VI. Anomaly and the Emergence of Scientific Discoveries

Normal science, the puzzle-solving activity we have just examined, is a highly cumulative enterprise, eminently successful in its aim, the steady extension of the scope and precision of scientific knowledge. In all these respects it fits with great precision the most usual image of scientific work. Yet one standard product of the scientific enterprise is missing. Normal science does not aim at novelties of fact or theory and, when successful, finds none. New and unsuspected phenomena are, however, repeatedly uncovered by scientific research, and radical new theories have again and again been invented by scientists. History even suggests that the scientific enterprise has developed a uniquely powerful technique for producing surprises of this sort. If this characteristic of science is to be reconciled with what has already been said, then research under a paradigm must be a particularly effective way of inducing paradigm change. That is what fundamental novelties of fact and theory do. Produced inadvertently by a game played under one set of rules, their assimilation requires the elaboration of another set. After they have become parts of science, the enterprise, at least of those specialists in whose particular field the novelties lie, is never quite the same again.

We must now ask how changes of this sort can come about, considering first discoveries, or novelties of fact, and then inventions, or novelties of theory. That distinction between discovery and invention or between fact and theory will, however, immediately prove to be exceedingly artificial. Its artificiality is an important clue to several of this essay's main theses. Examining selected discoveries in the rest of this section, we shall quickly find that they are not isolated events but extended episodes with a regularly recurrent structure. Discovery commences with the awareness of anomaly, i.e., with the recognition that nature has somehow violated the paradigm-induced expectations that govern normal science. It then continues with a more or less extended exploration of the area of anomaly. And it closes only when the paradigm theory has been adjusted so that the anomalous has become the expected. Assimilating a new sort of fact demands a more than additive adjustment of theory, and until that adjustment is completed—until the scientist has learned to see nature in a different way—the new fact is not quite a scientific fact at all.

To see how closely factual and theoretical novelty are intertwined in scientific discovery examine a particularly famous example, the discovery of oxygen. At least three different men have a legitimate claim to it, and several other chemists must, in the early 1770's, have had enriched air in a laboratory vessel without knowing it.[1] The progress of normal science, in this case of pneumatic chemistry, prepared the way to a breakthrough quite thoroughly. The earliest of the claimants to prepare a relatively pure sample of the gas was the Swedish apothecary, C. W. Scheele. We may, however, ignore his work since it was not published until oxygen's discovery had repeatedly been announced elsewhere and thus had no effect upon the historical pattern that most concerns us here.[2] The second in time to establish a claim was the British scientist and divine, Joseph Priestley, who collected the gas released by heated red oxide of mercury as one item in a prolonged normal investigation of the "airs" evolved by a large number of solid substances. In 1774 he identified the gas thus produced as nitrous oxide and in 1775, led by further tests, as common air with less than its usual quantity of phlogiston. The third claimant, Lavoisier, started the work that led him to oxygen after Priestley's experiments of 1774 and possibly as the result of a hint from Priestley. Early in

[1] For the still classic discussion of oxygen's discovery, see A. N. Meldrum, *The Eighteenth-Century Revolution in Science—the First Phase* (Calcutta, 1930), chap. v. An indispensable recent review, including an account of the priority controversy, is Maurice Daumas, *Lavoisier, théoricien et expérimentateur* (Paris, 1955), chaps. ii–iii. For a fuller account and bibliography, see also T. S. Kuhn, "The Historical Structure of Scientific Discovery," *Science*, CXXXVI (June 1, 1962), 760–64.

[2] See, however, Uno Bocklund, "A Lost Letter from Scheele to Lavoisier," *Lychnos*, 1957-58, pp. 39-62, for a different evaluation of Scheele's role.

1775 Lavoisier reported that the gas obtained by heating the red oxide of mercury was "air itself entire without alteration [except that] . . . it comes out more pure, more respirable."[3] By 1777, probably with the assistance of a second hint from Priestley, Lavoisier had concluded that the gas was a distinct species, one of the two main constituents of the atmosphere, a conclusion that Priestley was never able to accept.

