DISMANTLING
The Nature
of Scientific Discovery

THE UNIVERSE

Richard Morris
Chapter 5

THE PRIMACY OF THEORY

One of the best examples of conflict between observation and theory is provided by the story of the gradual acceptance of the theory of continental drift. The idea that continents can move across the face of the earth is one that has been suggested many times. Anyone who looks at a map of the world cannot help being struck by the fact that the South American and African continents look like two pieces of a jigsaw puzzle. It looks as though the east coast of Brazil would fit into Africa’s inward curve perfectly.

The striking fit between the two continents was noted as long ago as 1620 by Francis Bacon. Although Bacon failed to draw the conclusion that the two continents were once connected, scientists were making precisely that suggestion by the middle of the nineteenth century. In 1857, the English zoologist Richard Owen speculated that the Atlantic Ocean had been created when the two continents had drawn apart. A year later, in 1858, the American author Antonio Snider proposed a similar theory, suggesting that the Atlantic might have been created during the Biblical Deluge. So much rain had poured down during the Flood, Snider said, that it could have caused a primeval land mass to crack, and to separate into a number of different pieces.

Snider’s catastrophic theory was not taken very seriously by
professional geologists. Nevertheless, some of them speculated that there might exist natural forces which could cause continents to split and to gradually drift apart. In 1885, the Austrian geologist Eduard Suess published a map that showed how the southern continents could be fitted together into a single supercontinent, which Suess named Gondwanaland after India’s Gondwana province. In 1908, the American geologist Frank B. Taylor suggested that continental drift might explain the origin of mountains. Continental movements, he pointed out, would create forces that would compress sections of the earth’s crust, causing mountains to rise.

Although Owen, Snider, Suess and Taylor all suggested that it was possible that continents could drift apart, it is the German meteorologist Alfred Wegener who is generally credited with being the originator of the modern theory of continental drift. Where his predecessors generally confined themselves to making suggestions about the possible validity of the idea, it was Wegener who spent years amassing evidence in an attempt to demonstrate that continental drift was a reality.

When Wegener published his book The Origin of Continents and Oceans in 1915, it had long been known that there were striking similarities between fossils found on widely separated continents. In some cases the fossils seemed to be identical. For example, fossils of Mesosaurus, a reptile that lived near the end of the Paleozoic era, about 270 million years ago, were found in Brazil and in South Africa, and nowhere else.

Wegener added to these observations by noting that there were similarities between living species as well. The lemur, for example, was found in both India and Africa. The garden snail Helix pomatia also lived on continents that were separated by oceans; it was found in western Europe and North America alike.

Wegener was aware that continental drift was not the only possible explanation for the distribution of living species. A number of others were frequently cited by biologists. Charles Darwin had suggested in On the Origin of Species that the snail might have been carried across the Atlantic Ocean on the feet of birds. Wegener did not consider this idea to be very plausible, however. In any case, it didn’t explain how larger animals had migrated. Nor was Wegener impressed by the theory, proposed by many biologists, that there had once existed land bridges between the continents.

The trouble with land bridges was that there seemed to be no way to explain their subsidence. Wegener did not believe that such bridges could have disappeared without a trace. After all, the granite that made up the continental crust was lighter than the basaltic rock of the ocean floors. If land bridges had come into existence and then disappeared, a lighter material would have had to sink into a heavier one.

In fact, the sinking of land bridges contradicted the principle of isostasy, which had been formulated by Suess around the end of the nineteenth century. According to Suess, the continents floated on the heavier material that lay beneath them; in effect, the continents were giant rafts.

To provide evidence in support of this idea, Suess had pointed out that it had been known for some time that both Canada and Scandinavia were gradually rising at about the rate of one centimeter per year. According to Suess, this could be explained by the fact that both land masses had been covered with ice during the ice age that had ended 11,000 years ago. When the ice melted, the weight that had been holding Canada and Scandinavia down was removed, and they began to rise. The process was analogous to the rising of a ship in the water when its cargo was unloaded. The only difference was that since the rock which supported continents flowed more slowly, the rise would be much more gradual.

Wegener realized that the principle of isostasy not only made the sinking of land bridges improbable, it also provided an explanation for the similarities between the fossils and the plants and animals that were found in widely separated locations. If the material in the interior of the earth was fluid enough to allow continents to rise, however slowly, then there was no reason why there should not be horizontal movement also. If continents
drifted, Wegener realized, plants and animals would not have had to migrate over wide expanses of ocean.

The fit between opposite coastlines, similarities in fossils and living organisms, and the theory of isostasy made the idea of continental drift seem plausible. But they did not provide convincing evidence for the idea. Consequently, Wegener set to work to see if he could find some other kind of evidence for continental drift.

