Articles

When Do Anomalies Begin?

ALAN LIGHTMAN AND OWEN GINGERICH

An anomaly in science is an observed fact that is difficult to explain in terms of the existing conceptual framework. Anomalies often point to the inadequacy of the current theory and herald a new one. It is argued here that certain scientific anomalies are recognized as anomalies only after they are given compelling explanations within a new conceptual framework. Before this recognition, the peculiar facts are taken as givens or are ignored in the old framework. Such a "retrorecognition" phenomenon reveals not only a significant feature of the process of scientific discovery but also an important aspect of human psychology.

IN ANY EXAMINATION OF HOW SCIENTIFIC THEORIES CHANGE over time, "anomalies" enter the discussion. The word anomaly has a venerable astronomical usage, going back to the Greek, meaning a celestial motion that deviates from simple uniformity. In Latin, it frequently designated any deviation from a regular law of grammar. In English, the word gradually took on the meaning of any deviation from the expected natural order, well exemplified by the *Oxford English Dictionary*'s 1873 citation from Charles Darwin: "There is no greater anomaly in nature than a bird that cannot fly."

Anomalies are particularly helpful in understanding the scientific process, for they point to the inadequacies of an old model and emphasize the merits of the new. In these terms, an anomalous fact is one that is unexpected and difficult to explain within an existing conceptual framework. For example, the inadequacy of classical electrodynamics in the atomic domain was indicated by a number of anomalies found in the early 1900s, such as the behavior of electrons in metals, and the stability and emission of electron shells in Rutherford's nuclear model of the atom. These phenomena were later given compelling explanations by the new quantum theory.

In his seminal study of the scientific process, *The Structure of Scientific Revolutions*, Thomas Kuhn described scientific discovery as a complex process in which an "anomalous" fact of nature is recognized and then followed by a change in conceptual framework (paradigm) that makes the new fact no longer an anomaly. As Kuhn described it, "Discovery commences with the awareness of anomaly, that is, with the recognition that nature has somehow violated the pre-induced expectations that govern normal science" (*1*, pp. 52–53).

But when do anomalies begin? We will argue that certain scientific anomalies are recognized only after they are given compelling explanations within a new conceptual framework. In some cases, an anomalous fact may be unquestioned or accepted as a given in the old paradigm. In others, the anomaly may be noted by a small segment of the scientific community but not widely regarded as important or legitimized until a good explanation is at hand in a new paradigm. The development of this class of anomalies we call the "retrorecognition" phenomenon. We will give several examples of retrorecognition.

The Flatness Problem

According to the Big Bang model, the leading theory of modern cosmology, the universe began in an explosion about 10 billion years ago. Since that violent beginning, the universe has been expanding and cooling. As it expands, its parts attract each other gravitationally, and that attraction slows down the expansion. The competition between the outward motion of expansion and the inward pull of gravity leads to three possibilities. The universe may expand forever, with its outward motion always overwhelming the inward pull of gravity. Such a universe is called "open." A second possibility is that the inward force of gravity is sufficiently strong to halt and reverse the expansion. Such a universe is called "closed." The final possibility, a "flat" universe, lies exactly midway between a closed and open universe and is analogous to a rock thrown upward with precisely the minimum speed that ensures its escape from the pull of Earth. (In Einstein's theory of gravity, open and closed universes have curved, non-Euclidean geometries, whereas a flat universe has a noncurved, Euclidean geometry.)

The Big Bang model allows any of the three possibilities. Which one holds for our universe depends on the manner in which the cosmic expansion began, or, in particular, the initial gravity relative to the initial rate of expansion. In other terms, the fate of the universe was determined by its initial gravitational energy relative to its initial kinetic energy of expansion. Even without knowledge of these initial conditions, we can infer the fate of our universe by comparing its present gravitational energy with its present kinetic energy of expansion. If the magnitude of the first of these two energies is greater, the universe is closed, fated to collapse at some time in the future. If the second is greater, the universe is open, fated to expand forever. If the magnitudes of the two energies are precisely equal, the universe is flat. The ratio of magnitudes of the two energies is Ω = (gravitational energy)/(kinetic energy). Thus, the universe is closed, flat, or open depending on whether Ω is greater than one, equal to one, or less than one, respectively.

Current measurements of Ω give it a value of about 0.1 (*2*). Although the measurements are difficult and may be revised, cosmologists feel certain that the value of Ω lies between 0.1 and 10. As we will see, such a range is surprisingly close to unity.

