosmic phenomena. Astrobiology encompasses the search for habitable environments in our solar system and on planets around other stars; the search for evidence of prebiotic chemistry or life on solar system bodies such as Mars, Jupiter's moon Europa, and Saturn's moon Titan; and research into the origin, early evolution, and diversity of life on Earth. Astrobiologists address three fundamental questions: How does life begin and evolve? Is there life elsewhere in the Universe? What is the future of life on Earth and beyond?

The themes expressed in this quote are also depicted in Figure 1.1, which illustrates the topics overlapping with astrobiology.

To reiterate the points conveyed in the prior paragraph, astrobiology grapples with three overarching and truly fundamental questions, which may be colloquially expressed in the following fashion.

1. *Where did we come from?* [How does life begin and evolve?]
2. *Are we alone?* [Is there life elsewhere in the Universe?]
3. *Where are we going?* [What is the future of life on Earth and beyond?]

This trio of questions governs the organisation of the book. In Parts I and II, we address the astronomical, physical, chemical, and geological processes heralding the origin(s) and evolution of life on Earth. Next, in Parts III, IV, and V, we describe astrobiological targets in our solar system and beyond, their potential for supporting extraterrestrial life, and how to detect the latter. Finally, in Part VI, we touch on the future of humanity, especially in relation to space exploration. We do not tackle the third question in as much depth as the preceding duo because uncovering and knowing the past and the present is difficult, but feasible, whereas forecasting and knowing the future is deeply challenging.

As the previous definition of astrobiology implies, the Earth serves as a bedrock for this discipline. The rationale is simple: Earth is currently the only world unequivocally confirmed to host life, although this status might change in the future. Hence, to varying degrees, most attempts to assess the possibility of extraterrestrial life extrapolate from life on Earth. It may be argued that such an approach runs the risk of geocentrism (i.e., overly relying on data from Earth). While this objection is valid to an extent, in the absence of any other samples, the Earth remains the only (and therefore the best available) guide to resolving profound questions such as: how did the origin(s) of life occur? how would life and its environment coevolve together? what are the physical and chemical extremes tolerable by organisms?

Thus, it is apparent that the Earth is an essential ingredient of astrobiology. In this respect, astrobiology is distinguishable from the earlier term *exobiology*, which encapsulates the discipline that aims to infer the '*cosmic distribution of life*' (Lederberg, 1960, pg. 393). Given that exobiology emphasises the study of extraterrestrial life, it can be said to primarily focus on the second question (are we alone?), and might be regarded as a major subset of astrobiology. An early prominent advocate of pursuing exobiology was Joshua Lederberg (1925–2008), who received the 1958 Nobel Prize in Physiology or Medicine for his work on genetic transfer in bacteria.

Lederberg is widely credited with having coined the term exobiology, although this word might predate his seminal 1960 paper (Lederberg, 1960). As an interesting aside, the word 'astrobiology' was coined several decades earlier, towards the end of the nineteenth century; however, this field has much older roots, as outlined in Section 1.2. A short review of the etymology and twentieth-century history of astrobiology is furnished in Lingam and Loeb (2020b).

1.1.2 Habitability

If we inspect the NAI definition of astrobiology in Section 1.1.1, we encounter the word 'habitable', which brings up the question: how do we define 'habitability' or a 'habitable environment'? As per the 2015 NASA Astrobiology Strategy,[2] these two terms may be understood as follows:

> Habitability has been defined as the potential of an environment (past or present) to support life of any kind. . . . A habitable environment is one with the ability to generate life endogenously – solely using available resources – or support the survival of life that may arrive from elsewhere.

A closely related definition of habitability is furnished in a comprehensive review of this subject by Cockell et al. (2016, pg. 89):

> In this review on habitability, we define it as the ability of an environment to support the activity of at least one known organism.

---

[2] https://astrobiology.nasa.gov/nai/media/medialibrary/2015/10/NASA_Astrobiology_Strategy_2015_151008.pdf.