He soon found what he was seeking. He found that the rock formations in corresponding parts of South America and Africa were similar. Furthermore, mountain ranges on opposite continents would link up with each other if a map was drawn in which the continents had been brought together. The mountains in eastern Canada were a continuation of those in Norway and Scotland. The Sierra in Argentina linked up perfectly with the Cape Mountains in South Africa. When continents on opposite sides of the Atlantic were matched up with each other, it was as though a torn page of newspaper had been pieced back together again. The “lines of print” matched up, not just in one place, but in many different places from top to bottom.

One would think that geologists would have found Wegener’s evidence convincing. Indeed, his theory of continental drift seems to have been given a reasonably sympathetic reception at first. But the scientific reaction rapidly became hostile. By the mid-1920s, Wegener’s ideas began to encounter an intense antagonism. Geophysicists, in particular, subjected the theory to vehement attack. There were no known forces, they said, that could cause continents to move in the way that Wegener postulated; the entire idea was nonsense.

Wegener had suggested that perhaps tidal forces caused by the gravitational pull of the sun and the moon, and forces associated with the earth’s rotation, might account for the movement of the continents. The geophysicists replied that Wegener had exaggerated their strength. If tidal forces were strong enough to move continents, they said, the same forces would halt the earth’s rotation within a year. As for the rotational forces—they were too weak to have any effect whatsoever.

By the late 1920s, geologists were claiming that the supposed jigsaw-puzzle-like fit between the continents on the opposite sides of the Atlantic was imaginary. The rock formations, they said, were really not as similar as Wegener had claimed. In any case, the similarity was not sufficient to prove that the continents had formerly been connected.

Wegener’s critics also attacked his credentials. He was not a professional geologist, they pointed out. On the contrary, he was an amateur who took liberties with the globe. Wegener, they said, was guilty of selecting for presentation only those facts which favored his hypothesis.

After Wegener’s death in 1930, geologists and geophysicists became even more hostile to the theory. Disdain for Wegener’s ideas became especially intense in the United States. If an American geologist did so much as express interest in the continental-drift theory, he was risking his reputation. Although Wegener’s theory continued to have some supporters in Europe, they were in the minority.

By the 1940s, the theory was considered to have been entirely discredited. When it was mentioned at all, it was used as an example of a scientific blunder. In the view of most geologists, it was absurd to consider the idea that continents could move; all the evidence seemed to point to the conclusion that the earth was rigid.

The orthodox scientists could not, however, explain away all the evidence that Wegener had amassed. If they wanted to explain the distribution of fossils on the surface of the earth, they were forced either to resurrect the concept of land bridges or to resort to hypotheses that were even more implausible. Nor could they adequately explain why such geological features as mountain ranges should exist.

During the nineteenth century, geologists had attributed the existence of mountains to the wrinkling that had presumably taken place on the earth’s surface as the planet cooled from its original molten state. But then, during the first few decades of the twentieth century, geologists began to realize that there were difficulties associated with this idea. It had become apparent that
the energy released by radioactive elements adequately explained the heat in the earth's interior. There was no need to assume that the planet had once been a molten ball. Furthermore, modern theories of planetary formation indicated that the earth had been formed from cool clouds of dust. If it had changed in size at all, it could only have expanded as the radioactive energy heated it up.

In spite of its obvious deficiencies, the rigid-earth theory became geological dogma. Geologists found it easier to live with the existence of unexplained theoretical anomalies than to accept the seemingly preposterous idea that continents could drift across the face of the earth.

But then, during the 1950s, new data about the structure of the earth's crust began to appear. Before long, a revolution in geological thinking was under way.

The revolution began during the early years of the decade, when the British physicist Patrick Maynard Stuart Blackett perfected an instrument called the magnetometer that was capable of detecting magnetic fields only one ten-millionth as intense as that of the earth. Geologists soon realized that Blackett's device could be used to study the residual magnetism of rocks that had been formed millions of years ago.

The study of the magnetic properties of ancient rocks is a branch of geology called paleomagnetism. Paleomagnetic techniques can be used to determine the direction of the earth's magnetic field at the time that the rocks were formed. A magnetization that is called fossil magnetism or natural remanent magnetism is acquired by volcanic lavas when they cool and harden. The magnetization, which is aligned with the earth's field, is remarkably stable; it will remain the same if the rock is moved, or if the geomagnetic field changes. Remanent magnetism is also exhibited by sedimentary rocks. If previously magnetized particles are deposited in a sediment, they will align themselves with the earth's field. When the sediment hardens into rock, a magnetic imprint will be retained.

When geologists measured the remanent magnetism of layers of rock in the English countryside that had been formed some 200 million years ago, they discovered that England had been situated at a latitude of 30 degrees North at that time. England is presently located at a latitude of 65 degrees. Had it migrated from one spot on the earth's surface to another over the course of 200 million years?

