

VII. Crisis and the Emergence of Scientific Theories

All the discoveries considered in Section VI were causes of or contributors to paradigm change. Furthermore, the changes in which these discoveries were implicated were all destructive as well as constructive. After the discovery had been assimilated, scientists were able to account for a wider range of natural phenomena or to account with greater precision for some of those previously known. But that gain was achieved only by discarding some previously standard beliefs or procedures and, simultaneously, by replacing those components of the previous paradigm with others. Shifts of this sort are, I have argued, associated with all discoveries achieved through normal science, excepting only the unsurprising ones that had been anticipated in all but their details. Discoveries are not, however, the only sources of these destructive-constructive paradigm changes. In this section we shall begin to consider the similar, but usually far larger, shifts that result from the invention of new theories.

Having argued already that in the sciences fact and theory, discovery and invention, are not categorically and permanently distinct, we can anticipate overlap between this section and the last. (The impossible suggestion that Priestley first discovered oxygen and Lavoisier then invented it has its attractions. Oxygen has already been encountered as discovery; we shall shortly meet it again as invention.) In taking up the emergence of new theories we shall inevitably extend our understanding of discovery as well. Still, overlap is not identity. The sorts of discoveries considered in the last section were not, at least singly, responsible for such paradigm shifts as the Copernican, Newtonian, chemical, and Einsteinian revolutions. Nor were they responsible for the somewhat smaller, because more exclusively professional, changes in paradigm produced by the wave theory of light, the dynamical theory of heat, or Maxwell's electromagnetic theory. How can theories like these arise from normal

science, an activity even less directed to their pursuit than to that of discoveries?

If awareness of anomaly plays a role in the emergence of new sorts of phenomena, it should surprise no one that a similar but more profound awareness is prerequisite to all acceptable changes of theory. On this point historical evidence is, I think, entirely unequivocal. The state of Ptolemaic astronomy was a scandal before Copernicus' announcement.¹ Galileo's contributions to the study of motion depended closely upon difficulties discovered in Aristotle's theory by scholastic critics.² Newton's new theory of light and color originated in the discovery that none of the existing pre-paradigm theories would account for the length of the spectrum, and the wave theory that replaced Newton's was announced in the midst of growing concern about anomalies in the relation of diffraction and polarization effects to Newton's theory.³ Thermodynamics was born from the collision of two existing nineteenth-century physical theories, and quantum mechanics from a variety of difficulties surrounding black-body radiation, specific heats, and the photoelectric effect.⁴ Furthermore, in all these cases except that of Newton the awareness of anomaly had lasted so long and penetrated so deep that one can appropriately describe the fields affected by it as in a state of growing crisis. Because it demands large-scale paradigm destruction and major shifts in the problems and techniques of normal science, the emergence of new theories is generally preceded by a period of pronounced professional in-

¹ A. R. Hall, *The Scientific Revolution, 1500-1800* (London, 1954), p. 16.

² Marshall Clagett, *The Science of Mechanics in the Middle Ages* (Madison, Wis., 1959), Parts II-III. A. Koyré displays a number of medieval elements in Galileo's thought in his *Etudes Galiléennes* (Paris, 1939), particularly Vol. I.

³ For Newton, see T. S. Kuhn, "Newton's Optical Papers," in *Isaac Newton's Papers and Letters in Natural Philosophy*, ed. I. B. Cohen (Cambridge, Mass., 1958), pp. 27-45. For the prelude to the wave theory, see E. T. Whittaker, *A History of the Theories of Aether and Electricity*, I (2d ed.; London, 1951), 94-109; and W. Whewell, *History of the Inductive Sciences* (rev. ed.; London, 1847), II, 396-466.

⁴ For thermodynamics, see Silvanus P. Thompson, *Life of William Thomson Baron Kelvin of Largs* (London, 1910), I, 266-81. For the quantum theory, see Fritz Reiche, *The Quantum Theory*, trans. H. S. Hatfield and H. I. Brose (London, 1922), chaps. i-ii.

security. As one might expect, that insecurity is generated by the persistent failure of the puzzles of normal science to come out as they should. Failure of existing rules is the prelude to a search for new ones.

