Introduction

Is there life beyond Earth? Unlike most of the great cosmic questions pondered by anyone who has spent an evening of wonder beneath starry skies, this one seems accessible, perhaps even answerable. Other equally profound questions such as “Why does the universe exist?” and “How did life begin?” are perhaps more difficult to address and must have complex explanations. But when one asks, “Is there life beyond Earth?” the answer is “Yes” or “No”. Yet despite the apparent simplicity, either conclusion would have profound implications.

Few scientific discoveries have the power to reshape our sense of place in the cosmos. The Copernican Revolution, the first such discovery, marked the birth of modern science. Suddenly, the Earth was no longer the center of the universe. This revelation heralded a series of findings that further diminished our perceived self-importance: the cosmic distance scale (Bessel, 1838), the true size of our galaxy (Shapley, 1918), the existence of other galaxies (Hubble, 1925), and finally, the large-scale structure and evolution of the cosmos. As Carl Sagan put it, “The Earth is a very small stage in a vast cosmic arena” (Sagan, 1994, p. 6).

Darwin’s theory of evolution by natural selection was the next perspective-shifting discovery. By providing a scientific explanation for the complexity and diversity of life, the theory of evolution replaced the almost universal belief that each organism was designed by a creator. Every species, including our own, was a small twig in an immense and slowly changing tree of life, driven by variation and natural selection.

Answering the question “Is there life elsewhere?” would be a third shift. An affirmative answer would fuse the Copernican Revolution with Darwinian evolution. Earth would not be special. We would live on one of billions of planets teeming with life. Yet, with this new outlook, the universe would abruptly become immensely richer. The night sky would no longer be a theatre of sterile physics.
and chemistry, but would instead be full of living worlds, evolving creatures, and perhaps even conscious beings.

If no life exists elsewhere then we would likewise have reason to reconsider our place in the cosmos. Trillions of other planets in existence would be wholly barren, and if life were extinguished on Earth then it would be extinguished everywhere and perhaps forever. What greater value would an individual human life or a species have? Ongoing anthropogenic extinctions would become cosmic losses.

The profundity of such possible implications has fueled speculation about life beyond Earth for millennia. The first recorded musings are those of the Greek philosopher Anaximander whose arguments for a “plurality of worlds” predates our understanding of planets, and indeed precede scientific thought completely (Preus, 2001, p. 58). If we fast-forward to today, popular culture is full of caricatures of life elsewhere, from gray humanoids with big almond eyes and enormous fingers, to Hans Giger’s nightmarish, insectoid xenomorphs.

The purpose of this chapter is to convince you that we may not have to speculate about life elsewhere for much longer. We are on the verge of the aforementioned third shift in cosmic understanding. The science of exoplanets – planets around other stars – has exploded in recent decades. We now know that there are at least as many planets in our galaxy as there are stars (Cassan et al., 2012). Many of these planets are rocky, Earth-sized planets that orbit at just the right distance from their host star to permit liquid water on their surfaces. It will soon be possible to detect life on these habitable exoplanets. The necessary technology is well understood and could be employed on a space telescope within the next 10–20 years. Alternatively, clever techniques using ground-based extremely large telescopes might also detect signs of exoplanet life (Rodler & López-Morales, 2014; Snellen et al., 2015). Not only is the question “Is there life elsewhere?” answerable, it will likely be answered within the lifetime of readers of this book.

**Detecting Planets Around Other Stars**

Exoplanets are difficult to study. By 1995, astronomers had elucidated the structure and history of the universe: they knew that the universe began about 14 billion years ago in the fiery expansion of the Big Bang, they recognized that the universe has billions of galaxies each with tens to hundreds of billions of stars, and from the abnormal rotation of distant galaxies they deduced that most of universe was made of mysterious “dark matter” (Rubin et al., 1980). Astronomers had discovered how the elements were forged in the thermonuclear furnaces of stars, and the entire lifecycle of stars, from formation to supernova to black hole, was well understood. This was all common scientific knowledge by 1995, and yet nobody knew for sure
Joshua Krissansen-Totton and David C. Catling

if other stars had planets. In fact, some astronomers even speculated that the Sun might be the only star with planets (Dick, 1993; Jeans, 1942).

Why is it so difficult to learn anything about exoplanets? The problem is that exoplanets are extremely faint when compared with their host stars. Suppose a distant civilization pointed their telescopes at us. For every meager photon received from the Earth, their telescope would be swamped by ten billion photons from the Sun. It is difficult to separate the planetary light from this blinding glare because planets and their host stars are close together compared to the vast distances between the stars. One of us (DC) remembers being confidently told as an undergraduate by a professor that astronomers would never know anything about exoplanets!

Fortunately, several techniques avoid or overcome the problem of the relative faintness of exoplanets. Broadly speaking, exoplanets can be studied by using either indirect methods or direct imaging. The indirect methods exploit a planet’s subtle influences on its star to infer the properties of the planet. Almost everything that has been learned about exoplanets to date has made use of these indirect methods. The alternative approach is to design a telescope that can isolate the planet’s light from the host star to obtain an image of the planet. Direct imaging is technically challenging, but it is the best way to systematically search for signs of life around nearby stars.

**Indirect Methods for Detecting Exoplanets**

A simple view of planets orbiting an immobile central star is incorrect. Planets have mass, and so they exert a gravitational force on their star causing both planet and star to orbit their common center of mass. Because stars are much more massive than planets, the star’s orbit is a slow wobble around a point interior to, or slightly exterior to, the star.

Two indirect methods exploit the fact that the motion of a star due to the presence of a planet is much easier to detect than the planet itself. The radial velocity technique measures the star’s back-and-forth motion relative to the observer. This is done using Doppler spectroscopy: starlight is split into its component wavelengths, which shift with very slight changes in the star’s radial velocity. From the small swing in wavelengths, the speed of the star’s orbit and the presence of the planet can be inferred.