Although this seems to be the obvious explanation of the data, it was not the conclusion that was reached. The geologists assumed, instead, that England had remained in the same place, and that the positions of the earth's magnetic poles had shifted. At the time, it seemed easier to believe in a geomagnetic field that moved around than in shifting land masses.

Rocks of various other ages were subsequently studied, and migration paths for the North and South Poles were worked out. But the theory of magnetic pole wandering quickly ran into trouble. When paleomagnetic measurements were performed on rocks from other continents, the data did not quite agree with those obtained from the English rocks. For example, paleomagnetic measurements of rocks in North America produced a polar migration path that was somewhat different.

Geologists soon found that the only way to make the two paths coincide was to assume that England and North America had once been connected, and had gradually drifted apart. This suggested that there had never been any such thing as pole wandering at all. The relative motion that the rocks revealed was nothing other than continental drift.

Not all scientists were convinced. By this time, the continental-drift theory had been thought discredited for so long that many scientists were convinced that there must be some other explanation. It was suggested that paleomagnetic measurements might contain unknown sources of error. Or possibly the earth had had multiple poles at some time in the past; if it had, this could easily lead to misleading results.

The debate continued. When it began to seem that no firm conclusion could be reached, a seemingly outrageous new hypothesis was suddenly proposed. In 1962, Princeton University
The following year, in 1963, the British oceanographers Frederick J. Vine and Drummond H. Matthews devised a test for Hess’s theory. Vine and Matthews pointed out that there had been periodic reversals in the earth’s magnetic field. These reversals, they said, should leave an imprint on the rocks of the ocean floor. The lava that emerged from the ocean ridges would be magnetized in the direction of the field. As the material spread across the sea floor, this magnetization would be retained. But if the direction of the geomagnetic field suddenly reversed, the new material emerging from the ridges would be magnetized in the opposite direction. If the reversals took place a number of times before the material disappeared back into the earth’s interior, then the rocks that made up the ocean floor would exhibit a striped magnetic pattern. Bands of opposite magnetic polarity would be found between the ridges and the trenches. In addition, corresponding stripes would be found on both sides of a ridge. The theory of sea-floor spreading implied that the material moved away from a ridge in both directions. Therefore, the magnetization on opposite sides should exhibit a symmetrical pattern.

No one really knows why the magnetic field of the earth should periodically reverse itself. However, the fact that it does is well established. Paleomagnetic studies have shown that there have been 171 field reversals in the last 76 millions years. Once every 450,000 years or so—on the average—the north magnetic pole becomes a south pole, and vice versa. The reversals do not take place instantaneously. The magnetic field weakens to about 20 percent of its usual intensity, and remains at that level for 5,000 to 20,000 years before a reversal can be established.

But 5,000 and 20,000 years are brief periods on the geological time scale; they are small compared with the nearly half-million years that a field of a given orientation will persist. The finite amount of time required for the field to reverse itself does not blur the boundaries of the stripes appreciably.

Offhand, one might think that it would be practically impossible to measure the magnetization of the ocean beds. After all, it
would not be easy to drill through 600 meters of sediment that might lie miles below the ocean’s surface. Furthermore, if a sample could be taken to the surface, it would be hard to make sure that it had retained its original orientation.

Fortunately, nothing like this was necessary, thanks to Blackett’s magnetometer. The instrument was so sensitive that investigators could measure the magnetization simply by suspending magnetometers on long cables and towing them behind ships. The magnetic orientations of the rock could then be read by scientists on the surface. Nor did the magnetization of the sediment that covered the basaltic bedrock obscure the results. Although the sediment was magnetized too, its magnetization was only about one ten-thousandth that of the basalt—not nearly enough to interfere with the readings.

During the early 1960s, the magnetization of the floors of all the earth’s oceans was mapped. It was found that they were magnetized in the striped pattern that Vine and Matthews had predicted. The sea-floor-spreading hypothesis had been confirmed.

With the confirmation of Hess’s theory, Wegener’s concept of continental drift suddenly became respectable. By the late 1960s it was almost universally accepted. By the end of the decade, geologists took it for granted. Incorporating into it the concept of sea-floor spreading, they renamed it the theory of plate tectonics, and used it as a fundamental assumption when they did research.

This sounds like a story of a theory that was ahead of its time, which was finally vindicated when new evidence came to light. There is something very puzzling about the affair, however. Why should Wegener’s theory have been accepted so readily in the space of a few years, after it had been reviled for decades? How could scientists dismiss it as a crackpot idea, and then accept it without a murmur a few years later?

The confirmation of the sea-floor-spreading hypothesis tipped the balance in favor of the theory of continental drift. But this does not really answer the question that I have posed. After all, most of the evidence for continental drift that was available in 1970 already existed when Wegener published his book on the subject in 1915. In 1915, scientists were aware of the apparent fit between opposing continental coastlines. They had begun to realize that the theory of land bridges was not very plausible, and they could not dispute the fact that mountain ranges matched up when continents were brought together on a map. One would think that they would at least have admitted that continental drift was a possibility.

