Possibilities for life elsewhere in the Solar System

In our fifth supplement we have a brief survey of possibilities for life on the Moon, Mercury, Venus, and some of the outer moons in the Solar System. Odds aren't great, so here we'll focus on Jupiter's moons.

The Jovian moon system

Given that Jupiter is outside of our official habitable zone by a large factor (Jupiter is 5.2 AU away from the Sun, so it receives only 1/27 of the illumination that the Earth does), we might think that the prospects for life are bleak. However, surprisingly, some of the very best prospects for life in the Solar System outside Earth are in the Jovian system. The reason is that although Jupiter itself might not host life (although we should as always be open-minded), some of its moons seem promising.

Jupiter is the king of the planets, with two and a half times the mass of all the rest put together. One reason for its size is that it formed outside the "frost line" where water ices can form. These ice grains incorporated hydrogen (the most common element), hence a substantial amount of mass came together. Eventually the mass was large enough to start pulling in gas from the surrounding protoplanetary nebula and grow by direct gravitational accumulation instead of by the collisions between protoplanets that we think played a major role for terrestrials. One result of this is that whereas Earth is highly enriched in heavy elements such as iron and silicon, Jupiter's composition is close to that of the Sun and thus of the nebula.

Jupiter has four large moons and a host of much smaller ones (more than sixty total at last count). Galileo, who discovered the big moons, called them the "Medicean satellites" in an effort to curry favor with his patrons the Medicis. Posterity would have none of this, however, so they are collectively called the Galilean moons instead. In order from closest in to farther out, they are Io, Europa, Ganymede, and Callisto. Ganymede is the largest moon in the solar system, and in fact has a larger diameter than Mercury (although a smaller mass, due to Mercury's higher density). These four satellites are a microcosm of the solar system itself. Their average densities are greater closer in, suggesting that the heat was larger there in the protojovian disk and thus it was more difficult to capture things with lower melting points.

Of special relevance to our investigation is the orbital periods of the inner three of these. Io goes around Jupiter once every 1.77 days; Europa every 3.55 days; and Ganymede every 7.15 days. If you look at these numbers you will realize that these are very close to a 1:2:4 ratio. Coincidence? No! Indeed, this close ratio turns out to be the reason why Europa, and possibly Ganymede, are considered good prospects for life. This is due to the gravitational tidal effects of Jupiter on the satellites and the satellites on each other. This is important enough that we will now go into some detail about this system.

Tidal effects

Tides on the Earth are caused primarily by the gravity of the Moon, and secondarily by the gravity of the Sun. Basically, the part of the ocean that faces the Moon is closer to it than the Earth on average. Therefore, that part is pulled more strongly, leading to water that is slightly raised, i.e., a tide. What is less intuitive is that there is also a tide on the *opposite* side of the Earth. That water is *farther* from the Moon than the Earth on average, so the Earth on average is pulled more strongly than the far ocean. Thus the Earth is pulled away from the far part of the ocean, so the ocean has a bump there relative to the Earth.

If the Earth rotated at the same rate that the Moon orbits us, then those bulges would always be directly underneath the Moon. In reality, the Earth rotates more rapidly than the Moon orbits (a day is shorter than a month!). Therefore these bulges are carried a bit forward by the Earth's rotation. The bulge nearer the Moon will itself tug on the moon, and since that bulge is forward of the Moon's orbit the tug always increases the angular momentum of the Moon. Orbits with larger radii have larger angular momentum, meaning that this tug moves the Moon out and increases its orbital period. At the same time, Earth's rotation is slowing down (it has to, so that the whole system has a constant angular momentum). Tidal effects decrease in strength quite rapidly as distances increase (like one over the cube of the distance), so when the Earth-Moon system was formed and the Moon was closer than it is now, it moved away more rapidly than it is currently.

Now let's return to Jupiter's moons. Jupiter rotates in just 9.8 hours. That means that the tidal bulges raised by its moons always lead the moons. Therefore, just as with our Moon, the Galilean satellites are moving outwards with time.

Let's suppose that there were only one big moon around Jupiter. The tidal effects would push that moon away. They would also quickly make that moon's orbit circular, and would even more quickly make that moon's rotation period equal to its orbit, so that it always put the same face to Jupiter (as our Moon does to Earth, and indeed as Io and the gang do to Jupiter). But a circular orbit plus synchronous rotation (i.e., same face always to the host planet) means that the tidal effects would not be squeezing and kneading the moon, and hence would not produce any extra heat. As a result, we would expect Io, Europa, and Ganymede to be cold and dead because they are too small to hold onto any significant amount of their heat of formation.