The first exoplanet around a normal star, 51 Pegasi b, was discovered in 1995 by using the radial velocity method (Mayor & Queloz, 1995). This method is most sensitive to large planets that are close to their host stars because these planets induce the largest stellar wobble. Consequently, many of the first exoplanets discovered were “Hot Jupiters”: Jupiter-sized planets that orbit their host stars in several days or less and have atmospheric temperatures in excess of 1000 °C. Despite being easy
The most precise radial velocity instruments can currently measure variations in stellar velocities of about 1 m/s (Fig. 2.1). To detect an Earth-like planet around a Sun-like star requires precision of 0.1 m/s (e.g. Catling & Kasting, 2017, p. 433), and so no such planets have been found using radial velocity. However, the Eschelle SPectrograph for Rocky Exoplanet and Stable Spectroscopic Observations (ESPRESSO) at the European Southern Observatory’s Very Large Telescope (VLT) will soon begin operation (Pepe et al., 2014). This new instrument will have <0.1 m/s measurement precision and so can find nearby Earth-like planets around Sun-like stars.

An alternative approach that also makes use of a star’s wobble is astrometry. This technique makes precise measurements of a star’s position in the sky over time. This allows the complete orbital motion of planets to be calculated, rather than just the radial motion. Unfortunately, the full potential of astrometry in detecting Earth-sized planets has yet to be realized because NASA’s astrometry mission concept called the Space Interferometry Mission (SIM) was canceled due to budget constraints.

Not all indirect detection methods make use of a star’s motion. Gravitational microlensing takes advantage of the gravitational effect of a planet on the path of light, as predicted by Einstein’s theory of general relativity. When two stars align relative to Earth, the gravitational influence of the foreground star focuses the light from the background star like a lens (Fig. 2.1). Such an alignment temporarily magnifies the background star as seen from Earth. If the foreground star has a planet orbiting in the right position, then its gravitational influence may also distort the lensed light. The planet causes a pulse in background star magnification (Fig. 2.2). Einstein (1936) himself predicted that gravitational microlensing events involving stars would never be observed because stellar alignments are extremely improbable. However, modern technology allows large numbers of distant stars to be monitored, and numerous exoplanets have been detected using this method. In fact, microlensing reveals that the majority of stars in our galaxy have one or more planets (Cassan et al., 2012).

Finally, exoplanets can be detected when they transit their host stars (Fig. 2.1). If the orbital plane of an exoplanet aligns with the Earth, then astronomers can observe the exoplanet crossing the disk of its host star once per orbit. This primary transit obscures some starlight, and with sufficiently sensitive telescopes a periodic dimming of the host star can be detected. The chance of seeing any specific planet transiting its star is small (about 0.5% for Earth-like planets), but by continuously monitoring a large number of stars many planetary transits can be detected. NASA’s Kepler telescope and the European CoRoT spacecraft have been spectacularly
Fig. 2.1 Comparison of the main exoplanet detection methods, updated from a schematic by Shklovskii and Sagan (1966, p. 154). Current observational precision and the precision required to detect an Earth-like planet around a Sun-like star at 33 light years is contrasted. Techniques currently capable of detecting Earth-like planets are highlighted in green, whereas those not yet sufficiently precise to detect Earth-like planets are in red. The obscured stars in direct imaging represent starlight suppression. Gravitational microlensing is separated from the other methods because it cannot be used to target individual nearby stars. Instead, it is necessary to survey a very large field of stars for exceptionally rare stellar alignments. (A black and white version of this figure will appear in some formats. For the colour version, please refer to the plate section.)
Fig. 2.2 The detection of an exoplanet using gravitational microlensing. The graph shows the magnification of a background star as it aligns with a foreground star and is lensed by its gravitational field. If the foreground star has a planet, then this may cause a brief but sizeable magnification in the light curve.

successful in detecting planets using transits. Kepler has discovered over two thousand exoplanets and a few thousand more probable planet candidates. Moreover, transits reveal that Earth-sized planets are common: Sun-like stars have on average at least 0.78 planets with a radius between 0.75 and 2.5 times the Earth (Burke et al., 2015). Figure 2.1 summarizes all the different exoplanet detection methods.

It is worth pausing here to consider what can be learnt from indirect exoplanet detection. Speculative news articles frequently contain sweeping statements about the possible presence of life on exoplanets, and are accompanied by vivid artistic impressions of habitable worlds. These articles capture the imagination and convey the idea that exoplanets are not merely abstract; they are instead real places – worlds that could be mapped, explored, and maybe even settled. But such over-enthusiasm can be misleading.

In practice, indirect methods provide very limited information about exoplanets: radial velocities reveal planet mass, and transits reveal planet radius. For planets with both transit and radial velocity observations, mass and radius measurements can be combined to infer average density. Density provides crude information about whether the planet is gaseous, icy, or rocky. For rocky planets, the planet-star separation and the stellar radiation may tell you if it’s possible for liquid water to exist on the surface. But with the lone exception of transit transmission spectroscopy (see below), that is the extent of what can be deduced from indirect detection methods – anything more is speculation.
The indirect methods provide no way of knowing whether an exoplanet is truly Earth-like. We cannot say for sure if an exoplanet has an atmosphere, ocean, or continents, and there is no way of knowing whether a planet has a biosphere. Other observations are required to address these questions, which brings us to the subject of direct imaging.

**Direct Imaging of Exoplanets**

From the 1990s to mid-2000s radial velocity techniques produced stunning new exoplanet detections. Then there was about a decade of new discoveries with transits. Both techniques will continue to expand our knowledge. But we anticipate that startling new advances will soon be made with direct imaging. As mentioned previously, faint exoplanets are difficult to see amid the glare from their luminous host stars. But with impressive optical know-how, starlight can be suppressed to isolate the light from orbiting planets.