Now comes the flatness problem: Why is Ω so close to one so long after the universe began? It follows from the Big Bang model that, as time goes on, Ω differs more and more from one, unless it started out exactly one. In an open universe, Ω begins less than one and gets smaller in time; in a closed universe, Ω begins bigger than one and gets larger in time. Only in a flat universe does Ω begin and remain one. Finding the universe today with its gravitational energy so closely balanced with its kinetic energy of expansion is analogous to

A. Lightman is in the Program in Writing and Humanistic Studies and in the Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139. O. Gingerich is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138.

finding a rock thrown upward from Earth, far from Earth, still moving outward but at a tiny speed, having neither fallen back to Earth nor escaped Earth altogether. Such a situation would require the rock's initial kinetic energy of motion to have been extraordinarily close to its initial gravitational energy at launch.

The real issue behind the flatness problem is the value of Ω in the early universe. Physicists believe that the initial conditions of the cosmos were set when the universe was about 10^{-43} s old, the era of "quantum gravity." In order for the value of Ω to lie between 0.1 and 10 today, 10 billion years after the quantum era, after the universe has expanded in size by a factor of more than 10^{30}, the initial value of Ω had to lie between about $1 + 10^{-59}$ and $1 - 10^{-59}$. Equivalently, the kinetic energy of expansion and the gravitational energy of the cosmos had to be initially balanced to within one part in 10^{59}. It is important to add that the Big Bang model has nothing to say about the initial conditions of the universe. In particular, the model does not require any special value for the initial ratio of gravitational energy to kinetic energy. Yet to many scientists today, it seems unlikely that so fine an initial balance, as required by the observations, could have been merely an accident. Thus, there is no "natural" explanation for the balance in the Big Bang model. The extremely close balance of the two energies is an anomaly.

The flatness problem was first raised by Robert Dicke of Princeton University in 1969 (*3*). For a number of years afterward, however, few cosmologists considered the observed value of Ω a serious anomaly, an observed fact that required a physical explanation. Some scientists, for example, regarded the initial value of Ω as a given or accidental property of our universe and saw no difficulty with the near flatness of the cosmos; it was perhaps a philosophical enigma but certainly not a legitimate scientific problem. Typifying this viewpoint are Margaret Geller of the Harvard-Smithsonian Center for Astrophysics and Robert Wagoner of Stanford University. According to Geller, "the flatness problem has always seemed to me like an argument of religion rather than an argument of science. Because the universe is one realization. It's one system. So how can you talk about a priori probabilities?" (*4*, p. 368). Wagoner says, "I don't think any of these arguments [for or against the naturalness of Ω being so close to one] are relevant because I think they are philosophical. Let observation decide what Ω is" (*4*, p. 181). Other cosmologists paid no attention at all to the flatness problem, and some briefly considered it but then dismissed it because they had no good solutions to it.

The attitude of many scientists toward the flatness problem changed after 1981, when Alan Guth of Massachusetts Institute of Technology proposed a significant addition to the Big Bang model called the inflationary universe model (*5*). According to calculable physical processes described by new "grand unified" theories of physics, the matter and energy in the infant universe existed in a peculiar state, behaving as if they had repulsive gravity and resulting in a very brief period of extremely rapid cosmic expansion. One of the consequences of the inflationary epoch expansion was that, whatever its initial value, Ω would have been driven to a value extremely close to one. Thus, the inflationary universe model gives a natural solution to the flatness problem. The inflationary expansion, and the physics underlying it, provided a mechanism to achieve an extremely close balance between the kinetic and gravitational energies of the infant universe.

The questionable status of the flatness problem before the inflationary universe model is evident in Guth's paper, where he devotes an entire appendix to arguing that the problem is real and significant. According to astrophysicist Marc Davis of Berkeley, "I have to say that I was so impressed with the inflationary model because it had promoted the horizon [and flatness] problems to tractable problems. . . . The reason that the flatness problem wasn't wholly compelling [before the inflationary model] was that we couldn't really justify why Ω started off [near] one in the first place. . . . Unless you have a dynamical argument, you're arguing about nonphysical questions" (*4*, pp. 352 and 354). In the words of physicist Charles Misner at the University of Maryland, "I just couldn't see how to play with those equations, and so I didn't come on board thinking [the flatness problem] was serious until the inflationary models came out. Later, I developed a strong preference for the flat universe, feeling that the Dicke paradox [the flatness problem] suggested it. The key point for me was that inflation offers an explanation. . . . What was crucial was that the inflationary universe [model] provided an example that turned the Dicke paradox into a standard physics problem" (*4*, pp. 240–241).

