The Form and Function of Scientific Discoveries

Kenneth L. Caneva

Dibner Library Lecture, Smithsonian Institution Libraries, November 16, 2000

In the course of my work over the last twenty-five years, I have repeatedly noticed that what a scientist is typically credited with having discovered often differs significantly from the way in which the scientist himself characterized his work.

For example, everyone knows that Danish physicist Hans Christian Ørsted (1777-1851) discovered electromagnetism in 1820. Understanding electromagnetism to be the interaction between electricity and magnetism, I was startled, a few years ago, to discover that Ørsted had intended the term-which he introduced-to mean something much different, indeed something quite foreign to the way we now view things.[1]

One of the many researchers inspired by Ørsted's work was the Berlin academician, Thomas Johann Seebeck (1770-1831), whom the world of science and scholarship honors as the discoverer of thermoelectricity.[2] Hence in studying German work in electricity during the first half of the nineteenth century many years ago, I was startled to find that Seebeck's well-known discovery of thermoelectricity in 1821 did not correspond to what he himself thought he had found.[3] Failing to pay due attention to the significance of the title of his collective memoir, "On the Magnetic Polarization of Metals and Ores by Means of Temperature Difference," I followed others in mistakenly thinking that what Seebeck had thought he discovered was thermomagnetism. Although there is some truth to that, the full story is more complicated.

In a third and still more complicated case, a study of the work of German romantic physicist Johann Wilhelm Ritter (1776-1810), I came from the secondary literature with the naive idea that among Ritter's lasting achievements was his discovery of ultraviolet light. Even knowing something of Ritter's indebtedness to Schelling's Naturphilosophie I was startled to discover the overwhelming otherness of the terms in which he reported his findings, embedded as they were in a wide-ranging web of polarities.[4]

For the most part I tended to see such instances as examples of careless historiography, of the distortions of a story rewritten from a later perspective. My thinking began to change as I noticed that it wasn't just that historians were guilty of retrospective distortion, but that such recastings were an intrinsic part of the scientific enterprise. More recently, I have come to see how such instances can be exploited as examples of how what is generally accepted as scientific knowledge is essentially the outcome of a process by which knowledge is reshaped as it passes through the hands of people with different agendas using different language. Such a 'discovery' is not an atomized contribution to knowledge that others need merely recognize and accept, but rather represents a retrospective characterization coming at the end of a complex process of transformative negotiation. That characterization simultaneously formalizes the essential character of the discovery and confers upon it the stamp of objectivity as an aspect of the physical world that was there waiting to be 'discovered.' Before developing these interpretive aspects any further, let me pass in review some of the particulars upon which my interpretation rests.

Ørsted and electromagnetism

The electromagnetism that Ørsted is universally credited with having discovered is-to use an older vocabulary-typically understood to consist in the interaction between magnetism and so-called voltaic or galvanic electricity, or, more phenomenologically, in the interaction between the connecting wire of a voltaic pile and a magnetic needle.[5] (*Figure* 1. First page of the pamphlet in which Ørsted announced his discovery of electromagneticism (Ørsted 1820a). Courtesy Dibner Library, Smithsonian Institution Libraries).


EXPERIMENTA
CIRCA EFFECTUM
CONFLICTUS ELECTRICI IN ACUM
MAGNETICAM.

That was also the way his discovery was characterized by the commentators in Paris and Geneva who immediately and consequentially followed up on his work. So, too, in England. In Germany the situation was more complicated. For the most part, however, Ørsted himself had a different understanding of what the electromagnetism was that he had discovered and named. Our story is thus one of changing interpretations.

In a phrase quietly recalling his involvement with the conceptual schemes of Naturphilosophie, Ørsted had spoken initially of "the effect of the electrical conflict on the magnetic needle."[6] Ørsted did not thereafter much use the term conflictus electricus (in any language), and his contemporaries seem to have paid little attention to his declarations that the electrical conflict is not confined to the connecting wire, but also takes place in the circumjacent space, spiraling around the conductor.[7]

Hans Chrisian Ørsted in 1822. Portrait by Christopher Wilhelm Eckersberg (1788-1853). Courtesy Danmarks Tekniske Museum, Helsingor. A 1959 copy by Daniel Hvidt hangs in the library of the Dibner Institute in Cambridge, Massachusetts.

In a followup paper on his "Recent Electromagnetic Experiments," Ørsted introduced the adjective thenceforth favored to describe the new phenomena electromagnetic.[8] At this point it is necessary to review some of Ørsted's earlier concerns, which were to impress a peculiar stamp on his interpretation of the new phenomena.