This pattern of discovery raises a question that can be asked about every novel phenomenon that has ever entered the consciousness of scientists. Was it Priestley or Lavoisier, if either, who first discovered oxygen? In any case, when was oxygen discovered? In that form the question could be asked even if only one claimant had existed. As a ruling about priority and date, an answer does not at all concern us. Nevertheless, an attempt to produce one will illuminate the nature of discovery, because there is no answer of the kind that is sought. Discovery is not the sort of process about which the question is appropriately asked. The fact that it is asked—the priority for oxygen has repeatedly been contested since the 1780's—is a symptom of something askew in the image of science that gives discovery so fundamental a role. Look once more at our example. Priestley's claim to the discovery of oxygen is based upon his priority in isolating a gas that was later recognized as a distinct species. But Priestley's sample was not pure, and, if holding impure oxygen in one's hands is to discover it, that had been done by everyone who ever bottled atmospheric air. Besides, if Priestley was the discoverer, when was the discovery made? In 1774 he thought he had obtained nitrous oxide, a species he already knew; in 1775 he saw the gas as dephlogisticated air, which is still not oxygen or even, for phlogistic chemists, a quite unexpected sort of gas. Lavoisier's claim may be stronger, but it presents the same problems. If we refuse the palm to Priestley, we cannot award it to Lavoisier for the work of 1775 which led

[3] J. B. Conant, *The Overthrow of the Phlogiston Theory: The Chemical Revolution of 1775–1789* ("Harvard Case Histories in Experimental Science," Case 2; Cambridge, Mass., 1950), p. 23. This very useful pamphlet reprints many of the relevant documents.

him to identify the gas as the "air itself entire." Presumably we wait for the work of 1776 and 1777 which led Lavoisier to see not merely the gas but what the gas was. Yet even this award could be questioned, for in 1777 and to the end of his life Lavoisier insisted that oxygen was an atomic "principle of acidity" and that oxygen gas was formed only when that "principle" united with caloric, the matter of heat.[4] Shall we therefore say that oxygen had not yet been discovered in 1777? Some may be tempted to do so. But the principle of acidity was not banished from chemistry until after 1810, and caloric lingered until the 1860's. Oxygen had become a standard chemical substance before either of those dates.

Clearly we need a new vocabulary and concepts for analyzing events like the discovery of oxygen. Though undoubtedly correct, the sentence, "Oxygen was discovered," misleads by suggesting that discovering something is a single simple act assimilable to our usual (and also questionable) concept of seeing. That is why we so readily assume that discovering, like seeing or touching, should be unequivocally attributable to an individual and to a moment in time. But the latter attribution is always impossible, and the former often is as well. Ignoring Scheele, we can safely say that oxygen had not been discovered before 1774, and we would probably also say that it had been discovered by 1777 or shortly thereafter. But within those limits or others like them, any attempt to date the discovery must inevitably be arbitrary because discovering a new sort of phenomenon is necessarily a complex event, one which involves recognizing both *that* something is and *what* it is. Note, for example, that if oxygen were dephlogisticated air for us, we should insist without hesitation that Priestley had discovered it, though we would still not know quite when. But if both observation and conceptualization, fact and assimilation to theory, are inseparably linked in discovery, then discovery is a process and must take time. Only when all the relevant conceptual categories are prepared in advance, in which case the phenomenon would not

[4] H. Metzger, *La philosophie de la matière chez Lavoisier* (Paris, 1935); and Daumas, *op. cit.*, chap. vii.

be of a new sort, can discovering *that* and discovering *what* occur effortlessly, together, and in an instant.