These definitions appear straightforward at first glimpse, but in reality, a number of ambiguities persist. For example, is habitability discrete (e.g., binary) or continuous? On the one hand, it may be argued that an environment can either support one or more organisms or that it cannot do so, thereby conferring a binary basis (Cockell et al., 2019). Yet, on the other hand, the complexity of organisms and the associated environments, as well as their coevolution, could effectively impart continuity to the notion of habitability (Space Studies Board, 2019; Heller, 2020). Another subtlety pertaining to habitability is spelt out next because of its significance.

Broadly speaking, we can envision habitability as some set of characteristics that are imperative at any given instant in time to render an environment suitable for hosting life (Cockell et al., 2016; Domagal-Goldman et al., 2016). Such factors would contribute to what might be dubbed *instantaneous* habitability. The variables that enter the picture insofar as life-as-we-know-it is concerned would therefore include the likes of a solvent (specifically water), energy sources for metabolism, essential elements (in nutrient form), and appropriate physicochemical conditions. We shall delve into the multiple aspects that constitute instantaneous habitability in Chapter 7.

Alternatively, since life should necessitate a certain amount of time to originate, evolve, and create a biosphere, it is evident that environments must retain clement properties for life over a sufficiently long timescale. Hence, we must also engage with the notion of *continuous* habitability, which encapsulates the potential of a particular world (typically a planet) to sustain conditions amenable to life over an extended period of time. The variables that modulate continuous habitability are many and variegated, ranging from planetary characteristics such as size, axial tilt, and plate tectonics to stellar factors (e.g., winds and flares) and even galactic processes like gamma ray bursts and supermassive black hole activity. We will touch on these components, depicted in Figure 1.2 in Chapter 8.

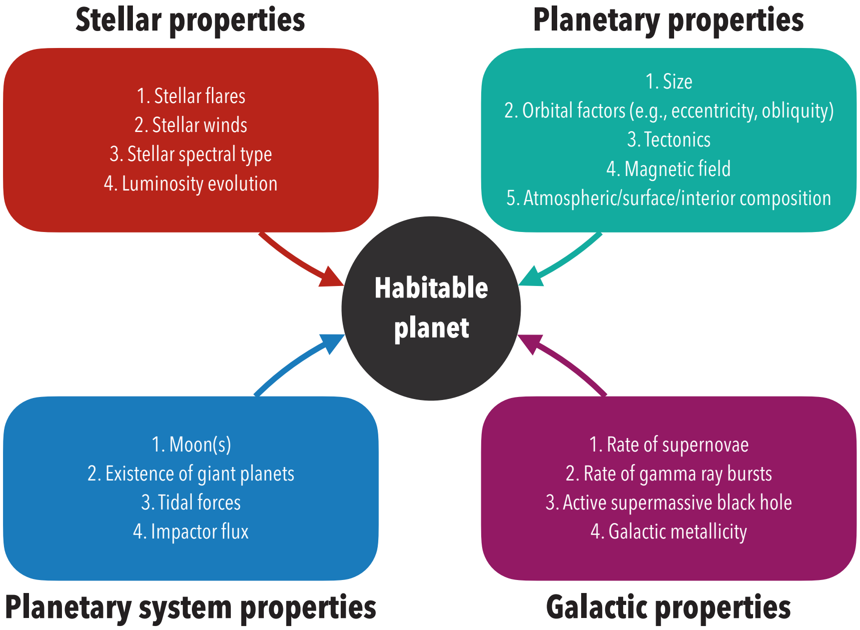

Figure 1.2 The myriad variables that influence the habitability of a planet; they are either properties intrinsic to the planet, host star(s), and planetary system, or regions of the Milky Way.

1.1.3 Life

Hitherto, we have often employed the term *life* without explicitly clarifying what we mean by it. *What is the definition of life?* As the reader may imagine, this question has engaged intellectuals for millennia, and still remains unresolved. This subject has become so extensive that it has been thoroughly examined in entire books, which can be consulted for more information (e.g., Schrödinger, 1944; Cleland, 2019; Smith and Mariscal, 2020).