From early on in his career, Ørsted had struggled to attain a systematic understanding of the relationship among magnetism, electricity, the so-called chemical process, heat, and light.[9] Following Schelling, in 1805 he speculated that magnetism, electricity, and the chemical process correspond to the three dimensions of space-magnetism to a line, electricity to a surface, and the chemical process to space.[10] More fruitfully for him, in 1812 Ørsted had begun to speak of the different "forms of action" in which the opposing fundamental forces of nature manifest themselves.[11] It was the pursuit of these speculations that led him to his discovery of the interaction between a magnetic needle and the connecting wire of a galvanic circuit.

In his first Danish-language discussion of that work, he interpreted the new relationships by invoking an echo of his notion of form of action: "What we here a moment ago called electricity is not so in the word's stricter meaning; for the force that in the open galvanic or electric circuit acted in a distinctive manner-under a different form-that we call the electric or galvanic, acts here under an entirely different form that we most appropriately call the magnetic; meanwhile, since magnetism acts under the form of a straight line …[while] the forces here … flow incessantly into each other and form a circular course [Kredslöb], the author has called the action dealt with here electromagnetism."[12] It thus appears that for Ørsted the principal need for a new term stemmed from the unprecedented circular form of the electromagnetic action and not so much from the fact that it represented an interaction between electricity and magnetism. But Ørsted early left the development of this new field to others, and his peculiar conceptualization of the phenomena died without issue.

If that was what Ørsted understood by electromagnetism, then we must reconsider what we mean by saying that 'Ørsted discovered electromagnetism.' He did indeed discover that a magnetic needle is deflected by some action present in the connecting wire, but that was not the meaning of the discovery for him. While he applied the term electromagnetism to his particular theory of the form of action of spiraling electromagnetic activity, others applied it simply to the new phenomena. And that, for the most part, is the meaning that has come down to us, sanctioned by longstanding consensus.[13]

Before the dust had settled, however, the issue over whether the phenomena were to be traced to an underlying electric current or an underlying magnetism gave rise to several theories of so-called transversal magnetism, which supposed the existence of small magnets running head-to-tail around the circumference of the conducting wire.[14] Such ideas never gained much currency outside the German-speaking scientific community, and their appeal quickly waned as Ampère's electrodynamic theory gained ground, but for a time they were a serious contender for the explanation of what virtually everyone recognized as electromagnetic phenomena.

Seebeck and thermoelectricity

Although most writers have continued to speak in terms of Seebeck's discovery of thermoelectricity, a minority voice-including my own-has insisted that Seebeck termed the new phenomena thermomagnetism, although without actually citing examples of that usage from Seebeck's own work.[15] In the event, it turned out that the identification of just what Seebeck had discovered was a complex process in which Seebeck himself played only a late and ineffectual role. Because of his nearly four-year delay in publishing his findings, conceptual and terminological possession of the new field fell to others as word of his experimental findings leaked out and then attracted the attention of a spate of investigators throughout Europe.

Thomas Johann Seebeck. Undated engraving, courtesy Deutsches Museum, Munich.

Seebeck reported his early findings at three meetings of the Academy of Sciences in Berlin in August and October of 1821.[16] The earliest published accounts reported his having discovered how to produce magnetic effects in metals other than iron by heating various metals connected together into a closed circuit.[17] The phenomenon was explicitly magnetic, sometimes more specifically electromagnetic. By the fall of 1822 it had become clear to the scientific public just what the essential phenomena consisted in experimentally. How they were to be properly characterized remained a topic of discussion through the 1830s, with a clear consensus in favor of a tacitly unproblematical "thermoelectric" not being achieved till after around 1840. The role of Seebeck's Danish colleague in this process was pivotal.[18]