Today, Misner and many other cosmologists consider the close balance of kinetic and gravitational energies to be one of the most significant observational facts of the universe, whether or not the inflationary universe model itself survives the test of time. Before the new paradigm of the inflationary universe model, only a handful of cosmologists considered the close balance of energies to be a serious anomaly in the standard Big Bang model.

The Perigee-Opposition Problem

The contemporary reactions to the flatness problem have a fascinating parallel with a cosmological revolution that took place four and a half centuries ago, when Nicholas Copernicus (1473–1543) introduced the heliocentric planetary system. The principal challenge for the astronomers of antiquity and the Renaissance was to account for the seemingly irregular motions of the planets among the stars, especially the so-called retrograde motion, in which a planet appears temporarily to reverse its eastward motion against the background of stars as seen from Earth. In the sun-centered system of Copernicus, this phenomenon is easily explained. When the swifter moving Earth bypasses the slower moving Mars, for example, Mars temporarily appears to move backward.

Precisely the same observed phenomenon was explained 1400 years earlier in the geocentric system of Claudius Ptolemy (A.D. 140). To account for the retrograde motion, Ptolemy proposed that each planet moved in a small circle, called the epicycle, which in turn rode on a larger circle centered on Earth (Fig. 1A). The compounded circles produced an occasional reverse motion.

But there is more. It is a basic observational fact, known since antiquity, that retrograde motion occurs only around the time when the sun is in a direct line with the planet. For the superior planets, Mars, Jupiter, and Saturn, the sun must lie opposite the planet in the sky, hence the designation "opposition." In particular, and this was especially obvious for Mars, the planet was observed to be brightest, and therefore presumably closest to Earth, during the time of retrogression.

In a sun-centered system, it is a simple geometrical truth that the middle of the retrograde motion, and the planet's closest approach, must coincide with opposition, when the sun, Earth, and planet lie in a straight line. But in an Earth-centered system, such a coincidence is not required by the geometry. A planet at the moment of opposition could, a priori, lie at any position on its epicycle (Fig. 1B). (Only at perigee, at the bottom of the epicycle, would the planet be in retrogression.) Alternatively, in the middle of retrograde motion, the planet-Earth line and the sun-Earth line could a priori form any angle at all (Fig. 1C). To explain the observations, Ptolemy had to assume that each superior planet revolved in its epicycle at just the right rate so that it reached perigee at the moment of opposition on every orbit (Fig. 1D). We know that pre-

Copernican astronomers were aware of these observational facts because the Alfonsine planetary tables, made early in the 14th century, took advantage of the solar connections, even though astronomers rarely mentioned the fact explicitly. Thus, a striking observational fact that would later have a completely natural explanation in the heliocentric system of Copernicus had to be accepted as a given, without explanation, in the geocentric system of Ptolemy.

For centuries, no one, not even Copernicus, remarked on the oddness of Ptolemy's tacit assumption regarding perigee and opposition. It was an astronomer in the generation after Copernicus, Gemma Frisius (1508–1555), who first recognized the assumption as a problem. Gemma wrote (*6*, p. 42):

> While at first glance the Ptolemaic hypotheses may seem more plausible than Copernicus', nevertheless the former are based on not a few absurdities, not only because the stars are understood to be moved nonuniformly in their circles, but also because they do not have explanations for the phenomena as clear as those of Copernicus. For example, Ptolemy assumes that the three superior planets in opposition—diametrically opposite the sun—are always in the perigees of their epicycles, that is, a "fact-in-itself." In contrast, the Copernican hypotheses necessarily infer the same thing, but they demonstrate a "reasoned fact."

The perigee-opposition phenomenon was recognized as an anomaly in the Earth-centered framework only after it was given a "reasoned" explanation in the new sun-centered framework.

The Continental-Fit Problem

As a third example of the retrorecognition phenomenon, consider the remarkable similarity of shapes of the opposite coasts on the two sides of the Atlantic. South America and Africa, in particular, are shaped as if they were two fitting pieces of a jigsaw puzzle. We believe today that the two continents were once joined and part of a single landmass, which subsequently split and drifted apart. In such a framework, the good fit of continents on opposite sides of the Atlantic is easy to explain. However, the fit is without explanation in