If we wish to discover extraterrestrial life or ascertain how the origin(s) of life may transpire, which are two prime goals of astrobiology, it is natural to contend that we must define life and demarcate it from non-life. Even though this stance seems straightforward, it faces several subtle drawbacks, as summarised in Cleland (2019). For each proposed definition of life, it is feasible to come up with counterexamples and/or exceptions; this issue is illustrated a few paragraphs hereafter. Hence, in place of 'universal' formulations of life, some authors have advocated for operational definitions and heuristics, while others have suggested that the category of 'life' itself is problematic (Smith and Mariscal, 2020).

One of the earliest recorded definitions of life was provided by Aristotle (384–322 BCE) in the fourth century BCE. In the famous *De Anima* (On the Soul), Aristotle (1907, pg. 49) postulated:

> Of natural bodies some possess life and some do not: where by life we mean the power of self-nourishment and of independent growth and decay.

On inspecting this definition, it is apparent that Aristotle highlighted the metabolic facets of life – to wit, its capacity to perform self-sustaining activities such as maintenance and growth by employing energy obtained from its environment. As we shall witness in Section 5.3.1, a prominent set of hypotheses for the origin(s) of life has, likewise, attempted to trace the steps comprising this transition from a metabolic perspective.

The above definition illustrates the pitfalls of subscribing to a single formulation of life without including any accompanying caveats. Fire and crystals, both of which are not living systems, are capable of maintenance and growth when they are granted access to suitable energy sources and/or raw materials. Now, let us turn our attention to the so-called NASA definition of life (Joyce, 1994, pg. xi), because it is perhaps the closest that we have to a consensus, although its acceptance is by no means universal.

> Life is a self-sustaining chemical system capable of Darwinian evolution.

If we scrutinise this definition carefully, we see that most of the terms therein are relatively straightforward, except for *Darwinian evolution*. Singling out this phrase is problematic due to the following three reasons, and additional objections could be raised (refer to Question 1.3).

- In referring to Darwinian evolution, multiple theories are clubbed together under the same umbrella (Mayr, 2004, Chapter 6), which makes it challenging to define this term. Darwinian evolution is often equated with the theory of natural selection, although they are not synonymous; the latter may be summarised as follows (Mayr, 2004, pg. 31):

  > It is rather a shock for some biologists to learn that natural selection, taken strictly, is not a selection process at all, but rather a process of elimination and differential reproduction. It is the least adapted individuals that in every generation are eliminated first, while those that are better adapted have a greater chance to survive and reproduce.

- While Darwinian evolution is conventionally held to be preeminent and predominant in the evolutionary history of Earth, there is growing recognition that alternative (i.e., non-Darwinian) modes of evolution might have played vital roles on our planet. For instance, Lamarckian evolution involves '*non-randomly acquired, beneficial phenotypic changes*' (Koonin and Wolf, 2009).[3] If other modes of evolution can operate in isolation, then the emphasis on Darwinian evolution may be misplaced.
- As per the NASA definition, it is conceivable that viruses and artificial intelligences (AIs) might need to be excluded: the former because they may not count as self-sustaining in the strict sense, and the latter because they are not necessarily subject to Darwinian evolution. Yet, persuasive arguments could be (and have been) made as to why both these entities are classifiable as 'living' systems, in which case the NASA definition would be rendered incomplete, if not incorrect.

As already mentioned, in place of concrete definitions, some authors have opted for identifying general characteristics of living systems. Such lists have proliferated over the years, owing to which we shall concentrate on just one of them. Koshland Jr. (2002) hypothesised that life may consist of seven pillars, collectively called PICERAS: (1) **P**rogram (a blueprint encoding the information pertaining to the organism); (2) **I**mprovisation (achieved via mutation and selection); (3) **C**ompartmentalisation (to mitigate dilution); (4) **E**nergy (to sustain the system); (5) **R**egeneration (transport of chemicals to replenish losses); (6) **A**daptability (behavioural responses to stimuli); and (7) **S**eclusion (allowing many reactions to occur in tandem).