Look first at a particularly famous case of paradigm change, the emergence of Copernican astronomy. When its predecessor, the Ptolemaic system, was first developed during the last two centuries before Christ and the first two after, it was admirably successful in predicting the changing positions of both stars and planets. No other ancient system had performed so well; for the stars, Ptolemaic astronomy is still widely used today as an engineering approximation; for the planets, Ptolemy's predictions were as good as Copernicus'. But to be admirably successful is never, for a scientific theory, to be completely successful. With respect both to planetary position and to precession of the equinoxes, predictions made with Ptolemy's system never quite conformed with the best available observations. Further reduction of those minor discrepancies constituted many of the principal problems of normal astronomical research for many of Ptolemy's successors, just as a similar attempt to bring celestial observation and Newtonian theory together provided normal research problems for Newton's eighteenth-century successors. For some time astronomers had every reason to suppose that these attempts would be as successful as those that had led to Ptolemy's system. Given a particular discrepancy, astronomers were invariably able to eliminate it by making some particular adjustment in Ptolemy's system of compounded circles. But as time went on, a man looking at the net result of the normal research effort of many astronomers could observe that astronomy's complexity was increasing far more rapidly than its accuracy and that a discrepancy corrected in one place was likely to show up in another.⁵

Because the astronomical tradition was repeatedly interrupted from outside and because, in the absence of printing, communication between astronomers was restricted, these dif-

⁵ J. L. E. Dreyer, *A History of Astronomy from Thales to Kepler* (2d ed.; New York, 1953), chaps. xi-xii.

ficulties were only slowly recognized. But awareness did come. By the thirteenth century Alfonso X could proclaim that if God had consulted him when creating the universe, he would have received good advice. In the sixteenth century, Copernicus' co-worker, Domenico da Novara, held that no system so cumbersome and inaccurate as the Ptolemaic had become could possibly be true of nature. And Copernicus himself wrote in the Preface to the *De Revolutionibus* that the astronomical tradition he inherited had finally created only a monster. By the early sixteenth century an increasing number of Europe's best astronomers were recognizing that the astronomical paradigm was failing in application to its own traditional problems. That recognition was prerequisite to Copernicus' rejection of the Ptolemaic paradigm and his search for a new one. His famous preface still provides one of the classic descriptions of a crisis state.⁶

Breakdown of the normal technical puzzle-solving activity is not, of course, the only ingredient of the astronomical crisis that faced Copernicus. An extended treatment would also discuss the social pressure for calendar reform, a pressure that made the puzzle of precession particularly urgent. In addition, a fuller account would consider medieval criticism of Aristotle, the rise of Renaissance Neoplatonism, and other significant historical elements besides. But technical breakdown would still remain the core of the crisis. In a mature science—and astronomy had become that in antiquity—external factors like those cited above are principally significant in determining the timing of breakdown, the ease with which it can be recognized, and the area in which, because it is given particular attention, the breakdown first occurs. Though immensely important, issues of that sort are out of bounds for this essay.

If that much is clear in the case of the Copernican revolution, let us turn from it to a second and rather different example, the crisis that preceded the emergence of Lavoisier's oxygen theory of combustion. In the 1770's many factors combined to generate

⁶ T. S. Kuhn, *The Copernican Revolution* (Cambridge, Mass., 1957), pp. 135-43.

a crisis in chemistry, and historians are not altogether agreed about either their nature or their relative importance. But two of them are generally accepted as of first-rate significance: the rise of pneumatic chemistry and the question of weight relations. The history of the first begins in the seventeenth century with development of the air pump and its deployment in chemical experimentation. During the following century, using that pump and a number of other pneumatic devices, chemists came increasingly to realize that air must be an active ingredient in chemical reactions. But with a few exceptions—so equivocal that they may not be exceptions at all—chemists continued to believe that air was the only sort of gas. Until 1756, when Joseph Black showed that fixed air (CO_2) was consistently distinguishable from normal air, two samples of gas were thought to be distinct only in their impurities.⁷

After Black's work the investigation of gases proceeded rapidly, most notably in the hands of Cavendish, Priestley, and Scheele, who together developed a number of new techniques capable of distinguishing one sample of gas from another. All these men, from Black through Scheele, believed in the phlogiston theory and often employed it in their design and interpretation of experiments. Scheele actually first produced oxygen by an elaborate chain of experiments designed to dephlogisticate heat. Yet the net result of their experiments was a variety of gas samples and gas properties so elaborate that the phlogiston theory proved increasingly little able to cope with laboratory experience. Though none of these chemists suggested that the theory should be replaced, they were unable to apply it consistently. By the time Lavoisier began his experiments on airs in the early 1770's, there were almost as many versions of the phlogiston theory as there were pneumatic chemists.⁸ That

⁷ J. R. Partington, *A Short History of Chemistry* (2d ed.; London, 1951), pp. 48-51, 73-85, 90-120.