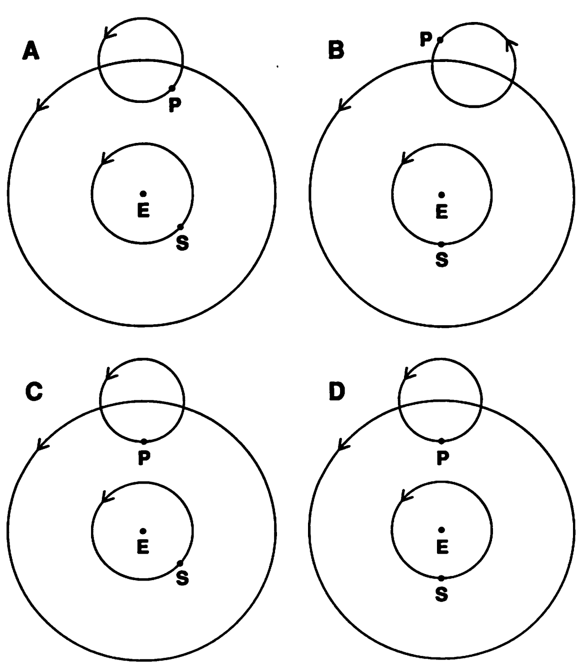

Fig. 1. (**A**) Generic case of the Earth-centered planetary system with epicycles. (**B**) Planet in opposition but not at the perigee of its epicycle. (**C**) Planet at the perigee of its epicycle but not at opposition. (**D**) Planet simultaneously at perigee and at opposition.

the previous conceptual framework, which held that landmasses could move only vertically.

The remarkable fit of the continents could have been noticed soon after the Atlantic Ocean had been mapped, certainly by the early 17th century (7). Around 1800, the German naturalist and geographer Alexander von Humboldt (1769–1859) proposed that the lands bordering the Atlantic were once joined. His suggestion was not taken seriously. Half a century later the French scientist Antonio Snider-Pellegrini, using fossil evidence as well as the fit of the shapes, claimed that the continents were once joined. Again, the proposal, which in this case was accompanied by a rather preposterous mechanism, was not taken seriously by the majority of scientists. In 1881, Reverend Osmond Fisher, English scientist and author of perhaps the earliest textbook on geophysics, discussed a geological mechanism to explain the good fit of the continents. He was largely ignored. Belief in the fixity of continents held fast.

In 1912, the German geophysicist Alfred Wegener (1880–1930) analyzed the situation much more carefully and included geological and fossil evidence to argue for an ancient continuity of the landmasses, which then broke apart and drifted away from each other (*8*). Wegener called his theory "continental drift." Although additional evidence for continental drift began accumulating, the hypothesis was highly controversial until the mid-1960s, when patterns of magnetism in rocks on the ocean floor became convincing. Then, in the late 1960s, the theory of plate tectonics was developed. This theory, for the first time, provided a persuasive mechanism by which the continents could move horizontally, namely, the existence of a series of "plates" on which the continents sit. The slow, convective flows within Earth's mantle force neighboring plates apart, carrying along the continents piggyback. Given the mechanism provided by the theory of plate tectonics and the evidence for that theory, the framework of continental drift has become accepted and has replaced the previous framework of the fixity of continents.

What was for Wegener a clear anomaly in need of a reasoned explanation had been for the great majority of geologists just a curiosity, scarcely even a puzzle awaiting a solution. Only after the paradigm changed was the fit of the continents seen as an anomaly pointing toward a major new way of looking at the stability of continental arrangements.

The Adaptation-of-Organism Problem

As a fourth example, we turn to biology. For centuries, naturalists have marveled at the exquisite specificity and adaptation of organisms to their environment. Camels carry their energy-storing fat all in one place, on their backs; thus, the rest of their bodies are not blanketed by a thick layer of fat and so can efficiently cool off in the arid deserts where camels live. The long necks of giraffes allow the animals to eat from the high trees in their environment. Pandas have a thumb-like sixth digit, which they use for stripping the leaves off the bamboo shoots in the mountains of western China. And so on.

Before the mid-19th century, most naturalists and many others took such adaptation as evidence of a grand design, evidence of an intelligent and powerful creator, and they explained the situation accordingly. For example, in his *The Wisdom of God Manifested in the Works of the Creation* British naturalist John Ray (1627–1705) wrote "because it is the great design of Providence to maintain and combine every Species, I shall take notice of the great Care and abundant Provision that is made in securing this End" (*9*, p. 133). A clear statement of this view can be also found in Jean Jacques Rousseau's (1712–1778) *Profession of Faith of a Savoyard Vicar* (*10*, pp. 259 and 261):

How much sophistry does it not require to disavow the harmony of created beings and that admirable order in which all the parts of the system concur to the preservation of each other? . . . it is impossible for me to conceive that a system of beings can be so wisely regulated, without the existence of some intelligent cause which effects such regulation. . . . I believe, therefore, that the world is governed by a wise and powerful Will.