![An image of four giant exoplanets around the star HR 8799 – 130 light years away – taken using the Keck telescope in Hawaii (Marois et al., 2008). The central star looks unusual because an internal coronagraph was used to reduce the brightness of the star and reveal the relatively faint planets. Reproduced with permission from Nature Publishing Group.](image-url)
Several space telescope designs can overcome the glare of starlight. The first is a starshade. A thin, opaque disc with a pattern of notched edges is positioned in front of a space telescope. The starshade blocks the starlight and its patterned edge compensates for the spread of starlight (diffraction), so that only light from the planet enters the telescope. However, the starshade must be placed thousands of kilometers away from the telescope to be effective (Stark et al., 2016). An alternative to the starshade is an internal coronagraph whereby the starlight and planet light are separated within the optics of the telescope itself. In principle, both designs could suppress visible starlight to take pictures of Earth-like planets around other stars. In fact, this has already been done for Jupiter-sized planets using ground-based telescopes. Figure 2.3 is a real image of a distant star system with four planets orbiting around a central star. A coronagraph reduced the brightness of the star so that the planets were revealed.

Direct imaging can do much more than take blurry pictures of exoplanets. As we’ll see, planetary light can be split into a spectrum of its component wavelengths, and the atmospheric composition and perhaps even surface properties of the planet can be deduced. This brings us to the subject of exoplanet characterization and habitability.

**Characterizing Exoplanets and Habitability**

**What Makes a Planet Habitable?**

Broadly speaking, a planet is habitable if it can support life. However, exoplanet habitability means something more specific: a rocky planet that can sustain liquid water on its surface. The reasoning behind this definition is that liquid water appears necessary for life as we know it. The habitable zone of a star is the range of distances between planet and star that permit liquid water to persist on a rocky planet’s surface. Planets closer than the habitable zone are unsuitable for life because they will undergo a hellish runaway greenhouse whereby oceans boil away and are lost to space. This was the fate of Venus, which today has only trace amounts of water in its atmosphere and a scorching surface at 460 °C. Planets beyond the habitable zone are also barren because surface water is permanently frozen, such as on Mars (although we will see the fate of Mars is more complex than merely being too far from the Sun).

In the late 1970s, some scientists believed Earth’s habitability was a fluke. Had Earth formed merely 5% closer to the Sun, they argued, it would have been rendered uninhabitable by a runaway greenhouse, and if it had formed just 1% further away from the Sun it would be permanently locked in a global ice age (Hart, 1978). Such a narrow habitable zone would imply that life is rare in the cosmos.

Fortunately, this picture of the habitable zone is incorrect. It turns out that the cycling of carbon between the atmosphere and the solid Earth acts a planetary
thermostat (Walker et al., 1981). Carbon dioxide (CO₂) is a greenhouse gas that warms the planet’s surface. CO₂ is continuously added to the atmosphere by volcanism and removed from the atmosphere through weathering of rocks and subsequent burial of carbon-bearing rocks. Specifically, atmospheric CO₂ and water react to form acid that gradually dissolves surface rocks into products that are carried to the ocean where they eventually form carbonate rocks. The rate of weathering and carbon burial depends on temperature – the warmer the atmosphere, the greater the rate of carbon removal. This feedback helps maintain the planet’s temperature in a habitable range. If the Earth plummets into a global ice age, as has occurred several times over Earth’s history, then the removal of CO₂ from the atmosphere will cease because conditions are too cold and dry for weathering. However, volcanoes will continue to erupt, causing CO₂ to build up in the atmosphere, thereby warming the planet until the ice melts. Conversely, if the Earth becomes warmer, then rates of weathering will be enhanced due to greater atmospheric water content, thereby reducing atmospheric CO₂ and cooling the climate. In fact, the Sun has increased in brightness by roughly 10% every billion years, but the Earth’s temperature has remained comparatively stable because of this planetary thermostat (unfortunately this feedback takes hundreds of thousands of years to take effect, and so it will not help mitigate anthropogenic global warming). Recent habitable zone estimates suggest the Earth would be safe anywhere between its current position and 70% further away from the Sun (Kopparapu et al., 2013).

By continuously observing about 150,000 stars for transits, the Kepler telescope was designed to find the fraction of stars that have an Earth-sized planet in their habitable zones. Kepler discovered that habitable Earth-sized planets are fairly common – by extrapolation, about 5%–10% of Sun-like stars have a rocky planet orbiting in their habitable zone (Silburt et al., 2015). Unfortunately Kepler broke down before enough data were gathered to pin down an exact number.

At this point you might be wondering why our notion of the habitable zone is so restrictive. Within our own Solar System, both Europa (a moon of Jupiter) and Enceladus (a moon of Saturn) have potentially habitable oceans of liquid water beneath their icy surfaces, but are far outside the Sun’s habitable zone. Their internal oceans are maintained by tidal forces: the immense gravitational forces from the gas giants they orbit, in combination with continuous tugging from other moons, create tides that heat the interiors of Europa and Enceladus. Indeed, it could be argued that there is no true outer edge of the habitable zone because tidally heated oceans can be maintained at any distance from a host star.

However, practical reasons restrict the habitable zone to the region where surface oceans can persist. Icy exoplanets with subsurface oceans may host life, but it would be extremely difficult to detect this life remotely because there are no clear atmospheric or surface biosignatures. Several spacecraft have visited Europa and Enceladus and we still do not know if these moons host life! In short, the
“habitable zone” is not intended to rigorously define where life is possible, but is instead a pragmatic guide to detectable life.