Instead, Io is the most volcanically active object in the solar system. Europa isn't that extreme, but cracks in its ice indicate that there is an ocean of liquid water starting a few miles down in the ice. It is thought that even Ganymede has liquid water, albeit under an even thicker ice layer. What gives?

This is where the 1:2:4 orbital period ratio comes in; it's not a coincidence at all. Remembering that tidal effects are much stronger for closer things, Io was pushed out much faster than Europa and Ganymede. Suppose that Io started out farther inside Europa than it is currently, maybe with a 1:2.6 orbital period ratio. As it moved out this gap closed, to 1:2.4, 1:2.2, and eventually to 1:2. At this stage, Io orbits twice for every single time Europa orbits. As a result, Io and Europa give each other regular gravitational kicks, always at the same phase.

This is the recipe for a *resonance*, by which a large number of small kicks that are at the same location can build up. A familiar example is of pushing a swing. By pushing at the same place each time you can get someone going rather high quite quickly. In contrast, if you pushed at random places during the cycle you would sometimes add energy and sometimes remove it, for not much of a net effect.

The key in the Jovian moon system is that one consequence of these kicks is to make the orbits slightly eccentric instead of exactly circular. These types of resonances, in which the orbits converge towards each other, also stick: once Io and Europa were in a 1:2 resonance, they stayed in that orbital period ratio even though both of them have moved out since. Likewise, when that pair had moved out far enough, they captured Ganymede and they are now in a nice 1:2:4 setup.

The eccentricities that are introduced in this way are crucial for heating the moons. Since Io and Europa are not always at the same distance from Jupiter, the rate at which they orbit also changes and thus they can't keep the same face to Jupiter all the time. Jupiter's tidal force therefore squeezes each of them and warms them up far beyond what they could sustain normally. The same is true, to a lesser extent, for Ganymede.

With this in mind, what can we say about the likelihood of life on the Galilean moons? Io is probably too hot. Europa has liquid water and a geothermal energy source, and is probably plenty stable. This leaves the question about the chemical composition. We know that the outer part is water ice, one of the consequences of being so far from the Sun. We suspect that the inner portion is rocky, but that there is plenty of carbon. If so, then in principle I see no obstacle to there being life on Europa. For all these reasons, Europa is considered to be second only to Mars in terms of potential for life elsewhere in the solar system.

Consequences for life in the universe

Whew! After that discussion, what does it mean for life elsewhere? We obviously don't know how common such moon systems are. In particular, if Io, Europa, or Ganymede were very small moons, the effects would be negligible. Nonetheless, I think the implications are profound. Jupiter is *way* outside of our nominal habitable zone, yet here is at least one moon, possibly two, with likely substantial oceans of liquid water! In addition, the squeezing and cracking of Europa is thought by some to imply that at the bottom of this ocean one would have hydrothermal vents. We know that on Earth these host ecosystems of extremophiles, and it is possibly even more extensive on Europa because the lack of plate tectonics (the temperature is too low for magma) could mean that the locations of these vents might be much more persistent than the locations on Earth.

This may well open up a vast new range of possibilities for life. Could intelligent life emerge? I don't know. There are macroscopic life forms near vents on Earth, so complexity is clearly possible. One could also envision scenarios in which intelligent life evolved but the thick covering of ice prevented them from communicating. There is some discussion of a mission to Europa to drill through the ice and extract samples of the water to search for life. These are fairly pie in the sky at the moment; I invite you, for example, to think about how best to drill through a several mile thick layer of ice (there are ways, but it is not clear how one would then be able to get samples back!).

Giant planets in the early solar system

Before departing the outer solar system for good we need to think about what it looked like early. When we discussed the terrestrial planets we stressed that larger planets retain their internal heat more. This is even more so for gas giants, and indeed even today Jupiter radiates more than twice as much energy as it receives from the Sun. The way to think of this is that Jupiter and the other giants are basically balls of gas that are slowly settling and releasing energy as a result.

If we project this back to the first hundred million years of the solar system, at that point Jupiter was radiating more energy and indeed its giant satellites might have received more energy from Jupiter than from the Sun. Could this have meant surface liquid water for Europa during this time?