Ørsted, in Berlin for a week or so during November and December of 1822, spent two mornings and an afternoon with Seebeck being shown the new experiments that Ørsted immediately regarded as continuations of his own discovery.[19] Having arrived in Paris in January 1823, Ørsted informed the French Academy in March of "Seebeck's new experiments on electromagnetic actions."[20] Its first sentence continued the subtle process by which Ørsted sought to control the interpretation of those experiments: "Seebeck … has discovered that one can establish an electric circuit in metals without the interposition of any liquid. One establishes the current in this circuit by disturbing the equilibrium of temperature."[21] In noting that "[o]ne can only discover these electric currents by means of the magnetized needle," he was already ignoring the issue of the legitimacy of speaking in terms of an electric current when what one observes is magnetic effects.[22] He went on to suggest several new coinages, which would eventually become standard: "It will from now on doubtless be necessary to distinguish this new class of electric circuits by an appropriate term; and as such I propose the expression thermoelectric circuits …; at the same time one would be able to distinguish the galvanic circuit by the name hydroelectric circuit."[23] His later Danish account already reported impersonally-if not entirely accurately-that "[o]ne now calls the Seebeckian circuit the thermoelectric circuit, and the Galvanic, in opposition thereto, the hydroelectric circuit."[24] By the end of the year most researchers in France and England had adopted the language of thermoelectricity and its attendant conceptualization of the phenomena as due to the generation of an electric current in the metallic circuit. In Germany things were a little more complicated.

Among the earliest to pursue Seebeck's lead was Julius Conrad von Yelin in Munich. Yelin had been investigating the relationship between magnetism and heat and light when Ørsted, passing through Munich, informed him of Seebeck's work. As soon as Ørsted left Munich, Yelin undertook his own experiments, reporting his first results to the Bavarian Academy in January 1823.[25] He referred to Seebeck's "thermelectromagnetic experiments" as he went on to claim for himself the discovery "that through unequal heating all bodies acquire magnetomotive properties," a fact he demonstrated by forming circuits composed of a single metal which, when heated at one place, produce a deflection of an appropriately placed magnetic needle.[26] (Figures 2 and 3.) Having further detected a magnetic effect-a magnetic polarization- by the appropriate heating of a bar of metal not part of a closed circuit, he noted that "[b]ismuth most strikingly exhibits the polarization occurring in this thermomagnetism of metals." A footnote explained the phrase "Thermo-Magnetismus der Metalle": "It appears to me that we will thus have to denote in a characteristic fashion this kind of magnetic action, in contradistinction to the already known Ørstedian electromagnetism, on account of its peculiar behavior."[27] Yelin's further discussion made clear that he did not intend thermomagnetism to apply to the full range of phenomena discovered by Seebeck, but only to a new class of actions he had discovered which did not appear to involve a closed electric circuit and in which no trace of free electricity could be detected.[28] That was not, however, the way the term came to be employed by others, who applied it without restriction to the range of phenomena discovered by Seebeck.

As of 1825 there was still no terminological or conceptual consensus. Were the phenomena to be described simply and generically as electromagnetic, in a more phenomenological sense of the word than Ørsted had intended with its coinage, or perhaps, more specifically, as thermelectromagnetic, as Yelin proposed? Or were they rather better characterized as thermoelectric, as Ørsted confidently urged? Others broadened the application of Yelin's more restricted concept of thermomagnetism to cover the same field as thermoelectricity. Behind such terminological matters lay uncertainty over whether the phenomena were fundamentally electric or magnetic, whether magnetism could be reduced to electricity (as Ampère said), and whether and where one was entitled to speak of an electric current. Nor were matters brought closer to settlement when Seebeck's long memoir finally made its appearance toward the end of 1825.

As mentioned earlier, the discovery Seebeck announced in the title of his memoir was the "Magnetic Polarization of Metals and Ores by Means of Temperature Difference."[29] In repeating and extending Ørsted's experiments, Seebeck came to suspect that any inequality of action on the metals used in the galvanic circuit might produce what he persisted in calling a "magnetic polarization" of the circuit, his choice of terms indicating his disinclination to trace Ørsted's electromagnetic phenomena to the generation of an electric current. A series of trials led him to the discovery that heat alone, applied to the junction of two metals, would produce that effect.

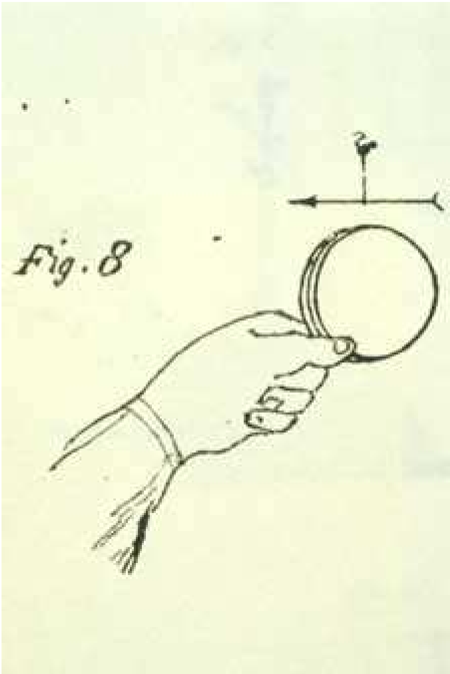

Figure 2. Demonstration of the deflection of a magnetic needle by the "thermoelectromagnetic" action of disks of two different metals held between the fingers. From Yelin 1823c.