Residing within the habitable zone is an important criterion for habitability, but it is certainly not the only one. Whether or not a planet’s surface can sustain liquid water depends on many other factors, such as size. By some estimates, Mars orbits within the Sun’s habitable zone, and yet it is a frigid, polar desert with no surface water. How is this possible? Mars’ small size and low gravity means that it quickly lost most of its atmosphere to space. Furthermore, Mars’ interior cooled rapidly due to being small. This cold interior does not generate enough volcanism to replenish the lost atmosphere. Consequently, Mars today has a very thin atmosphere with insufficient greenhouse warming for a habitable surface. Had Mars been larger it might have retained a thick atmosphere and still be habitable.

To appreciate the surprising variety of factors that influence habitability, consider Proxima Centauri b, an approximately Earth-mass planet that orbits within the habitable zone of its host star (Anglada-Escudé et al., 2016). The discovery of this planet generated lots of excitement because it’s only 4.24 light years from Earth, and will thus be comparatively easy to image with future telescopes. However, it is uncertain whether Proxima Centauri b is actually habitable because it orbits an M-dwarf star. After their formation, M-dwarfs undergo a long contraction phase whereby they are \( \sim 10–100 \) times more luminous than later. Consequently, planets that reside within the habitable zone of mature M-dwarfs were effectively within the inner edge of the habitable zone for \( \sim 100 \) million years after formation, and may have lost all their water during this time (Luger & Barnes, 2015). Whether or not Proxima Centauri b suffered this fate will depend on the details of planetary formation such as whether it formed in its current location and how much water it started with (Barnes et al., 2016; Ribas et al., 2016). Similar issues apply to a system of seven planets around the small TRAPPIST-1 ultracool M-dwarf star (Gillon et al., 2017).

In fact, many potential influences on planetary habitability exist, including stellar properties, orbital dynamics, galactic position, atmospheric properties, and interior properties. Finding a truly habitable planet is more complicated than merely finding a rocky planet in the habitable zone.

**Identifying Habitable Planets**

Telescope observations can reveal a lot about an exoplanet’s habitability. At this point we must make a brief digression into the physics of light, atoms, and molecules. Surprisingly, the physics of very small things is important on the scale of worlds. Every molecule absorbs electromagnetic radiation at a specific set of frequencies. These frequencies depend on the configuration of electrons and the ways in which a particular molecule rotates and vibrates, which in turn depend on
Fig. 2.4 Earth’s spectrum in both (a) reflected visible light and (b) emitted infrared light. Atmospheric gases absorb light at characteristic wavelengths, and these spectral fingerprints are labeled in the figures above. Gases that are potential biosignatures are highlighted in bold font. For example, the oxygen (O₂) absorption feature at 0.76 μm is a biosignature because on Earth almost all atmospheric oxygen is produced by photosynthesis. Similarly, ozone (O₃) is a byproduct of biogenic O₂. The signatures of water vapor and carbon dioxide are also clearly visible. Courtesy of Ty Robinson and the Virtual Planetary Laboratory.

Light is thus preferentially absorbed at particular wavelengths when it passes through a planet’s atmosphere, depending on the type and amount of gases present. The light that remains has diagnostic absorption features, which are the fingerprints of the constituent gases. From the size and position of these spectral features, it is possible to work backwards to derive the composition of gases present from direct imaging observations (Crossfield, 2015). Figure 2.4 shows Earth’s spectrum with the fingerprints of different atmospheric gases labeled.

It is even possible to deduce atmospheric composition without the need for starlight suppression by coronagraphs or starshades for exoplanets that transit their
The Search for Another Earth-Like Planet and Life Elsewhere

When a transit occurs, some starlight passes through the planet’s atmosphere. Gases in the atmosphere may absorb some of this planet-grazing light, so that the amount of blocked light depends on wavelength (Seager & Sasselov, 2000). Thus, transits observed at different wavelengths reveal the spectral fingerprints of gases. This “transmission spectroscopy” technique has already been used to identify gases in the atmospheres of transiting “Hot Jupiter” exoplanets. This method might be suitable for investigating transiting rocky planets with NASA’s upcoming James Webb Space Telescope (JWST).

A wealth of information can be gleaned about an exoplanet through direct imaging and spectroscopy. One can apply climate models and calculate a planet’s surface temperature once atmospheric composition is known. Subtle spectral fingerprints from the atmosphere can be used to infer total atmospheric pressure. Then from temperature, pressure, and the amount of atmospheric water vapor, we can paint a more complete picture of a planet’s habitability.

Although exoplanets appear as blurry blobs in telescope images, a surprising amount of information is revealed by variations in brightness over time. The length of a planet’s day and the presence of variable cloud cover could be evident. It may also be possible to detect the glint of a surface ocean. This effect is most clearly visible in an exoplanet’s crescent phase because the bright glint spot is large relative to the rest of the illuminated crescent surface (Robinson et al., 2010, 2014). By measuring changes in exoplanet brightness over several months, it may even be possible to crudely map the planet’s surface and differentiate continents and oceans. Figure 2.5 shows surface maps of Earth that could be obtained from nothing more than brightness and color observations over time.

Clearly, direct imaging could identify habitable exoplanets. However, simply finding a habitable planet will not necessarily change our cosmic perspective. If the origin of life was exceedingly improbable, billions of Earth-like exoplanets with oceans, clouds, and continents might be completely sterile, like global ghost towns for all life. Merely detecting a habitable world is not enough; we need to find life.

Detecting Life on Exoplanets

A back-of-the-envelope calculation reveals that any telescope capable of photographing whale-sized organisms on an exoplanet’s surface would need to be the size of the Solar System! Indeed, distant planets will appear as faint, unresolved blobs to the first telescopes capable of taking pictures of them (similar to Fig. 2.3). How then, will it be possible to detect life?