However, even transitioning from rigorous definitions to lists does not altogether solve the riddle of understanding life. In connection with the PICERAS scheme, it is possible that rudimentary (perhaps the first) life forms on Earth or elsewhere lacked well-defined compartments while retaining the looser property of spatial localisation; in such a scenario, they would be labelled as non-living, despite being the opposite. Moreover, the likes of viruses and AIs may not exactly fulfil all the criteria, though some scientists consider them living systems.

This brief foray into the realm of delineating life is obviously not meant to be comprehensive. Instead, it is meant to convey a handful of commonly employed formulations of life, and illustrate how they are limited in scope.

1.2 A brief history of astrobiology

As already remarked, the word 'astrobiology' was coined merely about 125 years ago, but the history of speculations about extraterrestrial life stretches back millennia. The excellent monographs by Dick (1982), Crowe (1986, 2008), and Weintraub (2014) may be consulted to obtain further details. In the subsequent pages, we will furnish a highly abbreviated and selective timeline of some notable developments in this arena.

It has become customary to attribute the concept of *plurality of worlds* and the existence of extraterrestrial life to Anaximander (sixth century BCE) and the Greek atomists, who believed that matter was composed of indivisible units, namely, atoms. In *Letter to Herodotus*, Epicurus (341–270 BCE) conjectured that the Universe is infinite and that other inhabited worlds abound, as evidenced by the following text (Bailey, 1957):

---

[3] Simply put, the phenotype encompasses the set of observable traits of a particular organism.

> Furthermore, there are infinite worlds both like and unlike this world of ours. . . . We must believe that in all worlds there are living creatures, and plants and other things that we see here in this world.

In Europe, the Epicurean standpoint was mostly relegated to the margins by the worldview of Aristotle, which held that the Earth is the only inhabited world in the Cosmos. The arguments that underpinned Aristotle's thesis are rather intricate, stemming from the notion that the elements constituting the Universe possess an intrinsic order, culminating in the Earth emerging as the centre of the Cosmos and the only world with life (Dick, 1982).

As the Aristotelian worldview and its variants held sway in Europe, most narratives tend to jump nearly 1,500 years ahead in time, that is a few decades or centuries prior to the Copernican revolution (which dethroned the Earth from its place as the centre of the Universe). Intellectuals such as Albertus Magnus (1200–1280), Hasdai Crescas (1340–1410), and Nicholas of Cusa (1401–1464) wrote persuasively about the possibility of extraterrestrial life. However, undertaking such a skip in time would manifestly paint an inaccurate picture because it ignores parallel musings unfolding in Asia and Africa; we shall concisely explore the former herein.

The Śānti Parva (book 12) of the renowned Indian epic *Mahābhārata* – whose precise date of (oral) composition is unclear, but might be roughly 2,000 years old – contains the following passage (Ray, 1891, pg. 34):

> The sky thou seest above is Infinite. It is the abode of persons crowned with ascetic success and of divine beings. It is delightful, and consists of various regions.

Additional examples of similar speculations from ancient India and China are furnished in Selin and Sun (2000) and Nazé (2009). Many of the major religions currently concentrated (or with origins) in South Asia, such as Hinduism, Buddhism, Jainism, and Sikhism, are distinguished by writings supporting the existence of extraterrestrial life (Weintraub, 2014). Moving to west Asia, multiple authors wrote about extraterrestrial life during the Islamic Golden Age. One of them was Muhammad al-Bāqir (676–733), who asserted that (Weintraub, 2014, pg. 165):

> Maybe you see that God created only this single world and that God did not create Homo sapiens besides you. Well, I swear by God that God created thousands and thousands of worlds and thousands and thousands of humankind.