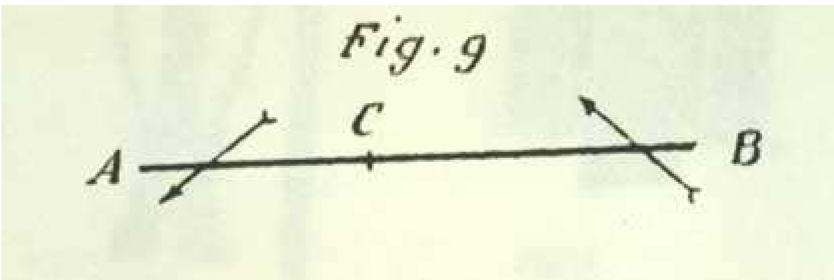

Figure 3. The "thermomagnetism" exhibited by a rod of a single metal heated along CB and cooled along AC, showing the opposite deflection of magnetic needles along each half. From Yelin 1823c.

He thereupon took pairwise combinations of twenty-eight different metals and ores joined together to form a ring, heating one junction and noting the direction and rough strength of the resulting "magnetic polarization" by means of the deflection of a suspended magnetic needle.[30] From these experiments he concluded that metals form "a peculiar magnetic series that does not correspond to any of the known series derived from other properties of metals"-in particular, the well- established electromotive series of metals governing the production of electricity in the galvanic circuit.[31] This fact underlay his rejection of the hypothesis of the identity of electricity and magnetism and, in particular, Ampère's derivation of magnetism from electricity.[32] Seebeck argued that the magnetic polarization he observed could not be solely due to any free electricity-such as could be detected with an electroscope-and that one was therefore not justified in calling such circuits "electromagnetic."[33]

What Seebeck thought was going on was the excitation of a "magnetic polarity" or "magnetic polarization" by means of a difference in temperature.[34] Electricity as such played no role. Nor in his original paper did Seebeck employ the language of thermomagnetism, though he did use it five times in subsequently added notes. That was the extent to which Seebeck retrospectively assimilated his discovery of magnetic polarization to the language of thermomagnetism in this, his first and last paper on the subject. To many of his contemporaries, however-especially among German contemporaries of his generation-that was what Seebeck was credited with having discovered until well into the 1830s.

Although time prevents me even from sketching the details of the community-wide reception of Seebeck's and others' work, by the 1840s it appears that, for the most part, the language of thermoelectricity had wholly driven out that of thermomagnetism. Few seem ever to have noticed that Seebeck himself had spoken of the "magnetic polarization" of metals by heat. In the end, what Seebeck discovered was decided for him by others.

Ritter and ultraviolet light

An even more striking example of this sort of recasting of meaning has attended what we have learned to see as Ritter's discovery of ultraviolet light, rays, or radiation in 1801.[35] The path from Ritter's work to our conceptualization of it is long and complicated. Our odyssey begins just before the earliest accounts of his discovery.

Prompted by William Herschel's discovery in 1800 of calorific rays beyond the red end of the solar spectrum and his own conviction that polarities underlie the principal phenomena of nature, Ritter thought to see if he could identify invisible solar rays also at the other end of the spectrum. As he wrote hastily to the editor of the Annalen der Physik, "On February 22nd [1801] I found solar rays-discovered by means of horn silver-also on the violet side of the spectrum of colors, outside of it. They reduce even more strongly than violet light itself, and the field of these rays is very large… . More on this soon."[36] He announced his discovery more fully in a note in an Erlangen periodical with the revealing title, "Chemical Polarity in Light," in which he summarized his findings under three headings:

> I. There are rays in sunlight that do not illuminate and of which one part is refracted more strongly, the other more weakly, than all those that illuminate.
>
> II. Sunlight in the undivided state is a neutralization of the two ultimate determinants of all chemical activity, oxygeneity and deoxygeneity (= hydrogeneity).
>
> III. By means of the prism the two diverge like poles. The red side of the spectrum and that which borders on it externally become the side of oxygeneity, the violet side, on the contrary, and that which borders on it become the side of hydrogeneity. The maxima of both fall outside the visible spectrum; their indifference, however [falls] inside it in the region of green.[37]

What most excited Ritter was the evidence his discovery provided for the thoroughgoing polarity of all the activities of nature and their complex interconnections. In his final paper on the subject Ritter emphasized the fundamental distinctiveness of the invisible chemical and thermal rays he and Herschel had discovered vis-à-vis the light of the visible spectrum.[38]

Johann Wilhelm Ritter. Undated woodcut, courtesy Deutsches Museum, Munich.
Reproduced in Ritter 1986.