What was it about the discovery of magnetic stripes in the ocean beds that caused them to change their minds? This hardly seems to be the most convincing piece of evidence in favor of the theory of continental drift. After all, any argument that uses the magnetic data to verify drift must proceed by a very roundabout route. One must first assume that the stripes are a record of magnetic reversals which took place millions of years ago. One must conclude, next, that this is evidence for sea-floor spreading. Finally, one must assume that evidence which implies that ocean floors move tells us that continents drift also.

The last link in the chain of arguments seems to be the weakest. After all, it would presumably be possible for ocean rock to form in one place and disappear in another without causing the continental masses to move. The evidence derived from sea-floor spreading seems to be somewhat less compelling than that which Wegener assembled during the early part of the twentieth century. If we view matters objectively, we must conclude that data derived from rock formations and studies of “pole wandering” provide more direct proof for his theory.

Nevertheless, the confirmation of sea-floor spreading seems to have represented a turning point. When this hypothesis was verified, geologists suddenly became willing to accept evidence that they had discounted for decades. Wegener was suddenly transformed from a crackpot into a scientist who was ahead of his time.

The reason that this happened was that the theory of sea-floor spreading showed how continental drift was possible. When Hess’s hypothesis was confirmed, geologists realized that there existed a mechanism which could cause continental masses to
move. In 1963, there was a theory that could explain the evidence that Wegener had amassed. During the 1930s and 1940s, there was not.

In all scientific fields, theory is frequently more important than experimental data. Scientists are generally reluctant to accept the existence of a phenomenon when they do not know how to explain it. On the other hand, they will often accept a theory that is especially plausible before there exist any data to support it. This is why Copernicus’ heliocentric hypothesis was accepted long before the experiments that confirmed it could be performed.

Around the beginning of the seventeenth century, Francis Bacon put forth the view that scientific laws were generalizations from observed facts. A scientist supposedly performed experiments, pondered his results and then deduced a theory from what he had observed. Bacon’s view of science was so influential that many people believe that this is the way science operates today.

But it isn’t. Bacon’s description of “scientific method” was not accurate in his own day, and it is not accurate in ours. More often than not, it is the theory that comes first. Without a theory to guide them, scientists would not know what experiments to perform. No one thought to look for magnetic stripes in the ocean floor until Hess proposed his hypothesis of sea-floor spreading. No one would have checked to see if starlight grazing the surface of the sun was bent if Einstein had not proposed his general theory of relativity. No one would have thought to perform electron-diffraction experiments if De Broglie had not put forth his theory of matter waves.

Sometimes a theory is so convincing that it is accepted even when the facts seem to contradict it. Einstein was not impressed when Kaufmann’s experiments on electrons gave results that were inconsistent with the special theory of relativity. Copernicus and Kepler were not bothered by the fact that the arguments against a moving earth were difficult to answer. If a theory seems to have the ability to explain a wide variety of phenomena, if it possesses the ability to bring a large number of disparate elements together into a coherent picture of the universe, scientists will generally accept it as true. They will indeed plan to perform experiments to test it, but they will do so with the expectation that experimental data will provide a verification of the theory’s validity.

On the other hand, they tend to be skeptical of observations that are not supported by theory. One of the reasons that they are so ready to discount the idea of extrasensory perception is that no one has been able to come up with a theory which can explain how ESP would operate. Admittedly, the experimental evidence in support of the existence of ESP is not very good. But that is only part of the story.

Wegener’s theory of continental drift was denigrated because Wegener was unable to come up with an adequate explanation of how continents were able to move across the surface of the earth. He did amass a great deal of evidence; but without an adequate theory, that evidence did not seem very convincing.

Theories are sometimes mistaken. As a result, science sometimes blunders. Worse, the effects of the blunders occasionally persist for decades. Nevertheless, few scientists would be willing to follow Bacon’s rules of “scientific method.” If scientists did research the way that Bacon said they should, significant discoveries would become extremely rare. Theory occupies a preeminent position not because scientists tend to have their heads in the clouds, but because the scientific endeavor works best when theory is dominant.

Those scientists who do follow Bacon’s prescription often fail to discover anything that is very significant. Even such scientists as Planck who try to see what experimental data imply would not get anywhere if they did not have the insight needed to see what theoretical assumptions are necessary.

On some occasions, a confrontation between theory and observation takes place. Under such circumstances, it is generally theory that wins. One such confrontation is currently happening in the field of astronomy. So far, it is theory which seems to have the upper hand.