The polymath Fakhr al-Dīn al-Rāzī (1149–1209) also presented counterarguments against the Aristotelian worldview, and advocated for the plurality of worlds, and even the potentiality of a multiverse.

Of the early post-Copernican thinkers, the most famous among them is Giordano Bruno (1548–1600). Much of Bruno's fame is attributable to the fact that he was burnt at the stake for religious heresy. It is often claimed that Bruno's gruesome fate was directly caused by his beliefs regarding the plurality of worlds and extraterrestrial life, but other matters of religious doctrine appear to have been partly (if not chiefly) responsible. In his treatise *De l'infinito, universo e mondi* (published in 1584), Bruno strikingly asserted (Boulting, 1914, pg. 144):

> There are countless suns and an infinity of planets which circle round their suns as our seven planets circle round our Sun . . . and there must be plants and minerals in the worlds of space like those of our Earth or different. We can attribute life to worlds with better reason than we can to our own Earth.

In the centuries after Bruno, the idea of extraterrestrial life slowly gained acceptance, as chronicled in Crowe (1986, 2008). By the time we reach the commencement of the twentieth century, extraterrestrial life was quite widely embedded in the public consciousness through science fiction books such as H. G. Wells' *The War of the Worlds* (1898) and the writings of Percival Lowell (1855–1916), who erroneously claimed that Mars' surface was crisscrossed by 'canals' constructed by extraterrestrial technological species.

It would be a mistake, however, to presume that astrobiology at this stage was confined only to mere speculations. One of the central concepts in habitability is the circumstellar *habitable zone*, which we shall introduce in Section 8.1, and employ in subsequent chapters. Recent research has established that the habitable zone was already cast into semi-modern (albeit qualitative) form by the start of the twentieth century, thanks to contributions from Sir Isaac Newton (1643–1727), William Whewell (1794–1866), and Alexander Winchell (1824–1891), among others (Lingam, 2021a).

As we have seen thus far, astrobiology was predominantly (yet not exclusively) confined to the realm of philosophical musings until the start of the twentieth century. In some respects, this situation started to change in the nineteenth century, but it was the first half of the twentieth century that birthed many crucial breakthroughs. We will underscore only a handful of them, using the trio of questions posed in Section 1.1.1 as our guide; we will encounter more such milestones in the upcoming chapters.

- In 1913, Edward Maunder (1851–1928) authored a book with the self-explanatory title *Are the planets inhabited?*, wherein he presented systematic calculations to demonstrate that the surface of Mars is unlikely to host long-lived bodies of liquid water. Furthermore, the habitable zone concept was delineated, and a heuristic estimate of the number of inhabited worlds in the Milky Way was provided.
- As described in Section 5.1.1, amino acids are the building blocks of proteins. Hence, producing these molecules from simple, widespread, inorganic compounds is relevant to the origin(s) of life. This synthesis was accomplished in 1913 independently by two chemists: Walther Löb (1872–1916) and Oskar Baudisch (1881–1950) (Lazcano and Bada, 2003).
- Identifying signatures generated by biology is vital since they can pave the way for the discovery of extraterrestrial life. Vladimir Martynovitch Artsikhovski (1876–1931) advocated in 1912 that chlorophylls (utilised in photosynthesis) may constitute biological indicators that could be detected by telescopes. Likewise, Sir James Jeans (1877–1946) suggested in 1930 that molecular oxygen – a product of oxygenic photosynthesis (i.e., of potentially biological origin) – might be discernible by telescopes.
- In 1935, Ary Shtérnfeld (1905–1970) authored an article in the science magazine *La Nature*, which was notable because he outlined one of the first modern definitions of astrobiology. In this publication, he predicted that Saturn's moon Titan hosts an atmosphere, and delved into topics as diverse as the origin(s) of life and organisms in extreme habitats.
- Gavriil Adrianovich Tikhov (1875–1960) ranks among the forgotten pioneers of astrobiology (Briot et al., 2004). He conducted experiments on plant physiology under extreme conditions, developed spectroscopic methods to detect chlorophylls and analogous pigments on the surface of Mars, and analysed light from Earth reflected back by the Moon.