Grant now that discovery involves an extended, though not necessarily long, process of conceptual assimilation. Can we also say that it involves a change in paradigm? To that question, no general answer can yet be given, but in this case at least, the answer must be yes. What Lavoisier announced in his papers from 1777 on was not so much the discovery of oxygen as the oxygen theory of combustion. That theory was the keystone for a reformulation of chemistry so vast that it is usually called the chemical revolution. Indeed, if the discovery of oxygen had not been an intimate part of the emergence of a new paradigm for chemistry, the question of priority from which we began would never have seemed so important. In this case as in others, the value placed upon a new phenomenon and thus upon its discoverer varies with our estimate of the extent to which the phenomenon violated paradigm-induced anticipations. Notice, however, since it will be important later, that the discovery of oxygen was not by itself the cause of the change in chemical theory. Long before he played any part in the discovery of the new gas, Lavoisier was convinced both that something was wrong with the phlogiston theory and that burning bodies absorbed some part of the atmosphere. That much he had recorded in a sealed note deposited with the Secretary of the French Academy in 1772.[5] What the work on oxygen did was to give much additional form and structure to Lavoisier's earlier sense that something was amiss. It told him a thing he was already prepared to discover—the nature of the substance that combustion removes from the atmosphere. That advance awareness of difficulties must be a significant part of what enabled Lavoisier to see in experiments like Priestley's a gas that Priestley had been unable to see there himself. Conversely, the fact that a major paradigm revision was needed to see what Lavoisier saw must be the principal reason why Priestley was, to the end of his long life, unable to see it.

[5] The most authoritative account of the origin of Lavoisier's discontent is Henry Guerlac, *Lavoisier—the Crucial Year: The Background and Origin of His First Experiments on Combustion in 1772* (Ithaca, N.Y., 1961).

Two other and far briefer examples will reinforce much that has just been said and simultaneously carry us from an elucidation of the nature of discoveries toward an understanding of the circumstances under which they emerge in science. In an effort to represent the main ways in which discoveries can come about, these examples are chosen to be different both from each other and from the discovery of oxygen. The first, X-rays, is a classic case of discovery through accident, a type that occurs more frequently than the impersonal standards of scientific reporting allow us easily to realize. Its story opens on the day that the physicist Roentgen interrupted a normal investigation of cathode rays because he had noticed that a barium platinocyanide screen at some distance from his shielded apparatus glowed when the discharge was in process. Further investigations—they required seven hectic weeks during which Roentgen rarely left the laboratory—indicated that the cause of the glow came in straight lines from the cathode ray tube, that the radiation cast shadows, could not be deflected by a magnet, and much else besides. Before announcing his discovery, Roentgen had convinced himself that his effect was not due to cathode rays but to an agent with at least some similarity to light.[6]

Even so brief an epitome reveals striking resemblances to the discovery of oxygen: before experimenting with red oxide of mercury, Lavoisier had performed experiments that did not produce the results anticipated under the phlogiston paradigm; Roentgen's discovery commenced with the recognition that his screen glowed when it should not. In both cases the perception of anomaly—of a phenomenon, that is, for which his paradigm had not readied the investigator—played an essential role in preparing the way for perception of novelty. But, again in both cases, the perception that something had gone wrong was only the prelude to discovery. Neither oxygen nor X-rays emerged without a further process of experimentation and assimilation. At what point in Roentgen's investigation, for example, ought we say that X-rays had actually been discovered? Not, in any

[6] L. W. Taylor, *Physics, the Pioneer Science* (Boston, 1941), pp. 790–94; and T. W. Chalmers, *Historic Researches* (London, 1949), pp. 218–19.

case, at the first instant, when all that had been noted was a glowing screen. At least one other investigator had seen that glow and, to his subsequent chagrin, discovered nothing at all.[7] Nor, it is almost as clear, can the moment of discovery be pushed forward to a point during the last week of investigation, by which time Roentgen was exploring the properties of the new radiation he had *already* discovered. We can only say that X-rays emerged in Würzburg between November 8 and December 28, 1895.

In a third area, however, the existence of significant parallels between the discoveries of oxygen and of X-rays is far less apparent. Unlike the discovery of oxygen, that of X-rays was not, at least for a decade after the event, implicated in any obvious upheaval in scientific theory. In what sense, then, can the assimilation of that discovery be said to have necessitated paradigm change? The case for denying such a change is very strong. To be sure, the paradigms subscribed to by Roentgen and his contemporaries could not have been used to predict X-rays. (Maxwell's electromagnetic theory had not yet been accepted everywhere, and the particulate theory of cathode rays was only one of several current speculations.) But neither did those paradigms, at least in any obvious sense, prohibit the existence of X-rays as the phlogiston theory had prohibited Lavoisier's interpretation of Priestley's gas. On the contrary, in 1895 accepted scientific theory and practice admitted a number of forms of radiation—visible, infrared, and ultraviolet. Why could not X-rays have been accepted as just one more form of a well-known class of natural phenomena? Why were they not, for example, received in the same way as the discovery of an additional chemical element? New elements to fill empty places in the periodic table were still being sought and found in Roentgen's day. Their pursuit was a standard project for normal science, and success was an occasion only for congratulations, not for surprise.

