Alan P. Boss

on the search for extrasolar planets

Astrobiology, the search for life's origins and its existence elsewhere in the universe, used to seem like a visionary dream. But in recent years, it has become a true science, thanks in part to new developments in the search for Earth-like planets outside our own solar system.

A new era in scientific discovery was initiated in 1995 with the announcement, by the Swiss astronomers Michel Mayor and Didier Queloz, of the first Jupiter-mass companion to a Sun-like star. Progress on the detection of planets outside our solar system has occurred at a breathtaking pace ever since; scarcely a month goes by without a new revelation of some sort or another. Already well over a hundred extrasolar planet candidates have been announced, and the pace of discovery promises to quicken as additional ground-based search programs swing into action. Meanwhile a number of powerful space-based observatories specifically designed to search for and characterize planets as small in mass as Earth are being planned for the next two decades.

These advances have fueled, in turn, furious theoretical work on the formation and migration mechanisms of planets inside and outside our solar system. All the extrasolar planets discovered to date appear to be gas giant planets, similar to Jupiter and Saturn, and the theory of gas giant planet formation is in flux as a result.

The amazingly short period of the first extrasolar, Jupiter-mass planet discovered brought the possibility of planet migration to the attention of theorists. The extrasolar planet orbits its host star, 51 Pegasi, in a mere 4.23 days, compared to Jupiter's leisurely 11.9-year orbit around the Sun – and, according to Kepler's third law, it orbits 51 Pegasi about a hundred times closer than Jupiter orbits the Sun. The formation of a Jupiter-mass planet so close to its parent star appears to be difficult, if not impossible, so theorists such as myself have hypothesized that some giant planets must form at larger distances and then migrate inward to their final orbital distances.

There are two very different ideas for how gas giant planets might form. Most astronomers favor the conventional theory of core accretion, where a solid core forms first and then accretes a gaseous envelope. In 1997 I proposed a very different mechanism, based on the hypothesis that a protoplanetary disk was likely to pass through a phase of marginal

Alan P. Boss, a Fellow of the American Academy since 2003, has been a research staff member at the Carnegie Institution of Washington since 1981. Internationally known for his theoretical work on the formation of stellar and planetary systems, he advises NASA on the search for extrasolar planets. He is the author of "Looking for Earths: The Race to Find New Solar Systems" (1998).

gravitational instability, where random density perturbations could lead rapidly to the growth of self-gravitating clumps of gas and dust in the disk that might survive to form giant planets. The two competing theories have very different implications for the formation environment of the solar system, and hence for the frequency of planetary systems similar to our own, for the number of habitable planets that may orbit nearby stars, and for our chances of finding another Earth-like planet outside our own solar system.

The theory of core accretion supposes the collisional accumulation of solid bodies, the process that is universally accepted as the formation mechanism of the terrestrial planets. Collisional accumulation simply means that when a swarm of particles is in orbit around a star, random collisions between these particles may lead to their sticking together to form a larger body, if they hit each other gently enough. This accumulation is thought to proceed through successively larger bodies – starting with submicron-sized dust grains, inherited from previous generations of stars, that stick together by intermolecular forces when they collide; to meter-sized boulders; on up to kilometer-sized planetesimals (comets), where self-gravity begins to become important; to lunar-sized planetary embryos; and finally to Earth-sized planets. The core accretion theory envisions this process as occurring in both the inner and outer regions of a star's planet-forming, rotationally flattened disk of gas and dust.

In the innermost region of the disk out of which our sun and solar system formed, collisional accumulation leads over the course of several tens of millions of years to the formation of Earth-sized rocky planets. In the outer region of the disk, beyond the asteroid belt, the same process is thought to lead to the formation of solid cores, equal in mass to roughly ten Earths, which may then acquire massive gaseous envelopes from the disk gas. These cores are said to form through runaway accretion, where the largest bodies grow the fastest because their self-gravity increases their collisional cross-sections; two bodies that would otherwise miss each other will hit because their mutual gravitational attraction deflects their orbits toward each other.

At an early phase, the accretion of disk gas falling onto the protoplanet causes an atmosphere to form on its growing core. As the protoplanet continues to grow by accreting disk gas and solid planetesimals, its atmosphere eventually can no longer be supported in hydrostatic equilibrium, and so it contracts. This contraction culminates in a brief period of atmospheric collapse, during which the protoplanet gains the bulk of its final mass.

At Jupiter's distance from the Sun, the timescale for the entire core accretion process is estimated to be on the order of several million years or more. Estimates of the lifetimes of planet-forming disks range from a few million years in quiescent regions of star formation, like the Taurus molecular cloud, to well under a million years in regions where the most massive stars form, such as the Orion Nebula cluster.

If there were only one solar system to explain, core accretion might be an attractive theory, because there are probably some disks that last long enough for this process to form gas giant planets. But unless the timescale for core accretion is significantly shorter than the prevailing estimates suggest, the theory seems unable to account for the observed abundance of gas giants elsewhere.