The controversy is one that revolves around the astronomical objects known as quasars. The American astronomer Halton Arp
has obtained observational evidence which, he claims, indicates that the accepted theoretical explanations of the quasars' brightness may be incorrect. But few astronomers accept Arp's interpretation. They tend to perceive him as a gadfly who wishes to dispute currently accepted ideas, but who is unable to come up with any good theoretical alternatives.

Quasars are bright objects that look very much like stars when they appear in astronomical photographs; the term "quasar" is short, in fact, for *quasi-stellar object*. When quasars were discovered in the early 1960s, astronomers found them very puzzling. The quasars had very large red shifts. As Hubble had emphasized in 1929, red shift and speed of recession were related. So it followed that the quasars were moving away from the earth at high velocities. This implied, in turn, that they must be very distant, billions of light-years away.

The farther away an astronomical object is, the dimmer it appears to be. This relationship holds for any luminous object; it can be applied to a candle or a light bulb as well as to a star or a galaxy. A searchlight can easily be seen at a distance of a mile, but a match flame cannot be perceived at all.

Astronomers concluded, therefore, that the quasars must be very bright. To be seen at such great distances, they had to be the most luminous objects in the universe. Some of them appeared to be shining with the brightness of a hundred galaxies. And yet the dimensions of quasars were very small. Study of astronomical data indicated that the diameter of a typical quasar was less than that of the solar system.

This discovery created problems. It was not easy to understand how something that was so small could be so bright. Some astronomers therefore proposed that perhaps the red shifts of the quasars had nothing to do with the expansion of the universe. If the quasars had been ejected from our own galaxy at high velocities, they pointed out, they could have large red shifts and yet not be very distant. The quasars, they said, looked bright only because they were nearby. This hypothetical situation would be roughly analogous to one in which a 25-watt light bulb that was very close to an observer would look brighter than a very distant searchlight.

But this theory quickly ran into trouble. Other astronomers pointed out that if quasars had been ejected from our galaxy, it was only reasonable to assume that similar objects were ejected from other galaxies also. If this was the case, it would be possible to observe quasars that were traveling toward us. These approaching quasars would exhibit blue shifts rather than red shifts. But quasar blue shifts were not observed.

Since this objection was not easily overcome, the majority of astronomers concluded that quasars were as bright and as distant as they seemed to be, and that they had existed early in the history of the universe. The last conclusion was a consequence of the fact that when astronomers look at objects that are very distant, they are also looking into the distant past. If an object is, say, 10 billion light-years away, the light that it emits must travel through space for 10 billion years before it reaches the earth.

Today astronomers generally agree that quasars are the luminous cores of young galaxies. They believe that it is possible that our own galaxy, and the galaxies near it, may have contained quasars at one stage in their evolution. A number of hypotheses have been proposed which would explain how quasars could be so bright. The most popular theory is one which proposes that quasars contain black holes which have masses a hundred million or a billion times greater than that of our sun. According to this theory, light and other kinds of radiation are given off by interstellar gas that falls into the black hole. There is every reason to believe that enormous quantities of such gas would exist in the center of a young galaxy. As this gas fell toward the event horizon of the central black hole, its gravitational energy would be transformed into heat. As it heated up, this energy would be radiated away.

This theory explains the luminosity of quasars very well. Detailed calculations show that such a mechanism could easily produce the amounts of energy that are required. The theory also
explains why quasars existed only billions of years ago. Sooner or later, the supplies of gas in the galactic core would be exhausted. When that happened, the quasar would become dimmer and finally die out. The central black hole would remain, of course, but there would be no obvious sign of its presence.

If our own galaxy once contained a quasar, then it should still have a supermassive black hole in its core. If such a black hole were found, astronomers would have some evidence in support of the theory. It would not be possible to say that the theory had been proved beyond any doubt, but at least there would be some evidence to confirm it.

Unfortunately, optical telescopes cannot see into the core of our galaxy, which is hidden by clouds of interstellar dust. And of course, black holes cannot be seen directly in any case; their presence can only be inferred from the effects they have on other matter.

Astronomers have studied radio waves and infrared radiation that comes from the center of our galaxy. But there is, as yet, no conclusive evidence for the existence of a supermassive black hole. Since it is just as difficult to observe conditions in the centers of other galaxies, the black-hole theory of quasar luminosity has not been confirmed.

However, the theory does seem very plausible. It is all the more appealing because the only other theory of quasar luminosity that seems at all credible is one that is very much like it. This is the spinar theory, which replaces the central black hole with a spinar—a supermassive star that is spinning so rapidly that centrifugal forces prevent collapse into a black hole.

Arp expresses no opinions about the plausibility of these theories. However, he does maintain that it has not been established that quasars are as far away as they seem to be. To prove his point, he has made astronomical photographs of quasars over a period of years. He has attempted to show that quasars are often linked to galaxies that have totally different red shifts.