⁸ Though their main concern is with a slightly later period, much relevant material is scattered throughout J. R. Partington and Douglas McKie's "Historical Studies on the Phlogiston Theory," *Annals of Science*, II (1937), 361-404; III (1938), 1-56, 337-71; and IV (1939), 337-71.

proliferation of versions of a theory is a very usual symptom of crisis. In his preface, Copernicus complained of it as well.

The increasing vagueness and decreasing utility of the phlogiston theory for pneumatic chemistry were not, however, the only source of the crisis that confronted Lavoisier. He was also much concerned to explain the gain in weight that most bodies experience when burned or roasted, and that again is a problem with a long prehistory. At least a few Islamic chemists had known that some metals gain weight when roasted. In the seventeenth century several investigators had concluded from this same fact that a roasted metal takes up some ingredient from the atmosphere. But in the seventeenth century that conclusion seemed unnecessary to most chemists. If chemical reactions could alter the volume, color, and texture of the ingredients, why should they not alter weight as well? Weight was not always taken to be the measure of quantity of matter. Besides, weight-gain on roasting remained an isolated phenomenon. Most natural bodies (e.g., wood) lose weight on roasting as the phlogiston theory was later to say they should.

During the eighteenth century, however, these initially adequate responses to the problem of weight-gain became increasingly difficult to maintain. Partly because the balance was increasingly used as a standard chemical tool and partly because the development of pneumatic chemistry made it possible and desirable to retain the gaseous products of reactions, chemists discovered more and more cases in which weight-gain accompanied roasting. Simultaneously, the gradual assimilation of Newton's gravitational theory led chemists to insist that gain in weight must mean gain in quantity of matter. Those conclusions did not result in rejection of the phlogiston theory, for that theory could be adjusted in many ways. Perhaps phlogiston had negative weight, or perhaps fire particles or something else entered the roasted body as phlogiston left it. There were other explanations besides. But if the problem of weight-gain did not lead to rejection, it did lead to an increasing number of special studies in which this problem bulked large. One of them, "On

phlogiston considered as a substance with weight and [analyzed] in terms of the weight changes it produces in bodies with which it unites," was read to the French Academy early in 1772, the year which closed with Lavoisier's delivery of his famous sealed note to the Academy's Secretary. Before that note was written a problem that had been at the edge of the chemist's consciousness for many years had become an outstanding unsolved puzzle.⁹ Many different versions of the phlogiston theory were being elaborated to meet it. Like the problems of pneumatic chemistry, those of weight-gain were making it harder and harder to know what the phlogiston theory was. Though still believed and trusted as a working tool, a paradigm of eighteenth-century chemistry was gradually losing its unique status. Increasingly, the research it guided resembled that conducted under the competing schools of the pre-paradigm period, another typical effect of crisis.

Consider now, as a third and final example, the late nineteenth century crisis in physics that prepared the way for the emergence of relativity theory. One root of that crisis can be traced to the late seventeenth century when a number of natural philosophers, most notably Leibniz, criticized Newton's retention of an updated version of the classic conception of absolute space.¹⁰ They were very nearly, though never quite, able to show that absolute positions and absolute motions were without any function at all in Newton's system; and they did succeed in hinting at the considerable aesthetic appeal a fully relativistic conception of space and motion would later come to display. But their critique was purely logical. Like the early Copernicans who criticized Aristotle's proofs of the earth's stability, they did not dream that transition to a relativistic system could have observational consequences. At no point did they relate their views to any problems that arose when applying Newtonian theory to nature. As a result, their views died with

⁹ H. Guerlac, *Lavoisier—the Crucial Year* (Ithaca, N.Y., 1961). The entire book documents the evolution and first recognition of a crisis. For a clear statement of the situation with respect to Lavoisier, see p. 35.