There seemed to be three distinct spectra, which an appropriate arrangement of prisms could separate out from the solar rays.

The light under which Ritter's work was seen was soon colored by the contemporaneous work of William Hyde Wollaston. Like Ritter, Wollaston was inspired by Herschel's discovery in 1800 of calorific rays less refrangible than red light and guided by Scheele's experiments with muriate of silver. Wollaston reported in June 1802 that "on the other [side of the solar spectrum] I have myself observed, (and the same remark has been made by Mr. Ritter,) that there are likewise invisible rays of another kind, that are more refracted than the violet. It is by their chemical effects alone that the existence of these can be discovered."[39] Finding that the blackening of the silver chloride extended far beyond the violet end of the spectrum and that "by narrowing the pencil of light received by the prism, the discoloration may be made to fall almost entirely beyond the violet," Wollaston concluded "that this and other effects usually attributed to light, are not in fact owing to any of the rays usually perceived, but to invisible rays that accompany them."[40] Although there were to be a few dissenters, Wollaston's insistence on the essential distinctiveness of chemically active and luminous rays was generally subscribed to until the 1840s. Indeed, commentators from the early 1840s through the early 1860s regularly remarked on the generality of the belief in the qualitative distinctiveness of the chemical, calorific, and luminous rays, a belief they were concerned to oppose.

Things began to change decisively in 1842 with the combined advocacy of Macedonio Melloni and Edmond Becquerel for the fundamental identity of all the variously named rays of the solar spectrum, and with Ludwig Moser's more limited advocacy of the identity of the chemical and luminous rays.[41] Their significantly different arguments all combined an appeal to the identity of the physical laws those rays obey, an acceptance of the wave theory of light-whereby the only relevant variable seemed to be the refrangibility or frequency of the ray-and a distinction between the latter as the "essential properties" of the rays versus the "accidental qualities" those rays exhibited in their various thermal, luminous, and chemical effects.[42] For a long time it had been known that different observers place the ends of the visible spectrum at different places; now the conclusion was decisively drawn that that phenomenon illustrates the fundamentally subjective nature of color and visibility. Relinquished now was the deep-seated conviction since Newton's day that color is an intrinsic property of light. Against earlier experimenters who had failed to detect any heat towards and beyond the violet end of the spectrum, Melloni claimed that it was now possible to measure the heat developed by all types of rays falling on a black surface. Where visibility had once been the defining character of light, Moser now spoke of "invisible light," implying that some criterion other than visibility is necessary to define what constitutes light.[43]

It was Becquerel, however, who provided what was to prove to be perhaps the most decisive bit of evidence in favor of the identity thesis as he exploited the recent development of photography for scientific ends. Exposing to sunlight a plate prepared "according to M. Daguerre's method," he obtained what he called "this curious result, that the chemical spectrum has the same lines as the luminous spectrum, provided we only consider the parts of the same degree of refrangibility of these two radiations."[44] Incorporating the results of several such exposures into a single diagram, he illustrated these spectral lines in a large and beautiful plate, pairing them line for line with the corresponding lines that Fraunhofer had first identified and named for the visible spectrum. (Figure 4.) It was hard, now, to avoid the conclusion that it is the same rays-as defined by their refrangibility-which produce either luminous or chemical effects, according to circumstances, and which make up a visually continuous spectrum indicating no qualitative difference as one passes from the visible into the invisible regions of the solar spectrum.

If, to be sure, the argument for the ontological identity of the various species of radiation depended greatly on the accumulation of good, hard evidence, it also derived much of its plausibility from the increasingly taken-for-granted implications of the wave theory of radiation. Thus Ernst Brücke, after a tortuous consideration of the experimental pros and cons of the still- disputed identity of the rays of light and heat, concluded as follows:

> If on the other hand one recognizes that the rays of light and heat are both polarizable (thus both are composed of transverse waves), [and] that both pass through empty space (thus both, if one does not wish to assume a second unknown medium besides the aether, must consist of oscillations of one and the same medium), then one perceives that there no longer exists any mechanical understanding for a difference between the two radiations. If one further recognizes that the invisibility of the rays beyond the red and beyond the violet-even if they differ from the luminous rays only in terms of wavelength-has absolutely nothing mysterious about it, then it must appear rash to give up the identity hypothesis completely and, with regard to radiant heat, to sink back into the former perplexity before one has tested in the most precise fashion the probative force of the counterexperiments.[45]