In some of Arp's photographs, there seem to be luminous connections between galaxy and quasar. In one case, the quasar appears to lie within the outer boundary of the galaxy itself. According to Arp, these photographs indicate that some of the current ideas about quasars should be abandoned.

His argument is a simple one. If the quasars and the galaxies are connected, then they must be located at roughly the same distance from the earth. If they are the same distance away, and they have different red shifts, then it is obviously wrong to assume that there is a relationship between quasar red shift and distance. Of course, it could be the galaxy, not the quasar, that exhibited the anomalous red shift. It is not very likely, but it is possible. However, in such a case, the validity of the red-shift/distance relationship for quasars would still be uncertain.

Arp offers no theory to explain why some quasars should have anomalous red shifts. He says that it is simply an observed fact. Most astronomers disagree. Although Arp has a few supporters, most notably Geoffrey Burbidge, director of the Kitt Peak National Observatory in Arizona, the majority refuse to believe that he has discovered any evidence that is very conclusive. They claim that the connections between the members of Arp's galaxy-quasar pairs are illusionary, and point out that perspective can make two objects seem to be very close together when they really are not.

If two astronomical objects are lined up with one in front of the other, they can seem to be associated, when the distance between them is actually very great. If a quasar is situated billions of light-years behind a galaxy, and if it lies only a fraction of a degree to one side, it can appear to be connected to the galaxy in an astronomical photograph. Arp says that statistical studies indicate that galaxies and quasars appear together too often for this interpretation to be correct. His opponents reply that Arp's statistical methods give misleading results.

Orthodox astronomers claim that the bridges of light that Arp has seen between galaxies and quasars are illusionary also. Those who have attempted to verify Arp's results say that they have often experienced difficulty in making the luminous connections visible in their own photographs. In at least one case, they add,
the bridge of light turned out to be nothing more than a background galaxy that was seen edge-on.

Geoffrey Burbidge has compared the reception given Arp’s ideas to the rejection of Wegener’s theory of continental drift by geologists. The comparison is very apt. The evidence that Wegener presented was rejected because he could suggest no plausible mechanism that could cause continental drift to take place. Arp’s evidence for the existence of anomalous red shifts has been scorned because he has no theory that would explain why such red shifts should be observed. On the other hand, Arp’s opponents do have a theory; they believe that the brightness of quasars can be explained by assuming the presence of spinars or supermassive black holes.

The evidence for the existence of such objects is no more conclusive than that which Arp has presented. However, whenever there is a contest between theory and observation, it is theory which has the edge. Admittedly, theories are often overturned. But this generally happens only when a new theory appears which is more plausible than the old one. Theories are rarely discarded because anomalies of one sort or another have been discovered.

In a recent letter to The Sciences, a periodical published by the New York Academy of Sciences, Arp asserted that “The business of a scientist . . . seems to me to be to go out and observe real objects, see how they behave, and then induce generalizing principles from the observations.” This was Francis Bacon’s view of science. However, as we have seen, this is not the way that science ordinarily works. More often than not, it is theory which determines what science can observe.

It may be that Arp will eventually be vindicated, just as Wegener was. But even if he is, there will be little reason to deplore the actions of his colleagues. In rejecting evidence that cannot be fitted into any compelling theoretical scheme, the orthodox astronomers are simply behaving as scientists always have. When they say that Arp’s continued search for anomalous red shifts serves no useful scientific purpose, they are undoubt-
edly right. Even if some of the anomalous red shifts are real, photographing more of them is not likely to lead to any increase in scientific knowledge. Only a theory that explains why anomalous red shifts would exist would tell one what to look for.

Recently, a committee of astronomers which allocates observing time on the Mount Wilson, Palomar and Las Campanas telescopes recommended that Arp be denied the use of these telescopes after 1952 unless he agreed to change the direction of his research. The members of the committee pointed out that Arp had been allocated generous amounts of observing time in the past because committee members wanted to avoid giving the appearance that they were attempting to suppress an unpopular idea. It would be possible to argue that they are suppressing Arp’s ideas now. However, even if they do, their feeling that continued observations are useless when there exists no hypothesis that would allow these observations to be interpreted is probably correct. Science does not progress by observing and cataloguing strange-looking objects.

It is theory that makes observations meaningful. It imposes order upon them, and makes it possible for experimental data to be interpreted. But theories, of course, are not always correct. The insights upon which they are based are sometimes mistaken ones. After all, the intuitions of a creative scientist are no more infallible than the hunches of a compulsive gambler. As a result, acceptance of an incorrect but plausible theory sometimes causes scientists to interpret data incorrectly.