In this prevailing "creationist" framework, which included belief in the fixity of species, the perfect adaptation of organisms to their environment was both natural and expected. However, some organisms are not so adapted. Charles Darwin (1809–1882), in *The Origin of the Species*, cited a number of examples. There are the ducks with feet designed for swimming that do not swim (*11*, p. 177):

He who believes that each being is created as we now see it must have occasionally felt surprise when he has met with an animal having habits and structure not in agreement. What can be plainer than that the webbed feet of ducks and geese are formed for swimming? Yet there are upland geese with webbed feet which rarely go near the water.

There are the many animals that live in dark caves and are blind. Why should these animals have eyes if they are not needed? The cave rat (*Neotoma*), for example, has (blind) eyes that are lustrous and large. There are the birds, like the 300-pound ostrich or the penguin, that do not fly. Why have wings and not fly?

And there are so many perfect habitats that are uninhabited (*11*, p. 401):

The general absence of frogs, toads, and newts on so many true oceanic islands cannot be accounted for by their physical conditions: indeed it seems that islands are peculiarly fitted for these animals; for frogs have been introduced into Madeira, the Azores, and Mauritius, and have multiplied so as to become a nuisance. . . . But why, on the theory of creation, they should not have been created there, it would be very difficult to explain.

At the end of the last passage, Darwin pointed out that nonadaptations are anomalies in the creationist framework. Yet, these anomalies went unrecognized until Darwin's new theory of adaptation, natural selection. Because natural selection requires the evolution of organisms, it explains both adaptation and nonadaptation. Organisms with traits suitable for survival in a particular environment live to yield offspring, continue their line, and produce a descendant population adapted to that environment. But organisms continue to evolve and change habitats, so that a particular trait that was formerly beneficial, like the webbed feet of upland ducks, may be no longer beneficial, although still inherited. Traits not important for survival are not as strongly subject to the forces of natural selection and thus may appear unsuited to a particular environment at a particular time.

The Equality of Inertial and Gravitational Mass

As our final example, we consider the equality of inertial and gravitational mass. The first mass resists a body's change in motion whereas the second determines its gravitational force. It is the equality of these two masses that causes bodies of different masses or different materials to fall with the same acceleration in a gravitational field, a long-observed fact. Indeed, in 1592 Galileo wrote in his *De Motu* (*12*, p. 48):

The variation of speed in air between balls of gold, lead, copper, porphyr, and other heavy material is so slight that in a fall of 100 cubits [about 46 m] a ball of gold would surely not outstrip one of copper by as much as four fingers. Having observed this, I came to the conclusion that in a medium totally void of resistance all bodies would fall with the same speed.

In Newtonian physics, the inertial mass and gravitational mass are regularly canceled against each other. Newton himself was perplexed by this extraordinary equality between quantities that seemed conceptually very different, and he went to considerable lengths to establish their experimental equivalence. For example, Newton recognized that a pendulum was a case in which both types of mass played a role and that the equality of swings of pendula with different bobs would measure the equality of the two masses to high accuracy. Referring to his experiments timing the periods of pendula of different materials, Newton says in his *System of the World* (*13*, p. 568):

I tried the thing in gold, silver, lead, glass, sand, common salt, wood, water, and wheat. I provided two equal wooden boxes. I filled the one with wood, and suspended an equal weight of gold (exactly as I could) in the center of oscillation of the other. The boxes, hung by equal threads of 11 feet, made a couple of pendulums perfectly equal in weight and figure, and equally exposed to the resistance of the air: and, placing the one by the other, I observed them to play together forwards and backwards for a long while, with equal vibrations. And therefore the quantity of matter [inertial mass] in the gold was to the quantity of matter in the wood as the action of the motive force [gravitational mass] upon all the gold to the action of the same upon all the wood; that is, as the weight of the one to the weight of the other.

In his law for the gravitational force, Newton simply equated the inertial and gravitational masses without anything other than observational justification. There was no essential reason within the theory itself as to why these two quite different masses should be equal. They were simply assumed to be so, much as Ptolemy had assumed that the epicyclic and orbital phases would be exactly synchronized for the three superior planets or modern cosmologists had assumed that the value of Ω started off extremely close to one.

After Newton, the equality of inertial and gravitational mass was verified with greater and greater accuracy. In the late 19th century, Lorant Eotvos, a Hungarian baron, announced that his studies with plumb bobs showed that the acceleration of gravity on different objects could not differ by more than a few parts in a billion (*14*). Despite the extraordinary accuracy with which the equality of the two masses was verified, scientists continued to accept that equality as a given, without recognizing it as an anomaly in Newton's theory of gravity.