In a further what-if, we could imagine another stellar system with gas giants a few times larger than Jupiter (many of these have actually been detected). Such a giant would be an even more effective heat source for its satellites, so if the satellites are close enough and big enough they would be warmed enough to sustain liquid water for billions of years. If the satellites themselves were a factor of a few larger, they could retain an atmosphere and possibly be even more favorable to life.

The take-away point is that even in our own solar system, and surely in the universe as a whole, there are many places where life *might* exist. We do not know yet whether life *does* exist anywhere but Earth, but there is reason to stay optimistic.

Extrasolar Planets

Thus far we have concentrated on our own Solar System. This is a natural starting spot. But what if planets are exceedingly rare? This would put a damper on the prospects for abundant life in the universe. We therefore need information about how common planets are, and the properties of their orbits and of the planets themselves. The distances involved, however, make this challenging. Indeed, the first planets outside our Solar System were not discovered until 1992, and the first extrasolar planets around a normal star were not seen until 1995.

We will therefore discuss many aspects of extrasolar planets. We will begin by considering how one detects them at all. We will then point out that the available detection techniques are strongly biased; in fact, only very recently have we been able to detect Earth-sized planets around Sun-like stars. We will then discuss the systems we have detected, including the rather surprising revelation that most systems are radically different from our own. We will finish by discussing some of the future missions aimed at planet detection.

Detection of extrasolar planets

One's first thought might be that detection is straightforward: just observe a system with a telescope, and look for planets. The difficulty with this approach is twofold. First, stars are extremely far away. The closest one to the Sun is about 260,000 times farther away than the Earth is. At that distance, an Earth-Sun radius would subtend just 0.8 arcseconds, which is the angle spanned by the head of a pin at half a kilometer. The stars of interest are typically at least tens of times farther away from us, putting this beyond the reach of almost all current day telescopes.

The second problem is that stars are vastly brighter than planets. Recalling that planets shine by reflected light instead of by their own energy, the Earth would appear about a billion times dimmer than the Sun. Therefore, even if you could angularly resolve the systems, the poor planet would be drowned out. The ingenuity of instrument designers is impressive, though, and there are plans for future instruments that will try to get around this by holding up disks to block out the stellar light, or by using devices called interferometers to have such exceptional angular resolution that the stellar spill-over can be minimized.

In current observations, though, we have to try other methods. A different possibility might occur to you for systems that are specially oriented. If, from our viewpoint, the planet happens to orbit so that it crosses in front of, then behind, the star then we have an example of a (very) partial eclipse. For instance, the Earth has a radius roughly 1/100 of the Sun's, so it would block out $(1/100)^2 = 10^{-4}$ of the light. If the star were close enough and bright enough we might be able to measure this dip in intensity. With enough orbits we would be able to establish that this was periodic and not merely the result of sunspots or natural variation in light. Indeed, the very successful Kepler mission has used this method to detect more than 2,000 planetary candidates. This are called candidates instead of detections because there are some other confounding effects that can come in (e.g., sometimes there will be a background binary star).

Thus the first method that was used is the one that we still use to certify that we have an actual planet. This involves the Doppler shift. This is an effect that is familiar to anyone who has heard a siren approach then recede. The pitch appears to be high when the source of the noise approaches us, then appears to be low when it recedes. The same thing happens to light. When the source is coming towards us, the frequency increases. For example, light that is red in the rest frame of the source could appear green or blue. The opposite would happen if the source was moving away from us.

The key to the detection of planets is that the shift is greater for greater relative speed. In addition, because gravity is universal and mutual, a planet pulls on its host star as well as the other way around. The force is the same either way but the star's inertia is much greater than the planet's, meaning that while the planet moves at great speed in a large orbit the star is moving in a tiny circle and rather slowly. For example, consider the Earth-Sun system. Earth is moving in a 1 AU radius orbit at about 30 km s^{-1}. The Sun is 300,000 times more massive than the Earth, so it moves in an orbit that is only about 500 km in radius, and at only 10 cm s^{-1}. Not much!

That kind of speed is still undetectable around ordinary stars. However, the Sun's motion in response to Jupiter is a hefty 4.5 meters per second, and that *is* measurable (albeit with a lot of cleverness in the observational setup!). However, this detection method means that it is a lot easier to discover some types of planets than others. This is called a selection effect. Selection effects bedevil astronomy in many ways, so we need to understand them a bit before we can properly appreciate the context of the planets that have been discovered.