[7] E. T. Whittaker, *A History of the Theories of Aether and Electricity*, I (2d ed.; London, 1951), 358, n. 1. Sir George Thomson has informed me of a second near miss. Alerted by unaccountably fogged photographic plates, Sir William Crookes was also on the track of the discovery.

X-rays, however, were greeted not only with surprise but with shock. Lord Kelvin at first pronounced them an elaborate hoax.[8] Others, though they could not doubt the evidence, were clearly staggered by it. Though X-rays were not prohibited by established theory, they violated deeply entrenched expectations. Those expectations, I suggest, were implicit in the design and interpretation of established laboratory procedures. By the 1890's cathode ray equipment was widely deployed in numerous European laboratories. If Roentgen's apparatus had produced X-rays, then a number of other experimentalists must for some time have been producing those rays without knowing it. Perhaps those rays, which might well have other unacknowledged sources too, were implicated in behavior previously explained without reference to them. At the very least, several sorts of long familiar apparatus would in the future have to be shielded with lead. Previously completed work on normal projects would now have to be done again because earlier scientists had failed to recognize and control a relevant variable. X-rays, to be sure, opened up a new field and thus added to the potential domain of normal science. But they also, and this is now the more important point, changed fields that had already existed. In the process they denied previously paradigmatic types of instrumentation their right to that title.

In short, consciously or not, the decision to employ a particular piece of apparatus and to use it in a particular way carries an assumption that only certain sorts of circumstances will arise. There are instrumental as well as theoretical expectations, and they have often played a decisive role in scientific development. One such expectation is, for example, part of the story of oxygen's belated discovery. Using a standard test for "the goodness of air," both Priestley and Lavoisier mixed two volumes of their gas with one volume of nitric oxide, shook the mixture over water, and measured the volume of the gaseous residue. The previous experience from which this standard procedure had evolved assured them that with atmospheric air the residue

[8] Silvanus P. Thompson, *The Life of Sir William Thomson Baron Kelvin of Largs* (London, 1910), II, 1125.

would be one volume and that for any other gas (or for polluted air) it would be greater. In the oxygen experiments both found a residue close to one volume and identified the gas accordingly. Only much later and in part through an accident did Priestley renounce the standard procedure and try mixing nitric oxide with his gas in other proportions. He then found that with quadruple the volume of nitric oxide there was almost no residue at all. His commitment to the original test procedure—a procedure sanctioned by much previous experience—had been simultaneously a commitment to the non-existence of gases that could behave as oxygen did.[9]

Illustrations of this sort could be multiplied by reference, for example, to the belated identification of uranium fission. One reason why that nuclear reaction proved especially difficult to recognize was that men who knew what to expect when bombarding uranium chose chemical tests aimed mainly at elements from the upper end of the periodic table.[10] Ought we conclude from the frequency with which such instrumental commitments prove misleading that science should abandon standard tests and standard instruments? That would result in an inconceivable method of research. Paradigm procedures and applications are as necessary to science as paradigm laws and theories, and they have the same effects. Inevitably they restrict the phenomenological field accessible for scientific investigation at any

[9] Conant, *op. cit.*, pp. 18–20.