¹⁰ Max Jammer, *Concepts of Space: The History of Theories of Space in Physics* (Cambridge, Mass., 1954), pp. 114–24.

them during the early decades of the eighteenth century to be resurrected only in the last decades of the nineteenth when they had a very different relation to the practice of physics.

The technical problems to which a relativistic philosophy of space was ultimately to be related began to enter normal science with the acceptance of the wave theory of light after about 1815, though they evoked no crisis until the 1890's. If light is wave motion propagated in a mechanical ether governed by Newton's Laws, then both celestial observation and terrestrial experiment become potentially capable of detecting drift through the ether. Of the celestial observations, only those of aberration promised sufficient accuracy to provide relevant information, and the detection of ether-drift by aberration measurements therefore became a recognized problem for normal research. Much special equipment was built to resolve it. That equipment, however, detected no observable drift, and the problem was therefore transferred from the experimentalists and observers to the theoreticians. During the central decades of the century Fresnel, Stokes, and others devised numerous articulations of the ether theory designed to explain the failure to observe drift. Each of these articulations assumed that a moving body drags some fraction of the ether with it. And each was sufficiently successful to explain the negative results not only of celestial observation but also of terrestrial experimentation, including the famous experiment of Michelson and Morley.¹¹ There was still no conflict excepting that between the various articulations. In the absence of relevant experimental techniques, that conflict never became acute.

The situation changed again only with the gradual acceptance of Maxwell's electromagnetic theory in the last two decades of the nineteenth century. Maxwell himself was a Newtonian who believed that light and electromagnetism in general were due to variable displacements of the particles of a mechanical ether. His earliest versions of a theory for electricity and

¹¹ Joseph Larmor, *Aether and Matter . . . Including a Discussion of the Influence of the Earth's Motion on Optical Phenomena* (Cambridge, 1900), pp. 6–20, 320–22.

magnetism made direct use of hypothetical properties with which he endowed this medium. These were dropped from his final version, but he still believed his electromagnetic theory compatible with some articulation of the Newtonian mechanical view.¹² Developing a suitable articulation was a challenge for him and his successors. In practice, however, as has happened again and again in scientific development, the required articulation proved immensely difficult to produce. Just as Copernicus' astronomical proposal, despite the optimism of its author, created an increasing crisis for existing theories of motion, so Maxwell's theory, despite its Newtonian origin, ultimately produced a crisis for the paradigm from which it had sprung.¹³ Furthermore, the locus at which that crisis became most acute was provided by the problems we have just been considering, those of motion with respect to the ether.

Maxwell's discussion of the electromagnetic behavior of bodies in motion had made no reference to ether drag, and it proved very difficult to introduce such drag into his theory. As a result, a whole series of earlier observations designed to detect drift through the ether became anomalous. The years after 1890 therefore witnessed a long series of attempts, both experimental and theoretical, to detect motion with respect to the ether and to work ether drag into Maxwell's theory. The former were uniformly unsuccessful, though some analysts thought their results equivocal. The latter produced a number of promising starts, particularly those of Lorentz and Fitzgerald, but they also disclosed still other puzzles and finally resulted in just that proliferation of competing theories that we have previously found to be the concomitant of crisis.¹⁴ It is against that historical setting that Einstein's special theory of relativity emerged in 1905.

These three examples are almost entirely typical. In each case a novel theory emerged only after a pronounced failure in the

¹² R. T. Glazebrook, *James Clerk Maxwell and Modern Physics* (London, 1896), chap. ix. For Maxwell's final attitude, see his own book, *A Treatise on Electricity and Magnetism* (3d ed.; Oxford, 1892), p. 470.

¹³ For astronomy's role in the development of mechanics, see Kuhn, *op. cit.*, chap. vii.

¹⁴ Whittaker, *op. cit.*, I, 386-410; and II (London, 1953), 27-40.

normal problem-solving activity. Furthermore, except for the case of Copernicus in which factors external to science played a particularly large role, that breakdown and the proliferation of theories that is its sign occurred no more than a decade or two before the new theory's enunciation. The novel theory seems a direct response to crisis. Note also, though this may not be quite so typical, that the problems with respect to which breakdown occurred were all of a type that had long been recognized. Previous practice of normal science had given every reason to consider them solved or all but solved, which helps to explain why the sense of failure, when it came, could be so acute. Failure with a new sort of problem is often disappointing but never surprising. Neither problems nor puzzles yield often to the first attack. Finally, these examples share another characteristic that may help to make the case for the role of crisis impressive: the solution to each of them had been at least partially anticipated during a period when there was no crisis in the corresponding science; and in the absence of crisis those anticipations had been ignored.