It was not until Albert Einstein's new theory of gravity, general relativity, that a fundamental explanation was given for the equality of inertial and gravitational mass. Indeed, Einstein saw this equality, which was a part of his "equivalence principle," as a profound statement about the nature of gravity, and he constructed his entire theory around it. In the resulting theory, gravity is understood as a geometrical phenomenon, with the equality of the two masses a fundamental and necessary part of that picture. General relativity was an entirely new theory, with new predictions. For example, as a consequence of the equivalence principle, the bending of light by a gravitating body may be quantitatively explained. And, for the first time, it was realized that Newton's theory, and indeed all previous theories, had failed to account adequately for the equality of inertial and gravitational mass. As Einstein wrote in 1911, while struggling to develop his new theory of gravity (*15*, p. 100),

This experience, of the equal falling of all bodies in the gravitational field, is one of the most universal which the observation of nature has yielded; but in spite of that the law has not found any place in the foundations of our edifice of the physical universe.

Characterization of the Retrorecognition Phenomenon

The five examples given above follow a similar pattern:

1) A fact of nature is observed in the context of an existing explanatory framework.

2) The fact does not have a logical explanation in the existing framework but is nevertheless unquestioned and ignored, or accept-

ed as a given property of the world, or simply postulated to be true.

3) A new theory or model is advanced in which the observed fact now has a compelling and reasoned explanation. At the same time, the fact is retroactively recognized as an anomaly in the context of the old theory or model.

We might borrow the language of Gemma Frisius (*6*) by referring to facts taken as givens as "facts-in-themselves" and to facts logically explained as "reasoned facts." In this language, step 2 involves understanding the observed fact as a fact-in-itself, whereas in step 3, with the emergence of a new paradigm, the fact-in-itself is transformed into a reasoned fact. For the class of anomalies that we are considering, it is only in step 3 that the anomaly is recognized. Of course, in the new paradigm, the fact in question is no longer an anomaly.

The terms "fact-in-itself" and "reasoned fact" used by Gemma Frisius were actually taken from Aristotle's system of logic, the *Posterior Analytics*, where Aristotle distinguishes between the *to oti* (fact-in-itself) and the *di oti* (reasoned fact) (*16*). The assumptions that the coincidence of retrograde motion and opposition of planets is an accident or that the fit of the continents is an accident might be regarded as "explanations" of these observed facts. But these assumptions are not reasoned explanations—they do not have the logical force of the explanations easily provided by the sun-centered astronomical system or the principle of natural selection. And the anomaly in the old framework is not recognized as an anomaly until the reasoned explanation of the new. The term "retrorecognition" actually stands for recognition after a reasoned explanation.

We have described a special class of scientific anomalies. In fact, there is a continuum of kinds of scientific anomalies, ranging from those that initially draw no concern whatever, like the perigee-opposition problem, to those that are soon recognized as serious and perhaps fatal to the existing model, such as Ernest Rutherford's discovery that alpha particles shot at atoms sometimes scatter backwards, thus demolishing the "plum pudding" atomic model in which the positive and negative charges are distributed diffusely throughout the same volume.

Even within the class of anomalies discussed here, the situations are not identical. No explanation at all was initially proposed for the perigee-opposition problem or for the equality of intertial and gravitational mass. For the continental-fit problem, between 1800 and 1960 some scientists proposed various theories of continental drift, but in the absence of a mechanism the proposals were not taken seriously. Not surprisingly, scientists strongly prefer explanations that are mechanistic, logical, and calculable.

The flatness problem is perhaps the most complicated of the examples we have considered. Unlike the other examples, the new paradigm, the inflationary universe model, is by no means universally accepted among practicing cosmologists, nor is the legitimacy of the flatness problem. However, since the inflationary universe model was proposed, many more cosmologists recognize the peculiarity of the observational facts.

Discussion

There are several factors at work in the retrorecognition phenomenon, their relative importance varying with the specific example and the particular group of scientists reacting to that example: (i) the intellectual difficulty of recognizing anomalies initially, (ii) the tendency to ignore a problem when one has no idea how to solve it, and (iii) the conservatism of science. By definition, many retrorecognition anomalies go unnoticed initially, are not seen as requiring explanation, and are not appreciated as anomalous. Thus, it is hard to document them.

In the case of the flatness problem, for example, some scientists (exemplified by the comments of Misner) did not regard the problem as serious because they had no good ideas about how to solve it. By contrast, the perigee-opposition problem was not recognized as a problem to begin with.