One of the most notorious examples of misinterpretation of evidence was the affair of Piltdown Man. Piltdown Man was a manufactured “fossil” that consisted of a modern human skull and the jaw of an orangutan. Both parts of the “fossil” were artificially stained to give them the appearance of great antiquity. Although the combination was a rather ludicrous one, paleontologists accepted Piltdown as a genuine human ancestor for decades. Although Piltdown Man was “discovered” in 1912, no one realized that a hoax had been perpetrated until 1953.
The Piltdown "fossil" was discovered by the English solicitor and amateur geologist Charles Dawson. To this day, no one knows whether Dawson was the perpetrator of the deception, or the unwitting dupe of the hoaxer or hoaxers. There have been a number of suggestions as to who the guilty party may have been, but no one has been able to find enough evidence to convict any of the suspects.

If Dawson was not involved in the fabrication of the fossil, he cannot really be blamed for considering it an important find. After all, he was not a professional scientist. It is somewhat more surprising that some of the most eminent paleontologists of the day should have accepted the find as genuine. After all, the Piltdown "fossil" was not even a very good forgery.

When the hoax was exposed, it was discovered that the "fossil" had been constructed from 600-year-old human skull fragments and the 500-year-old jaw of an ape. The bone had been stained with paint or a paintlike substance (Britain's National Gallery suggested that it might have been ordinary Vandyke brown). The teeth in the jaw had been filed down to give them a human appearance. But this had not been done very expertly. One tooth had been filed down a little too far. When the hoaxter discovered his mistake, he had plugged the gap with a plug of plastic material that looked remarkably like chewing gum. And when the teeth were examined under a microscope, signs of artificial abrasion could easily be seen.

Once it had been proved that the "fossil" was a forgery, it became apparent that there were numerous signs of fakery that could be seen by anyone who was looking for them. For example, two of the molars in the jaw had been filed down in such a way that they were slightly out of alignment with each other. Another tooth, a canine, showed signs of having erupted only a short time before the specimen died. It had been filed down also. As a result, it exhibited signs of wear that were obviously inconsistent with its juvenile provenance.

The suspicious character of the canine had been pointed out by a dentist shortly after Piltdown was discovered. But paleontologists paid no attention to this observation. Nor were their suspicions aroused by the fact that the jaw looked like that of a modern ape. On the contrary, they unhesitatingly placed Piltdown in the human ancestral line, and made pronouncements about his evolutionary significance.

It has been suggested that the perpetrator of the hoax never expected that his little joke would be taken so seriously, and that he was afraid to reveal his deception when it was. It may never be known whether the forgery was intended as a joke or not. However, the scientists who studied the specimen have now proved astonishingly credulous.

There are undoubtedly a number of reasons why the paleontologists of the day should have accepted the find as genuine. First, although the hoax was an inexpert one in some respects, it was carried out in an elaborate manner. Genuine fossils of extinct animals were planted in the gravel bed in which the Piltdown specimen was found in order to make the skull and jaw seem to be of great antiquity. A second Piltdown specimen—a fragmentary one consisting of parts of a human skull and a molar tooth similar to the ones in the original jaw—were planted and discovered in 1915.

International rivalry and professional jealousy also played a role. At the time of the Piltdown discovery, no human fossils of any significance had been discovered in Britain. On the other hand, numerous Neanderthal specimens had turned up in France and in Belgium, and a very primitive-looking jaw had been discovered near Heidelberg in Germany. As a result, the British paleontologists had been forced to play second fiddle to their colleagues on the Continent.

When Piltdown was discovered, all this changed. British scientists stressed the fact that Piltdown, with its apelike jaw, obviously had to be much older than the "missing links" that had been discovered in France, Belgium and Germany. It appeared that the first humans had been Englishmen after all.

However, it seems apparent that these were really not the decisive factors. It is not likely that Piltdown would have been
accepted as genuine if the "fossil" had not seemed to give striking confirmation to a theory of human evolution that was quite influential at the time. Shortly before Piltdown was discovered, the English anatomist Grafton Elliot Smith propounded a theory to the effect that the brain had been a kind of driving force in human evolution. According to Elliot Smith, the human brain had begun to evolve before human ancestors had acquired an erect posture, or lost their simian characteristics. Once the brain did begin to increase in size and in complexity, other modern human characteristics rapidly began to appear. When the Piltdown "fossil" was found, Elliot Smith was quite naturally delighted. He attempted to dispel doubts by insisting that the association of an apelike jaw with a very human-looking skull should not be surprising "to anyone familiar with recent research upon the evolution of man.

Today it is known that human ancestors began to walk upright while their brains were no larger than that of a chimpanzee. But in 1912, the "brain first" theory seemed a very attractive one. After all, the complexity of the human brain seems to be what most distinguishes us from other mammals. With no evidence to the contrary, what would be more natural than to believe that the brain somehow led the way in evolution?

When Piltdown came to light, some scientists were skeptical at first. They suggested that the skull and jaw were not fossils of the same individual. They must have come together accidentally, they said. But then Elliot Smith threw his considerable scientific weight behind the paleontologists who had pronounced Piltdown to be genuine. Before long, other important members of the British scientific establishment had followed. By the time it was discovered that they had been wrong, the experts who had insisted that the specimen be accepted as genuine had all died, or retired.