The only complete anticipation is also the most famous, that of Copernicus by Aristarchus in the third century B.C. It is often said that if Greek science had been less deductive and less ridden by dogma, heliocentric astronomy might have begun its development eighteen centuries earlier than it did.¹⁵ But that is to ignore all historical context. When Aristarchus' suggestion was made, the vastly more reasonable geocentric system had no needs that a heliocentric system might even conceivably have fulfilled. The whole development of Ptolemaic astronomy, both its triumphs and its breakdown, falls in the centuries after Aristarchus' proposal. Besides, there were no obvious reasons for taking Aristarchus seriously. Even Copernicus' more elaborate proposal was neither simpler nor more accurate than Ptolemy's system. Available observational tests, as we shall see more clear-

¹⁵ For Aristarchus' work, see T. L. Heath, *Aristarchus of Samos: The Ancient Copernicus* (Oxford, 1913), Part II. For an extreme statement of the traditional position about the neglect of Aristarchus' achievement, see Arthur Koestler, *The Sleepwalkers: A History of Man's Changing Vision of the Universe* (London, 1959), p. 50.

ly below, provided no basis for a choice between them. Under those circumstances, one of the factors that led astronomers to Copernicus (and one that could not have led them to Aristarchus) was the recognized crisis that had been responsible for innovation in the first place. Ptolemaic astronomy had failed to solve its problems; the time had come to give a competitor a chance. Our other two examples provide no similarly full anticipations. But surely one reason why the theories of combustion by absorption from the atmosphere—theories developed in the seventeenth century by Rey, Hooke, and Mayow—failed to get a sufficient hearing was that they made no contact with a recognized trouble spot in normal scientific practice.¹⁶ And the long neglect by eighteenth- and nineteenth-century scientists of Newton's relativistic critics must largely have been due to a similar failure in confrontation.

Philosophers of science have repeatedly demonstrated that more than one theoretical construction can always be placed upon a given collection of data. History of science indicates that, particularly in the early developmental stages of a new paradigm, it is not even very difficult to invent such alternates. But that invention of alternates is just what scientists seldom undertake except during the pre-paradigm stage of their science's development and at very special occasions during its subsequent evolution. So long as the tools a paradigm supplies continue to prove capable of solving the problems it defines, science moves fastest and penetrates most deeply through confident employment of those tools. The reason is clear. As in manufacture so in science—retooling is an extravagance to be reserved for the occasion that demands it. The significance of crises is the indication they provide that an occasion for retooling has arrived.

¹⁶ Partington, *op. cit.*, pp. 78-85.

VIII. The Response to Crisis

Let us then assume that crises are a necessary precondition for the emergence of novel theories and ask next how scientists respond to their existence. Part of the answer, as obvious as it is important, can be discovered by noting first what scientists never do when confronted by even severe and prolonged anomalies. Though they may begin to lose faith and then to consider alternatives, they do not renounce the paradigm that has led them into crisis. They do not, that is, treat anomalies as counterinstances, though in the vocabulary of philosophy of science that is what they are. In part this generalization is simply a statement from historic fact, based upon examples like those given above and, more extensively, below. These hint what our later examination of paradigm rejection will disclose more fully: once it has achieved the status of paradigm, a scientific theory is declared invalid only if an alternate candidate is available to take its place. No process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature. That remark does not mean that scientists do not reject scientific theories, or that experience and experiment are not essential to the process in which they do so. But it does mean—what will ultimately be a central point—that the act of judgment that leads scientists to reject a previously accepted theory is always based upon more than a comparison of that theory with the world. The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgment leading to that decision involves the comparison of both paradigms with nature *and* with each other.

There is, in addition, a second reason for doubting that scientists reject paradigms because confronted with anomalies or counterinstances. In developing it my argument will itself foreshadow another of this essay's main theses. The reasons for doubt sketched above were purely factual; they were, that is,