In other words, the inconceivability of radiant heat's being in any way essentially different from other forms of solar radiation, as judged from the standpoint of the generally accepted undulatory theory, should outweigh any experimental quibbles that might suggest some underlying difference. There seems to be a tipping point with regard to the believability of a theory, beyond which potential counterevidence no longer easily counts as such, but is rather generously reinterpreted so as to be in likely accord with the now-accepted view of things.

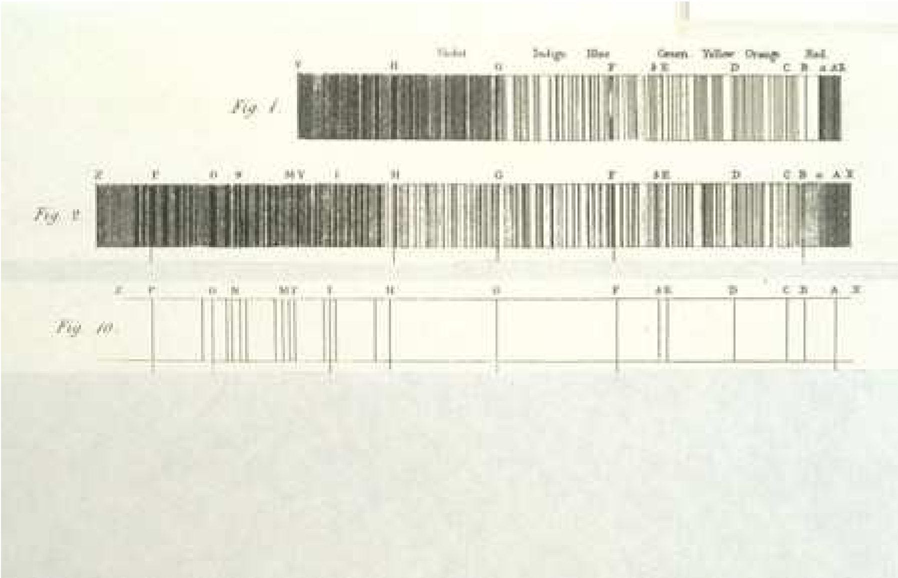

Figure 4. Edmond Becquerel's demonstration of the correspindence between the dark lines in the luminous and chemical solar spectra. Fig. 1 represents the luminous spectrum as drawn by Fraunhofer. Fig. 2 represents a composite chemical spectrum from a number of photographic exposures. Fig. 10 is a schematic representation of the dark lines common to all solar spectra. From E. Becquerel 1843 (edited).

The explanatory function of attachment to the wave theory of light is nicely illustrated by the response of George Gabriel Stokes to the phenomenon of internal dispersion he named "fluorescence" in 1852.[46] It was obvious to him that the phenomenon could only be explained by one or the other of the only two things that define the nature of light, its period of vibration and its state of polarization. To its period of vibration, of course, corresponds its refrangibility. Aware of the long tradition that assigned qualitative differences to the rays responsible for producing different effects, Stokes went out of his way to emphasize that there are not and cannot be any such differences: visibility is a function of the eye depending solely on the rays' refrangibility, not a property of some special class of rays. The last of his formal conclusions was of universal scope:

> The phenomena of internal dispersion oppose fresh difficulties to the supposition of a difference of nature in luminous, chemical, and phosphorogenic rays, but are perfectly conformable to the supposition that the production of light, of chemical changes, and of phosphoric excitement, are merely different effects of the same cause. The phosphorogenic rays of an electric spark … appear to be nothing more than invisible rays of excessively high refrangibility, which there is no reason for supposing to be of a different nature from rays of light.[47]

Embracing this new conceptualization of light, Wilhelm Friedrich Eisenlohr soon thereafter introduced the term that defined the region beyond the violet end of the spectrum not in terms of its chemical action but simply in terms of its location along a spectrum that was now explicitly all light.[48] He opened his 1854 paper, "On the Action of Violet and Ultraviolet Invisible Light," with a reference to Stokes' recent work:

> The phenomenon to which Stokes gave the name fluorescence has led me to the conjecture that it is produced by interference of the shorter systems of waves, blue-violet and ultraviolet (thus can one call for brevity's sake the chemically active invisible light alongside the violet in the spectrum). … Light itself consists accordingly of the visible systems of waves and, in addition, of waves that are longer than red and of such that are shorter than violet. … Ultraviolet light … whose existence before the wonderful discovery of Stokes could only be demonstrated by its chemical activity, consists of innumerable systems of waves, whose mutually different lengths all have a shorter period of oscillation than violet light.[49]

For Eisenlohr, the apparent transformation of highly refrangible and invisible ultraviolet light-the erstwhile chemical rays-into less refrangible and thus visible light meant that all are of essentially the same nature. Other Germans quickly adopted the new terminology of "ultraviolet light," "ultraviolet rays," and "the ultraviolet."[50] (Figure 5.)

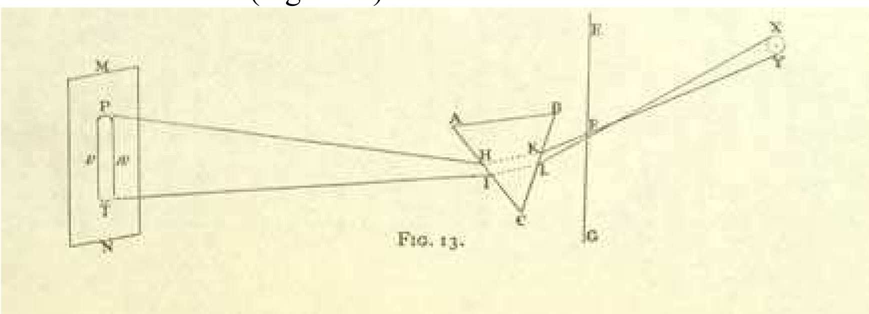

Figure 5. Newton's illustration - from the first edition (1704) of his *Opticks* - of the theorem that "The Light of the Sun consists of Rays differently Refrangible." His arrangement of the prism and the resulting orientation of the solar spectrum -- violet above and red below -- became canonical and probably encouraged the terminology of "ultraviolet" and "infrared." Courtesy Dibner Library, Smithsonian Institution Libraries.

Although the more detailed tracking of this usage into the standard vocabularies of European scientists is still to be done, we are finally at the point where it becomes possible even to speak of Ritter's having discovered ultraviolet light, regardless of whether or not that misrepresents what he thought he had done. It should be clear that major conceptual transformations had to take place with regard to the common understanding of light-entailing both general acceptance of the wave theory of light and the concomitant conclusion that solar rays simply could not have any qualitatively distinguishing features-before such terminological rephrasing was possible. That change in understanding itself very much depended both on the cumulative evidence pointing to the analogous physical behavior of the variously named rays-same laws of refraction, interference, etc.-and, in particular, on the evidence provided by Becquerel's photograph of the broad chemical spectrum, showing the exact correspondence between its dark lines and those Fraunhofer had identified in the luminous spectrum. That and, ten years later, Stokes' forceful advocacy of the undulatory identity of the entire range of solar radiations, coupled with his experimental demonstration of the transformation of invisible into visible rays, seem to have been the most important events in bringing about a general shift in the way scientists thought about light. The changes in vocabulary that followed soon after finally made it possible to speak without circumlocution of invisible light rays beyond the violet end of the spectrum. Their defining characteristic was now their location along a spectrum of varying refrangibilities, not the chemical means by which their presence was detected, let alone a presumed qualitative difference between them and other species of solar rays regarded as representing distinct and separable spectra. If, after all this, we still do not want to say that Ritter discovered ultraviolet light, we might nevertheless be happy with saying that Ritter became the discoverer of ultraviolet light.[51]

Reflections on the form and function of scientific discoveries

The origins of this paper lay with the recognition that the conventional characterization of what Ørsted, Seebeck, and Ritter discovered does not correspond to what each man thought he had discovered. The bulk of the historical material adduced here has been chosen to demonstrate that fact and to sketch the reasons for and route by which that conventional characterization was attained. That essentially descriptive enterprise must now give way to an analysis of the general significance of that process, a process one might describe as the collective construction of scientific knowledge.[52]

Perhaps the most important point about the characterization of any discovery is that, in order to be intelligible, it must be phrased in language understood by the intended audience, in language that typically implicates the taken-for-granted reality of that audience. It thus verges on the unintelligible to say that