To answer this question we must go back to the dawn of the Space Age, and the search for life on Solar System planets. Before the 1970s, life on Earth’s surface –
Fig. 2.5 Surface maps like this one could be obtained for Earth-like exoplanets using only observations of planet brightness. The position of continents and oceans can be extracted from precise observations of brightness variations over a planet’s rotation and orbit (Kawahara & Fujii, 2010). Reproduced with permission from IOP publishing. (A black and white version of this figure will appear in some formats. For the colour version, please refer to the plate section.)

or on any planet for that matter – was believed to be an inconsequential veneer of organic scum. The idea that biological processes could substantially modify a planet was rarely discussed, with some exceptions such as the work of Vladimir Vernadsky (1926). James Lovelock and Lynn Margulis broke with orthodoxy by popularizing their Gaia hypothesis – the idea that life modifies its environment to maintain conditions suitable for life (Lovelock & Margulis, 1974). Gaian ideas have had a mixed reception in the scientific literature (e.g. Kirchner, 2003). Some posit that the stability of Earth’s temperature over geological time can be explained by Gaian climatic regulation by organisms, but as we saw earlier, a key planetary thermostat is the inorganic carbon cycle, which would still operate (albeit differently) in the absence of life. However, the idea that life modifies its environment on a planetary scale turned out to be prescient.
Nowhere is this concept more evident than in Earth’s atmosphere. Most life on Earth produces waste gases as byproducts of metabolism and these gases accumulate and change atmospheric composition. In fact, every gas in Earth’s atmosphere except the noble gases is cycled by life. In the 1960s and 1970s, it was argued that the presence of life on other Solar System planets could perhaps be deduced from biogenic gases in atmospheres (Lovelock, 1975). Spectroscopic detection of these so-called biosignature gases could also be a sign of life on exoplanets if non-biological sources like geological activity or photochemistry can be ruled out.

The most promising biosignature gas is molecular oxygen ($O_2$). Almost all the $O_2$ in Earth’s atmosphere has been produced by photosynthesis. Indeed, it is difficult to make large quantities of oxygen without life because lots of energy is required to break apart water molecules and liberate oxygen atoms. Life overcomes this energy barrier by combining the energy of many photons from the Sun in complex biomolecular machinery. It is highly unlikely that any naturally occurring mineral could produce much molecular oxygen from sunlight (Léger et al., 2011). In the upper atmosphere, high-energy UV photons can break apart molecules and produce small amounts of oxygen when the accompanying hydrogen escapes to space. But oxygen typically can’t accumulate this way because reactions with volcanic gases mop-up the oxygen, leaving only trace amounts.

With that said, there are a few unusual scenarios whereby large amounts of oxygen could accumulate in the absence of life. Planets in the process of losing their oceans, planets too small to sustain volcanism, or even planets around unusually UV-luminous stars might all build up oxygen in their atmospheres (Harman et al., 2015). These ambiguities have led some to suggest that a better indicator of life would be the coexistence of atmospheric oxygen and some other biosignature gas that we wouldn’t expect to find in the absence of life (Krissansen-Totton et al., 2016; Lovelock, 1965, 1975). For example the coexistence of oxygen and methane is a strong signature of life in Earth’s atmosphere. Without life, all the methane in Earth’s atmosphere would be destroyed by chemical reactions with oxygen in about 10 years. Methane persists in the atmosphere because it is continuously replenished by biology. In fact, the coexistence of oxygen and methane implies both gases are being generated in large quantities, which is almost impossible to explain without life.

There is another way that life may change its environment on a planetary scale. If organisms cover a significant fraction of a planet’s surface then their color might be detectable in the reflected light from that planet. The chlorophyll pigment that plants use to carry out photosynthesis has a distinctive reflectance spectrum; the leaves of plants reflect a lot more infrared light than visible light, and so there is a large increase in reflectivity between red and infrared wavelengths. This so-called “red edge” is visible in spectra of Earth from space (Arnold et al., 2002; Sagan
et al., 1993). Other biological pigments, such as those found in many prokaryotes, might also be visible in an exoplanet’s reflected light (Schwieterman et al., 2015).

Geocentrism in Astrobiology?

At this point the skeptical reader might argue that our methods for life detection are myopically focused on Earth-like life. What if life on exoplanets is very different? Naturally, it is impossible to rule out weird forms of life that we cannot imagine. But there are good reasons to focus on the habitability criteria and biosignatures described above. For instance, liquid water is arguably necessary for all life in the universe (Pohorille & Pratt, 2012). Any life based on chemistry will require a liquid solvent to mediate its chemical reactions. Solid phase life is unlikely because atoms are fixed in lattices, making chemical reactions extremely slow. Life-like reactions could occur in the gas phase, but high temperatures are required to vaporize large molecules, and heat tends to cause such molecules to react and decompose. This is problematic since large molecules are necessary for chemical life.

The next question to consider is whether there are any viable alternatives to water as the liquid solvent. Even if we knew nothing about Earth life, we might suspect water as a likely candidate for life’s solvent based purely on cosmic abundances. Hydrogen is the most abundant element in the universe whereas oxygen is third (helium is second). Water is also a liquid over a much wider temperature range than most other substances.

Crucially, the properties of water are ideal for sustaining information processing in life. Even if extraterrestrial life has very strange biochemistry, it must have complex molecular machinery capable of inheritance and Darwinian evolution. Non-polar solvents like hydrocarbons are poorly suited to this because the chemical bonds formed in solution are incredibly hard to break, and rapid making-and-breaking of bonds is necessary for biological information processing such as replication, transcription, and translation (Pohorille & Pratt, 2012). These arguments do not definitively rule out unconventional solvents – indeed some astrobiologists have proposed exotic alternatives (Baross et al., 2007) – but they suggest that liquid water is likely a commonly used solvent for extraterrestrial life.