Selection effects

A selection effect is anything that makes it easier to observe one category of thing than another. It isn't limited to astronomy, either. There was a famous case of selection bias during the presidential election of 1936. Literary Digest, a well-known magazine with millions of subscribers, polled a list of 10 million car and phone owners about their presidential preferences, and got 2.3 million responses. Based on this massive sample, they confidently projected that Alf Landon was going to be the next president. They were very surprised when Franklin Delano Roosevelt wiped the floor with him in one of the most lopsided elections ever, but in retrospect they should have understood that their readership was fairly wealthy and thus not at all representative. Incidentally, the election was called correctly by young George Gallop, who used the technique of "stratified random sampling" of only 50,000 voters, by which a much smaller sample of representative voters gave a better idea of the tendencies of the population as a whole.

In astronomy we have lots of such problems. For example, pretty obviously, bright things are easier to see than dim things. In particular, they can be seen farther away. Without realizing this you might draw the conclusion that most stars are brighter than the Sun, because a simple count of the stars you can see with your naked eye indicates that this is the case. However, the much more numerous low-mass stars are so dim that you can't see them far away.

When it comes to planet detection using Doppler shifts, the issue is that large shifts are easier to detect. On top of that, it is much easier to be confident of your detection if you have many orbital periods. One cycle isn't enough to be sure that something else isn't going on, but ten regular cycles is plenty.

Combined, these effects mean that Doppler shifts are hugely skewed towards those coming from high-mass planets that orbit close to their host star. There are also some much more subtle effects. The Doppler shifts are established by noting the shifts in spectral lines. Therefore, stars that have more lines are more likely to allow planetary discoveries. Since the lines are associated with heavy elements (e.g., iron or titanium), stars with more heavy elements are more likely to have planets *discovered* around them, even if *in reality* they are no more likely to have planets around them. Keep these caveats in mind as we discuss the discoveries.

The first extrasolar planets

I am very pleased to report that the first accepted extrasolar planets were *not* discovered around an ordinary star like our Sun. Instead, their host star was a neutron star, one of my favorite objects. Neutron stars pack maybe 50% more mass than our Sun into a region that would fit inside the Capital Beltway: just 10 km in radius. They are formed from the collapsed cores of stars that begin their lives with at least eight times the mass of our Sun. Neutron stars often spin rapidly and have strong magnetic fields, with the result that they send out beams of radio waves and are called pulsars. These stars maintain their spin frequencies with incredible precision, and over short times they are a million times better clocks than the best clocks on Earth.

This high precision means that even a tiny tug from a planet is easily detectable. In fact, the pulsar in question, PSR 1257+12, has the three lowest-mass planets ever detected outside our solar system. Two of them have about 3-4 times the mass of the Earth, and one has about the mass of our Moon! These systems seem rare, though. Only two other pulsars are known that have planetary-mass objects around them. One of the planets is in a globular cluster, so the pulsar probably got the planet though complicated pilfering from another object. The other pulsar planet appears to have originally been a white dwarf that was gradually evaporated by the pulsar.

After this discovery in 1992, interest naturally turned to systems around ordinary stars. In 1995, they found one. Michel Mayor and Didier Queloz discovered that the otherwise anonymous star 51 Pegasi had a planet half the mass of Jupiter... and had an orbital period of four days!!! Mercury, the closest planet to our Sun, has an orbital period of 88 days, so the 51 Peg planet is ridiculously close. There was some skepticism initially, because models of planetary formation based on our Solar System seemed to indicate that such a system was impossible. However, with such a short period and such large Doppler shifts this discovery was confirmed quickly.

At that point the race was on. As of January 1, 2014, more than 1000 extrasolar planets have been confirmed (with at least as many more candidates from Kepler). Tens of these are multiple-planet systems, with the current record holder (the system around 55 Cancri) having five planets so far detected. We will now consider the properties of these systems compared to our Solar System, and what this might tell us about the prospects for life.

Properties of extrasolar planet systems

Let's first remind ourselves of the properties of our Solar System. We have eight official planets (ever since Pluto was demoted). The closest planet, Mercury, has an orbital period of 88 days. The next ones are Venus and the Earth, with orbital periods of respectively 225 days and 365 days. The only planets with large mass are very far away, with orbital periods of many years. The eccentricities of the orbits are also small. Mercury has the highest eccentricity, $e = 0.21$, but none of the others have eccentricities greater than $e = 0.09$ (Mars), and two planets (Venus and Neptune) have eccentricities less than 0.01. All of their orbits, even Mercury's, look like circles to the naked eye; if you viewed Mercury's orbit it would appear to be a circle in which the Sun was somewhat offset from the center.