[10] K. K. Darrow, "Nuclear Fission," *Bell System Technical Journal*, XIX (1940), 267–89. Krypton, one of the two main fission products, seems not to have been identified by chemical means until after the reaction was well understood. Barium, the other product, was almost identified chemically at a late stage of the investigation because, as it happened, that element had to be added to the radioactive solution to precipitate the heavy element for which nuclear chemists were looking. Failure to separate that added barium from the radioactive product finally led, after the reaction had been repeatedly investigated for almost five years, to the following report: "As chemists we should be led by this research . . . to change all the names in the preceding [reaction] schema and thus write Ba, La, Ce instead of Ra, Ac, Th. But as 'nuclear chemists,' with close affiliations to physics, we cannot bring ourselves to this leap which would contradict all previous experience of nuclear physics. It may be that a series of strange accidents renders our results deceptive" (Otto Hahn and Fritz Strassman, "Über den Nachweis und das Verhalten der bei der Bestrahlung des Urans mittels Neutronen entstehended Erdalkalimetalle," *Die Naturwissenschaften*, XXVII [1939], 15).

given time. Recognizing that much, we may simultaneously see an essential sense in which a discovery like X-rays necessitates paradigm change—and therefore change in both procedures and expectations—for a special segment of the scientific community. As a result, we may also understand how the discovery of X-rays could seem to open a strange new world to many scientists and could thus participate so effectively in the crisis that led to twentieth-century physics.

Our final example of scientific discovery, that of the Leyden jar, belongs to a class that may be described as theory-induced. Initially, the term may seem paradoxical. Much that has been said so far suggests that discoveries predicted by theory in advance are parts of normal science and result in no *new sort* of fact. I have, for example, previously referred to the discoveries of new chemical elements during the second half of the nineteenth century as proceeding from normal science in that way. But not all theories are paradigm theories. Both during preparadigm periods and during the crises that lead to large-scale changes of paradigm, scientists usually develop many speculative and unarticulated theories that can themselves point the way to discovery. Often, however, that discovery is not quite the one anticipated by the speculative and tentative hypothesis. Only as experiment and tentative theory are together articulated to a match does the discovery emerge and the theory become a paradigm.

The discovery of the Leyden jar displays all these features as well as the others we have observed before. When it began, there was no single paradigm for electrical research. Instead, a number of theories, all derived from relatively accessible phenomena, were in competition. None of them succeeded in ordering the whole variety of electrical phenomena very well. That failure is the source of several of the anomalies that provide background for the discovery of the Leyden jar. One of the competing schools of electricians took electricity to be a fluid, and that conception led a number of men to attempt bottling the fluid by holding a water-filled glass vial in their hands and touching the water to a conductor suspended from an active electrostatic generator. On removing the jar from the machine and touching the water (or a conductor connected to it) with his free hand, each of these investigators experienced a severe shock. Those first experiments did not, however, provide electricians with the Leyden jar. That device emerged more slowly, and it is again impossible to say just when its discovery was completed. The initial attempts to store electrical fluid worked only because investigators held the vial in their hands while standing upon the ground. Electricians had still to learn that the jar required an outer as well as an inner conducting coating and that the fluid is not really stored in the jar at all. Somewhere in the course of the investigations that showed them this, and which introduced them to several other anomalous effects, the device that we call the Leyden jar emerged. Furthermore, the experiments that led to its emergence, many of them performed by Franklin, were also the ones that necessitated the drastic revision of the fluid theory and thus provided the first full paradigm for electricity.[11]

To a greater or lesser extent (corresponding to the continuum from the shocking to the anticipated result), the characteristics common to the three examples above are characteristic of all discoveries from which new sorts of phenomena emerge. Those characteristics include: the previous awareness of anomaly, the gradual and simultaneous emergence of both observational and conceptual recognition, and the consequent change of paradigm categories and procedures often accompanied by resistance. There is even evidence that these same characteristics are built into the nature of the perceptual process itself. In a psychological experiment that deserves to be far better known outside the trade, Bruner and Postman asked experimental subjects to identify on short and controlled exposure a series of playing cards. Many of the cards were normal, but some were made anoma-

[11] For various stages in the Leyden jar's evolution, see I. B. Cohen, *Franklin and Newton: An Inquiry into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof* (Philadelphia, 1956), pp. 385–86, 400–406, 452–67, 506–7. The last stage is described by Whittaker, *op. cit.*, pp. 50–52.

lous, e.g., a red six of spades and a black four of hearts. Each experimental run was constituted by the display of a single card to a single subject in a series of gradually increased exposures. After each exposure the subject was asked what he had seen, and the run was terminated by two successive correct identifications.[12]