Eliminating this timescale problem is one of the main attractions of my alternative to the theory of core accretion – the theory of disk instability. The disk instability theory envisions a rapid process somewhat the opposite of core accretion. In disk instability, a clump of disk gas and dust forms first, and then the dust grains settle to the center of that clump to form a solid core. This mechanism requires a marginally gravitationally unstable disk, a disk cool enough to be on the verge of breaking up into self-gravitating spiral arms and clumps. (The inner regions of the disk gas rotate faster than the outer regions, shearing the growing clumps of gas into spiral arms.) While early theoretical models of marginally gravitationally unstable disks suggested that such disks would only form spiral arms, more recent computer models have indicated that self-gravitating clumps may form within the spiral arms and survive their subsequent orbital evolution to form gaseous protoplanets. A planet-forming disk with a mass about one tenth of the Sun's (which is the sort of disk mass that core accretion models typically assume) will be marginally gravitationally unstable, provided that the disk temperatures in the giant planet region are on the order of 50° K or less. Observations of planet-forming disks and of delicate molecular species in long-period comets seem to support such temperatures.

As a clump forms within the spiral arms, the dust grains within it begin to sediment down toward this center of the burgeoning protoplanet. This process is hastened by the coagulation of the dust grains as they migrate to the center. As a result, disk instability may be capable of forming a self-gravitating protoplanet within a time period as short as about a thousand years. For a Jupiter-mass (318 Earth masses) protoplanet containing the solar abundance (2 percent by mass) of elements heavier than hydrogen or helium, this central core could be as massive as 6 Earth masses.

The disk instability mechanism requires the presence of a strong flux of ultraviolet (UV) light to explain the formation of ice giant planets like Uranus and Neptune. (Strong fluxes of UV light occur in regions of high-mass star formation, such as the Orion and Carina Nebulas.) Intense UV light can heat up and photoevaporate the disk gas outside a critical radius from the parent star; for a solar-mass star, the critical radius, which depends on the mass of the parent star, is roughly equivalent to Saturn's. Giant gaseous protoplanets that form from clumps outside this critical orbital radius – stripped by the UV light of the bulk of their gaseous envelopes, reduced down to their solid rock and ice cores, with only thin veneers of remaining gas – will be turned into ice giants. Protoplanets inside this critical radius, meanwhile, will be largely unaffected by the UV light. This scenario explains the bulk compositions of Jupiter, Uranus, and Neptune, as well as Saturn's retention of most of its once much larger gaseous envelope. The formation of habitable planets in the inner region of the planet-forming disk would proceed more or less unfazed by this searing experience in the disk's outer region.

Earth-like planets are thought to be able to form with just about equal probability whether the gas giant planets form quickly (as in disk instability) or slowly (as in core accretion). In either case, the collisional growth of Earth-sized planets on orbits similar to Earth's requires tens of millions of years to run to completion, so the events in the first few thousand or million years are not necessarily critical to the formation of such planets. Furthermore, in either case, habitable planets should be able to form along with the gas giants. Gas giant planets are important to have around, because they shield the habitable planets from constant bombardment by residual icy planetesimals that might otherwise frustrate the origin and evolution of life.

This helps explain the importance of recent theoretical developments in the emerging field of astrobiology. Estimates of the number of technological civilizations in our galaxy are commonly based on the equation first presented by Frank Drake in 1961. Two of the many factors in the Drake equation are *fp* and *ne* – the fraction of stars with planetary systems (presumably similar to the solar system, the only known example in 1961) and the number of habitable planets per planetary system, respectively.

If my heretical theory of disk instability is correct, then *fp* can be considerably larger than conventional wisdom holds. Conventional wisdom would seem to limit the formation of solar-like planetary systems to stars formed in regions of low-mass star formation like Taurus. Core accretion presumably could not form gas giant planets in a region like Orion, because the lifetimes of the Orion disks are even shorter than those in Taurus. But if the heretical approach is correct, then solar-like planetary systems can form essentially everywhere Sun-like stars form – even in Orion. Roughly 90 percent of stars are thought to form in regions of high-mass star formation like Orion. This implies a difference in *fp* as large as about a factor of ten between the orthodoxy and the heresy.

Knowing the prevalence of habitable planets in our region of the galaxy is important for our search for other Earths. If Earths are rare, telescopes built to detect them will need to be designed differently. NASA is now in the process of designing several such telescopes, called the Terrestrial Planet Finders, intended for launch around 2015 and 2020. If all goes well with this extremely ambitious, difficult project, in a little over a decade we will know if we have any neighboring planets that are capable of supporting life – or, indeed, are actually supporting life right now. I, for one, hope that we heretics are right – that the prevalence of life elsewhere in the universe could be much greater than the conventional theory predicts.