The first surprise is how many planets there are with extremely short periods. Of the ¿1000 confirmed planets, 14 have orbital periods less than one day! Another 41 have orbital periods between one and two days. These are amazingly small. In fact, they are so close that the tides they raise on their host stars are probably bringing the planets *closer* with time. That is, the stars are probably rotating more slowly than the planets orbit, so the tidal bulge lags the planet and takes angular momentum away from it. The time for inspiral, however, is many billions of years.

The second surprise is that these very close-in planets are *big*. For example, five of the 14 with periods less than one day have masses greater than Jupiter. Doesn't this completely contradict what we said earlier in class, that big planets could only form beyond the frost line? Oddly enough, there is no contradiction. We still have strong reasons to believe that large planets can only *form* far enough away that ices can condense. However, these planets can drift inwards. They could do this by being dragged along as the disk of gas around the star moves towards the star. The currently unresolved question is what would *stop* them from drifting all the way to the star itself. Does something naturally stop them? Or could it be that many other high-mass planets were dragged in, and we only see the last one or few, which survived because by that time there was very little mass left in the disk and thus the dragging was ineffective?

The close-in planets all have very small or unmeasurable eccentricities, which is expected from the effects of tidal forces. The more distant ones, though, are a different story. Hundreds of the planets have eccentricities greater than 0.5, which really would look elliptical to your eyes. Three have eccentricities more than 0.9! Remarkably, this is a lot more consistent with numerical simulations of planet formation than our Solar System is. The reason is that towards the end, when there are just a few large protoplanets around, their mutual gravity tends to perturb each other, leading to higher eccentricities. Could it be that the relatively small eccentricities in our system are favorable to or even essential for life?

Lessons

Let us now think about these properties in light of the selection effects we discussed earlier. High-mass, short-period planets are much easier to detect than low-mass or long-period planets. None of the terrestrial planets (Mercury through Mars) could be seen, for example. It is therefore not surprising that we have this bias. As the time of observation has lengthened, so has the maximum orbital period of detected planets. We have even detected what is basically a clone of Jupiter: 0.947 Jupiter masses, orbital radius of 4.2 AU (Jupiter is at 5.2 AU), and eccentricity of 0.044 (Jupiter's is 0.048). However, Saturn, Uranus, and Neptune would not be currently detectable because their orbital periods are longer than we have been looking with sufficient precision. We can therefore safely assume that many of the systems that have only one known planet really have a lot more.

As time has gone on and detections have mounted, the fraction of stars with planets has increased. In fact, according to a study reported by Petigura et al. in the November 4, 2013 Proceedings of the National Academy of Sciences, $5.7^{+1.7}_{-2.2}$% of sun-like stars have an Earth-like planet with an orbital period between 200 and 400 days. These are thus very much in the habitable zone. The fraction of stars with slightly smaller planets could be even greater. This sure gives us some hope! In addition, as we mentioned with Europa, it may be that moons of giant planets are excellent candidates to host life.

The interest in extrasolar life has exploded, and there are many additional detection methods and planned missions. One example has to do with a cool effect called microlensing. The path of a light ray is bent by gravity. Therefore, light from a distant star can be deflected by a star in the way. If that "lens" star has a planet, the light that we see varies in a characteristic way. This has been used to detect several planets, and efforts are underway to increase the number with enhanced sensitivity.

Another method is to track the location of the star very carefully. We have emphasized Doppler shifts, but these only tell us the component of the stellar motion towards or away from us. If we had sufficiently good information about the apparent location of the star, we could see it moving in tiny circles on the sky. Indeed, this was how the very dim companion to Sirius (the brightest star in the sky) was first suspected in 1844. The technology for this is in development, and is likely to be productive within several years.

In the meantime, other firsts have been achieved. Perhaps the most exciting is that a few years ago we had a report of the spectrum of an extrasolar planet. This was possible because a close-in planet had an orbit that crossed the star from our point of view, so it was possible to see absorption lines that indicated silicate clouds. More instances of this will occur with time, although with current technology it is unlikely that we will get much more detail.

In summary, in your lifetimes we have gone from not knowing of any planets outside of the Solar System to having a vast and rapidly-expanding set of exoplanets. This tells us that planets are very common indeed. What we don't know, and what you can think about, is whether planets suitable for life are also common or whether we are very unusual in that respect.