If we accept water as a likely solvent for life, then searching for biosignature gases such as oxygen, which comes from the biological decomposition of water, is credible. Organisms that carry out oxygen-producing photosynthesis (and organisms that eat photosynthetic life or its dead remnants) dominate the Earth’s biosphere because the necessary materials are virtually unlimited. Earth’s surface is covered in life because oxygen-producing photosynthesis requires only water, carbon dioxide, and visible light.
Available materials limit other known metabolisms. Anoxygenic photosyntheticizers—organisms that use sunlight to get energy but don’t release oxygen—are limited by the amount of iron and sulfur in their local environment, whereas chemoautotrophs—organisms that get their energy from materials in their environment rather than from sunlight—are limited by the rate at which volcanic gases are released. In contrast, the materials required for oxygenic photosynthesis will all be readily available on habitable exoplanets. Therefore if organisms evolve oxygenic photosynthesis on such planets, then we would expect these organisms to similarly dominate the biosphere.

With that said, oxygenic photosynthesis probably only evolved once on Earth (Knoll, 2008; Lane, 2002, p. 145). If this innovation is highly improbable, then perhaps oxygenic photosynthesis is rare elsewhere. Additionally, Earth’s atmosphere has only contained appreciable levels of oxygen for about half its history, and oxygen levels have only been comparatively high (between about 10% and 30% of atmospheric composition) for the most recent eighth of Earth’s history. Scientists are still debating why it took such a long time for oxygen levels to rise because oxygenic photosynthesis probably evolved long before atmospheric oxygen increased (Catling, 2014; Kasting, 2013; Lyons et al., 2014).

In light of the early Earth having negligible oxygen, it might be wise to consider alternative biosignatures. It may be possible to detect sulfur and organic compounds produced by life in atmospheres like that of the early Earth (Domagal-Goldman et al., 2011). Another exotic possibility is the detection of ammonia (NH₃) on planets with N₂- and H₂-rich atmospheres. The formation of NH₃ from N₂ and H₂ occurs in the Haber process: the industrial reaction used to make fertilizer. If organisms could evolve catalysts to carry out the Haber reaction, then we might be able to detect them remotely (Seager et al., 2013).

The Future

Suppose a future direct imaging mission finds a habitable Earth-like planet. Upon closer investigation it is revealed that this planet has both oxygen and methane in its atmosphere, a clear indicator of biological activity. Perhaps there are also hints of a “red edge” in the reflected light, subtle features suggestive of a surface covered in photosynthetic pigments. Once the effect of champagne and excitement from the discovery has subsided, the most obvious next question is “How do we learn more?”

The next generation of direct imaging space telescopes will not reveal the surface features of exoplanets—habitable exoplanets will initially be imaged as unresolved pale blue dots. However, improved telescope designs have been proposed. The hyper-telescope (Labeyrie, 1999) would involve a flotilla of space telescopes
dispersed over 100 km. Such an array of telescopes could image an exoplanet’s surface. Figure 2.6 shows what Earth would look like from a distance of 10 light years using this telescope. Continents, oceans, vegetation, desert, weather patterns, and mountain ranges could all be revealed.

With large space telescopes, more subtle signs of life would also be detectable. For example, the CO\textsubscript{2} concentration in Earth’s atmosphere oscillates annually due to the asymmetric distribution of continents. Most of the Earth’s landmass is in the Northern Hemisphere, and so the annual uptake of CO\textsubscript{2} by plants in the northern spring and release of CO\textsubscript{2} by plants in the northern autumn dominates, producing an annual oscillation superimposed upon the steady increase due to anthropogenic carbon emissions (Fig. 2.7). Annually fluctuating carbon dioxide on an exoplanet would be a compelling biosignature. Large telescopes could also detect trace biosignature gases such as nitrous oxide (N\textsubscript{2}O). In short, with a large space
Fig. 2.7 Atmospheric CO$_2$ since 1960 measured at Mauna Loa observatory in Hawaii. The atmospheric CO$_2$ concentration oscillates every year due to the annual growth and die-off of vegetation in the Northern hemisphere, which has most of Earth’s landmass. If habitable exoplanets also have asymmetric landmass then the seasonal variation of CO$_2$ could be a potential biosignature. Courtesy of NOAA.

telescope it would be possible to confirm the presence of life beyond reasonable doubt.

Ultimately, the best way to learn more about a planet is to visit it with a spacecraft. Although plausible concepts have been proposed for small interstellar probes (Long et al., 2010; Martin, 1978), with current technology the cost would be prohibitively large. But perhaps the discovery of a nearby exoplanet teeming with life would be sufficient impetus to seriously consider such a mission.

The detection of life on an exoplanet with telescopes is a feasible, albeit technologically challenging, approach to answering the question “Is there life elsewhere?”. However, there is a complementary approach that could bypass the difficulties described above.

**The Search for Extraterrestrial Intelligence (SETI)**

Rather than search exoplanets for biosignatures, we could instead look for signs of extraterrestrial intelligence. More precisely, we could search the stars for evidence of alien technology. The Drake equation is a useful conceptual framework for “organizing our ignorance” on SETI. The astronomer Frank Drake originally formulated the equation in 1961 as a way of stimulating discussion on SETI. The equation, written below, quantifies the number of communicating technological
Joshua Krissansen-Totton and David C. Catling

civilizations if we knew the values of all the terms:

\[ N = R_\ast \times f_p \times n_e \times f_l \times f_i \times f_c \times L. \]  

(2.1)

Here, \( N \) is the number of technological civilizations in the galaxy that we could communicate with, which is the quantity we want to calculate. \( R_\ast \) is the rate of star formation in the galaxy, \( f_p \) is the fraction of stars with planets, and \( n_e \) is average number of habitable planets per star with planets. Astronomers have measured all three of these terms, although there is still some uncertainty in \( n_e \). The remaining terms are \( f_l \), the fraction of habitable planets upon which life emerges, \( f_i \), the fraction of those planets upon which intelligence evolves, \( f_c \), the fraction of intelligence civilizations that develop the technology to communicate with us, and \( L \), the average lifetime of a technological civilization. There are no hard constraints on any of these last four variables, and depending on the values you choose, the number of communicating technological civilizations in the galaxy could be anywhere from close to zero, suggesting we’re the only civilization in the observable universe, to millions of civilizations. If we were to detect biosignatures on nearby exoplanets this would tell us that \( f_l \) is close to 1, but there is probably no way of definitively constraining the remaining three terms without a SETI programme.