Science is a conservative activity, and scientists are reluctant to change their explanatory frameworks. As discussed by sociologist Bernard Barber, there are a variety of social and cultural factors that lead to conservatism in science, including commitment to particular physical concepts, commitment to particular methodological conceptions, professional standing, and investment in particular scientific organizations (*17*). Although such conservatism may seem inflexible and ultimately destructive, it has the short-term asset of allowing each current conceptual framework to be articulated so clearly that it is well understood and can serve as an organizing principle for the multitude of facts that scientists observe. Furthermore, it may be intellectually difficult to recognize the importance of each of these multitude of facts and to spot the one peculiar fact that heralds a fundamental flaw with the current theory.

Scientists may also be reluctant to change paradigms for the purely psychological reasons that the familiar is often more comfortable than the unfamiliar and that inconsistencies in belief are uncomfortable. In his *Theory of Cognitive Dissonance*, psychologist Leon Festinger says that "the existence of dissonance [inconsistency], being psychologically uncomfortable, will motivate the person to try to reduce the dissonance and achieve consonance [consistency]. When dissonance is present, in addition to reducing it, the person will actively avoid situations and information which would likely increase the dissonance" (*18*, p. 3).

We suggest that the phenomenon discussed here—the recognition of some anomalies only after they are given reasoned explanations by a new conceptual framework—is in some cases an extreme example of the conservatism of science. At times, scientists may be so resistant to replacing their current paradigm that they cannot acknowledge certain facts as anomalous. To be sure, such facts are observed and recorded. The ancient Greeks duly noted that the superior planets were in retrograde motion and brightest at opposition; naturalists cataloged the many varied characteristics of animals and plants; astronomers in this century carefully measured the close balance between expansion energy and gravitational energy of the cosmos; geographers noted the remarkable fit of the continents; physicists measured the equal rates of acceleration of falling bodies. But these anomalous facts, and others like them, were not initially recognized as anomalies. If unexplained facts can be glossed over or reduced in importance or simply accepted as givens, the possible inadequacy of the current theory does not have to be confronted. Then, when a new theory gives a compelling explanation of the previously unexplained facts, it is "safe" to recognize them for what they are.

REFERENCES AND NOTES

1. T. S. Kuhn, *The Structure of Scientific Revolutions* (Univ. of Chicago Press, Chicago, ed. 2, 1970).
2. For a discussion of the standard Big Bang model, see S. Weinberg, *The First Three Minutes* (Basic Books, New York, 1977). For a review of new ideas in cosmology, including recent observational results, discussions of Ω, the flatness problem, and the inflationary universe model, see A. Lightman, *Ancient Light* (Harvard Univ. Press, Cambridge, MA, 1991).
3. R. H. Dicke, *Gravitation and the Universe: The Jayne Lectures for 1969* (American Philosophical Society, Philadelphia, 1970), p. 62.
4. A. Lightman and R. Brawer, *Origins: The Lives and Worlds of Modern Cosmologists* (Harvard Univ. Press, Cambridge, MA, 1990).
5. A. Guth, *Phys. Rev. D* **23**, 347 (1981).
6. Reiner Gemma Frisius, in Johannes Stadius, *Ephemerides Novae et Auctae* (Cologne, 1560), si. b3-b3v; trans. by O. Gingerich and R. S. Westman, *The Wittich Connection: Conflict and Priority in Late Sixteenth-Century Cosmology, Trans. Am. Philos. Soc.* **78**, part 7, 42 (1988).

7. Francis Bacon is often cited as having pointed out the unusually good fit of the continents, in a passage in *The New Organon* (1620), but a close examination of the passage suggests that Bacon was probably referring only to the similarity of shapes of two western coasts, rather than to the fit of an east coast with a west coast.
8. For a comprehensive history of the theory of continental drift, see U. B. Marvin, *Continental Drift* (Smithsonian Institution Press, Washington, DC, 1973). For a discussion of Wegener, see N. Oreskes, *Hist. Stud. Phys. Biol. Sci.* **18**, 311 (1988).
9. J. Ray, *The Wisdom of God Manifested in the Works of the Creation* (1704) (Samuel Smith, London, 1704).
10. J. J. Rousseau, *Profession of Faith of a Savoyard Vicar* (1765), trans. in *Harvard Classics*, C. W. Eliot, Ed. (P. F. Collier and Son, New York, 1910), vol. 34.
11. C. Darwin, *The Origin of Species* (1859) (Collier Books, New York, 1962).
12. Galileo, *De Motu* (1592), trans. in *Galileo on Motion and on Mechanics* by I. E. Drabkin and S. Drake (Univ. of Wisconsin Press, Madison, 1960).
13. I. Newton, *System of the World*, F. Cajori, Ed. (Univ. of California Press, Berkeley, 1934), sect. 19.
14. R. V. Eotvos, *Math. Naturw. Ber. Ungarn* **8**, 65 (1889).
15. A. Einstein, *Ann. Phys.* **35**, 898 (1911), trans. in *The Principle of Relativity* by H. A. Lorentz, A. Einstein, H. Minkowski, and H. Weyl (Dover, New York, 1952).
16. Aristotle, *Posterior Analytics*, 1.13. In the Middle Ages the fact-in-itself was called a *quia* and the reasoned fact was called a *propter quid*.
17. B. Barber, *Science* **134**, 596 (1961).
18. L. Festinger, *A Theory of Cognitive Dissonance* (Stanford Univ. Press, Stanford, CA, 1957), p. 3.
19. For helpful discussion and comments we thank R. Brawer, S. Brush, P. Galison, S. J. Gould, G. Holton, H. Margolis, U. Marvin, A. Pickering, H. Ritvo, and F. Sulloway.