Ørsted discovered the circular form of the electromagnetic action of the electrical conflict, or that Seebeck discovered the magnetic polarization of metals by heat, or that Ritter discovered the chemical polarity of light-in his words, that "[s]unlight in the undivided state is a neutralization of the two ultimate determinants of all chemical activity, oxygeneity and deoxygeneity."[53] To be sure, each man's discovery can be rendered in largely phenomenological terms: do this, and you will observe this. Appropriately suspend a magnetized needle near the connecting wire of a closed galvanic circuit, and you will observe the needle to move in a certain way. Indeed, in each case contemporaries' first order of business was precisely to convince themselves that the phenomena were in fact as described, and in every case the establishment of at least phenomenological consensus was quick and lasting: the reported phenomena are real and repeatable, constituted by generally understood material practices. In that sense, too, these discoveries have retained their legitimacy and integrity as discoveries of phenomena that have an objective existence in physical reality even as their characterization has changed.

But, then as now, scientists do not typically content themselves with purely phenomenological descriptions-nor did any of our three principals. Where Ørsted saw the effect of an electrical conflict extending beyond the confines of the connecting wire, others saw the effect of a transversal magnetism within the wire. Where Seebeck saw an unambiguously magnetic effect, others, like Ørsted, saw that same effect as the clear and direct manifestation of an electric current. What Ritter actually observed was the local darkening of silver chloride beyond the violet end of the visible solar spectrum, but he, like everyone else, saw that as resulting from the action of invisible rays. All such characterizations entail the acceptance of the existence of certain theoretical entities-entities which, however, are commonly not seen as theoretical constructs but as unproblematic statements about things in the world. Or at least they sooner or later become that. The discoveries we've been concerned with became 'facts' only by being recast in terms corresponding to accepted theory in congruence with other accepted facts and theories.[54] As people's ideas change about the underlying nature of physical reality, so too must their characterizations change of what it was that someone discovered. The process by which a discovery acquires meaning is thus not effectively over until a locally stable consensus has been attained with regard to its characterization in up-to-date terms.

Which in turn brings us to one of the most important functions of anachronistically recast discovery accounts: they validate as real current views both about the nature of reality and the nature of science.[55] More than that, by saying that so-and-so discovered such-and-such, one implies that such-and-such was there all along waiting to be discovered, an objective part of physical reality whose nature is given by the very structure of the world, not by a labored process of consensus formation.[56] Acknowledging an historical element in the characterization of basic scientific phenomena threatens their status as objective facts.[57] Indeed, the perceived danger of the so-called social construction of scientific knowledge is that, by historicizing knowledge claims to a contingent context, it threatens to strip them of any claim to objective truth and thus to undercut the authority of science.[58] Recast discovery accounts help avert this danger, especially since conventionality of representation itself seems to imply a kind of objectivity. De facto terminological consensus is necessary before a phenomenon can be taken for granted as simply 'the way the world is.'

The concept of scientific discovery is thus intimately connected with the concept of scientific fact, and the transformations that discovery accounts undergo all reinforce the objectivity, the facticity, of that which was discovered. Although the name of the discoverer survives as an identifying icon, all that was personal, historically contingent, and not in accord with presently employed concepts and language is erased from the public record. Discoveries tell us the way things are.

From a sociological perspective, discoveries function as part of the intellectual reward structure of science. Indeed, to some extent-especially in the nineteenth century-making a discovery was what it meant to be an original and productive scientist. Hence canonized discovery accounts reinforce a particular conception of the role of the individual scientist as the originator of the facts, concepts, and phenomena that make up the abiding edifice of science.[59]

A final function of anachronistically simplified attributions of discovery cuts closer to home. It will not have escaped the notice of many historians of science how difficult it can be in the classroom not to give in to the simplifying formula, so-and-so discovered such-and-such in some particular year. Even if the complex stories one might tell concerning the kinds of 'discoveries' dealt with here would in fact enable us to make other important points about the history of science and the nature of the processes by which scientific knowledge is produced, we usually don't have the time to elucidate the contexts within which the work of-for example-Ørsted, Seebeck, and Ritter must be situated in order to be understood properly. We typically don't even know the full story behind the perhaps decades-long process by which such 'discoveries' acquired their canonical form. And if we did, it would usually be too complicated to tell students. Hence we often collude in the perpetuation of an epistemological fiction as we struggle to achieve order-i.e., certainty, simplicity, and vividness, to use Ludwik Fleck's terms.[60] The only alternative, both pedagogically and professionally, would seem to be to refuse to separate knowledge as a product from the process of its production. In both historical and scientific practice, however, such a solution proves very difficult to sustain both because it is intrinsically difficult and because its denial offers such tangible rewards.

~top of page~