Motivated by this need for observations, the conventional approach to SETI – popularized by the film Contact – is to search for narrowband radio waves that could only be produced by a radio transmitter. The observer can only guess what frequency extraterrestrial intelligences would use, and early SETI searches focused on a region of the radio spectrum dubbed the “waterhole” (Fig. 2.8). Within the waterhole are frequencies corresponding to the spectral fingerprints of hydrogen (H) and the hydroxyl molecule (OH). Taken together these form water, and because water is cosmically abundant and perhaps even necessary for life, extraterrestrial civilizations might choose to broadcast at this universally recognized frequency (Oliver, 1979). The lack of galactic noise or atmospheric absorption around this range of frequencies thus makes the appropriately named “waterhole” region an ideal place for interstellar correspondence.

Modern computing can simultaneously search millions of radio frequencies, which eliminates the need to guess the precise broadcast frequency of other civilizations. Even so, the task is daunting. Searching for extraterrestrial broadcasts is sometimes described as a multi-dimensional cosmic haystack. The observer must find the right frequency, spatial location, moment in time (signals may be pulsed rather than continuous), bandwidth, modulation, and perhaps even polarization; the cosmic haystack has at least eight dimensions! Most of the dedicated SETI radio searches to date have only thoroughly surveyed the nearest few hundred stars for a specific type of narrowband radio signal (Tarter, 2001). Other SETI surveys
Fig. 2.8 Most of the electromagnetic spectrum does not make it through the Earth’s atmosphere and so is unsuitable for ground-based SETI. Radio wavelengths and visible light are ideal for SETI since both penetrate the atmosphere. Within the radio spectrum there is a region around 10 cm where natural background radio noise is minimal. The “water hole” lies within this window. Figures adapted from NASA and Wikimedia commons. (A black and white version of this figure will appear in some formats. For the colour version, please refer to the plate section.)

have surveyed broader swaths of sky, but these searches are limited to detecting extremely powerful signals from deep space.

Of course there is no guarantee that extraterrestrial civilizations would use radio waves to signal their presence. Nonetheless, there are good reasons to focus on radio waves because Earth’s atmosphere and the atmospheres of other habitable planets block X-rays, gamma radiation, and most of the infrared spectrum (Fig. 2.8). Earth’s oxygenated atmosphere is transparent to visible light, and so extraterrestrial civilizations might choose to signal their presence with pulsed optical lasers, assuming they live under similarly oxic, haze-free atmosphere.

Even if extraterrestrial civilizations choose not to deliberately broadcast, it might be possible to detect their presence through other technosignatures because all technology – no matter how advanced – produces waste heat. If a civilization decided
Joshua Krissansen-Totton and David C. Catling

to use the majority of their planet’s stellar energy, then the waste heat might be visible to infrared telescopes (Kuhn & Berdyugina, 2015). In particular, if they built a so-called Dyson sphere – a spherical structure to harness the power output of an entire star – then the waste heat could be detectable. Despite searches, no unambiguous Dyson sphere has been found (Carrigan Jr, 2009). Alternatively, if extraterrestrial civilizations construct large, non-circular structures in orbit around their host stars, then these may be detectable in unusual transits. In fact, a recent analysis of *Kepler* data revealed strange, aperiodic transits that were speculated to be a possible extraterrestrial megastructure (Wright *et al*., 2015); however, the break-up of an exocomet turns out to be a more plausible hypothesis for a variety of reasons, including the high orbital inclination of the transiting material (Boyajian *et al*., 2016).

Given the size of the cosmic haystack, it is unsurprising that SETI has not found unambiguous extraterrestrial broadcasts. There have, however, been a small number of ambiguous detections. The most intriguing of these is the appropriately named “Wow! signal”, which was a narrowband radio signal detected in 1971 by the Big Ear radio telescope in Ohio. The signal was in the ‘waterhole’ region, a frequency that terrestrial broadcasters are prohibited from using. The location of the signal in the sky does not fit with the position of any known satellite or asteroid at the time, and is unlikely to be a reflection off space debris because the debris would have had to be perfectly stationary and well beyond low Earth orbit (Ehman, 2010). The signal was recorded for 72 s, but has never been detected again in follow-up observations. The source of the signal remains a mystery.

The physicist Enrico Fermi formulated a thought experiment related to SETI. In this so-called *Fermi paradox* we suppose that a technological civilization emerged somewhere in the galaxy. If this civilization decided to colonize other stellar systems, then the time required to colonize the entire galaxy is very short compared to the age of the universe – this is true even if we assume colonization of a new star system takes a long time and traveling between the stars is slow. Perhaps not every civilization in the galaxy develops the technology or has the desire to do this, but it only takes one sufficiently motivated civilization to colonize the galaxy. The question posed by Fermi is that if intelligent life is common in the cosmos, then why aren’t they already here? Of course it is easy to dream up speculative solutions to the Fermi paradox that don’t preclude the existence of extraterrestrial civilizations (Webb, 2015): perhaps interstellar travel is prohibitively resource-intensive, perhaps the Earth is the equivalent of a cosmic zoo, or perhaps spacecraft have visited the Earth at some point in its 4.57 billion year history! Naturally, it is hard to draw any firm conclusions about the Fermi paradox.