From the second half of the twentieth century (i.e., 1950s and later), a multitude of experimental and theoretical advances propelled astrobiology into the mainstream, despite some pushback from detractors. Many of the subfields that comprise astrobiology – such as origin-of-life studies, microbiology in extreme environments, habitability of worlds in our solar system, and the search for extraterrestrial technological intelligences (ETIs) – were characterised by such rapid progress that even handpicking a select few examples from each domain would significantly expand our narrative; some of the salient milestones are, instead, covered in the appropriate chapters.

It is tempting to suppose that astrobiology entered its ‘mature’ phase with the discovery of planets beyond our solar system (exoplanets) and the establishment of NAI in the 1990s. In reality, however, astrobiology has been growing in complexity and depth since the twentieth century, as already remarked. With that said, the modern period of astrobiology (which is ongoing) might be credibly dated to the 1990s on account of the aforementioned two reasons. The state of funding, educational programmes, data, modelling, and personnel is more robust and substantial than ever before, owing to which the future seems bright, as summarised next.

- We are combining sophisticated physical and chemical models with state-of-the-art experiments and field analyses to ascertain how, where, and when the origin(s) of life was actualised on Earth, and perhaps other worlds. This combined approach may enable us to tackle the first fundamental question: where did we come from?
- We are performing laboratory experiments and field studies to gauge the physicochemical limits tolerable by organisms, as well as the signatures and markers they generate. Future missions to promising targets in our solar system (e.g., Mars) will inform us about their habitability, and might uncover traces of life. In concurrence, forthcoming space- and ground-based telescopes may help us detect atmospheric and surface signatures of biological activity on exoplanets. These avenues could collectively help us resolve the second major question: are we alone?
- We are developing a deeper global awareness of the direct impact of human activity on the Earth system (including, but not limited to, climate change), and of the existential threats that humanity will face in the immediate and distant future. Considerable attention is now devoted to addressing such challenges, and devising sustainable trajectories for prolonging the lifespan of human civilisation. In parallel, a renewed interest in space exploration, promoted not only by national space agencies but also by private entrepreneurs, is perceived by some as the dawn of human expansion beyond our planet. Exploring such issues might help us gain insights into the final question: where are we going?

1.3 Problems

Question 1.1: The website you will be utilising for this problem is Google Scholar.[4] Enter the word ‘astrobiology’ and use the *Custom Range* feature to determine how many times this term appears in 10-year intervals over the past 100 years, that is quantify how many instances of this word are recorded in 1920–1930, 1930–1940, and so on until 2010–2020. Report your results in tabular form and/or generate a histogram. What do these results suggest concerning the progress of astrobiology in the last 100 years?

Question 1.2: After consulting the references cited in Section 1.1.2, discuss whether you would regard habitability as binary, discrete but multi-valued, continuous, or none of these categories. Make sure to explain your answer by drawing on peer-reviewed sources.

---

[4] https://scholar.google.com/.

Question 1.3: In Section 1.1.3, some of the limitations associated with including 'Darwinian evolution' in the NASA definition of life were addressed. Identify at least one more reason as to why the NASA definition is either incomplete or incorrect. Draw judiciously on internet resources, and restrict your sources to peer-reviewed publications.

Question 1.4: To begin with, familiarise yourself with the review of viruses in an astrobiological context by Berliner et al. (2018). Next, equipped with this information, peruse the three formulations of life furnished in Section 1.1.3. Would viruses count as 'life' as per any of this trio? Justify your reasoning and back it up with peer-reviewed sources, if and when necessary.

Question 1.5: Aside from the proponents of the plurality of worlds (also called cosmic pluralism) in Section 1.2, select another individual from $\gtrsim$500 years ago who subscribed to cosmic pluralism and the existence of extraterrestrial life. Provide a short biography of this person, and mention which of their work(s) explicitly conveyed their belief in extraterrestrial life.