Even on the shortest exposures many subjects identified most of the cards, and after a small increase all the subjects identified them all. For the normal cards these identifications were usually correct, but the anomalous cards were almost always identified, without apparent hesitation or puzzlement, as normal. The black four of hearts might, for example, be identified as the four of either spades or hearts. Without any awareness of trouble, it was immediately fitted to one of the conceptual categories prepared by prior experience. One would not even like to say that the subjects had seen something different from what they identified. With a further increase of exposure to the anomalous cards, subjects did begin to hesitate and to display awareness of anomaly. Exposed, for example, to the red six of spades, some would say: That's the six of spades, but there's something wrong with it—the black has a red border. Further increase of exposure resulted in still more hesitation and confusion until finally, and sometimes quite suddenly, most subjects would produce the correct identification without hesitation. Moreover, after doing this with two or three of the anomalous cards, they would have little further difficulty with the others. A few subjects, however, were never able to make the requisite adjustment of their categories. Even at forty times the average exposure required to recognize normal cards for what they were, more than 10 per cent of the anomalous cards were not correctly identified. And the subjects who then failed often experienced acute personal distress. One of them exclaimed: "I can't make the suit out, whatever it is. It didn't even look like a card that time. I don't know what color it is now or whether it's a spade or a heart. I'm

[12] J. S. Bruner and Leo Postman, "On the Perception of Incongruity: A Paradigm," *Journal of Personality*, XVIII (1949), 206–23.

not even sure now what a spade looks like. My God!"[13] In the next section we shall occasionally see scientists behaving this way too.

Either as a metaphor or because it reflects the nature of the mind, that psychological experiment provides a wonderfully simple and cogent schema for the process of scientific discovery. In science, as in the playing card experiment, novelty emerges only with difficulty, manifested by resistance, against a background provided by expectation. Initially, only the anticipated and usual are experienced even under circumstances where anomaly is later to be observed. Further acquaintance, however, does result in awareness of something wrong or does relate the effect to something that has gone wrong before. That awareness of anomaly opens a period in which conceptual categories are adjusted until the initially anomalous has become the anticipated. At this point the discovery has been completed. I have already urged that that process or one very much like it is involved in the emergence of all fundamental scientific novelties. Let me now point out that, recognizing the process, we can at last begin to see why normal science, a pursuit not directed to novelties and tending at first to suppress them, should nevertheless be so effective in causing them to arise.

In the development of any science, the first received paradigm is usually felt to account quite successfully for most of the observations and experiments easily accessible to that science's practitioners. Further development, therefore, ordinarily calls for the construction of elaborate equipment, the development of an esoteric vocabulary and skills, and a refinement of concepts that increasingly lessens their resemblance to their usual common-sense prototypes. That professionalization leads, on the one hand, to an immense restriction of the scientist's vision and to a considerable resistance to paradigm change. The science has become increasingly rigid. On the other hand, within those areas to which the paradigm directs the attention of the

[13] *Ibid.*, p. 218. My colleague Postman tells me that, though knowing all about the apparatus and display in advance, he nevertheless found looking at the incongruous cards acutely uncomfortable.

group, normal science leads to a detail of information and to a precision of the observation-theory match that could be achieved in no other way. Furthermore, that detail and precision-of-match have a value that transcends their not always very high intrinsic interest. Without the special apparatus that is constructed mainly for anticipated functions, the results that lead ultimately to novelty could not occur. And even when the apparatus exists, novelty ordinarily emerges only for the man who, knowing *with precision* what he should expect, is able to recognize that something has gone wrong. Anomaly appears only against the background provided by the paradigm. The more precise and far-reaching that paradigm is, the more sensitive an indicator it provides of anomaly and hence of an occasion for paradigm change. In the normal mode of discovery, even resistance to change has a use that will be explored more fully in the next section. By ensuring that the paradigm will not be too easily surrendered, resistance guarantees that scientists will not be lightly distracted and that the anomalies that lead to paradigm change will penetrate existing knowledge to the core. The very fact that a significant scientific novelty so often emerges simultaneously from several laboratories is an index both to the strongly traditional nature of normal science and to the completeness with which that traditional pursuit prepares the way for its own change.

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. L. 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), 1–58, 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 dis closed 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.