The Piltdown discovery sent science blundering off in the wrong direction. Since a hoax was involved, it could be argued that this was not a typical case. It is obvious, however, that scientists are perfectly capable of making monumental mistakes even when there is no trickery involved. As one might expect, it is normally theory that causes the blunders to be made. A good example is provided by the discovery, in 1859, of an imaginary planet.

In 1859, it had long been known that the planets in the solar system did not travel along orbits that were perfectly elliptical. By this time, it had been realized that Kepler's law was an approximation, that the gravitational attraction of one planet for another created perturbations that caused the orbits to become slightly irregular. This effect is particularly noticeable in the case of the planet Mercury. Mercury's orbit swivels around the sun so that the point of closest approach, the perihelion, is not always in the same place.

Mercury's perihelion swings around the sun at the rate of $1^\circ33'20''$ (1 degree, 33 minutes and 20 seconds) per century. Of this, $1^\circ32'37''$ can be explained by Newton's law of gravitation; calculations show that the other planets should perturb Mercury's orbit by this amount. The remaining 43 seconds of arc are due to relativistic effects; this discrepancy is explained by Einstein's general theory of relativity.

In 1859, the French astronomer Urbain Jean Joseph Leverrier used Newton's theory to calculate the effects that the other planets would have on Mercury's orbit. His calculations indicated that there existed a discrepancy of 38 seconds of arc, which is very close to the modern value of 43 seconds. The error was only a little more than 10 percent.

There was no way that Leverrier could have dreamed that Mercury's deviation from its expected orbit would eventually be explained by Einstein. As a result, he made what seemed to be a very natural assumption. He assumed that the deviation was caused by an as-yet-undiscovered planet. Leverrier called this planet "Vulcan," and hypothesized that it lay inside the orbit of Mercury. He calculated that a planet halfway between Mercury and the sun would have just the right effect if it had two-thirds of Mercury's mass.

Leverrier had every reason to believe that his methods would lead to a correct result. In 1846, he had performed similar calcu-
lations on the orbit of the planet Uranus. His predictions had led to the discovery of Neptune in the very same year.

When Leverrier announced that he had deduced the existence of an inter-Mercurial planet, a French physician and amateur astronomer named Lescarbault announced that he had sighted the planet some months earlier. When viewed in his telescope, it had seemed to be a dark spot moving across the disk of the sun.

During the next few years, Dr. Lescarbault’s discovery was confirmed by famous astronomers of a number of different nationalities. They apparently had little difficulty seeing a planet which did not exist. Furthermore, searches of astronomical records revealed that other observers had sighted the planet as early as 1802. Soon the existence of Vulcan was considered to be so well confirmed that textbooks on astronomy began to assign it a place in the solar system that was as secure as that of Mercury itself.

But astronomers soon found that it was not easy to work out the details of Vulcan’s mass and orbit. Sometimes it didn’t even seem to be there when they looked for it. At first, this didn’t trouble them much. No one had thought that observations of Vulcan would be easy. It was to be expected that a planet so close to the sun would often be blotted out by the sun’s light. After all, Mercury itself was difficult to observe at times.

As additional searches were made for Vulcan, the situation rapidly became even more confused. Astronomers looked for Vulcan during solar eclipses in the hope that the blotting out of the sun’s light would allow them to observe it clearly. Sometimes they saw (or thought they saw) it, and sometimes they didn’t. And when Vulcan was observed by different astronomers, the results did not always agree. In 1878, two noted astronomers reported that there existed not one, but two planets inside Mercury’s orbit. However, the two planets reported by one astronomer did not seem to follow the same paths as the two reported by the other. Meanwhile, other astronomers who had been observing at the same time saw nothing at all.

The problem of the orbit of Vulcan attracted even greater interest when the perfection of dry-plate photography provided astronomers with a tool that was easier to use than those which had previously been available. After the problems associated with photographing faint bodies near the sun were solved in 1900, photographic searches for Vulcan were systematically conducted whenever an eclipse of the sun occurred. But observations during the eclipses of 1900, 1901 and 1908 revealed nothing. Although reports of visual sightings did not cease entirely, most astronomers came to the conclusion that Vulcan did not exist.

But the question was not completely settled until 1916, when Einstein published his general theory of relativity, and explained the discrepancy of 43 minutes of arc. Within a few years, speculation about the possible existence of Vulcan, which had lingered on up to this time, ceased entirely. Theory had explained the irregularities in Mercury’s orbit, and had made the planet Vulcan unnecessary. The theory of 1859 had suggested that Vulcan might exist, and it was the theory of 1916 which proclaimed that the case was closed.