Conventional SETI targets may be too restrictive if extraterrestrial intelligence is not biological. Most experts in the field of artificial intelligence (AI) research believe that AI will surpass human-level intelligence by the end of the century
(Baum et al., 2011). If this is true, then the length of time that biological intelligence is the dominant form of intelligence on Earth will be brief, and so any intelligent lifeforms we detect are likely to be machines (Schneider, 2016; Shostak, 2015). Machine intelligence may not need environments required by biological life, i.e. terrestrial planets in the habitable zone with surface liquid water. For example, machines might choose to live on airless planets close to their host stars to maximize solar energy. On the other hand, maintaining complex machinery may demand similar environments to biological life: abundant liquid water for manufacturing parts and a thick atmosphere to protect sensitive electronics from cosmic rays.

Despite it compelling nature, SETI has largely lagged behind other growth areas of astrobiology such as exoplanet research. The lack of government funding is a key reason. In the United States, Congress defunded NASA’s SETI program in 1993, which was an arbitrary decision given that NASA spends a significant amount of money on searching for non-intelligent life. There is no reason to exclude intelligent life, which, after all, is the most interesting kind. Perhaps very few biospheres evolve intelligence, perhaps the origin of life itself is improbable, or perhaps intelligence civilizations are quite common. Presently, all of the last four terms in the Drake equation are unknown. For that reason, we should not handicap ourselves in the search for life beyond the Earth by ignoring intelligent life.

**Conclusion**

How likely is it that we will find life beyond Earth in the coming decades? *Hubble’s* successor, the *James Webb Space Telescope (JWST)*, is scheduled for launch in 2019. This telescope is not designed specifically to study exoplanets, and so it will probably observe only one to two potentially habitable worlds for signs of life, assuming that nearby targets can be found. If life is ubiquitous then we might get lucky and detect atmospheric biosignatures with *JWST* in the early 2020s. Upcoming ground-based telescopes such as the European Extremely Large Telescope or Giant Magellan Telescope will also begin to directly image Earth-sized planets around the nearest stars in the mid-2020s. Proxima Centauri b – a habitable zone planet around our nearest star – will be a prime target for these telescopes. However, a null result with *JWST* and ground-based observations will mean little because they will only sample a handful of planets.

A likely outcome is that we will have to wait for a dedicated direct imaging mission sometime after 2025. A large space-telescope with starlight suppression technology could survey the nearest 1000 stars for planets with signs of life (Postman et al., 2008). *Kepler* data suggest that approximately 5%–10% of Sun-like stars have planets in the habitable zone (Silburt et al., 2015), and so this future telescope could survey around 50–100 habitable planets for biosignatures. If life is common
in the universe then we will find it with such a telescope. Conversely, if we find no signs of life, the inescapable conclusion is that \( f_h \), the fraction of habitable planets upon which life emerges, is very low. In fact, if we discovered that all the nearest planets were devoid of life then we might begin to suspect that the origin of life is hard.

What about the prospects for finding intelligent life in the coming decades? The first three terms in the Drake equation are known to be approximately \( R = 7 \) stars/year (Diehl et al., 2006), \( f_p = 1 \) (Cassan et al., 2012), and \( n_e = 0.1 \) (Silburt et al., 2015). Let us suppose for the moment that life is common, perhaps \( f_l = 0.5 \). Additionally, let us optimistically assume that both intelligent life and communicating intelligent life are inevitable wherever life emerges (i.e. \( f_c = f_l = 1 \)). For the final term in the Drake equation, the average lifetime of communicating civilizations, let us optimistically assume 1 million years, which is roughly the average lifetime of an animal species in the fossil record. Then it follows that the number of communicating civilizations that exist in the galaxy today is: \( N = 7 \times 1 \times 0.1 \times 0.5 \times 1 \times 1 \times 1 \times 1000000 = 350000 \). This may seem like a large number, but there are around 300 billion stars in our galaxy, meaning that even in this extremely optimistic scenario only one in a million stars hosts an intelligent, communicating civilization. Current SETI searches have surveyed the nearest few thousand stars, not nearly enough to rule out even this best-case scenario. However, the Square Kilometer Array, a huge radio telescope scheduled to begin operation in 2020, will perhaps survey up to a million nearby stars for radio broadcasts (Siemion et al., 2014). If intelligent life is common in the galaxy, there is a good chance we will find it in the coming decades. If we find nothing, then the absence of evidence will start to provide evidence of absence.

Whatever we find, future generations will have one less cosmic question to ponder when they look up at the night sky. We may find that the universe is a desolate and sterile place. This knowledge will give the stars a very different character. Despite the incomprehensible vastness of the cosmos, we will know that Earth is exceptional, that terrestrial life is unique, and that the future of all life is our responsibility. But perhaps, we will instead discover exoplanets teeming with life. If so, this will become another fact to tell children about the night sky: the stars are distant suns, they all have planets just like our Solar System, and millions of them have life just like Earth. It may even be possible to point to an individual star in the sky and say that one, that star has a planet just like Earth. It is hard to imagine another scientific discovery that would instill more wonder than this.

Notes

1 JWST can only characterize transiting Earth-like planets because its starlight suppressing coronagraph will be too inefficient to image habitable planets. Therefore it will be necessary to find a nearby, transiting habitable planet in the right position in the sky to be frequently observed.
with JWST, such as the TRAPPIST-1 system. The upcoming *Transiting Exoplanet Survey Satellite* (TESS) mission scheduled to launch in 2018 will search nearby stars for further JWST targets.

2 Strictly speaking, $n_s$ is not equal to the fraction of Sun-like stars with habitable planets, which is the quantity Silburt et al. (2015) estimate. However, for the purposes of this approximate calculation they can be assumed to be equivalent.

References


Oliver, B. (1979). Rationale for the water hole. Communication with Extraterrestrial Intelligence, 6(1–2), 71.