Deformational Mass Transport and Invasive Processes in Soil Evolution

GEORGE H BRIMHALL,* OLIVER A. CHADWICK, CHRIS J. LEWIS, WILLIAM COMPSTON, IAN S. WILLIAMS, KATHY J. DANTI, WILLIAM E. DIETRICH, MARY E. POWER, DAVID HENDRICKS, JAMES BRATT

Soils are differentiated vertically by coupled chemical, mechanical, and biological transport processes. Soil properties vary with depth, depending on the subsurface stresses, the extent of mixing, and the balance between mass removal in solution or suspension and mass accumulation near the surface. Channels left by decayed roots and burrowing animals allow organic and inorganic detritus and precipitates to move through the soil from above. Accumulation occurs at depths where small pores restrict further passage. Consecutive phases of translocation and root growth stir the soil; these processes constitute an invasive dilatational process that leads to positive cumulative strains. In contrast, below the depth of root penetration and mass additions, mineral dissolution by descending organic acids leads to internal collapse under overburden load. This softened and condensed precursor horizon is transformed into soil by biological activity, which stirs and expands the evolving residuum by invasion by roots and macropore networks that allows mixing of materials from above.

SOILS AND WEATHERED BEDROCK FORM THE BASAL PORTION of open biogeochemical ecosystems at the interface of the atmosphere, biosphere, hydrosphere, and subaerial lithosphere (*1*). In an effort to define the role of soils in global change, researchers have considered soils to act as the earth's geomembrane (*2*), with some behavior analogous to that of biomembranes of living organisms (*3*). The term geomembrane implies that interfacial transport processes are regulated and raises the central questions of how transport occurs in soils, how organisms in the soil community (*4, 5*) influence transport processes, and how soils decompose rocks. In this article we examine the theory that vertical evolution of soil formed in place (not on steep hillslopes) is a consequence of spontaneous dynamic interaction of the biota and waste-product organic acids with rock minerals derived both from the underlying rock and from eolian sources. This interaction is governed by nonequilibrium thermodynamics, mechanics, mass transport, and ground-water flow in the rock-soil-plant system. We quantify and interpret modes and rates of interaction among organisms, transported materials, and the minerals that compose weathered rocks in these invasive interfacial systems.

Mass fluxes between different portions of chemical weathering and soil forming systems are particularly useful monitors of near surface transport processes. They serve as natural chemical tracers indicative of the extent of erosion (*6*), source, pathway, and reservoir regions (*7*), and can be related directly to observed soil features (*8–11*). Because mass fluxes in soils are computed from mass conservation volume-density-composition relations, it is imperative to evaluate the effects of volume change. Deformation and buckling of concrete sidewalks by root growth provide vivid and familiar evidence that volume changes attributable to common stresses do occur near the surface in soils.

Mass balance modeling techniques yielding chemical gains and losses that attend chemical weathering and soil formation have evolved considerably in their treatment of volume change. Initially, and until recently, isovolumetric reactions were assumed (*6, 12–14*)

G. H Brimhall, C. J. Lewis, K. J. Danti, and W. E. Dietrich are with the Department of Geology and Geophysics, University of California, Berkeley, CA 94720. O. A. Chadwick is with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109. W. Compston and I. S. Williams are with the Research School of Earth Sciences, Australian National University, Canberra ACT 2601, Australia. M. E. Power is in the Department of Integrative Biology, University of California, Berkeley, CA 94720. D. Hendricks is in the Department of Soil and Water Science, University of Arizona, Tucson, AZ 85721. J. Bratt is with BHP-Utah International Metals, 550 California Street, San Francisco, CA 94104.

*To whom correspondence should be addressed.